Good Data and Bad Data: The Welfare Effects of Price Discrimination

Maryam Farboodi* Nima Haghpanah† Ali Shourideh‡

November 18, 2025§

Abstract

We study how a monopolist's use of consumer data for price discrimination affects welfare. To answer this question, we develop a model of market segmentation subject to residual uncertainty. We fully characterize when data usage monotonically increases or decreases welfare or when the effect is non-monotone. The characterization reduces the problem to one with only two demand curves, and gives a condition for the two-demand-curves case that highlights that information affects welfare in three distinct ways. In the non-monotone case, we provide tight bounds on the welfare effects of information and identify the best local direction for providing additional information.

*MIT Sloan, NBER and CEPR. Email: farboodi@mit.edu.

†Yale School of Management. Email: nima.haghpanah@gmail.com.

‡Tepper School of Business, Carnegie Mellon University. Email: ashourid@andrew.cmu.edu.

§We thank Nageeb Ali, Dirk Bergemann, Ben Brooks, Alessandro Bonatti, Roberto Corrao, Stephen Morris, Jonathan Parker, Maryam Saeedi, Roland Strausz, Alexander Wolitzky, Jidong Zhou, and audiences at various seminars and conferences for their thoughts and comments.

1 Introduction

The rise of big data technologies, allowing firms to collect detailed consumer data to estimate their willingness to pay, has reignited the longstanding debate on the welfare implications of price discrimination. The prevalence of these practices has raised concerns among policymakers about big tech's exploitation of consumer data. A letter that followed a Senate hearing on May 2, 2024, reflects these concerns:

> As more consumers shop online, large tech platforms have access to vast stores of personal data [...] that can be exploited by corporations to set prices based on the time of day, location, or even the electronic device used by a consumer.[1]

The Federal Trade Commission also recently issued an order seeking information from several companies about their "segmentation solutions" that categorize consumers based on location, demographics, and credit history and set different prices for the same good or service.[2]

We study the welfare effects of price discrimination through the lens of information economics. There is a given set of consumer types, each associated with a downward-sloping demand curve representing a population of consumers with heterogeneous willingness to pay. The seller has access to some information structure that maps types to signal realizations, allowing her to segment the market and charge a profit-maximizing price for each segment. Importantly, the information is only about consumer types and cannot distinguish between consumers of the same type. Even if the seller perfectly observes types, she faces residual uncertainty in the form of a downward-sloping demand curve for each type and cannot implement first-degree price discrimination. This residual uncertainty reflects the practical limitations, legal and technological, that sellers face in perfectly predicting individual willingness to pay.

By providing a framework to study all market segmentations based on consumer types, our work bridges the classical and modern approaches to price discrimination. The classical approach to this problem, pioneered by Pigou (1920)'s foundational work, starts with the same primitives but, translated to our setting, compares only the two extreme cases where the seller is either fully informed or fully uninformed about the types. Because this literature compares two extremes, it only provides sufficient conditions for price discrimination to affect welfare positively or negatively, whereas our focus on the smooth class of all information structures allows us to obtain sharp results and complete characterizations. On the other hand, the modern approach to this problem, pioneered by the seminal work of Bergemann et al. (2015), studies all market segmentations in a setting where the seller has access to signals about consumer *willingness to pay*, which may overstate the seller's ability to extract surplus in scenarios with constraints on market segmentation. Our framework merges these

[1] See the hearing's follow-up letter to Amazon for details.
[2] See the order for details.

two approaches and allows us to study all market segmentations that are subject to residual uncertainty.

Within our framework, we ask two questions. First, does refining the seller's segmentation monotonically increase or decrease welfare (a convex combination of consumer and producer surplus), or is the effect non-monotone? In the first two cases, information is "per se" good or bad: Simply knowing the fact that the seller is collecting information is enough to specify whether its effect is positive or negative, without the need for knowing *what* the information is. This distinction is useful because it is often difficult to perfectly monitor and control how firms use consumer data, so it is desirable to know when doing so is necessary. When the effect is non-monotone, information has the *potential* to be beneficial or harmful, and further analysis might be needed. Our second question delves deeper into the non-monotone case: How good or bad can information be, and what are the best and the worst ways to provide additional information?

First result. We first characterize when welfare is monotonically increasing, monotonically decreasing, or non-monotone in information, using three conditions. The first condition, which we call partial inclusion, says the demand curves are not too far apart in the sense that the optimal monopoly price of each of them is in the interior of any other's domain of prices. The second condition is a spanning one that says the set of all demand curves is decomposable into at most two basis demand curves. If either of the first two conditions do not hold, welfare is not monotone in information. If they both hold, the third condition gives an expression on the two basis demand curves that specifies whether welfare is monotonically increasing, monotonically decreasing, or non-monotone.

Examining this expression reveals a novel insight, namely that information affects welfare in three ways. To understand these three effects, consider an example. Suppose there are two types, type L and type H, representing low and high income consumers. Suppose half of the consumers have type L, the other half have type H, and if the seller could perfectly price discriminate based on the type, she would choose a higher price for type H consumers. Without any information, the seller offers some uniform price p to all consumers.

Now suppose the seller can use geography as a proxy for income to price discriminate. There are two locations, each representing a segment, where the first location contains $\frac{2}{3}$ of type L consumers and $\frac{1}{3}$ of type H consumers, and the second location contains the remaining $\frac{1}{3}$ of type L consumers and $\frac{2}{3}$ of type H consumers. The seller lowers the price to some $p_1 < p$ for the first location, and increases the price to some $p_2 > p$ for the second location. Table 1 shows, for each type and each price, the fraction of consumers of that type that face that price, before and after information is provided.

The three effects of information are reflected in the table. The first effect is the *within-type price change* effect: For each type, information disperses prices, with some consumers facing a higher price and some facing a lower price. The second effect is the *cross-types price change*

		p_1	p	p_2
Before	Type L	0	1	0
	Type H	0	1	0
After	Type L	$\frac{2}{3}$	0	$\frac{1}{3}$
	Type H	$\frac{1}{3}$	0	$\frac{2}{3}$

Table 1: The fraction of consumers of each type facing each price before and after information is provided.

effect: The price drop applies asymmetrically, to more type L consumers than type H ones. The third effect is the *price curvature* effect: The size of the price drop might not be equal to the size of the price increase. The larger the price drop compared to the price increase, the larger the positive effect of providing information.

Our three effects of information are related to those identified in the classical literature that compares the total surplus of full information to that of no information. This literature points out that a sufficient condition for price discrimination to decrease total surplus is that the "output effect" is negative, that is, the quantity produced by the seller decreases (Robinson, 1969; Schmalensee, 1981; Varian, 1985). This is because the "misallocation effect", that is, the fact that the quantity is sold at non-uniform prices, further distorts the allocation and decreases total surplus. Our price-curvature effect corresponds to the output effect because they are both about how price discrimination affects the aggregate decisions of the firm (average price versus total quantity). Similarly, our cross-types price change effect corresponds to the misallocation effect because they are both about how price discrimination treats types non-uniformly. A distinction, however, is that our effects are expressed in terms of prices (and not quantities), which is crucial for obtaining a unified characterization for arbitrary welfare functions (and not just total surplus). A second distinction is our within-type price change effect, which has no counter-part and arises because we study partial information: Because types are not fully separated, even consumers of a given type are affected by price discrimination non-uniformly. Finally, our characterization specifies how to add up the three effects by giving a single expression that identifies the net effect, whereas the classical argument only specifies sufficient conditions for total surplus to decrease.[3]

Another insight of our characterization is that price discrimination generally has the potential to be beneficial even without "opening new markets". To elaborate, suppose the optimal monopoly price of type H is higher than the highest price in the support of type L so our partial inclusion condition is violated. If almost all consumers are of type H, the seller

[3] Aguirre et al. (2010) uses a different logic to also obtain sufficient conditions under which price discrimination increases welfare.

will optimally exclude type L consumers. Then partially separating the two types opens a new market and prompts the seller to lower the price to one segment without increasing it in the other. This is indeed an important force in Bergemann et al. (2015). They demonstrate the power of this force by showing that as long as there is any exclusion, some information can achieve full efficiency while giving all the gains to consumers. The classical literature, on the other hand, imposes a non-exclusion assumption to ask if price discrimination can be beneficial even if this force is shut down. Our result shows that it generally can. When there is no exclusion, so no new markets are opened, as long as one of the two other conditions of the result are violated, information does not affect welfare monotonically so there exists some information that increases welfare.

Second result. Our second result studies the non-monotone case further. It provides bounds on welfare gains and losses from information and identifies the best and the worst directions for providing additional information. A key role is played by the value function that specifies the welfare of each posterior belief (once the seller prices optimally for that belief), and the Hessian matrix of this value function. The largest possible gain of information is proportional to the largest eigenvalue of the Hessian matrix, and the best possible direction to provide additional information is in the direction of the corresponding eigenvector. The largest possible loss and the worst directions are similarly related to the smallest eigenvalue and its eigenvector.

Notably, although eigenvalues and eigenvectors of arbitrary matrices are often complex objects that cannot be expressed in closed form, we give closed forms for them in our setting. This is because the Hessian itself in our setting is a low-rank matrix with a separable form. In fact, the Hessian has at most two non-zero eigenvalues. Examining the Hessian and its eigenvalues reveals and generalizes the same three effects of information from our two-type analysis.

The result has a simple intuition. An observation based on Kamenica and Gentzkow (2011) is that information monotonically increases welfare if and only if the value function mentioned above is convex. So it is natural to expect that if the value function is almost but not exactly convex, then the potential harms of information are small. Eigenvalues of the Hessian of the value function measure its curvature along their corresponding eigenvectors. The largest eigenvector represents the direction along which the function is "most convex", and providing information along that direction provides the largest value proportional to the curvature of the value function which is measured by the eigenvalue. A similar argument applies to the lowest eigenvalue and its connection to the largest possible harm of information.[4]

The bounds in our second result provide a *quantitative* analogue to our first, qualitative result, which characterizes when these bounds are both positive, both negative, or have op-

[4]This intuition more generally applies to Bayesian persuasion problems in which the receiver's action is specified by a first-order condition (for example, Kolotilin et al., 2024).

posite signs. These bounds could therefore inform a quantitative framework to regulate the use of consumer data in price discrimination. Analogous to the guidelines commonly used in antitrust and merger cases, such a framework could specify thresholds for the potential gains and losses from data usage, and use these bounds to determine whether such practices should be encouraged or discouraged regardless of their form, or whether additional scrutiny, monitoring, or restrictions are warranted. In this latter case, our result on the best and worst ways to provide additional information helps guide the design of appropriate restrictions on data collection and usage.

The rest of the paper is organized as follows. Section 1.1 reviews the literature. Section 2 describes our model. Section 3 defines our welfare-monotonicity properties and gives our first main result, Theorem 1, which provides the necessary and sufficient conditions for information to be monotonically good or bad for welfare, or for the effect to be non-monotone. Section 4 states our second main result, Theorem 2, which provides a tight bound on the impact of information on welfare when the impact is globally non-monotone and identifies the marginally optimal segmentation. Section 5 concludes.

1.1 Related Literature

Our work is related to the modern literature on price discrimination that studies all segmentations, pioneered by Bergemann et al. (2015). Our analysis does not directly apply to the setting of Bergemann et al. (2015) because we make smoothness assumptions to replace the seller's profit-maximization problem with its first-order condition. Nonetheless, we show that our results are consistent with Bergemann et al. (2015) as we approach their setting within ours. Bergemann et al. (2015) and Kartik and Zhong (2023) identify all possible pairs of consumer and producer surplus. Yang (2022) studies how a profit-maximizing intermediary sells segmentations to a producer for price discrimination. Hidir and Vellodi (2021), Haghpanah and Siegel (2022), Haghpanah and Siegel (2023), Asseyer (2024), and Bergemann et al. (2024) study market segmentation with a multi-product seller. In a general framework, Bonatti et al. (2025) study when consumer surplus is convex or concave in a market's belief, which allows them to characterize when consumers are better or worse off if the market receives additional information about their types. Our conceptual contribution to this literature is to introduce a framework in which segmentations are constrained by some residual uncertainty.[5]

Our work also relates to the large classical literature on monopolistic third-degree price discrimination. As discussed earlier, the classical approach to this problem compares uniform pricing to full segmentation of a given set of demand curves. These papers typically either

[5]A working paper version Bergemann et al. (2013) of Bergemann et al. (2015) studies consumer- and producer-surplus pairs in some examples with partial information about types. Strack and Yang (2025) also study a setting in which there is a separation between a consumer's type (which they call characteristic) and willingness to pay. They study segmentations in which different types are offered the same distribution of prices. Because we ask whether refining segmentations monotonically affects welfare, our work is also related to those that study information orders in persuasion (Curello and Sinander, 2022, Brooks et al., 2024).

focus on studying total surplus (Varian, 1985; Aguirre et al., 2010; Cowan, 2016) or consumer surplus (Cowan, 2012; Aguirre and Cowan, 2015). Our general approach allows us to characterize welfare-monotonicity for any weighted combination of consumer and producer surplus, including as special cases consumer and total surplus. Additionally, our framework allows us to study intermediate forms of price discrimination where the types are partially separated. Most importantly, whereas the papers in the literature give only sufficient conditions for welfare to rise or fall, we give a complete characterization of our welfare-monotonicity conditions. As we show with examples, our conditions do not imply nor are implied by those in the literature. On one hand, our conditions are stronger because we seek to characterize a stronger property, but on the other hand, the conditions in the literature are stronger because they are only sufficient for a ranking of the welfare of full segmentation versus uniform pricing. Nonetheless, the conditions can be qualitatively compared, as we outline in the introduction.

Methodologically, our work builds on the recent literature that uses duality in Bayesian persuasion (Kolotilin, 2018; Dworczak and Martini, 2019; Dworczak and Kolotilin, 2024; Kolotilin et al., 2024).[6] In particular, we convert our welfare-monotonicity properties to a class of Bayesian persuasion problems that seek to identify when no-information maximizes or minimizes welfare *for all* prior distributions over the given set of demand curves. Because the seller's profit-maximization problem in our setting is identified by its first-order condition, we take the strong-duality results of Kolotilin (2018) and Kolotilin et al. (2024) off-the-shelf to solve these Bayesian persuasion problem. We then show that for no-information to be a solution to the entire class of Bayesian-persuasion problems, our spanning condition must hold. The main technical result in Kolotilin et al. (2024) gives a sufficient optimality condition, the "twist" condition, that is reminiscent of, but different from, our spanning condition. Our characterization identifies an additional condition on two demands capturing the three effects of information, which has no counter-part in Kolotilin et al. (2024), that together give necessary *and sufficient* conditions for our welfare-monotonicity properties.

2 Model

A seller produces a product at a constant marginal cost that is normalized to zero. There is a unit mass of unit-demand consumers, where each consumer has a type and a willingness to pay for the product. The set of types Θ is a compact subset of a linear normed vector space (which holds trivially if Θ is finite). The types are distributed according to a full-support prior distribution $\mu_0 \in \Delta\Theta$. Each type $\theta \in \Theta$ represents a group of consumers with heterogeneous willingness to pay. The demand curve $D(p, \theta)$ specifies the fraction of type θ consumers whose willingness to pay for the product is at least p.[7] We denote the family of demand curves by

[6] See also Immorlica et al. (2022), Smolin and Yamashita (2022), Saeedi and Shourideh (2024), and Saeedi et al. (2024).

[7] Our model has other interpretations. One is that the seller faces a single consumer with a random type drawn from μ. The consumer has multi-unit demands with non-linear utility for quantity and her utility-

$\mathcal{D} = \{D(p,\theta)\}_{\theta\in\Theta}$. The primitive of our model is a pair $(\mathcal{D}, \mu_0)$, that is, the family of demand curves and the prior distribution over them.

We make the following assumption on the demand curves throughout the paper.

Assumption 1. *For all types $\theta \in \Theta$, the demand curve $D(\cdot,\theta) : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies the following properties:*

1. *There exists an interval $I(\theta) = \left[\underline{p}(\theta), \overline{p}(\theta)\right]$, $0 \leq \underline{p}(\theta) < \overline{p}(\theta) \leq \infty$, such that $D(p,\theta)$ is differentiable and strictly decreasing in p, $D_p(p,\theta) < 0$, for values $p \in I(\theta)$. These intervals need not be identical across types.*

2. *For values of $p < \underline{p}(\theta)$, $D(p,\theta) = D(\underline{p}(\theta),\theta)$, and for values of $p > \overline{p}(\theta)$, $D(p,\theta) = 0$.*

3. *For values of $p \in I(\theta)$, the revenue function $R(p,\theta) = pD(p,\theta)$ is strictly concave.*[8]

4. *There exists $p(\theta) \in I(\theta)$ such that $R_p(p(\theta),\theta) = 0$.*

The above assumption implies that $p(\theta)$ is the optimal (revenue-maximizing) price set by the seller when she faces only type θ consumers.

Demand curves that are step functions, each representing a population of unit-demand consumers who all have the same willingness to pay, corresponds to the model in Bergemann et al. (2015) but violates Assumption 1 (because such a demand curve is not strictly decreasing over a non-empty interval). Later we show how our analysis confirms their results by studying demand curves that pointwise approach step functions while maintaining Assumption 1.

A market $\mu \in \Delta\Theta$ is a probability distribution over types. A segmentation $\sigma \in \Delta\Delta\Theta$ is a Bayes-plausible distribution over markets, that is, $\mathbb{E}_\sigma[\mu] = \mu_0$, and $S(\mu_0)$ denotes the set of all segmentations. The interpretation is that the seller has access to some information structure that reveals a signal about the type of the buyer. The seller then forms a posterior μ after each signal realization and chooses a profit-maximizing price for that posterior. The seller uses a pricing rule $p : \Delta\Theta \to \mathbb{R}_+$ that specifies an optimal price for every possible posterior, breaking ties if necessary,

$$p(\mu) \in \arg\max_{p\in\mathbb{R}_+} \int_\Theta R(p,\theta)\,\mathrm{d}\mu(\theta), \forall \mu \in \Delta\Theta.$$

maximization problem for a given price induces a demand curve for each type. One can also combine the two interpretations and think about each demand as representing a population of consumers with multi-unit demands.

[8] This assumption allows us to replace the seller's profit-maximization problem by its first-order condition. Kolotilin et al. (2024) show that the first-order approach remains valid under a slight relaxation of strict concavity, roughly assuming concavity up to a normalization, so our results go through with this weaker assumption. We state our assumption with strict concavity because it is more straightforward and is commonly used in the literature.

We study the welfare that each segmentation induces. For this, let

$$CS(p,\theta) = \int_{p}^{\overline{p}(\theta)} D(z,\theta)\,\mathrm{d}z$$

denote the surplus of type θ consumers from facing price p. The α-surplus of price p for type θ consumers is a weighted average of consumer surplus and producer surplus with weights $\alpha \in (0,1]$ and $1-\alpha$,

$$V^{\alpha}(p,\theta) := \alpha \cdot CS(p,\theta) + (1-\alpha)\cdot R(p,\theta).$$

The special case of $\alpha = 1$ corresponds to consumer surplus and the special case of $\alpha = \frac{1}{2}$ corresponds to total surplus.[9] The α-surplus of a segmentation is the expectation of the α-surplus of all price and type pairs in the segmentation, given by

$$V^{\alpha}(\sigma) := \int_{\Delta\Theta}\int_{\Theta} V^{\alpha}(p(\mu),\theta)\,\mathrm{d}\mu(\theta)\,\mathrm{d}\sigma(\mu).$$

3 Welfare-Monotonicity Characterization and Implications

Our first result characterizes when allowing the monopolist to refine segmentations monotonically affects welfare, and when the effect is non-monotone. We first define these monotonicity properties and then give the result, its proof sketch, interpretation, and examples.

Consider two segmentations σ and σ'. We say σ is a mean-preserving spread of σ', and write

$$\sigma \underset{\text{MPS}}{\succeq} \sigma',$$

if there exists a joint distribution over pairs of markets $\nu \in \Delta(\Delta\Theta \times \Delta\Theta)$ that induces marginals σ and σ', i.e., $\nu(\cdot, \Delta\Theta) = \sigma(\cdot)$ and $\nu(\Delta\Theta, \cdot) = \sigma'(\cdot)$, and random markets (μ, μ') drawn from ν satisfy $\mathbb{E}[\mu|\mu'] = \mu'$ almost surely. We call σ a "refinement" of σ' because it can be obtained by splitting each market in segmentation σ' into possibly multiple markets in a mean-preserving way. The *full-information segmentation*, in which each market μ in the support of the segmentation has only a single type in its support, is finer than any segmentation, and any segmentation is finer than the *no-information segmentation* that assigns probability 1 to the prior market μ_0.

Given this definition of refinement, we define welfare-monotonicity:

Definition 1 (Welfare-monotonicity in information). *Consider a given* $(\mathcal{D}, \mu_0)$.

[9] We assume $\alpha > 0$ because our properties of interest are trivial if the entire weight is on producer surplus (we explain why after defining the properties).

1. *Information is monotonically bad for α-surplus, "α-IMB holds", if*

$$V^\alpha(\sigma) \leq V^\alpha(\sigma'), \qquad \forall \sigma, \sigma' \in S(\mu_0) \text{ such that } \sigma \underset{MPS}{\succeq} \sigma'.$$

2. *Information is monotonically good for α-surplus, "α-IMG holds", if*

$$V^\alpha(\sigma) \geq V^\alpha(\sigma'), \qquad \forall \sigma, \sigma' \in S(\mu_0) \text{ such that } \sigma \underset{MPS}{\succeq} \sigma'.$$

We refer to these properties jointly as "welfare-monotonicity" properties.

As stated in the introduction, our motivation for studying these properties is to delineate when the act of information usage by the seller is "per se" beneficial or harmful versus when the answer depends on the context (i.e., what information is collected). This classification is useful because it is often costly to monitor and control what data the firms collect and how they use it, so it is desirable to know when such monitoring and control is necessary. When welfare-monotonicity holds, depending on its direction, price discrimination should be encouraged or discouraged without the need for any additional context. When welfare-monotonicity does not hold, further analysis is needed, to which we return later.

Our first main result completely delineates the three possible cases.[10] The result starts by a simplifying step: It shows that the welfare-monotonicity properties are prior-free, allowing us to drop the prior distribution through the rest of the statement. Formally, prior-freeness means that for any two full-support distributions μ, μ', α-IMG (α-IMB) holds for $(\mathcal{D}, \mu)$ if and only if α-IMG (α-IMB) holds for $(\mathcal{D}, \mu')$.

The main content of the result are statements (i) and (ii). Statement (i) gives a reduction of the problem from any number of demand curves to *binary* families of demand curves. It gives conditions under which we only need to verify welfare-monotonicity for the binary family $\{D(\cdot, \theta)\}_{\theta \in \{\theta_L, \theta_H\}}$ that consists of the two demand curves with the lowest and the highest optimal monopoly price, $p(\theta_L) = \min_\theta p(\theta)$ and $p(\theta_H) = \max_\theta p(\theta)$ (ties can be broken arbitrarily). Statement (ii) provides the characterization for binary families.

A condition used in the characterization is the *partial inclusion* condition, which means $\underline{p}(\theta') \leq p(\theta) \leq \overline{p}(\theta')$ for all $\theta, \theta' \in \Theta$. We call this the partial inclusion condition because it says that if type θ' consumers are offered the optimal price for type θ, some but not all of them will be served.

We now state the result and then go over its two main statements in more detail.

Theorem 1 (Welfare-monotonicity). *The welfare-monotonicity properties are prior-free, and we can therefore refer to them as properties of $\mathcal{D}$. Let $\theta_L, \theta_H \in \Theta$ be the two types with the lowest and the highest optimal monopoly prices in $\mathcal{D}$. The following two statements hold:*

[10] Blackwell (1953) implies that information is monotonically good for producer surplus, which is why we focus on $\alpha > 0$ throughout the paper.

(i) α*-IMB (*α*-IMG) holds for* $\mathcal{D}$ *if and only if*

(A) there is partial inclusion, and

(B) there exist two functions $f_1, f_2 : \Theta \to \mathbb{R}_+$ *such that*

$$D(p,\theta) = f_1(\theta)D(p,\theta_L) + f_2(\theta)D(p,\theta_H)$$

for all θ *and* $p \in [p(\theta_L), p(\theta_H)]$*, and*

(C) α*-IMB (*α*-IMG) holds for the binary family* $\{D(\cdot,\theta)\}_{\theta\in\{\theta_L,\theta_H\}}$.

(ii) α*-IMB (*α*-IMG) holds for the binary family* $\{D(\cdot,\theta)\}_{\theta\in\{\theta_L,\theta_H\}}$ *if and only if there is partial inclusion and*

$$V^\alpha(p,\theta_H) - V^\alpha(p,\theta_L) + \frac{V^\alpha_p(p,\theta_L) - \frac{R_p(p,\theta_L)}{R_p(p,\theta_H)}V^\alpha_p(p,\theta_H)}{R_{pp}(p,\theta_L) - \frac{R_p(p,\theta_L)}{R_p(p,\theta_H)}R_{pp}(p,\theta_H)}(R_p(p,\theta_L) - R_p(p,\theta_H)) \qquad (1)$$

is decreasing (increasing) on $(p(\theta_L), p(\theta_H))$.

Let us now explain in more detail what the theorem says. Consider statement (i). Condition (A) of the statement, partial inclusion, says that the monopoly price $p(\theta)$ for any type θ cannot be less than the lowest price $\underline{p}(\theta')$ in the support of another type θ' or higher than the largest price $\overline{p}(\theta')$ of type θ'. Roughly speaking, this partial-inclusion condition means that the demand curves cannot be too far from each other.[11]

Condition (B) of statement (i) is a spanning condition that says that all demand curves in the family must be decomposable into a linear combination of at most two "basis" demand curves $D(\cdot,\theta_L), D(\cdot,\theta_H)$ with possibly varying weights, where these basis demand curves are those with the lowest and the highest optimal monopoly price. The only possible source of heterogeneity among the demand curves is the pair of weights $f_1(\theta)$ and $f_2(\theta)$, so heterogeneity must be reducible to a two-dimensional sufficient statistic. Notice that this condition must also hold regardless of what the weight α is or which one of the welfare-monotonicity properties we are characterizing.

Condition (C) of statement (i) puts additional constraints on the basis demand curves. The binary family of demand curves that consists only of the two basis demand curves $D(\cdot,\theta_L), D(\cdot,\theta_H)$ must itself satisfy the corresponding welfare-monotonicity property. This condition therefore reduces the problem of characterizing the welfare-monotonicity properties for an arbitrary class of distributions $\mathcal{D}$ to characterizing them with binary distributions.

Statement (ii) of the theorem characterizes the welfare-monotonicity properties for a binary family of demand curves in terms of partial inclusion and monotonicity of an expression

[11] This partial inclusion condition is related to a non-exclusion assumption that is typically made in the classical price discrimination literature. That assumption only entails one part of our partial inclusion condition, namely that $p(\theta) \leq \overline{p}(\theta')$ for all $\theta, \theta' \in \Theta$. We review the connection between our finding and those in that literature in Section 3.3.

that depends on the weight α. The direction of the monotonicity condition depends on which one of the welfare-monotonicity properties we are characterizing. We re-visit and interpret this expression shortly.

One intuitive implication of the result, that we use throughout the paper, is that information gets more beneficial as α decreases, that is, more weight is put on the seller's payoff.[12]

Corollary 1. *If α-IMG (α-IMB) holds, then α'-IMG (α'-IMB) holds for any $\alpha' \leq \alpha$ ($\alpha' \geq \alpha$).*

In the rest of this section, we visit each of the two main statements of Theorem 1 and explain the logic behind them, starting by the case of two demand curves because its analysis reveals a novel insight on the effects of price discrimination. We then relate our findings to those in the classical and modern literature on price discrimination.

3.1 Two Demand Curves

Let us start with two demand curves and establish and interpret statement (ii) of Theorem 1. In order to focus on how welfare-monotonicity is related to the monotonicity of Equation (1), which allows us to provide intuition for the effects of price discrimination, we assume that there is partial inclusion. In Section 3.2 we illustrate why partial inclusion is in fact necessary for both of the welfare-monotonicity properties.

3.1.1 Proof Sketch

It is useful to begin by describing how the result in the binary case is proved. The proof is based on the concavification approach of Kamenica and Gentzkow (2011). To state the result, let us denote by μ a market in which the probability of type θ_H, the type with the higher monopoly price, is μ. Consider the weighted-surplus function $W^\alpha : [0,1] \to R$ that specifies the expected α-surplus of a market μ,

$$W^\alpha(\mu) = \mathbb{E}_{\theta\sim\mu}[V^\alpha(p(\mu),\theta)], \tag{2}$$

where $p(\mu) \in [p(\theta_L), p(\theta_H)]$ is the profit-maximizing price for this market and is uniquely identified by the first-order condition

$$\mathbb{E}[R_p(p(\mu),\theta)] = 0. \tag{3}$$

Information is monotonically bad for α-surplus *if and only if* W^α is concave. Indeed, if W^α is concave, then splitting any market μ in the support of a segmentation into multiple markets with the same mean would only (weakly) decrease weighted surplus. And if W^α is not concave, that is, if $W^\alpha(\mu)$ is below the concavification of W^α for some market μ, then we

[12] This result can also be directly shown using Blackwell (1953)'s characterization without using ours. We call it a corollary as a "sanity check" of our characterization and to document it for later use.

can take any segmentation with market μ in its support and split μ into two markets in a way that increases weighted surplus. A similar argument shows that information is monotonically good for α-surplus if and only if W^α is convex.

To see how the concavity (convexity) of W^α relates to the monotonicity condition in Equation (1), consider the derivative of W^α with respect to μ,

$$W^\alpha_\mu(\mu) = -V^\alpha(p(\mu), \theta_L) + V^\alpha(p(\mu), \theta_H) + p_\mu(\mu)\mathbb{E}[V^\alpha_p(p(\mu), \theta)].$$

The weighted surplus function W^α is concave if and only if its derivative is decreasing. The expression above contains $p(\mu)$ and $p_\mu(\mu)$, both of which are defined implicitly given the seller's profit-maximization condition in Equation (3). In the proof of the result we eliminate this implicit dependence by observing that because $p(\mu)$ is increasing in μ, W^α_μ is a monotone function of μ if and only if $W^\alpha_\mu(\mu(p))$ is a monotone function of p, where $\mu(p)$ is the inverse of the price function, that is, $\mu(p(\mu)) = \mu$. Indeed, the expression in Equation (1) is exactly the above expression evaluated at $\mu(p)$.

3.1.2 Interpretation: The Three Effects of Information

To interpret the expression in Equation (1), let us keep the optimal price function $p(\mu)$ (and its derivatives) implicitly defined and take a second derivative of W^α. This second derivative $W^\alpha_{\mu\mu}(\mu)$ represents the value of providing a small amount of information starting from a market μ. Examining this second derivative reveals that providing information has three effects on weighted surplus that are combined into one formula. The second derivative of W^α is

$$\begin{aligned} W^\alpha_{\mu\mu}(\mu) = {} & (p_\mu(\mu))^2\, \mathbb{E}\Big[V^\alpha_{pp}(p(\mu), \theta)\Big] \\ & +2p_\mu(\mu) \quad \Big[V^\alpha_p(p(\mu), \theta_H) - V^\alpha_p(p(\mu), \theta_L)\Big] \\ & +p_{\mu\mu}(\mu)\, \mathbb{E}\Big[V^\alpha_p(p(\mu), \theta)\Big]. \end{aligned} \tag{4}$$

To understand the three effects, consider what happens if we take a market μ with optimal price p and split it into two markets $\mu_1 = \mu - \delta$ and $\mu_2 = \mu + \delta$ with optimal prices $p_1 < p < p_2$, each with probability one-half. Table 2 shows, for each type and each price, the fraction of consumers of that type that face that price, before and after information is provided.

The three effects of information are reflected in Table 2, each represented in one of the terms in Equation (4).

1. The *within-type price change* effect. Price discrimination disperses prices within each type, where some consumers face a lower price and some others a higher price. The sign of this effect depends on how V^α is affected by a price dispersion, which depends on the curvature of V^α. If the function is convex, then this effect is positive, and if it is concave, the effect is negative. This effect corresponds to the first term in Equation (4).

		p_1	p	p_2
before	θ_L	0	1	0
	θ_H	0	1	0
after	θ_L	$\frac{1}{2}(1+\frac{\delta}{1-\mu})$	0	$\frac{1}{2}(1-\frac{\delta}{1-\mu})$
	θ_H	$\frac{1}{2}(1-\frac{\delta}{\mu})$	0	$\frac{1}{2}(1+\frac{\delta}{\mu})$

Table 2: The fraction of consumers of each type facing each price before and after information is provided.

2. The *cross-types price change* effect. Relatively more type θ_L consumers face a price drop than do type θ_H consumers. In fact, more than half of type θ_L consumers but less than half of type θ_H consumers face a price drop. In the extreme case of full information, all type θ_L consumers face a price drop and all type θ_H consumers face a price increase. This effect is positive if the marginal benefit of decreasing the price for type θ_L is larger than increasing it for type θ_H, and is negative otherwise. This effect corresponds to the second term in Equation (4).

3. The *price curvature* effect. The size of price drop $p - p_1$ might not be equal to the size of the price increase $p_2 - p$. The comparison depends on the curvature of the price function $p(\mu)$. This effect corresponds to the third term in Equation (4). Notice that the seller's profit-maximization condition means that the expected marginal revenue is zero and therefore

$$\mathbb{E}\Big[V_p^{\alpha}(p(\mu),\theta)\Big] = \alpha\mathbb{E}\Big[CS_p(p(\mu),\theta)\Big] \leq 0,$$

and so the sign of the third effect only depends on the curvature of the price function. This effect is positive if $p(\mu)$ is concave, in which case the price drop is larger than the price increase, benefiting consumers overall. Similarly, this effect is negative if the price function is convex, so the price increase is larger than the price drop.

3.1.3 Sufficient Conditions

The overall effect of information depends on the *aggregation* of the three effects identified above, and each of the three effects might be positive or negative. However, if all three terms have the same sign, then their summation has that sign too, and this observation allows us to obtain sufficient conditions for welfare-monotonicity. The following corollary formalizes this discussion by directly considering each term in Equation (4).

Corollary 2. *α-IMB (respectively α-IMG) holds if there is partial inclusion and the following hold over the range of prices $(p(\theta_L), p(\theta_H))$.*

1. *The within-type price change effect is negative (positive): $V^\alpha(p,\theta)$ is concave (convex) for each type $\theta \in \{\theta_L, \theta_H\}$.*

2. *The cross-types price change effect is negative (positive): $V^\alpha_p(p,\theta_H) \leq (\geq) V^\alpha_p(p,\theta_L)$.*

3. *The price curvature effect is negative (positive): $p(\mu)$ is convex (concave). A sufficient condition for this is $R_{ppp}(p,\theta) \geq (\leq)\ 0$ and $R_{pp}(p,\theta_H) \geq (\leq) R_{pp}(p,\theta_L)$.*

To explain these conditions, let us refer to the demand curve with a lower monopoly price as the "more elastic" demand curve. Indeed, over the interval of prices $[p(\theta_L), p(\theta_H)]$, the elasticity of the demand curve θ_L is more than 1 (because its marginal revenue is negative), but the elasticity of the demand curve θ_H is less than 1 (because its marginal revenue is positive).

Corollary 2 implies that information is monotonically good for consumer surplus (and all α) if

1. $R_{ppp}(p,\theta) \leq 0$, that is, the marginal revenue curves are concave,

and the more elastic demand curve, θ_L, has

2. a higher level, $D(p,\theta_L) \geq D(p,\theta_H)$, and,

3. a less concave revenue curve, $R_{pp}(p,\theta_H) \leq R_{pp}(p,\theta_L)$, or equivalently a less steep marginal revenue curve.[13]

These conditions are shown in Figure 1. A parallel set of conditions, which we give in Appendix B.2, identify monotonically bad information.

To conclude our discussion of two demand curves, we apply Corollary 2 to an example. In this example, there is a range of parameters for which all of the effects are positive, implying information is monotonically good, and another range where they are all negative negative, implying information is monotonically bad.

Example 1 (Sufficient Condition for α-IMG and α-IMB). *Consider two demand curves $D(p,\theta_i) = a_i - p + \frac{c_i}{p}$ for $i \in \{1,2\}$ and $a_i, c_i \geq 0$ with supports $[\delta, a_i]$ for small enough $\delta > 0$ (to ensure demands are bounded). Without loss of generality, assume $a_1 \leq a_2$. Then α-IMG holds for all α if*

$$c_1 - c_2 \geq (a_2 - a_1)\frac{a_2}{2}.$$

[13] Consider $\alpha = 1$. The within-group price change effect is positive because consumer surplus is convex in price. The cross-groups price change effect is positive because $D(p,\theta_L) \geq D(p,\theta_H)$. The remaining conditions ensure the price curvature effect is positive by Corollary 2.

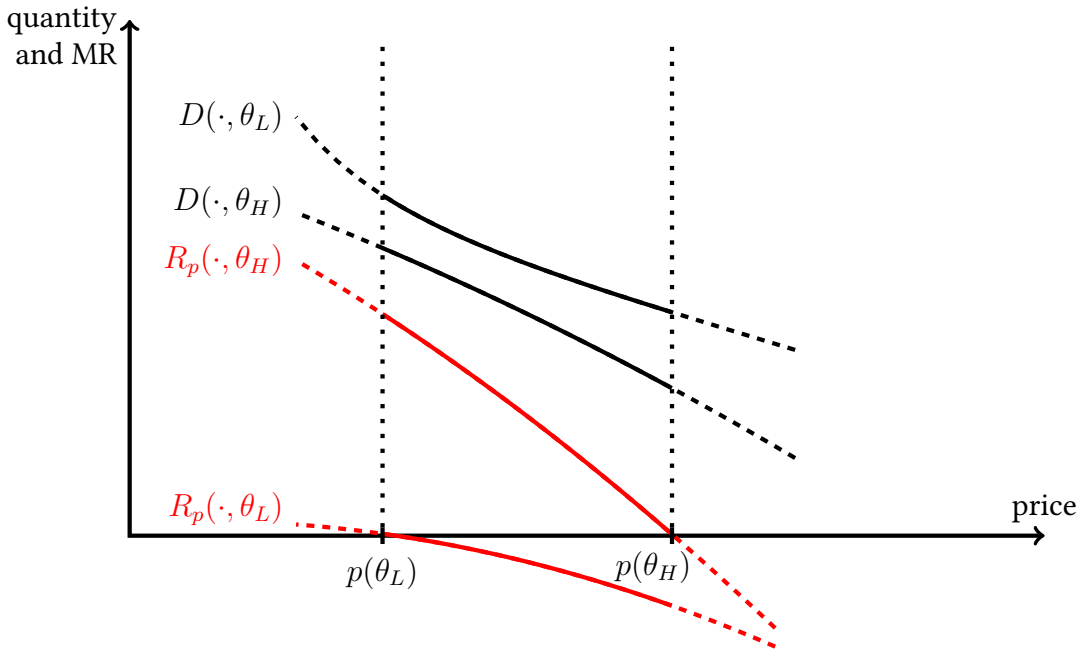


Figure 1: Information is monotonically good when both marginal revenue curves are concave and the more elastic demand θ_L has a higher level and a less steep marginal revenue curve.

α-IMB holds for all $\alpha \geq \frac{1}{2}$ if

$$c_1 \leq c_2 \leq \frac{a_1^2}{4}.$$

3.2 Reduction from Many Demand Curves to Two

We now sketch the proof of the rest of Theorem 1 that shows how to reduce a general family of demand curves to two. The reduction has three main steps. We first show how to transform the problem to an optimization problem and establish prior-freeness. We then show why partial inclusion is necessary for welfare-montonicity and how it enables us to use the first-order approach. Finally, we use strong duality to solve the optimization problem.

3.2.1 Step 1: Transformation to an Optimization Problem

If α-IMB holds for $(\mathcal{D}, \mu_0)$, then no-information segmentation maximizes V^α over all segmentations in $S(\mu_0)$, but not vice versa. We observe, however, that α-IMB for $(\mathcal{D}, \mu_0)$ is *equivalent* to the property that, for *every* μ, the no-information segmentation is a maximizer of V^α over all segmentations in $S(\mu)$. This equivalence follows from the standard observation that both of these two properties are equivalent to the concavity of the value function $W^\alpha : \Delta\Theta \to R$,

$$W^\alpha(\mu) = \int_\Theta V^\alpha(p(\mu), \theta)\, \mathrm{d}\mu(\theta), \tag{5}$$

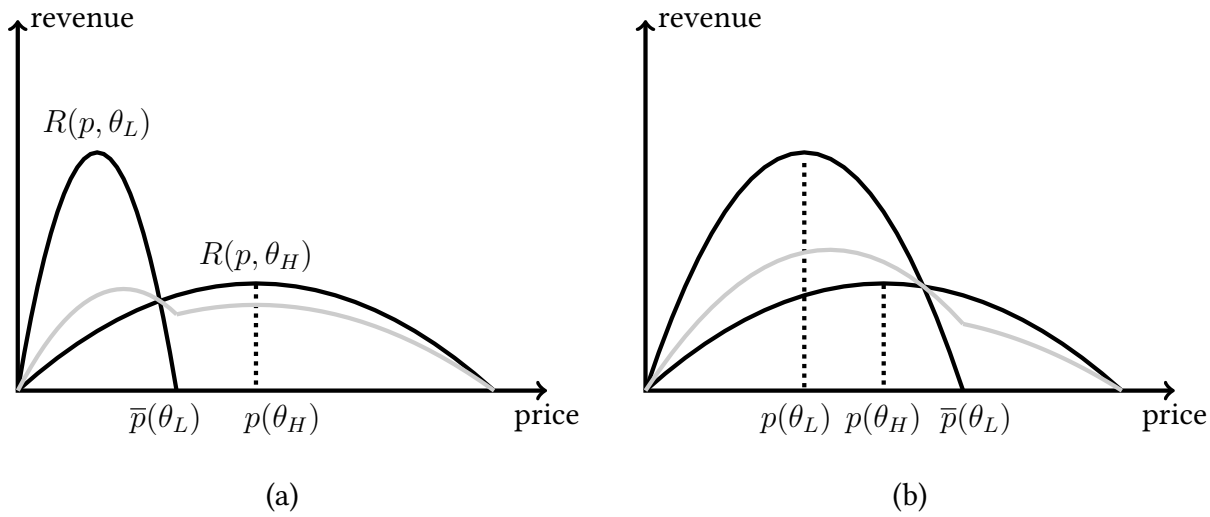


Figure 2: (a) The gray curve, which represents the revenue curve of market μ, is not concave. (b) The optimal price for market μ is the price that satisfies the first-order condition over $[p(\theta_L), p(\theta_H)]$.

that specifies the expected α-surplus of each market μ. Conversely, α-IMG is equivalent to the no-information segmentation being a *minimizer* of V^α, or a maximizer of $-V^\alpha$, *for all* priors. This observation transform our problem to a class of optimization problems and allows us to give a unified proof for the two welfare-monotonicity properties.

3.2.2 Step 2: Partial Inclusion and the First-Order Condition

The optimization problem derived above nests another one in which the seller chooses an optimal price for each market. We show that the partial-inclusion assumption allows us to replace the seller's optimal pricing problem with a first-order condition.

To understand the connection between partial inclusion and the first-order condition, consider some market μ, and suppose for simplicity that it has binary support over two demands θ_L, θ_H. Suppose there is full exclusion, i.e., $\overline{p}(\theta_L) < p(\theta_H)$, as shown in Panel (a) of Figure 2. Even though each type has a revenue curve that is concave over its support, the revenue curve associated with market μ need not be concave, as shown in the figure. As a result, the first-order condition does not pin down the optimal price. We therefore proceed with two arguments.

First, we show that if partial inclusion is violated, then welfare-monotonicity does not hold, without relying on the first-order approach. To do this, we consider the two cases of full exclusion, $\overline{p}(\theta_L) < p(\theta_H)$, and full inclusion, $p(\theta_L) < \underline{p}(\theta_H)$, separately and in each case construct an information structure that is good and one that is bad. One of these four constructions is standard: If there is full exclusion, as is the case in the figure, then information has the potential to be beneficial by "opening new markets". The other three cases require new arguments that we discuss in detail in the proof.

Second, we show that under partial inclusion, the first-order approach is indeed valid. The idea, as shown in Panel (b) of Figure 2, is that even though the revenue curve for a market is not necessarily concave, it is concave over the range $[\min_\theta p(\theta), \max_\theta p(\theta)]$ of lowest and highest optimal price of the types in the market, which are the only prices we need to consider to find the seller's profit-maximizing price.

3.2.3 Step 3: Applying Duality

We next apply the duality framework of Kolotilin (2018), Dworczak and Martini (2019), Dworczak and Kolotilin (2024), and Kolotilin et al. (2024). They key is to use those duality results to obtain optimality condition of the no-information segmentation for *every* prior.

Proposition 1. *Suppose there is partial inclusion. Welfare-monotonicity holds if and only if there exists a function ζ such that for every $p \in I$ and every type θ,*

$$p \in \arg\max_{p' \in I} U(p', \theta) - \zeta(p, p') R_p(p', \theta), \tag{6}$$

where $U = V^\alpha$ characterizes α-IMB and $U = -V^\alpha$ characterizes α-IMG.

Given the above result, the "sufficiency" direction of the proof is straightforward. Suppose welfare-monotonicity holds for the binary family $\{\theta_L, \theta_H\}$, so by Proposition 1, Equation (6) holds for θ_L and θ_H. Suppose also that the demand curves of θ_L, θ_H are bases that span some set $\mathcal{D}$ of demand curves, so every demand curve in the set is a linear combination of that of θ_L and θ_H. Then Equation (6) holds for all demand curves in $\mathcal{D}$, and welfare-monotonicity holds for the entire set.

For the "necessity" part of the proof, again, one part of the argument is straightforward. If welfare-monotonicity holds for $\mathcal{D}$, then Equation (6) is satisfied for every type, and in particular it must hold for the two types θ_L, θ_H that have the lowest and the highest optimal monopoly price in the family, so welfare-monotonicity must hold for the binary family that contains only the demand curves associated with θ_L, θ_H as well.

To argue that the spanning condition of the theorem must be satisfied, we argue first that *each* pair θ, θ' of types pin down the function ζ over the interval $(p(\theta), p(\theta'))$ of prices that are in between their two optimal monopoly prices. We then argue that if the spanning condition of the theorem is violated, the ζ functions that are pinned down with different pairs do not agree, and therefore a single function ζ does not exist.[14] Appendix A.2 provides detail of the proof.

[14]The observation that the spanning condition is necessary for welfare-monotonicity is consistent with the finding in Corollary 1 of Kolotilin et al. (2024) that no-information is suboptimal for generic utility functions that the receiver might have in their Bayesian persuasion setting. A difference is that the value functions of the receiver and the sender in their setting are unrelated whereas in our setting the value function U and the seller's objective R are both pinned down given the demand curves. We use this property to give a complete characterization of our welfare-monotonicity properties.

3.3 Discussion and Connections to Classical and Modern Literature

Having seen our characterization and the logic behind it, let us now compare our finding to those in the classical and modern literature on price discrimination.

3.3.1 The Classical Literature

Comparison of conditions. Our welfare-monotonicity conditions neither imply nor are implied by those in the classical literature that compare, in our language, only full information to no information. On the one hand, our conditions are stronger because we characterize stronger properties. On the other hand, the conditions in the literature are stronger because they provide only sufficient conditions. For instance, as far as we are aware, existing results do not apply to constant-elasticity demand curves for which we show welfare-monotonicity.[15]

Example 2 (Constant-elasticity demand: $\frac{1}{2}$-IMB). *Consider two demand curves $D(p, \theta_i) = (c+p)^{-\theta_i}$ for $i \in \{1, 2\}$ and constant $c > 0$. Assume $1 < \theta_2 < \theta_1 \leq 2\theta_2 - 1$ and truncate the demand curves at $\frac{2c}{\theta_2 - 1}$.[16] Then information is monotonically bad for total surplus if and only if $\theta_1 \leq \theta_2 + \frac{1}{2}$. Under this condition, information is monotonically bad for α-surplus for any $\alpha \geq \frac{1}{2}$.*

These demand curves have constant elasticities θ_1 and θ_2.[17] The example shows that providing information to the seller monotonically decreases total surplus exactly when the two demand elasticities are not too far apart. In other words, unless consumers have widely different elasticities with respect to price, price discrimination is bad for total (and consumer) surplus.

Our three effects versus the two classical ones. Although our conditions are different from those in the literature, they are in fact related in spirit. For this, recall the two effects from the classical literature discussed in the introduction: Price discrimination affects the total quantity produced, the "output" effect, and allows selling that total quantity at non-uniform prices, the "misallocation" effect. Because the misallocation effect is a distortion that reduces total surplus, price discrimination decreases total surplus whenever the output effect is also negative. This logic therefore provides a sufficient condition for total surplus to fall, but cannot

[15] In particular, this example violates the conditions of Aguirre et al. (2010) that give the most general conditions for price discrimination to decrease total surplus. Namely, their condition that the demand curve with a lower monopoly price has a lower curvature (is "more concave") than the demand curve with a higher monopoly price, is violated (it holds the opposite way) in this example.

[16] These conditions ensure that Assumption 1 holds. The revenue curve associated with demand $(c+p)^{-\theta}$ is maximized at $\frac{c}{\theta-1}$, is concave below $\frac{2c}{\theta-1}$, and is convex above that threshold. So by truncating the two demand curves at $\frac{2c}{\theta_2-1}$, both revenue curves are concave over the truncated interval. The extra assumption $\theta_1 \leq 2\theta_2 - 1$ guarantees that optimal prices are in the interior.

[17] To be precise, these demand curves are a *normalization* of constant-elasticity demands curves. Starting from demand curve $p^{-\theta_i}$, which has constant elasticity θ_i, and constant marginal cost $c > 0$, if we normalize costs to zero by shifting prices by c, we obtain the demand functions used in this example.

yield conditions for total surplus to rise because it offers no way to trade off the two effects when the output effect is positive (the misallocation effect is always negative).[18]

Our result addresses precisely this limitation. It identifies three distinct effects of information and combines them into a single expression, giving a complete characterization of welfare-monotonicity. Our effects are conceptually related to the classical ones but differ in important ways. Our price-curvature effect parallels the classical output effect, as both concern how price discrimination affects the seller's aggregate decisions (average price versus total quantity). Similarly, our cross-groups price change effect parallels the classical misallocation effect, as both concern how price discrimination affects types non-uniformly. A first difference between our effects and the classical ones, however, is that all of our effects are measured in prices, not quantities. This is important because we allow for general objectives, and not just total surplus.[19] A second difference is that our within-groups price change effect has no counter-part and arises because we study partial segmentations: Even consumers of one type are treated differently by price discrimination.

Flexible welfare measures. The flexibility of our characterization, which allows for arbitrary weights on consumer and producer surplus, helps paint a more comprehensive picture of the effects of price discrimination than fixing any particular objective would. To elaborate, consider the following corollary of our result.

Corollary 3. *Consider the family of demand curves*

$$\mathcal{D} = \Big\{a(\theta)D(p) + b(\theta)\Big\}_\theta$$

for some demand curve D. Then α-IMB (α-IMG) holds if and only if

$$(2\alpha - 1)p + \alpha(\frac{pD'(p)}{R''(p)}) \tag{7}$$

is increasing (decreasing) over $[\min_\theta p(\theta), \max_\theta p(\theta)]$ and there is partial inclusion.[20]

Corollary 3 implies that information is monotonically bad for total surplus if and only if $\frac{pD'(p)}{R''(p)}$ is increasing. This finding is consistent with those of Cowan (2007) and Aguirre et al. (2010) who show that if this expression is increasing, then total surplus is lower under full segmentation than under uniform pricing. Additionally, our result shows that when the expression is decreasing, then information is monotonically good for total surplus. In this

[18] Not all papers in the literature focus on this logic. Aguirre et al. (2010) give sufficient conditions for total surplus to increase or decrease with price discrimination using a different but related logic.

[19] Suppose we want to calculate consumer surplus in a market by taking the average of consumer surplus across all types in the market. The quantity sold to the market is not a sufficient statistic for doing so because one has to calculate the price in order to determine how that quantity is divided across different types in the market.

[20] We denote the first and second derivative of $R(p)$ by $R'(p)$ and $R''(p)$, and similarly $D'(p)$ is the derivative of $D(p)$.

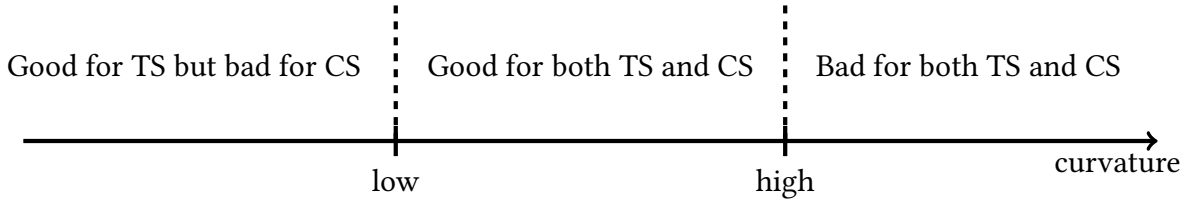


Figure 3: The three possible cases for how information affects total surplus (TS) and consumer surplus (CS) in a parametric example.

case, does total surplus increase because consumer surplus also increases, or because any loss in consumer surplus is offset by a larger gain in producer surplus? The corollary clarifies: Consumer surplus, and therefore total surplus, increases when Equation (7) is decreasing for $\alpha = 1$. Otherwise, consumer surplus may decrease, but any losses are offset by an increase in producer surplus when Equation (7) is decreasing for $\alpha = \frac{1}{2}$. In Appendix B.3.1, we give an example demonstrating each possibility based on the curvature of a parametric demand function, qualitatively shown in Figure 3.

3.3.2 Approaching Step Functions

Having seen the connection between our conditions and those in the classical price discrimination literature, let us compare our findings to those in Bergemann et al. (2015). Because the demand curves in our setting must be downward-sloping over a non-empty interval, they cannot be step functions, which are considered in Bergemann et al. (2015). One can nonetheless ask what happens when we *approach* step functions within our framework. Because we allow different types to have different supports of prices, it is in fact possible to approach each step function with a strictly decreasing demand curve that has a strictly concave revenue curve over a vanishingly small interval. But as these intervals get small, the partial inclusion condition will eventually be violated, in which case our result says that welfare-monotonicity must be violated, consistent with what Bergemann et al. (2015) find.[21] The next Corollary formalizes this observation.

Corollary 4. *Consider a family of demand curves $\mathcal{D}^\epsilon$ that uniformly converges to a family of unit-demand curves as ϵ goes to zero while satisfying our Assumption 1 for every $\epsilon > 0$. Then, for small enough ϵ, the partial-inclusion condition is violated and therefore information is neither monotonically good nor bad.*

[21] Even though the findings of Bergemann et al. (2015) are not stated in terms of welfare-monotonicity, they imply that both welfare-monotonicity properties are violated. In particular, with two types representing step functions where $p(\theta_L) < p(\theta_H)$, any segmentation in which some market has price $p(\theta_H)$ can be refined in a way that increases consumer surplus and therefore weighted surplus. And any segmentation in which some market has price $p(\theta_L)$ can be refined in a way that reduces consumer surplus while keeping producer surplus fixed, reducing weighted surplus for any weight.

3.3.3 Beneficial Information without Opening New Markets

Another insight from our result, that relates both to the classical and the modern literature, is that beneficial information generally exists even without opening new markets.

A folklore wisdom in the classical price discrimination literature is that information has the potential to be beneficial if it opens new markets, which happens if there is exclusion, $\overline{p}(\theta_L) < p(\theta_H)$, because then separating the demand curves might prompt the seller to lower the price for type θ_L consumers without increasing it for type θ_H consumers.

Following this observation, the modern and the classical literature then take different directions. Bergemann et al. (2015) focus on this force, asking what can happen when information is used to open new markets and demonstrating its power.[22] The classical literature imposes a non-exclusion assumption in order to examine whether, without opening new markets, price discrimination can improve welfare. Our result shows that it generally can. Indeed, when there is partial inclusion (so no new markets are opened) but as long as one of the two other conditions of the characterization is violated (for example, the spanning condition, which is a demanding one), information does not affect welfare monotonically, so there is some information that is beneficial. We pursue this case further in the next section.

4 Beyond Monotonicity: Bounds and Improvements

Theorem 1 immediately leads to two followup questions: In cases where information does not affect welfare monotonically, what are the bounds on the effects, and what are the best and worst ways to provide additional information? Our second main result answers these two questions.

To simplify the analysis we assume that there is a finite set of types, where n denotes the number of types. In parallel to our analysis of the binary case where we represented a market by a single scalar, we represent a market by a $n-1$-dimensional vector $\mu = (\mu_2, \ldots, \mu_n)$, where μ_i is the probability of each type θ_i for $i \neq 1$ and $1 - \sum_{i=2}^{n} \mu_i$ is the probability of type 1.

The first part of the result provides tight bounds on the marginal value of information. To formalize, for any two segmentation σ, σ' where σ is a mean-preserving spread of σ', define

$$\Delta V^{\alpha}(\sigma, \sigma') := \frac{V^{\alpha}(\sigma) - V^{\alpha}(\sigma')}{\mathbb{E}_{\sigma}\left[\|\mu\|_2^2\right] - \mathbb{E}_{\sigma'}\left[\|\mu\|_2^2\right]}, \tag{8}$$

where $||\cdot||$ is the Euclidean norm. The numerator is the difference between the weighted surplus of the two segmentations. The denominator is a measure of the additional information contained in σ compared to σ'. As a result, $\Delta V^{\alpha}(\sigma, \sigma')$ measures the *rate* of surplus change

[22] In their setting, simply separating the types does not increase consumer surplus, so more careful constructions are needed. The main force, however, is the ability to create segmentations that decrease prices for some consumers without increasing it for others.

with respect to the amount of additional information. The theorem identifies the lowest and the highest values that $\Delta V(\sigma, \sigma')$ can take over all pairs of segmentation $\sigma \underset{\text{MPS}}{\succeq} \sigma'$. This result is therefore a *quantitative* counterpart to our (qualitative) first result, which is concerned with identifying when the bounds are both (weakly) positive (α-IMG), both (weakly) negative (α-IMB), or have different signs (the non-montone case).

The second part of the result identifies the best and the worst ways to provide additional information. We particularly focus on the case where a small amount of additional information is provided, and ask what the best and worst "direction" for providing information is. To formalize this, for any segmentation σ and $\epsilon \leq 1$, let σ_ϵ be a segmentation that moves σ towards the prior market by replacing each market μ in the support of σ with $\epsilon\mu + (1-\epsilon)\mu_0$.[23] When $\epsilon = 1$, the segmentation is equal to σ, and as ϵ decreases, more weight is put on the prior market, so less information is provided in some formal sense. We say that a segmentation σ is marginally optimal if for every segmentation σ' of the same size, i.e., $\mathbb{E}_\sigma\left[\|\mu\|_2^2\right] = \mathbb{E}_{\sigma'}\left[\|\mu\|_2^2\right]$, we have $V^\alpha(\sigma_\epsilon) \geq V^\alpha(\sigma'_\epsilon)$ for small enough ϵ, and is marginally pessimal if the opposite holds, $V^\alpha(\sigma_\epsilon) \leq V^\alpha(\sigma'_\epsilon)$. For a given vector $\nu \in \mathbb{R}^{n-1}$, we say that a segmentation spreads the prior market in the direction ν if for every market μ in the segmentation, $\mu - \mu_0$ is proportional to ν. The result shows that marginally optimal (pessimal) segmentations are those that spreads the prior in a particular direction and identifies that direction.

For both the bounds and the best and worst directions, an important role is played by the lowest and the highest eigenvalues and eigenvectors of the Hessian matrix $\nabla^2 W^\alpha(\mu) = (\partial_{\mu_i}\partial_{\mu_j} W^\alpha(\mu))_{i,j=2,\ldots,n}$ of the value function $W^\alpha(\mu)$ (defined in Equation (5)). The largest possible benefit of providing information is proportional to the largest eigenvalue, and the best direction to provide information is in the direction of the associated eigenvector, which we simply call the "largest" eigenvector (and analogously for the worst benefit and direction).

These eigenvalues and eigenvectors exist and can be specified in closed form. Indeed, they directly reflect and generalize the three effects of information from our two-type analysis. Because exploring the connection and giving the expressions explicitly requires some additional notation, we defer doing so to after the statement of the result, and in the statement only note that these values can be identified in closed form.

Theorem 2 (Bounds and Improvements). *Suppose $\mathcal{D}$ contains a finite number of demand curves and there is partial inclusion.*

(i) For any market μ, the Hessian $\nabla^2 W^\alpha(\mu)$ has at most two non-zero eigenvalues. The lowest and the highest eigenvalues of the Hessian, $\underline{\lambda}^\alpha(\mu)$ and $\overline{\lambda}^\alpha(\mu)$, and their eigenvectors $\underline{v}^\alpha(\mu)$ and $\overline{v}^\alpha(\mu)$, exist and can be expressed in closed form. The two eigenvalues have weakly opposite signs, $\underline{\lambda}^\alpha(\mu) \leq 0 \leq \overline{\lambda}^\alpha(\mu)$.

[23] Formally the support of σ_ϵ is $\{\epsilon\mu + (1-\epsilon)\mu_0 \mid \mu \in supp(\sigma)\}$, and $\sigma_\epsilon(M) = \sigma(\{\frac{\mu-(1-\epsilon)\mu_0}{\epsilon} \mid \mu \in M\})$ for every set of markets M in the support of σ_ϵ.

(ii) The lower and upper bounds on the marginal value of information are proportional to the lowest and the highest eigenvalues,

$$\inf_{\sigma \underset{MPS}{\succeq} \sigma'} \Delta V^{\alpha}(\sigma, \sigma') = \frac{1}{2} \min_{\mu \in \Delta\Theta} \underline{\lambda}^{\alpha}(\mu), \tag{9}$$

$$\sup_{\sigma \underset{MPS}{\succeq} \sigma'} \Delta V^{\alpha}(\sigma, \sigma') = \frac{1}{2} \max_{\mu \in \Delta\Theta} \overline{\lambda}^{\alpha}(\mu). \tag{10}$$

(iii) A segmentation is marginally optimal (pessimal) if and only if it spreads the prior market μ_0 in the direction of the largest (smallest) eigenvector $\overline{v}^{\alpha}(\mu_0)$.

We prove Theorem 2 by using the Taylor expansion of the difference between surplus at σ and σ', with random markets μ and μ', to approximate it in a quadratic form

$$\Delta := V^{\alpha}(\sigma) - V^{\alpha}(\sigma') \approx \frac{1}{2}\mathbb{E}\left[(\mu' - \mu)^{T} \nabla^2 W^{\alpha}(\tilde{\mu})(\mu' - \mu)\right],$$

where $\tilde{\mu}$ is some market that lies on the line that connects μ and μ'. We then apply the Courant–Fischer–Weyl min-max principle to bound the above expression using eigenvalues.[24] In particular, we show

$$\Delta \leq \frac{1}{2}\mathbb{E}\left[(\mu' - \mu)^{T}(\mu' - \mu)\overline{\lambda}^{\alpha}(\tilde{\mu})\right] \leq \frac{1}{2}\left(\mathbb{E}_{\sigma}\left[\|\mu\|_2^2\right] - \mathbb{E}_{\sigma'}\left[\|\mu\|_2^2\right]\right) \max_{\mu \in \Delta\Theta} \overline{\lambda}^{\alpha}(\mu).$$

A similar application of the Taylor expansion, applied to small perturbations of the prior market, gives the marginally optimal and pessimal directions.

The result has a simple intuition. Recall that information is monotonically bad if the value function W^{α} is concave, and is monotonically good if it is convex. So it is natural that how bad or good information is depends on how concave or convex the value function is. Eigenvalues measure curvature along their eigenvectors: The largest eigenvector identifies a direction in which the value function is "most convex", and the eigenvalue measures how convex it is in that direction. The best information to provide is along that most convex direction, and the value of providing information in that direction is captured by the eigenvalue, the degree of convexity. The extreme cases of welfare-monotonicity are when the value function is concave or convex, in which case both the eigenvalues of the Hessian are negative or positive, respectively.

Because Theorem 2 is a quantitative result, it allows us to formalize the statement that if a family of demand curves approximately but not exactly satisfies the conditions of Theorem 1, then welfare-monotonicity approximately holds. This is simply because the bounds of Theorem 2, and in particular the eigenvalues of the matrix $\nabla^2 W^{\alpha}(\mu)$, change continuously as we perturb a family of demand curves, so if one of the bounds is zero for a family of demand

[24] See Theorem 1.3.2 in Tao (2023).

curves, then the corresponding bound approaches zero as we approach that family of demand curves.

Theorem 2 can also be used to provide bounds on the magnitude of the change in weighted surplus $V^{\alpha}(\sigma) - V^{\alpha}(\sigma')$, as opposed to the *rate* of surplus change as measured in Equation (8). Indeed, for any two segmentations $\sigma, \sigma' \in S(\mu_0)$ where σ is a mean-preserving spread of σ', we can bound the amount of additional information that σ contains compared to σ' as follows,

$$\mathbb{E}_{\sigma}\left[\|\mu\|_2^2\right] - \mathbb{E}_{\sigma'}\left[\|\mu\|_2^2\right] \leq 1 - \|\mu_0\|_2^2,$$

because the amount of additional information is highest when σ is the full-information segmentation and σ' is the no-information segmentation. As a result, substituting the above bound into that of Theorem 2, we get the following bounds on the magnitude of the change in surplus,

$$\frac{1 - \|\mu_0\|_2^2}{2} \min_{\mu \in \Delta\Theta} \underline{\lambda}^{\alpha}(\mu) \leq V^{\alpha}(\sigma) - V^{\alpha}(\sigma') \leq \frac{1 - \|\mu_0\|_2^2}{2} \max_{\mu \in \Delta\Theta} \overline{\lambda}^{\alpha}(\mu). \tag{11}$$

Whereas the bound in Equation (11) measures how much surplus increases or decreases, the bound in Equation (10) is on the rate of change in surplus. As a result, Equation (11) is useful in situations where we know nothing beyond the fact that the seller is collecting additional information, whereas Equation (10) gives tighter bounds when we also know (or can bound) *how much* additional information the seller is collecting. In particular, the bounds in Equation (11) are loose when the amount of additional information is small, because in such cases $1 - \|\mu_0\|_2^2$ is a loose bound on the amount of additional information that σ contains compared to σ'.

Even though the definition of what a marginally optimal segmentation is seems to start from the prior distribution, corresponding to no information, it can also be used to identify how to improve any segmentation. That is, suppose we start from some arbitrary segmentation, and want to optimally provide a small amount of additional information. The additional information, that spreads each posterior μ in the support of the segmentation in a mean-preserving way, must do so in the direction that is marginally optimal for that posterior, which is in the direction of the largest eigenvector $\overline{\lambda}^{\alpha}(\mu)$.

Finally, let us point out that appropriate generalization of the theorem applies to Bayesian persuasion problems in which the receiver's action is specified by a first-order condition (for example, Kolotilin et al., 2024).

4.1 Closed-form Expressions and the Three Effects of Information

As we mentioned above, the lowest and highest eigenvalues and eigenvectors of the Hessian matrix of the value function, which represent the magnitude and directions of best information, can be expressed in closed form. We provide these closed-form expressions here. An

examination of these expressions reveals that they generalize the three effects of information we discussed in our two-type analysis.

Before discussing the eigenvalues and the eigenvectors of the Hessian of W^α, let us first examine the Hessian itself. This Hessian can be written as a sum of three terms. In particular, each element $W^\alpha_{i,j}(\mu)$ of the Hessian matrix, which is a cross-partial derivatives of the value function with respect to μ_i, μ_j can be written as

$$\begin{aligned} W^\alpha_{i,j}(\mu) = {} & p_i(\mu)p_j(\mu)\mathbb{E}\left[V^\alpha_{pp}(\theta, p(\mu))\right] \\ & + p_j(\mu)\Big(V^\alpha_p(\theta_i, p(\mu)) - V^\alpha_p(\theta_1, p(\mu))\Big) + p_i(\mu)\Big(V^\alpha_p(\theta_j, p(\mu)) - V^\alpha_p(\theta_1, p(\mu))\Big) \\ & + p_{i,j}(\mu)\mathbb{E}\left[V^\alpha_p(\theta, p(\mu))\right]. \end{aligned} \tag{12}$$

This cross-partial derivative is a generalization of Equation (4) and represents the three effects of information. The first term represents the within-group price change effect, the second one the cross-groups price change effect, and the third one the price curvature effect.

The key observation that allows us to obtain closed-form expressions for the eigenvalues and eigenvectors of the Hessian is that the Hessian itself is a low rank matrix with a separable form. For this, let us first write the Hessian compactly in matrix form using Equation (12),

$$\begin{aligned} \nabla^2 W^\alpha(\mu) = {} & \nabla p(\mu)^T \nabla p(\mu) \mathbb{E}\left[V^\alpha_{pp}(\theta, p(\mu))\right] \\ & + (\Delta V^\alpha_p(\mu))^T \nabla p(\mu) + \nabla p(\mu)^T \Delta V^\alpha_p(\mu) \\ & + \nabla^2 p(\mu) \mathbb{E}\left[V^\alpha_p(\theta, p(\mu))\right], \end{aligned}$$

where $\nabla p(\mu) = (p_2(\mu), \ldots, p_n(\mu))$ is the gradient vector of optimal price, $\nabla^2 p(\mu)$ is the Hessian of the optimal price, the vector

$$\Delta V^\alpha_p(\mu) = \left(V^\alpha_p(\theta_i, p(\mu)) - V^\alpha_p(\theta_1, p(\mu))\right)_{i=2,\ldots,n}$$

is one that compares the derivative of the value function at every type $i \neq 1$ to that of type 1, and the superscript T transposes each row vector to a column vector.

Notice that the first to third term on the right hand side of the Hessian are already separated as products of a column vector and a row vector (multiplied possibly by a scalar). Only the last term, involving the Hessian $\nabla^2 p(\mu)$ of the optimal price, is not already separated as products of two vectors. We show that this term can also be separated, which allows us to write

$$\nabla^2 W^\alpha(\mu) = x(\mu)^T \nabla p(\mu) + \nabla p(\mu)^T x(\mu), \tag{13}$$

for some $x(\mu)$ that is defined as follows,

$$x(\mu) = \frac{1}{2}\mathbb{E}[V_{pp}^{\alpha}(\theta, p(\mu))]\nabla p(\mu) + \Delta V_{p}^{\alpha}(\mu) - \frac{\mathbb{E}[V_{p}^{\alpha}(\theta, p(\mu))]}{\mathbb{E}[R_{pp}(\theta, p(\mu))]}\left(\Delta R_{pp}(\mu) + \frac{1}{2}\mathbb{E}[R_{ppp}(\theta, p(\mu))]\nabla p(\mu)\right),$$

where the vector

$$\Delta R_{pp}(\mu) = \Big(R_{pp}(\theta_i, p(\mu)) - R_{pp}(\theta_1, p(\mu))\Big)_{i=2,\ldots,n}$$

is one that compares the second derivative of the revenue function at every type $i \neq 1$ to that of type 1.

The vector x is a direct generalization of the expression in Equation (1) that characterizes welfare-monotonicity for binary types. Recall that welfare-monotonicity is equivalent to the monotonicity of Equation (1) as a function of price. If we take the derivative of Equation (1), and evaluate it at the optimal price of a market, we obtain the equivalent of vector x in the binary case.

Given the separable form of the Hessian, its two eigenvalues and eigenvectors can be written in closed form using the same two vectors ∇p and x,

$$\begin{aligned}
\overline{\lambda}(\mu) =& \nabla p(\mu)\, x(\mu)^T \quad + \|\nabla p(\mu)\|_2 \|x(\mu)\|_2 \\
\underline{\lambda}(\mu) =& \nabla p(\mu)\, x(\mu)^T \quad - \|\nabla p(\mu)\|_2 \|x(\mu)\|_2 \\
\overline{v}(\mu) =& \nabla p(\mu) \|x(\mu)\|_2 + \|\nabla p(\mu)\|_2 x(\mu) \\
\underline{v}(\mu) =& \nabla p(\mu) \|x(\mu)\|_2 - \|\nabla p(\mu)\|_2 x(\mu).
\end{aligned}$$

4.2 Two Examples

To conclude this section, let us apply Theorem 2 to two examples. We will show that even though the first example does not satisfy welfare-monotonicity, we can think of it as an instance where information is essentially bad because the potential benefit is much smaller than the potential harm. To identify best directions in an example where the possible gains are potentially large, we study a second example.

Recall from Example 2 that information is monotonically bad for total surplus for the binary family of constant-elasticity demands $\{(1+p)^{-\theta_1}, (1+p)^{-\theta_2}\}$ when $1 < \theta_2 < \theta_1 \leq \theta_2 + \frac{1}{2}$, for example when $\theta_1 = 2$ and $\theta_2 = \frac{3}{2}$. However, information is *not* monotonically bad for total surplus when we add a third constant-elasticity demand curve, such as when the family is $\left\{(1+p)^{-3/2}, (1+p)^{-\theta}, (1+p)^{-2}\right\}$, $\frac{3}{2} < \theta < 2$, because the spanning condition of Theorem 1 is violated for this family. Let us now apply Theorem 2 to this family to obtain tight welfare bounds.

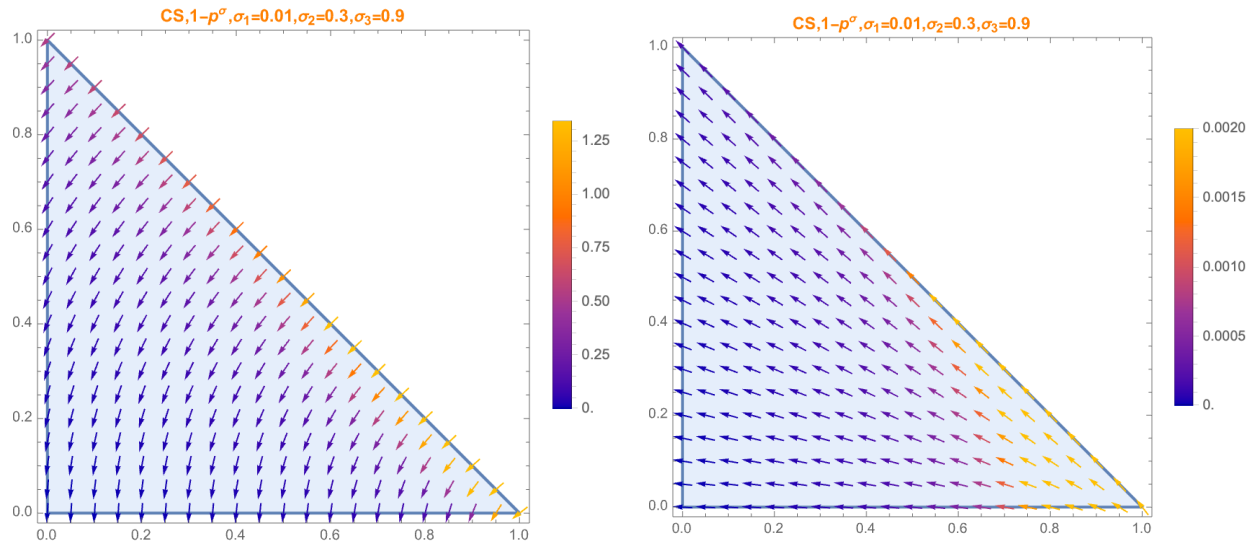
Table 3 contains the results of the numerical calculations of the lower and the upper bounds of Theorem 2 associated with different values of θ, for the case of total surplus. As

can be seen in the table, the upper bound on the marginal value of information is at most in the order of magnitude of 10^{-4}, whereas the lower bound is much larger (in absolute value), between -0.28 and -0.46. Recall also that these bounds are tight in the sense formalized in Theorem 2. We interpret these observations as saying that while information is not monotonically bad for this family, the possible benefit from segmentation is much smaller compared the to potential harm, and information is essentially monotonically bad for total surplus (and therefore also for consumer surplus).

θ	Lower bound	Upper bound
1.9	-0.460	4.6×10^{-5}
1.8	-0.395	1.47×10^{-4}
1.7	-0.332	2.19×10^{-4}
1.6	-0.282	1.48×10^{-4}

Table 3: The lower and upper bounds of Theorem 2 for a parametric family of three constant-elasticity demand curves.

In order to illustrate the second part of the theorem that identifies best directions, let us move on to another example in which the possible benefits of information are significant relative to its potential harms. Figure 4 shows the results of numerical calculations for the family of demand curves $\{1 - p^{\theta}\}_{\theta=0.01,0.3,0.9}$, for the case of consumer surplus.



(a) (b)

Figure 4: The (a) best and (b) worst directions for information.

Each triangle in the figure shows the set of all markets, with the market containing only the "low" type (the one with $\theta = 0.01$) as the right corner, the one with only the "middle" type

($\theta = 0.3$) as the top corner, and the one with only the "high" type ($\theta = 0.9$) as the bottom left corner. For each market in Panel (a), the arrow shows the direction of best information, that is, how the market should be split to maximize the gains from segmentation, and the color shows the magnitude of improvement (from blue for lower to yellow for higher). Similarly, the arrows in Panel (b) represent the worst direction and its magnitude. Comparing the scales to the right of the two panels shows that the potential gains are significant relative to potential harms for consumer surplus (1.25 compared to 0.002), and therefore also for total surplus. So let us focus on Panel (a) and discuss the best directions.

Roughly speaking, for markets towards the northeast, that is, those in which the probability of the high type is small, information has a large benefit and the best direction is diagonal. Such a segmentation keeps the difference between the probabilities of the low and the middle types constant across segments. That is, it splits the prior market to two identical segments, and then takes an equal measure of the low and middle type consumers from one segment and moves them to the other one. For markets towards the southwest, that is, those in which the probability of the high type is large, information has a small benefit and the best direction is vertical. Such a segmentation keeps the probability of the low type fixed, providing no information about the low type, but separates the middle type from the high type.

5 Conclusions

We develop a model of price discrimination based on information structures that connects the classical literature to the modern one. Like the classical literature, the primitive of our model is a set of demand curves and a distribution over them. Like the modern literature, we study the class of all information structures.

Our first main result gives conditions under which welfare increases monotonically, decreases monotonically, or changes non-monotonically when segmentations are refined. A main insight of the characterization is that information affects welfare in three ways: the within-group price change effect, the cross-groups price change effect, and the price curvature effect. While two of these effects are related in spirit to the output and misallocation effects identified in the classical literature (but with a main difference that they are concerned with price, not quantity), the third one, the within-group price change effect, has no counterpart and arises because even consumers of a given type are treated non-uniformly by partial information. Another insight of the first result is that information generally has the potential to increase welfare even without opening new markets.

Our second main result studies the non-monotone case further. It provides bounds on the effects of information and identifies the best direction to provide additional information. The result is based on a simple yet novel insight that eigenvalues of the Hessian matrix of the value function measure how good and bad information can be in the directions specified by their corresponding eigenvectors. Notably, in our setting these eigenvalues (and their eigenvectors)

can be identified in closed form, reflecting and generalizing the three effects of information from our two-type analysis.

References

Aguirre, Iñaki and Simon G Cowan, "Monopoly price discrimination with constant elasticity demand," *Economic Theory Bulletin*, 2015, *3*, 329–340. 7

Aguirre, Inaki, Simon Cowan, and John Vickers, "Monopoly price discrimination and demand curvature," *American Economic Review*, 2010, *100* (4), 1601–1615. 4, 7, 19, 20, 58

Asseyer, Andreas, "Information Orders in Screening Problems," *Available at SSRN 4192468*, 2024. 6

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris, "The Limits of Price Discrimination," *Cowles Foundation Discussion Paper No. 1896R*, 2013. 6

_ , _ , and _ , "The limits of price discrimination," *American Economic Review*, 2015, *105* (3), 921–957. 2, 5, 6, 8, 21, 22

_ , Tibor Heumann, and Michael C Wang, "A unified approach to second and third degree price discrimination," *arXiv preprint arXiv:2401.12366*, 2024. 6

Blackwell, David, "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 1953, pp. 265–272. 10, 12

Bonatti, Alessandro, Yunhao Huang, and J Miguel Villas-Boas, "A theory of the effects of privacy," *Management Science*, 2025. 6

Brooks, Benjamin, Alexander Frankel, and Emir Kamenica, "Comparisons of signals," *American Economic Review*, 2024, *114* (9), 2981–3006. 6

Cowan, Simon, "The welfare effects of third-degree price discrimination with nonlinear demand functions," *The RAND Journal of Economics*, 2007, *38* (2), 419–428. 20

_ , "Third-degree price discrimination and consumer surplus," *The Journal of Industrial Economics*, 2012, *60* (2), 333–345. 7

_ , "Welfare-increasing third-degree price discrimination," *The RAND Journal of Economics*, 2016, *47* (2), 326–340. 7

Curello, Gregorio and Ludvig Sinander, "The comparative statics of persuasion," *arXiv preprint arXiv:2204.07474*, 2022. 6

Dworczak, Piotr and Anton Kolotilin, "The persuasion duality," *Theoretical Economics*, 2024, *19* (4), 1701–1755. 7, 18

_ **and Giorgio Martini**, "The simple economics of optimal persuasion," *Journal of Political Economy*, 2019, *127* (5), 1993–2048. 7, 18

Haghpanah, Nima and Ron Siegel, "The limits of multiproduct price discrimination," *American Economic Review: Insights*, 2022, *4* (4), 443–458. 6

_ **and** _ , "Pareto-improving segmentation of multiproduct markets," *Journal of Political Economy*, 2023, *131* (6), 1546–1575. 6

Hidir, Sinem and Nikhil Vellodi, "Privacy, personalization, and price discrimination," *Journal of the European Economic Association*, 2021, *19* (2), 1342–1363. 6

Immorlica, Nicole, Andre M Sztutman, and Robert M Townsend, "Optimal Credit Scores Under Adverse Selection," in "Proceedings of the 23rd ACM Conference on Economics and Computation" 2022, pp. 737–738. 7

Kamenica, Emir and Matthew Gentzkow, "Bayesian persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615. 5, 12, 42

Kartik, Navin and Weijie Zhong, "Lemonade from lemons: Information design and adverse selection," *arXiv preprint arXiv:2305.02994*, 2023. 6

Kolotilin, Anton, "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 2018, *13* (2), 607–635. 7, 18, 32

_ **, Roberto Corrao, and Alexander Wolitzky**, "Persuasion and Matching: Optimal Productive Transport," *Journal or Political Ecconomy*, 2024. 5, 7, 8, 18, 25, 32, 33

Pigou, Arthur C., *The economics of welfare*, London: Macmillan, 1920. 2

Robinson, Joan, *The economics of imperfect competition*, Springer, 1969. 4

Saeedi, Maryam and Ali Shourideh, "Optimal Rating Design under Moral Hazard," *arXiv preprint arXiv:2008.09529*, 2024. 7

_ **, Yikang Shen, and Ali Shourideh**, "Getting the Agent to Wait," *arXiv preprint arXiv*, 2024, *2407*. 7

Schmalensee, Richard, "Output and welfare implications of monopolistic third-degree price discrimination," *The American Economic Review*, 1981, *71* (1), 242–247. 4

Smolin, Alex and Takuro Yamashita, "Information Design in Smooth Games," *arXiv e-prints*, 2022, pp. arXiv–2202. 7

Strack, Philipp and Kai Hao Yang, "Non-Discriminatory Personalized Pricing," *arXiv preprint arXiv:2506.20925*, 2025. 6

Tao, Terence, *Topics in random matrix theory*, Vol. 132, American Mathematical Society, 2023. 24

Varian, Hal R, "Price discrimination and social welfare," *The American Economic Review*, 1985, *75* (4), 870–875. 4, 7

Yang, Kai Hao, "Selling consumer data for profit: Optimal market-segmentation design and its consequences," *American Economic Review*, 2022, *112* (4), 1364–1393. 6

Appendix

A Proofs

A.1 Proof of Proposition 1

First notice that the optimal price $p(\mu)$ for any market μ must be at least $\min_\theta p(\theta)$ and at most $\max_\theta p(\theta)$. Because of partial inclusion, all revenue curves are strictly concave in this range. As a result, the optimal price $p(\mu) \in [\min_\theta p(\theta), \max_\theta p(\theta)]$ is unique and is characterized by

$$\int_\Theta R_p(p(x), \theta)\, \mathrm{d}\mu(\theta) = 0.$$

It is convenient to reformulate the problem as one of choosing a joint distribution G over types and prices subject to the constraint that the marginal of G over the types agrees with μ and the obedience constraint that conditioned on a given price recommendation, the expected marginal revenue of the seller must be zero.

$$\begin{aligned} \max_{G \in \Delta(I\times\Theta)} & \int_{I\times\Theta} U(p,\theta)\, \mathrm{d}G(p,\theta) \qquad \text{(P)} \\ \text{s.t. } & G(I\times\Theta') = \mu(\Theta'), \forall \Theta' \subset \Theta \\ & \int_{I'\times\Theta} R_p(p,\theta)\, \mathrm{d}G(p,\theta) = 0, \forall I' \subset I \end{aligned}$$

We use the following complementary slackness conditions that follow from the strong duality results of Kolotilin (2018) and Kolotilin et al. (2024), which relates problem (P) to the following dual problem.

$$\begin{aligned} \min_{\lambda,\zeta} & \int_\Theta \lambda(\theta)\, \mathrm{d}\mu(\theta) \qquad \text{(D)} \\ \text{s.t. } & \lambda(\theta) + \zeta(p) R_p(p,\theta) \geq U(p,\theta), \forall (p,\theta) \in I\times\Theta. \qquad (14) \end{aligned}$$

Lemma 1 (Kolotilin et al., 2024). *Optimal solutions G and λ, ζ to the primal (P) and the dual (D) exist and their values are equal,*

$$\int_{I\times\Theta} U(p,\theta)\,\mathrm{d}G(p,\theta) = \int_{\Theta} \lambda(\theta)\,\mathrm{d}\mu(\theta). \tag{15}$$

We now use this strong duality result to establish Proposition 1. By Lemma 1, a feasible G is an optimal solution to the primal problem (P) if and only if there exists a feasible solution λ, ζ to the dual problem (D) such that Equation (15) holds. For any feasible G, we have

$$\int_{I\times\Theta} \zeta(p) R_p(p,\theta)\,\mathrm{d}G(p,\theta) = 0,$$

so

$$\int_{I\times\Theta} U(p,\theta)\,\mathrm{d}G(p,\theta) = \int_{I\times\Theta} [U(p,\theta) - \zeta(p)R_p(p,\theta)]\,\mathrm{d}G(p,\theta),$$

and

$$\int_{\Theta} \lambda(\theta)\,\mathrm{d}\mu(\theta) = \int_{I\times\Theta} \lambda(\theta)\,\mathrm{d}G(p,\theta).$$

So Equation (15) is equivalent to

$$\int_{I\times\Theta} [U(p,\theta) - \zeta(p)R_p(p,\theta)]\,\mathrm{d}G(p,\theta) = \int_{I\times\Theta} \lambda(\theta)\,\mathrm{d}G(p,\theta),$$

which can be written as

$$\lambda(\theta) + \zeta(p)R_p(p,\theta) = U(p,\theta), G\text{-almost surely} \tag{16}$$

To summarize, we have shown that a probability measure G is an optimal solution to the problem (P) if and only if there exists functions $\lambda(\theta), \zeta(p)$ such that Equation (14) and Equation (16) hold. Recall that welfare-monotonicity is equivalent to the optimality of the no-information segmentation for *all* prior distributions. The no-information segmentation corresponds to a measure G that assigns probability 1 to the prior market and its optimal price. The no-information segmentation corresponds to a distribution G that assigns probability 1 to price p that is optimal for μ (and induces marginal μ over types). So suppose that such a distribution G is optimal for problem (P). This observation pins down the function λ, because given Equation (16) for every θ we must have

$$\lambda(\theta) = U(p,\theta) - \zeta(p)R_p(p,\theta).$$

Substituting the definition of λ into Equation (14), we conclude that the no-information seg-

mentation is optimal for a prior μ with optimal price p if and only if there exists a function ζ such that for every type θ,

$$U(p,\theta) - \zeta(p)R_p(p,\theta) \geq U(p',\theta) - \zeta(p')R_p(p',\theta), \forall p',$$

which we can write as

$$p \in \arg\max_{p'} U(p',\theta) - \zeta(p')R_p(p',\theta). \tag{17}$$

Recall that welfare-monotonicity holds if the no-information segmentation is optimal for *every* prior μ. Notice that any price in I is optimal for *some* prior μ. Welfare-monotonicity therefore holds if and only if for every p in I, there exists ζ such that Equation (17) holds. This function ζ may depend on p. To make this dependence explicit, let us use $\zeta(p,\cdot)$ as a function that is indexed by p. Using this notation, welfare-monotonicity holds if and only if there exists a single function ζ such that for every $p \in I$ and every type θ,

$$p \in \arg\max_{p'} U(p',\theta) - \zeta(p,p')R_p(p',\theta).$$

A.2 Proof of Theorem 1

A.2.1 Necessity of Partial Inclusion

A simplifying step is that we can focus on two demands. In particular, for a family of demands $\mathcal{D}$, and any two types θ_L, θ_H in it, if α-IMB (α-IMG) does not hold for the binary family $\mathcal{D}' = \{D(p,\theta_L), D(p,\theta_H)\}$, then it does not hold for $\mathcal{D}$ either.

Suppose first that there is full exclusion, that is, $\overline{p}(\theta_L) < p(\theta_H)$. We want to show that both α-IMB and α-IMG must be violated. It is sufficient to prove this claim for a certain (not every) prior market μ on θ_L, θ_H. This is because for any prior market μ' we can first construct a segmentation that contains market μ in its support, and then show that segmenting μ' further does not increase or decrease weighted surplus (depending on which one of IMG or IMB we are proving).

Consider the set $p(\mu)$ of optimal prices in a market in which θ has probability $1-\mu$ and θ_H has probability μ,

$$p(\mu) = \arg\max_{p} (1-\mu)R(p,\theta_L) + \mu R(p,\theta_H).$$

Notice that $p(\mu)$ never includes a price in $(\overline{p}(\theta_L), p(\theta_H))$ because any such price, which excludes θ_L entirely, leads to a lower revenue than $p(\theta_H)$. Moreover, $p(\mu)$ never includes a price less than $p(\theta_L)$ because any such price is also lower than $p(\theta_H)$ and increasing it increases revenue in both types. Therefore, $p(\mu)$ may only contain prices in $[p(\theta_L), \overline{p}(\theta_L)] \cup \{p(\theta_H)\}$. At $\mu = 0$, the only optimal price is $p(\theta_L)$, and at $\mu = 1$, the only optimal price is $p(\theta_H)$. Be-

cause for each p, revenue changes linearly in μ, there exists a threshold $\widehat{\mu} \in (0,1)$ such that $p(\theta_H) \in p(\mu)$ if and only if $\mu \geq \widehat{\mu}$, and $p(\theta_H)$ is the unique optimal price for all $\mu > \widehat{\mu}$.

To see that α-IMB does not hold, consider any segmentation of the prior market that contains a segment in which the probability of θ_H in the prior market is $\mu' > \widehat{\mu}$ (such a segmentation exists because the prior market has full support over θ_L, θ_H). The unique optimal price for μ' is $p(\theta_H)$. Consider segmenting μ' further into two segments $\mu_1 < \widehat{\mu} < \mu'$ and $\mu_2 > \mu'$. Because any optimal price for market μ_1 is strictly less than $p(\theta_H)$ and the unique optimal price for market μ_2 is $p(\theta_H)$, this segmentation increases consumer surplus. Because any segmentation weakly increases producer surplus, it also strictly increases weighted surplus.

To see that α-IMG does not hold, consider any segmentation of the prior market that contains a segment in which the probability of θ_H in the prior market is $\mu' = \widehat{\mu} - \epsilon$ for some small ϵ. Consider segmenting μ' into two segments $\mu_1 = \mu' - 2\epsilon$ and $\mu_2 = \mu' + 2\epsilon = \widehat{\mu} + \epsilon$, with probability $\frac{1}{2}$ each. As ϵ goes to zero, the seller's revenue approaches the optimal revenue in market μ' continuously by the Maximum theorem. But we argue that the consumer surplus decreases discontinuously, that it, it decreases by at least some $\delta > 0$, and therefore weighted surplus decreases for small enough ϵ.

To see the discontinuity in consumer surplus, consider optimal prices for markets μ_1 and μ_2. Because $\mu_1 < \widehat{\mu}$, optimal prices for μ_1 are those that maximize revenue over $[p(\theta_L), \overline{p}(\theta_L)]$. Because the revenue curve is strictly concave over this range, the optimal price and therefore consumer surplus changes continuously. But for market μ_2 the only optimal price is $p(\theta_H)$ and the consumer surplus of this market is bounded away from that of μ', so the consumer surplus and total surplus decrease for small enough ϵ.

Now suppose there is full inclusion, $p(\theta_L) < \underline{p}(\theta_H)$. We argue that there exist μ_1, μ_2 such that $\mu_1 < \mu_2$ and price $\underline{p}(\theta_H)$ is the unique optimal price for any market in (μ_1, μ_2). For this, we first show that note that the revenue curve associated with θ_H has a kink at price $\underline{p}(\theta_H)$. The revenue of type θ_H at any price $p < \underline{p}(\theta_H)$ is

$$R(p, \theta_H) = pD(\underline{p}(\theta_H), \theta_H)$$

so the left derivative of $R(p, \theta_H)$ at $p = \underline{p}(\theta)$ is

$$R_p(p, \theta_H) = D(\underline{p}(\theta), \theta_H).$$

The revenue of type θ_H at any price $p > \underline{p}(\theta_H)$ is

$$R(p, \theta_H) = pD(p, \theta_H)$$

so marginal revenue is

$$R_p(p, \theta_H) = D(p, \theta_H) + pD_p(p, \theta_H).$$

As p converges to $\underline{p}(\theta_H)$ from above, marginal revenue converges to

$$\lim_{p \to^+ \underline{p}(\theta_H)} R_p(p, \theta_H) = D(\underline{p}(\theta_H), \theta_H) + pD_p(\underline{p}(\theta_H), \theta_H) < D(\underline{p}(\theta_H), \theta_H).$$

So the right derivative of the revenue curve at $\underline{p}(\theta_H)$ is strictly less than its left derivative. Let $\delta^- > \delta^+$ be the left and the right derivatives of the revenue curve of type θ_H at $\underline{p}(\theta_H)$. Price $\underline{p}(\theta_H)$ is optimal in any market μ such that

$$(1-\mu)R_p(\underline{p}(\theta_H), \theta_L) + \mu\delta^- > 0, \text{ and } (1-\mu)R_p(\underline{p}(\theta_H), \theta_L) + \mu\delta^+ < 0,$$

that is

$$\mu \in (\mu_1, \mu_2) := \left(\frac{-R_p(\underline{p}(\theta_H), \theta_L)}{\delta^+ - R_p(\underline{p}(\theta_H), \theta_L)}, \frac{-R_p(\underline{p}(\theta_H), \theta_L)}{\delta^- - R_p(\underline{p}(\theta_H), \theta_L)}\right).$$

Now consider IMB. consider any segmentation of the prior market that contains a segment in which the probability of θ_H in the prior market is $\mu' \in (\mu_1, \mu_2)$. Consider segmenting μ' further into two segments μ_1' and μ_2' such that $\mu_1' < \mu_1 < \mu'$ and $\mu' < \mu_2' < \mu_2$. Any optimal price for μ_1' is less than $\underline{p}(\theta_H)$ and the optimal price for μ_2' is $\underline{p}(\theta_H)$. This segmentation therefore increases consumer surplus. Because any segmentation weakly increases producer surplus, it also increases weighted surplus.

Finally, consider IMG. For this, we examine the value function $W^\alpha(\mu)$ around μ_2 and show that the value function is locally concave, which means that providing a small amount of information reduces weighted surplus. For this, consider the left derivative of $W^\alpha(\mu)$ at $\mu = \mu_2$. Recall that price $\underline{p}(\theta_H)$ is optimal for all markets in (μ_1, μ_2), which means that the left derivative of $W^\alpha(\mu)$ at $\mu < \mu_2$ is

$$\begin{aligned} W^\alpha_\mu(\mu) &= V^\alpha(p(\mu), \theta_H) - V^\alpha(p(\mu), \theta_L) + \Big((1-\mu)V^\alpha_p(p(\mu), \theta_L) + \mu V^\alpha_p(p(\mu), \theta_H)\Big)p_\mu(\mu), \\ &= V^\alpha(p(\mu), \theta_H) - V^\alpha(p(\mu), \theta_L), \end{aligned}$$

where the equality follows because $p_\mu(\mu) = 0$ for μ in (μ_1, μ_2) as $p(\mu) = \underline{p}(\theta_H)$. As μ converges to μ_2 from below, this derivative converges to

$$V^\alpha(\underline{p}(\theta_H)), \theta_H) - V^\alpha(\underline{p}(\theta_H), \theta_L).$$

Now consider what happens as μ converges to μ_2 from above. The derivative of W^α is

$$V^\alpha(p(\mu),\theta_H) - V^\alpha(p(\mu),\theta_L) + \Big((1-\mu)V_p^\alpha(p(\mu),\theta_L) + \mu V_p^\alpha(p(\mu),\theta_H)\Big)p_\mu(\mu).$$

Given the first-order condition of the seller's problem, we can write this derivative as

$$V^\alpha(p(\mu),\theta_H) - V^\alpha(p(\mu),\theta_L) + \alpha\Big((1-\mu)CS_p(p(\mu),\theta_L) + \mu CS_p(p(\mu),\theta_H)\Big)p_\mu(\mu).$$

which, as μ goes to μ_2 from above, converges to

$$V^\alpha(\underline{p}(\theta_H),\theta_H) - V^\alpha(\underline{p}(\theta_H),\theta_L) + \alpha\Big((1-\mu)CS_p(\underline{p}(\theta_H),\theta_L) + \mu CS_p(\underline{p}(\theta_H),\theta_H)\Big)p_\mu(\mu_2)$$
$$< V^\alpha(\underline{p}(\theta_H),\theta_H) - V^\alpha(\underline{p}(\theta_H),\theta_L),$$

where the inequality follows because $p(\mu)$ is strictly increasing at $\mu \geq \mu_2$, and $CS_p < 0$. We conclude that W^α is not concave, which means that IMG does not hold.

A.2.2 Statement (i)

Let us start by an implication of the spanning property that will be later used in the proof.

Lemma 2. *If the spanning condition of Theorem 1 is violated, then there exists some θ, $p \in I$, and a, b such that $R_p(p,\theta) \neq 0$ and*

$$R_p(p,\theta) = aR_p(p,\theta_L) + bR_p(p,\theta_H)$$
$$R_{pp}(p,\theta) = aR_{pp}(p,\theta_L) + bR_{pp}(p,\theta_H)$$
$$U_p(p,\theta) \neq aU_p(p,\theta_L) + bU_p(p,\theta_H).$$

Proof. We will show that there is an open interval of prices $I' \subset (p(\theta_L), p(\theta_H))$ such that the above three equations hold for all $p \in I'$. Because marginal revenue is decreasing in p, there exists such a price that furthermore satisfies $R_p(p,\theta) \neq 0$.

For each p there exists a pair a, b of constants for which the first two equalities hold, that is, writing a, b as functions of p to make this dependence explicit,

$$R_p(p,\theta) = a(p)R_p(p,\theta_L) + b(p)R_p(p,\theta_H) \tag{18}$$
$$R_{pp}(p,\theta) = a(p)R_{pp}(p,\theta_L) + b(p)R_{pp}(p,\theta_H), \tag{19}$$

which is obtained by solving the system of two equations and two variables,

$$a(p) = \frac{R_{pp}(p,\theta)R_p(p,\theta_H) - R_p(p,\theta)R_{pp}(p,\theta_H)}{R_p(p,\theta_H)R_{pp}(p,\theta_H) - R_p(p,\theta_H)R_{pp}(p,\theta_H)}$$
$$b(p) = \frac{-R_{pp}(p,\theta)R_p(p,\theta_H) + R_p(p,\theta)R_{pp}(p,\theta_H)}{R_p(p,\theta_H)R_{pp}(p,\theta_H) - R_p(p,\theta_H)R_{pp}(p,\theta_H)}.$$

This solution is well-defined because $R_{pp} < 0$ and $R_p(p,\theta_H) < 0 < R_p(p,\theta_H)$.

Similarly, let us define functions c, d as follows

$$R(p,\theta) = c(p)R(p,\theta_H) + d(p)R(p,\theta_H)$$
$$R_p(p,\theta) = c(p)R_p(p,\theta_H) + d(p)R_p(p,\theta_H)$$

Using a similar argument to above, the pair $c(p), d(p)$ is also unique for each p. Violation of the spanning condition implies that c, d are not constant functions.

Taking a derivative of the above, we have

$$\begin{aligned} R_p(p,\theta) =& c'(p)R(p,\theta_H) + d'(p)R(p,\theta_H) + c(p)R_p(p,\theta_H) + d(p)R_p(p,\theta_H) \\ =& c(p)R_p(p,\theta_H) + d(p)R_p(p,\theta_H), \end{aligned} \tag{20}$$

which means

$$c'(p)R(p,\theta_H) = -d'(p)R(p,\theta_H),$$

and

$$\begin{aligned} R_{pp}(p,\theta) =& c'(p)R_p(p,\theta_H) + d'(p)R_p(p,\theta_H) + c(p)R_{pp}(p,\theta_H) + d(p)R_{pp}(p,\theta_H) \\ =& (\frac{R_p(p,\theta_H)}{R(p,\theta_H)} - \frac{R_p(p,\theta_H)}{R(p,\theta_H)})c'(p)R(p,\theta_H) + c(p)R_{pp}(p,\theta_H) + d(p)R_{pp}(p,\theta_H) \\ =& \Delta(p) + c(p)R_{pp}(p,\theta_H) + d(p)R_{pp}(p,\theta_H), \end{aligned} \tag{21}$$

where $\Delta(p)$ is defined as

$$\Delta(p) := (\frac{R_p(p,\theta_H)}{R(p,\theta_H)} - \frac{R_p(p,\theta_H)}{R(p,\theta_H)})c'(p)R(p,\theta_H).$$

Let I' be an interval for which $c'(p) \neq 0$, which means that $\Delta(p) \neq 0$ for the same interval.

Combining Equation (18) and Equation (19) with Equation (20) and Equation (21), we have

$$\begin{aligned} a(p) &= \frac{R_{pp}(p,\theta)R_p(p,\theta_H) - R_p(p,\theta)R_{pp}(p,\theta_H)}{R_p(p,\theta_H)R_{pp}(p,\theta_H) - R_p(p,\theta_H)R_{pp}(p,\theta_H)} \\ &= c(p) + \frac{\Delta(p)R_p(p,\theta_H)}{R_p(p,\theta_H)R_{pp}(p,\theta_L) - R_p(p,\theta_L)R_{pp}(p,\theta_H)} \qquad (22) \\ b(p) &= \frac{-R_{pp}(p,\theta)R_p(p,\theta_H) + R_p(p,\theta)R_{pp}(p,\theta_H)}{R_p(p,\theta_H)R_{pp}(p,\theta_H) - R_p(p,\theta_H)R_{pp}(p,\theta_H)} \\ &= d(p) - \frac{\Delta(p)R_p(p,\theta_L)}{R_p(p,\theta_H)R_{pp}(p,\theta_L) - R_p(p,\theta_L)R_{pp}(p,\theta_H)}. \qquad (23) \end{aligned}$$

Now let us evaluate

$$U_p(p,\theta) - a(p)U_p(p,\theta_L) - b(p)U_p(p,\theta_H) \qquad (24)$$

using the above two equations and show that it is not equal to zero. From the definition of U, when characterizing α-IMG the above expression is

$$\begin{aligned} U_p(p,\theta) - aU_p(p,\theta_L) - bU_p(p,\theta_H) &= \alpha\Big(CS_p(p,\theta) - a(p)CS_p(p,\theta_L) - b(p)CS_p(p,\theta_H)\Big) \\ &\quad + (1-\alpha)\Big(R_p(p,\theta) - a(p)R_p(p,\theta_L) - b(p)R_p(p,\theta_H)\Big) \\ &= -\alpha\Big(D(p,\theta) - a(p)D(p,\theta_L) - b(p)D(p,\theta_H)\Big) \\ &= \frac{-\alpha}{p}\Big(R(p,\theta) - a(p)R(p,\theta_L) - b(p)R(p,\theta_H)\Big). \end{aligned}$$

When characterizing α-IMB, we want to show that the negative of the above expression is non-zero. In either case, because $\alpha > 0$, to show that Equation (24) is not zero, it is sufficient to show that

$$R(p,\theta) - a(p)R(p,\theta_L) - b(p)R(p,\theta_H) \neq 0.$$

For this, use Equation (22) and Equation (23) to write

$$\begin{aligned} & R(p,\theta) - a(p)R(p,\theta_L) - b(p)R(p,\theta_H) \\ =& R(p,\theta) - c(p)R(p,\theta_L) - d(p)R(p,\theta_H) - \Delta(p)\frac{R_p(p,\theta_H)R(p,\theta_L) - R_p(p,\theta_L)R(p,\theta_H)}{R_p(p,\theta_H)R_{pp}(p,\theta_L) - R_p(p,\theta_L)R_{pp}(p,\theta_H)} \\ =& -\Delta(p)\frac{R_p(p,\theta_H)R(p,\theta_L) - R_p(p,\theta_L)R(p,\theta_H)}{R_p(p,\theta_H)R_{pp}(p,\theta_L) - R_p(p,\theta_L)R_{pp}(p,\theta_H)}. \end{aligned}$$

For any $p \in I'$, this expression is non-zero because $\Delta(p) \neq 0$, which proves the claim. □

Given Lemma 2, we now complete the proof of statement (i) of Theorem 1.

Proof of statement (i). We have already shown in Appendix A.2.1 that partial inclusion is necessary for the welfare-monotonicity properties. We therefore assume partial inclusion here and establish the necessity and sufficiency of the reduction.

The case where $\min_\theta p(\theta) = \max_\theta p(\theta)$ is straightforward. In this case, both welfare-monotonicity properties as well as the conditions of the statement hold trivially. So suppose for the rest of the proof that $\min_\theta p(\theta) < \max_\theta p(\theta)$.

Necessity. Suppose there is partial inclusion. Recall from Proposition 1 that welfare-monotonicity holds if and only if there exists a function ζ such that for every $p \in I$ and every type θ,

$$p \in \arg\max_{p' \in I} U(p', \theta) - \zeta(p, p') R_p(p', \theta). \tag{25}$$

First, notice that if Equation (25) is satisfied for all $\theta \in \Theta$, then it must be satisfied for θ_L, θ_H that have the lowest and the highest monopoly price in Θ, and therefore welfare-monotonicity holds for the binary family that consists only of θ_L, θ_H.

Now consider any pair θ, θ' of types for which $p(\theta) < p(\theta')$ and any $p \in (p(\theta), p(\theta'))$. Equation (25) implies

$$U(p', \theta) - \zeta(p, p') R_p(p', \theta) \leq U(p, \theta) - \zeta(p, p) R_p(p, \theta), \forall p' \in (p(\theta), p(\theta')).$$

Because $R_p(p', \theta) < 0$, we can divide the above inequality by $R_p(p', \theta)$ and write it as

$$\zeta(p, p') \leq \frac{U(p', \theta) - U(p, \theta) + \zeta(p, p) R_p(p, \theta)}{R_p(p', \theta)}.$$

A similar argument repeated for θ', but with the difference that $R_p(p', \theta') > 0$, implies

$$\zeta(p, p') \geq \frac{U(p', \theta') - U(p, \theta') + \zeta(p, p) R_p(p, \theta')}{R_p(p', \theta')}.$$

Therefore, we must have that for all $p, p' \in (p(\theta), p(\theta'))$:

$$\frac{U(p', \theta) - U(p, \theta) + \zeta(p, p) R_p(p, \theta)}{R_p(p', \theta)} \geq \frac{U(p', \theta') - U(p, \theta') + \zeta(p, p) R_p(p, \theta')}{R_p(p', \theta')}.$$

Note that evaluated at $p' = p$, both sides of the above are equal to $\zeta(p, p)$. Because both sides are continuously differentiable, it has to be that they are tangent at $p' = p$. Therefore,

$$\frac{U_p(p, \theta)}{R_p(p, \theta)} - \zeta(p, p) \frac{R_{pp}(p, \theta)}{R_p(p, \theta)} = \frac{U_p(p, \theta')}{R_p(p, \theta')} - \zeta(p, p) \frac{R_{pp}(p, \theta')}{R_p(p, \theta')}$$

which pins down $\zeta(p,p)$,

$$\zeta(p,p) = \frac{\frac{U_p(p,\theta)}{R_p(p,\theta)} - \frac{U_p(p,\theta')}{R_p(p,\theta')}}{\frac{R_{pp}(p,\theta)}{R_p(p,\theta)} - \frac{R_{pp}(p,\theta')}{R_p(p,\theta')}} = \frac{U_p(p,\theta)R_p(p,\theta') - U_p(p,\theta')R_p(p,\theta)}{R_{pp}(p,\theta)R_p(p,\theta') - R_{pp}(p,\theta')R_p(p,\theta)}. \tag{26}$$

Now suppose the spanning condition of the theorem is violated, so there exists some θ whose demand curve cannot be written as a linear combination of $D(\cdot,\theta_L)$ and $D(\cdot,\theta_H)$ over the interval I. Lemma 2 implies that that there is price $p \in I$ and a, b such that $R_p(p,\theta) \neq 0$ and

$$\begin{aligned} R_p(p,\theta) &= aR_p(p,\theta_L) + bR_p(p,\theta_H) \\ R_{pp}(p,\theta) &= aR_{pp}(p,\theta_L) + bR_{pp}(p,\theta_H) \\ U_p(p,\theta) &\neq aU_p(p,\theta_L) + bU_p(p,\theta_H). \end{aligned}$$

Because $R_p(p,\theta_L) < 0 < R_p(p,\theta_H)$ and $R_p(p,\theta) \neq 0$, the marginal revenue of θ has a different sign with either θ_L, θ_H. Suppose $0 < R_p(p,\theta)$ (the other case is similar). Then our discussion above pins down $\zeta(p,p)$ in two different ways

$$\zeta(p,p) = \frac{U_p(p,\theta_L)R_p(p,\theta_H) - U_p(p,\theta_H)R_p(p,\theta_L)}{R_{pp}(p,\theta_L)R_p(p,\theta_H) - R_{pp}(p,\theta_H)R_p(p,\theta_L)}$$

and

$$\zeta(p,p) = \frac{U_p(p,\theta_L)R_p(p,\theta) - U_p(p,\theta)R_p(p,\theta_L)}{R_{pp}(p,\theta_L)R_p(p,\theta) - R_{pp}(p,\theta)R_p(p,\theta_L)}.$$

But these two expressions cannot be equal by Lemma 2, implying that ζ satisfying Equation (25) does not exist. To see this, let us use Lemma 2 and the above two equalities to write

$$\begin{aligned} \zeta(p,p) &= \frac{U_p(p,\theta_L)R_p(p,\theta) - U_p(p,\theta)R_p(p,\theta_L)}{R_{pp}(p,\theta_L)R_p(p,\theta) - R_{pp}(p,\theta)R_p(p,\theta_L)} \\ &\neq \frac{U_p(p,\theta_L)\Big(aR_p(p,\theta_L) + bR_p(p,\theta_H)\Big) - \Big(aU_p(p,\theta_L) + bU_p(p,\theta_H)\Big)R_p(p,\theta_L)}{R_{pp}(p,\theta_L)\Big(aR_p(p,\theta_L) + bR_p(p,\theta_H)\Big) - \Big(aR_{pp}(p,\theta_L) + bR_{pp}(p,\theta_H)\Big)R_p(p,\theta_L)} \\ &= \frac{b\Big(U_p(p,\theta_L)R_p(p,\theta_H) - U_p(p,\theta_H)R_p(p,\theta_L)\Big)}{b\Big(R_{pp}(p,\theta_L)R_p(p,\theta_H) - R_{pp}(p,\theta_H)R_p(p,\theta_L)\Big)} \\ &= \zeta(p,p), \end{aligned}$$

which is a contradiction.

Sufficiency. We now prove the sufficiency part of the reduction. Suppose a family of demand curves $\mathcal{D}$ can be decomposed into two demands that satisfy α-IMG (α-IMB),

$$D(p,\theta) = f_1(\theta)D(p,\theta_L) + f_2(\theta)D(p,\theta_H), \forall p, \theta.$$

We want to show that $\mathcal{D}$ satisfies α-IMG (α-IMB). That is, there exists a function ζ such that for every price $p \in I$, we have

$$p \in \arg\max_{p' \in I} U(p',\theta) - \zeta(p,p')R_p(p',\theta),$$

for $W = V^\alpha$ ($W = -V^\alpha$).

Notice that because each demand is a linear combination of the two basis demands, each value, revenue, and marginal revenue function can also be written using a linear combination of the corresponding objects for the basis demand curves. Formally,

$$\begin{aligned} U(p,\theta) &= f_1(\theta)U(p,\theta_L) + f_2(\theta)U(p,\theta_H), \\ R(p,\theta) &= f_1(\theta)R(p,\theta_L) + f_2(\theta)R(p,\theta_H), \\ R_p(p,\theta) &= f_1(\theta)R_p(p,\theta_L) + f_2(\theta)R_p(p,\theta_H). \end{aligned}$$

Because the family $\{D(p,\theta_L), D(p,\theta_H)\}$ satisfies α-IMG (α-IMB), for each $p \in I$ we have

$$\begin{aligned} p &\in \arg\max_{p' \in I} U(p',\theta_L) - \zeta(p,p')R_p(p',\theta_L), \\ p &\in \arg\max_{p' \in I} U(p',\theta_H) - \zeta(p,p')R_p(p',\theta_H). \end{aligned}$$

Because p maximizes each of the above two expressions, it also maximizes their linear combination,

$$\begin{aligned} p \in \arg\max_{p' \in I} & f_1(\theta)\Big(U(p',\theta_L) - \zeta(p,p')R_p(p',\theta_L)\Big) + f_2(\theta)\Big(U(p',\theta_H) - \zeta(p,p')R_p(p',\theta_H)\Big) \\ &= \arg\max_{p' \in I} U(p',\theta) - \zeta(p,p')R_p(p',\theta) \end{aligned}$$

as claimed, completing the proof.

A.2.3 Statement (ii)

We have already shown in Appendix A.2.1 that partial inclusion is necessary for the welfare-monotonicity properties. So suppose there is partial inclusion. By the concavification result of Kamenica and Gentzkow (2011), α-IMB (α-IMG) holds if and only if value as a function of

μ,

$$W^\alpha(\mu) = (1-\mu)V^\alpha(p(\mu), \theta_L) + \mu V^\alpha(p(\mu), \theta_H),$$

is concave where $p(\mu)$ is the profit-maximizing price for the seller. The partial inclusion condition implies that this price is uniquely defined and $p(\mu) \in [p(\theta_L), p(\theta_H)]$ is the profit-maximizing price for the seller that solves the seller's first-order condition

$$(1-\mu)R_p(p(\mu), \theta_L) + \mu R_p(p(\mu), \theta_H) = 0.$$

We identify three consequences of the first-order condition for future use. First, re-arranging the first-order condition, we have

$$\frac{\mu}{1-\mu} = \frac{-R_p(p(\mu), \theta_L)}{R_p(p(\mu), \theta_H)}. \tag{27}$$

Second, taking the derivative of the first-order condition with respect to μ, we have

$$-R_p(p(\mu), \theta_L) + R_p(p(\mu), \theta_H) + \Big((1-\mu)R_{pp}(p(\mu), \theta_L) + \mu R_{pp}(p(\mu), \theta_H)\Big)p_\mu(\mu) = 0,$$

which, after re-arranging, means

$$p_\mu(\mu) = \frac{R_p(p(\mu), \theta_L) - R_p(p(\mu), \theta_H)}{(1-\mu)R_{pp}(p(\mu), \theta_L) + \mu R_{pp}(p(\mu), \theta_H)}. \tag{28}$$

Third, $p(\mu)$ is increasing in μ. For this, notice that because $p(\theta_L) \leq p(\mu) \leq p(\theta_H)$ and each revenue curve is concave, the first term in the numerator, $R_p(p(\mu), \theta_L)$, is negative and the second term, $R_p(p(\mu), \theta_H)$, is positive, so the numerator is negative. Further, because both revenue curves are strictly concave, the denominator is negative.

Now $W^\alpha(\mu)$ is concave if and only if its derivative $W^\alpha_\mu(\mu)$ is decreasing. From the definition of $W^\alpha(\mu)$, and using Equation (27) and Equation (28), we can write its derivative $W^\alpha_\mu(\mu)$ as

$$\begin{aligned} W^\alpha_\mu(\mu) =& V^\alpha(p(\mu), \theta_H) - V^\alpha(p(\mu), \theta_L) \\ &+ \frac{V^\alpha_p(p(\mu), \theta_L) - \frac{R_p(p(\mu), \theta_L)}{R_p(p(\mu), \theta_H)} V^\alpha_p(p(\mu), \theta_H)}{R_{pp}(p(\mu), \theta_L) - \frac{R_p(p(\mu), \theta_L)}{R_p(p(\mu), \theta_H)} R_{pp}(p(\mu), \theta_H)} \Big(R_p(p(\mu), \theta_L) - R_p(p(\mu), \theta_H)\Big). \end{aligned}$$

Because $p(\mu)$ is increasing in μ, $W^\alpha_\mu(\mu)$ is decreasing if and only if $W^\alpha_\mu(\mu(p))$ is decreasing,

where μ is the inverse of p, that is, $p(\mu(p)) = p$. So we want

$$V^\alpha(p,\theta_H) - V^\alpha(p,\theta_L) + \frac{V_p^\alpha(p,\theta_L) - \frac{R_p(p,\theta_L)}{R_p(p,\theta_H)} V_p^\alpha(p,\theta_H)}{R_{pp}(p,\theta_L) - \frac{R_p(p,\theta_L)}{R_p(p,\theta_H)} R_{pp}(p,\theta_H)} \Big(R_p(p,\theta_L) - R_p(p(\mu),\theta_H) \Big)$$

to be decreasing in p, as claimed.

A.3 Proof of Theorem 2

A.3.1 Statement (i)

In addition to proving statement (i) of the theorem, stating that the eigenvalues and eigenvectors of the Hessian of the value function exist and can be calculated in closed form, we explicitly derive the closed forms as discussed in Section 4.1.

The value function is

$$\begin{aligned} W^\alpha(\mu) &= \sum_{k=2}^{n} V^\alpha(\theta_k, p(\mu))\mu_k + (1 - \sum_{k=2}^{n} \mu_k) V^\alpha(\theta_L, p(\mu)) \\ &= \sum_{k=2}^{n} \Delta V^\alpha(\theta_k, p(\mu))\mu_k + V^\alpha(\theta_1, p(\mu)), \end{aligned}$$

The derivative with respect to μ_i and μ_j is

$$\begin{aligned} W_{i,j}^\alpha(\mu) = &\sum_{k=1}^{n} V_{pp}^\alpha(\theta_k, p(\mu)) p_j(\mu) p_i(\mu) \mu_k \\ &+ \sum_{k=1}^{n} V_p^\alpha(\theta_k, p(\mu)) p_{i,j}(\mu) \mu_k \\ &+ \Delta V_p^\alpha(\theta_j, p(\mu)) p_i(\mu) + \Delta V_p^\alpha(\theta_i, p(\mu)) p_j(\mu). \end{aligned}$$

In matrix form, the Hessian is

$$\begin{aligned} \nabla^2 W^\alpha(\mu) &= \mathbb{E}[V_{pp}^\alpha(\theta, p(\mu))] \nabla p(\mu)^T \nabla p(\mu) \\ &\quad + \mathbb{E}[V_p^\alpha(\theta, p(\mu))] \nabla^2 p(\mu) \\ &= (\Delta V_p^\alpha(\mu))^T \nabla p(\mu) + \nabla p(\mu)^T \Delta V_p^\alpha(\mu). \end{aligned} \tag{29}$$

To see that $p_{i,j}(\mu)$ can be written in a separable form, recall the first order condition for the optimal price,

$$0 = \sum_{k=2}^{n} R_p(\theta_k, p(\mu))\mu_k + (1 - \sum_{k=2}^{n} \mu_k) R_p(\theta_1, p(\mu)).$$

Taking the derivative with respect to μ_i and then μ_j gives,

$$\begin{aligned}0 = &\sum_{k=1}^{n} R_{ppp}(\theta_k, p(\mu))p_j(\mu)p_i(\mu)\mu_k \\ &+ \sum_{k=1}^{n} R_{pp}(\theta_k, p(\mu))p_{i,j}(\mu)\mu_k \\ &+ \Delta R_{pp}(\theta_j, p(\mu))p_i(\mu) + \Delta R_{pp}(\theta_i, p(\mu))p_j(\mu),\end{aligned}$$

and therefore, in matrix form,

$$\nabla^2 p(\mu) = -\frac{\mathbb{E}[R_{ppp}(\theta, p(\mu))]\nabla p(\mu)^T \nabla p(\mu) + (\Delta R_{pp}(\mu))^T \nabla p(\mu) + \nabla p(\mu)^T \Delta R_{pp}(\mu)}{\mathbb{E}[R_{pp}(\theta, p(\mu))]} \tag{30}$$

Combining Equation (29) and Equation (30) gives the separable form of the Hessian in Equation (13). The Hessian matrix has rank at most 2, which means it has at most two non-zero eigenvalues. We have given the explicit expression for two eigenvalues, which proves that they are the lowest and the highest eigenvalues.

A.3.2 Statement (ii) and (iii)

Consider segmentations σ, σ' with random markets μ, μ' where σ is a mean-preserving spread of σ'. By Blackwell's theorem, there exists a probability distribution function $Q(\cdot|\mu') \in \Delta\Delta\Theta$ such that

$$\begin{aligned}\sigma(A) &= \int Q(A|\mu')\,\mathrm{d}\sigma', \forall \text{ Borel } A \subset \Delta\Theta \\ \mu' &= \int \mu\,\mathrm{d}Q(\mu|\mu'), \forall \mu' \in \Delta\Theta.\end{aligned}$$

This implies that

$$\begin{aligned}V^\alpha(\sigma) - V^\alpha(\sigma') &= \int W^\alpha(\mu)\,\mathrm{d}\sigma - \int W^\alpha(\mu')\,\mathrm{d}\sigma' \\ &= \int\int W^\alpha(\mu)\,\mathrm{d}Q(\mu|\mu')\,\mathrm{d}\sigma' - \int W^\alpha(\mu')\,\mathrm{d}\sigma' \\ &= \int\int [W^\alpha(\mu) - W^\alpha(\mu')]\,\mathrm{d}Q(\mu|\mu')\,\mathrm{d}\sigma'.\end{aligned}$$

Since W^α is a twice differentiable function, we can use the Cauchy form of the Taylor expansion of W and write

$$W^\alpha(\mu) - W^\alpha(\mu') = (\mu - \mu')^T \nabla W^\alpha(\mu) + \frac{1}{2}(\mu - \mu')^T \nabla^2 W^\alpha(\tilde{\mu})(\mu - \mu)$$

where $\tilde{\mu}$ is some belief on the line that connects μ to μ'. Integrating over the above for a μ',

we must have that

$$\begin{aligned}\int \left[W^{\alpha}(\mu) - W^{\alpha}(\mu')\right] \mathrm{d}Q(\mu|\mu') &= \int (\mu-\mu')^{T} \nabla W^{\alpha}(\mu)\, \mathrm{d}Q(\mu|\mu') \\ &+ \int \frac{1}{2}(\mu-\mu')^{T} \nabla^{2} W^{\alpha}(\tilde{\mu})(\mu-\mu')\, \mathrm{d}Q(\mu|\mu') \\ &= \int \frac{1}{2}(\mu-\mu')^{T} \nabla^{2} W^{\alpha}(\tilde{\mu})(\mu-\mu')\, \mathrm{d}Q(\mu|\mu')\end{aligned}$$

where in the above we have used the fact that the average value of μ under $Q(\cdot|\mu')$ is μ'. Because the Hessian matrix is symmetric, we can use the Courant–Fischer–Weyl min-max principle to bound the integrand using the highest eigenvalue $\overline{\lambda}(\tilde{\mu})$ of $\nabla^{2} W(\tilde{\mu})$,

$$\begin{aligned}(\mu-\mu')^{T} \nabla^{2} W^{\alpha}(\tilde{\mu})(\mu-\mu') &\leq \overline{\lambda}(\tilde{\mu})(\mu-\mu')^{T}(\mu-\mu') && (31)\\ &\leq \max_{\mu'' \in \Delta\Theta} \overline{\lambda}(\mu'')(\mu-\mu')^{T}(\mu-\mu'), && (32)\end{aligned}$$

Using the above inequality, we must have that

$$\begin{aligned}\int \left[W^{\alpha}(\mu) - W^{\alpha}(\mu')\right] \mathrm{d}Q(\mu|\mu') &\leq \frac{1}{2} \max_{\mu'' \in \Delta\Theta} \overline{\lambda}(\mu'') \int (\mu-\mu')^{T}(\mu-\mu')\, \mathrm{d}Q(\mu|\mu') \\ &= \frac{1}{2} \max_{\mu'' \in \Delta\Theta} \overline{\lambda}(\mu'') \int \sum_{i \geq 2} \left[(\mu_i)^2 - (\mu'_i)^2\right] \mathrm{d}Q(\mu|\mu')\end{aligned}$$

Integrating over the above establishes the upper bound. The derivation of the lower bound follows the same logic, inverting the two inequalities Equation (31) and Equation (32) and replacing max with min.

To see that the bounds are tight, note that we can always get arbitrarily close to $\max_{\mu'' \in \Delta\Theta} \overline{\lambda}(\mu'')$ with full-support markets. Let $\widehat{\mu}$ be a full-support market that achieves that maximum minus ϵ. If we let σ' be the no information segmentation associated with $\widehat{\mu}$, as σ converges to σ', $\tilde{\mu}$ in the Taylor expansion of W^{α} converges to $\widehat{\mu}$. Since $\widehat{\mu}$ is interior, it can be perturbed in all directions including the direction that achieves the maximum value. This proves that the bound is tight with a gap of at most ϵ for any arbitrary ϵ.

Finally, we prove statement (iii) regarding the optimal direction. For any segmentation σ_{ϵ}, the Cauchy form of the Taylor expansion of W implies that

$$\begin{aligned}V^{\alpha}(\sigma_{\epsilon}) - W^{\alpha}(\mu_0) &= \int \left[W^{\alpha}(\mu) - W^{\alpha}(\mu_0)\right] \mathrm{d}\sigma_{\epsilon} \\ &= \int \left[W^{\alpha}((1-\epsilon)\mu_0 + \epsilon\mu) - W^{\alpha}(\mu_0)\right] \mathrm{d}\sigma \\ &= \int \frac{\epsilon^2}{2}(\mu-\mu_0)^{T} \nabla^{2} W^{\alpha}(\tilde{\mu}_{\epsilon})(\mu-\mu_0)\, \mathrm{d}\sigma && (33)\end{aligned}$$

where $\tilde{\mu}_{\varepsilon}$ is a market on the line that connects $(1-\epsilon)\mu_0 + \epsilon\mu$ to μ_0. Note that in the above as

$\varepsilon \to 0$, $\tilde{\mu}_\varepsilon \to \mu_0$ and hence, we can write

$$\begin{aligned}\lim_{\varepsilon\to 0} \frac{V^\alpha(\sigma_\epsilon) - W^\alpha(\mu_0)}{\varepsilon^2} &= \frac{1}{2}\int (\mu-\mu_0)^T \nabla^2 W(\mu_0)(\mu-\mu_0)\,\mathrm{d}\sigma \\ &\leq \frac{1}{2}\int \overline{\lambda}(\mu_0)(\mu-\mu_0)^T(\mu-\mu_0)\,\mathrm{d}\sigma \\ &= \frac{1}{2}\overline{\lambda}(\mu_0)\left[\mathbb{E}_\sigma\left[\|\mu\|_2^2\right] - \|\mu_0\|_2^2\right] \end{aligned} \tag{34}$$

where the inequality follows from the Courant–Fischer–Weyl min-max principle, with equality if every market μ' in the support of σ is an eigenvector associated with $\overline{\lambda}(\mu_0)$.

Now consider any segmentation σ. From Equation (33), notice that if we divide the difference in values $V^\alpha(\sigma_\epsilon) - W^\alpha(\mu_0)$ by ε^2 instead of ε, the limit is zero,

$$\lim_{\varepsilon\to 0} \frac{V^\alpha(\sigma_\epsilon) - W^\alpha(\mu_0)}{\varepsilon} = 0.$$

As a result, if a segmentation achieves the upper bound Equation (34) with equality, there for any other segmentation σ' of the same magnitude, $\mathbb{E}_\sigma\left[\|\mu\|_2^2\right] = \mathbb{E}_{\sigma'}\left[\|\mu\|_2^2\right]$, there exists ε small enough such that

$$V^\alpha(\sigma_\epsilon) - W^\alpha(\mu_0) \geq V^\alpha(\sigma'_\epsilon) - W^\alpha(\mu_0),$$

and therefore, $V^\alpha(\sigma_\epsilon) \geq V^\alpha(\sigma'_\epsilon)$.

Online Appendix

B Proofs of Corollaries

B.1 Proof of Corollary 1

Proof. Separating weighted surplus into consumer surplus and revenue, the expression from Equation (1) is equivalent to

$$\alpha\left[CS(p,\theta_H)-CS(p,\theta_L)+\frac{-\frac{R_p(p,\theta_L)}{R_p(p,\theta_H)}CS_p(p,\theta_H)+CS_p(p,\theta_L)}{-\frac{R_p(p,\theta_L)}{R_p(p,\theta_H)}R_{pp}(p,\theta_H)+R_{pp}(p,\theta_L)}\left(R_p(p,\theta_H)-R_p(p,\theta_L)\right)\right]$$
$$+(1-\alpha)\left[R(p,\theta_H)-R(p,\theta_L)\right].$$

The above expression is a convex combination of two functions with weights α and $1-\alpha$. The second function $R(p,\theta_H)-R(p,\theta_L)$ is increasing over $(p(\theta_L),p(\theta_H))$ because $R(p,\theta_H)$ is increasing for any price below the optimal monopoly price $p(\theta_H)$ for type θ_H and $R(p,\theta_L)$ is decreasing for any price above the optimal monopoly price $p(\theta_L)$ for type θ_L. Now suppose the convex combination is decreasing, so α-IMB holds, and consider a larger weight $\alpha'\geq\alpha$. Because the convex combination is decreasing but the second term is increasing, the first term must be decreasing. So if we replace α with the larger α', we are increasing the weight of the decreasing function and decreasing the weight of the increasing function, so the resulting combination must be decreasing as well, implying information is monotonically bad for α'. A similar argument applies for the case of monotonically good information. □

B.2 Proof of Corollary 2

Proof. Consider each of the terms in Equation (4).

The first term corresponds to the within-type price change effect. It is

$$(p_\mu(\mu))^2\Big[(1-\mu)V^\alpha_{pp}(p(\mu),\theta_L)+\mu V^\alpha_{pp}(p(\mu),\theta_H)\Big],$$

which is positive (negative) whenever

$$(1-\mu)V^\alpha_{pp}(p(\mu),\theta_L)+\mu V^\alpha_{pp}(p(\mu),\theta_H) \tag{35}$$

is positive (negative). A sufficient condition is that V^α is convex (concave), as claimed.

The second term corresponds to the cross-types price change effect. It is

$$2p_\mu(\mu)\Big[V^\alpha_p(p(\mu),\theta_H)-V^\alpha_p(p(\mu),\theta_L)\Big],$$

which is positive (negative) whenever

$$V_p^\alpha(p(\mu),\theta_H) - V_p^\alpha(p(\mu),\theta_L)$$

is positive (negative) because $p(\mu)$ is increasing in μ.

The third term corresponds to the price curvature effect. It is

$$p_{\mu\mu}(\mu)\Big[(1-\mu)V_p^\alpha(p(\mu),\theta_L) + \mu V_p^\alpha(p(\mu),\theta_H)\Big].$$

Notice that from the definition of V^α,

$$\begin{aligned}(1-\mu)V_p^\alpha(p(\mu),\theta_L) + \mu V_p^\alpha(p(\mu),\theta_H) &= \alpha\Big((1-\mu)CS_p(p(\mu),\theta_L) + \mu CS_p(p(\mu),\theta_H)\Big)\\ &\quad + (1-\alpha)\Big((1-\mu)R_p(p(\mu),\theta_L) + \mu R_p(p(\mu),\theta_H)\Big)\\ &= \alpha\Big((1-\mu)CS_p(p(\mu),\theta_L) + \mu CS_p(p(\mu),\theta_H)\Big),\\ &\leq 0,\end{aligned}$$

where the second equality follows because of the first-order condition of the seller's profit-maximization problem, and the inequality follows because consumer surplus is decreasing in price. So the third term is positive (negative) whenever $p_{\mu\mu}(\mu)$ is negative (positive).

To complete the proof, we relate the curvature of p to the two sufficient conditions in the third bullet of the corollary. For this, let us take two derivatives of the seller's profit-maximization condition,

$$(1-\mu)R_p(p(\mu),\theta_L) + \mu R_p(p(\mu),\theta_H) = 0.$$

The first derivative implies

$$-R_p(p(\mu),\theta_L) + R_p(p(\mu),\theta_H) + p_\mu(\mu)\Big[(1-\mu)R_{pp}(p(\mu),\theta_L) + \mu R_{pp}(p(\mu),\theta_H)\Big] = 0.$$

The second derivative implies

$$\begin{aligned}(p_\mu(\mu))^2 \quad &\Big[(1-\mu)R_{ppp}(p(\mu),\theta_L) + \mu R_{ppp}(p(\mu),\theta_H)\Big]\\ +2p_\mu(\mu) \quad &\Big[R_{pp}(p(\mu),\theta_H) - R_{pp}(p(\mu),\theta_L)\Big]\\ +p_{\mu\mu}(\mu) \quad &\Big[(1-\mu)R_{pp}(p(\mu),\theta_L) + \mu R_{pp}(p(\mu),\theta_H)\Big] = 0.\end{aligned} \tag{36}$$

Note that $p_{\mu\mu}(\mu)$ is multiplied by a negative term because both revenue curves are concave.

So $p_{\mu\mu}(\mu)$ is positive (negative) if the sum of the other two terms are positive (negative). The first term is positive (negative) whenever

$$(1-\mu)R_{ppp}(p(\mu),\theta_L)+\mu R_{ppp}(p(\mu),\theta_H)$$

is positive (negative), which is the case if $R_{ppp}(p,\theta)$ is positive (negative) for all p and both θ_L and θ_H. The second term is positive (negative) whenever

$$R_{pp}(p(\mu),\theta_H)-R_{pp}(p(\mu),\theta_L)$$

is positive (negative), completing the proof. □

Sufficient conditions for monotonically bad information. Corollary 2 implies that assuming $R_{ppp}(p,\theta)\geq 0$, that is, the marginal revenue curves are convex, and assuming an additional condition that each $V^{1/2}$ is concave, $\frac{1}{2}$-IMB (and therefore α-IMB for any $\alpha\geq\frac{1}{2}$) holds if the more elastic demand curve, θ_L, has

1. a higher derivative, $D_p(p,\theta_L)\geq D_p(p,\theta_H)$, and,
2. a more convex revenue curve, $R_{pp}(p,\theta_H)\geq R_{pp}(p,\theta_L)$, or equivalently a more steep marginal revenue curve.

The cross-groups price change effect is negative because $D_p(p,\theta_L)\geq D_p(p,\theta_H)$. The assumption that $V^{1/2}$ is concave is concave implies that the within-group price change effect is negative. The conditions that $R_{pp}(p,\theta_H)\geq R_{pp}(p,\theta_L)$ and $R_{ppp}(p,\theta)\geq 0$ ensure the price curvature effect is negative by Corollary 2.

B.3 Proof of Corollary 3

Proof. Let us without loss of generality parameterize the family as $\mathcal{D}=\{a(\theta)(D(p)+b(\theta))\}_\theta$. The first step is to show that $\mathcal{D}$ can be separated as in the theorem.

Let θ_L,θ_H be the demands in the family with the lowest and highest b, that is, $b(\theta_L)=\underline{b}$, $b(\theta_H)=\overline{b}$) (choosing arbitrarily if there are multiple candidates). Define

$$f_1(\theta)=\frac{a(\theta)(\overline{b}-b(\theta))}{a(\theta_L)(\overline{b}-\underline{b})}$$
$$f_2(\theta)=\frac{a(\theta)(b(\theta)-\underline{b})}{a(\theta_H)(\overline{b}-\underline{b})}.$$

We need to check that for all $\theta=(a,b)$,

$$f_1(\theta)D(p,\theta_L)+f_2(\theta)D(p,\theta_H)=D(p,\theta),$$

That is,

$$f_1(\theta)a(\theta_L) + f_2(\theta)a(\theta_H) = a,$$
$$f_1(\theta)a(\theta_L)b(\theta_L) + f_2(\theta)a(\theta_H)b(\theta_H) = ab.$$

These two equations hold because

$$f_1(\theta)a(\theta_L) + f_2(\theta)a(\theta_H) = \frac{a(\theta)}{\overline{b} - \underline{b}}((\overline{b} - b(\theta)) + (b(\theta) - \underline{b})) = a.$$
$$f_1(\theta)a(\theta_L)b(\theta_L) + f_2(\theta)a(\theta_H)b(\theta_H) = \frac{a(\theta)}{\overline{b} - \underline{b}}((\overline{b} - b(\theta))\underline{b} + (b(\theta) - \underline{b})\overline{b}) = ab.$$

The second step is to characterize the welfare-monotonicity properties for the binary family $\{D(p, \theta_L), D(p, \theta_H)\}$. In order to save on notation, suppose without loss of generality that $a(\theta_H) = b(\theta_H) = 1$, that is, $D(p, \theta_H) = D(p)$, and let $a = a(\theta_L), b = b(\theta_L)$. We want to show that the family $\{a(D(p) + b), D(p)\}$, satisfies α-IMB (α-IMG) if and only if there is partial inclusion and

$$(2\alpha - 1)p + \alpha(\frac{pD'(p)}{R''(p)})$$

is increasing (decreasing) over $[(R')^{-1}(-b), (R')^{-1}(0)]$.

We have

$$\begin{aligned}
D_1 &= a(D_2 + b) \\
R_1 &= aR_2 + apb \\
R_1' &= aR_2' + ab \\
R_1'' &= aR_2'' \\
V_1 &= aV_2 + ab(\alpha\overline{p} + (1 - 2\alpha)p) \\
V_1' &= aV_2' + ab(1 - 2\alpha).
\end{aligned}$$

So

$$\begin{aligned}
&V_2(p) - V_1(p) + \frac{-\frac{R_1'(p)}{R_2'(p)}V_2' + V_1'}{-\frac{R_1'(p)}{R_2'(p)}R_2'' + R_1''}(R_1'(p) - R_2'(p)) \\
=&V_2(p) - V_1(p) + \frac{-\frac{aR_2'(p)+ab}{R_2'(p)}V_2' + aV_2' + ab(1 - 2\alpha)}{-\frac{aR_2'+ab}{R_2'(p)}R_2'' + aR_2''}(R_2'(p)(a - 1) + ab) \\
=&V_2(p)(1 - a) - ab(\alpha\overline{p} + (1 - 2\alpha)p) + \frac{V_2'(p) - (1 - 2\alpha)R_2'(p)}{R_2''(p)}(R_2'(p)(a - 1) + ab)
\end{aligned}$$

The derivative of the expression is

$$\begin{aligned}
&V_2'(p)(1-a) - ab(1-2\alpha) + \frac{V_2'(p) - (1-2\alpha)R_2'(p)}{R_2''(p)} R_2''(p)(a-1) \\
&+ (\frac{V_2'(p) - (1-2\alpha)R_2'(p)}{R_2''(p)})'(R_2'(p)(a-1) + ab) \\
=&-(1-2\alpha)R_2'(p)(a-1) - ab(1-2\alpha) + (\frac{V_2'(p) - (1-2\alpha)R_2'(p)}{R_2''(p)})'(R_2'(p)(a-1) + ab) \\
=&(R_2'(p)(a-1) + ab)(-(1-2\alpha) + (\frac{V_2'(p) - (1-2\alpha)R_2'(p)}{R_2''(p)})').
\end{aligned}$$

Notice that $R_2'(p)(a-1) + ab = R_1'(p) - R_2'(p) \leq 0$,

so we want to show that

$$2\alpha - 1 + (\frac{V_2'(p) - (1-2\alpha)R_2'(p)}{R_2''(p)})' \geq 0.$$

Notice also that $V_2'(p) - (1-2\alpha)R_2'(p) = \alpha p D'(p)$. So we want to show that

$$2\alpha - 1 + (\frac{\alpha p D_2'(p)}{R_2''(p)})' \geq 0.$$

In other words we want to show that

$$(2\alpha - 1)p + \alpha(\frac{pD_2'(p)}{R_2''(p)})$$

to be increasing in p over $(p(\theta_L), p(\theta_H)) = [(R_2^{-1})'(-b), (R_2^{-1})'(0)]$, as claimed.

To see that log-concavity of f is sufficient when $\alpha = 1/2$, note that it is sufficient for

$$\frac{pD'(p)}{R_2''(p)}$$

to be increasing everywhere, or for

$$\frac{R_2''(p)}{pD'(p)} = \frac{2D'(p) + pD''(p)}{pD'(p)} = \frac{2}{p} + \frac{-D''(p)}{-D'(p)} = \frac{2}{p} + (log - D'(p))' = \frac{2}{p} + (\log f(p))'$$

to be decreasing. It is sufficient for $(\log f(p))'$ to be decreasing, that is, f is log-concave. □

B.3.1 Three Cases for How Information Affects CS and TS

Consider

$$f(p) = \frac{c_1(c_2 + c_3 p)^{c_4}}{p^2}.$$

This density function is positive if $c_1 \geq 0$ and $c_2 + c_3 p \geq 0$ for all $p \in [\underline{p}, \overline{p}]$. Then

$$\log p^2 f(p) = \log c_1 + c_4 \log(c_2 + c_3 p).$$

Because $D'(p) = -f(p)$ and $R''(p) = 2D'(p) + pD''(p) = -2f(p) - pf'(p)$, we have

$$\frac{pD'(p)}{R''(p)} = \frac{pf(p)}{2f(p) + pf'(p)} = (\frac{2}{p} + \frac{f'(p)}{f(p)})^{-1} = (\frac{d}{dp} \log p^2 f(p))^{-1} = \frac{c_2 + c_3 p}{c_3 c_4} = \frac{c_2}{c_3 c_4} + \frac{p}{c_4}.$$

Therefore, the expression in Equation (7) is

$$(2\alpha - 1)p + \alpha(\frac{pD'(p)}{R''(p)}) = p(2\alpha - 1 + \frac{\alpha}{c_4}) + \frac{\alpha c_2}{c_3 c_4},$$

which is increasing (decreasing) whenever the multiplier of p is positive (negative), that is, when

$$\alpha(2 + \frac{1}{c_4}) \geq (\leq)\ 1.$$

Now consider three cases for c_4. When $2 + \frac{1}{c_4} \leq 1$, that is, when $-1 \leq c_4 \leq 0$, then $\alpha(2 + \frac{1}{c_4}) \leq 1$ for all α and therefore information is monotonically good regardless of α. When $2 + \frac{1}{c_4} \geq 1$, that is, when $c_4 \leq -1$ or $c_4 \geq 0$, α-IMB (α-IMG) holds whenever

$$\alpha \geq (\leq)\ \widehat{\alpha} := \frac{1}{2 + \frac{1}{c_4}} = \frac{c_4}{2c_4 + 1}. \tag{37}$$

This threshold $\widehat{\alpha} \in (0, 1]$ is such that information is monotonically bad for all α above the threshold, is monotonically good for all α below the threshold, and has no effect on $\widehat{\alpha}$-surplus. When $c_4 \leq -1$, $\widehat{\alpha}$ ranges from $\frac{1}{2}$ to 1. This means that information is monotonically good for total surplus but bad for consumer surplus. When $c_4 \geq 0$, $\widehat{\alpha}$ ranges from 0 to $\frac{1}{2}$. This means that information is monotonically good for both total surplus and consumer surplus.

Concavity of revenue. We next verify that this example indeed has a concave revenue function. For this, let us calculate the derivative of the density,

$$f'(p) = \frac{c_1 c_3 c_4 (c_2 + c_3 p)^{c_4 - 1}}{p^2} - \frac{2c_1 (c_2 + c_3 p)^{c_4}}{p^3}.$$

Therefore,

$$\begin{aligned} R''(p) = -2f - pf' &= \frac{c_1 (c_2 + c_3 p)^{c_4 - 1}}{p^2} \Big(-2(c_2 + c_3 p) - p c_3 c_4 + 2(c_2 + c_3 p) \Big) \\ &= \frac{-c_1 c_3 c_4 (c_2 + c_3 p)^{c_4 - 1}}{p}, \end{aligned}$$

which is negative because $c_1 \geq 0$, $c_2 + c_3 p \geq 0$ for all $p \in [\underline{p}, \overline{p}]$, and c_3, c_4 have the same signs.

C Derivation for Examples

C.1 Example 1

First consider IMG. In this example, both marginal revenue curves are linear because $R_{ppp}(p, \theta) = 0$. Additionally, the two marginal revenue curves have the same slope, -2. Therefore, the only condition left to check is $D(p, \theta_1) \geq D(p, \theta_2)$ to ensure that the cross-types price change effect is positive. This condition is

$$c_1 - c_2 \geq (a_2 - a_1)p$$

for all prices in $[\frac{a_1}{2}, \frac{a_2}{2}]$. Because $a_2 \geq a_1$, the right hand side of the above inequality is increasing in p and takes its highest value at $p = p(\theta_2) = \frac{a_2}{2}$, so the above condition becomes

$$c_1 - c_2 \geq (a_2 - a_1)\frac{a_2}{2}.$$

Now consider IMB. Again, the only conditions to check are regarding the ranking of the derivatives and the technical condition $D_p(p, \theta) + pD_{pp}(p, \theta) \leq 0$ that implies that the within-type price change effect is negative. The condition $D_p(p, \theta_2) \leq D_p(p, \theta_1)$ becomes

$$c_1 \leq c_2.$$

The condition $D_p(p, \theta) + pD_{pp}(p, \theta) \leq 0$ becomes

$$c_i \leq p^2$$

for all prices in $[\frac{a_1}{2}, \frac{a_2}{2}]$ and each $i \in \{1, 2\}$, which given the condition $c_1 \leq c_2$ can be summarized as $c_2 \leq \frac{a_1^2}{4}$.

C.2 Example 2

For a constant-elasticity (CE) demand curve $D(p, \theta) = (c + p)^{-\theta}$, the optimal monopoly price is given by

$$p(\theta) = \frac{c}{\theta - 1}.$$

Thus, we require $c > 0$ and $\theta > 1$ to ensure that a finite monopoly price exists when the monopolist faces each demand curve. To make the exposition of the proof more intuitive, we

define $\theta_1 = \theta_H$ and $\theta_2 = \theta_L$, which means $\theta_H > \theta_L$, and use (θ_H, θ_L) instead of (θ_1, θ_2). Thus, the statement of the example is

$$|\theta_H - \theta_L| \leq \frac{1}{2} \Rightarrow \alpha - \text{IMB holds for } \alpha \geq \frac{1}{2}$$

Let $x = \theta_H - \theta_L, 0 < x \leq \frac{1}{2}$. The derivative of the expression in Equation (1) simplifies to

$$-\frac{(c+p)^{-\theta_L - x - 1}\left((c - \theta_L p + p)(c+p)^x - c + \theta_L p + px - p\right)}{2\left(2c^2 - cp(2\theta_L + x - 3) + (\theta_L - 1)p^2(\theta_L + x - 1)\right)^2} \times$$
$$\Big(2c^4 + 8c^3 p - c^2 p^2(\theta_L(3\theta_L + 2) + 3\theta_L x + x - 11) + cp^3(2\theta_L + x - 3)(\theta_L(\theta_L + x) - 2)$$
$$- (\theta_L - 1)p^4(\theta_L + x - 1)(\theta_L(\theta_L + x) - 1)\Big)$$

The denominator is positive. Thus, in order to show that the derivative is negative for all p, we need to show that the numerator is positive.

1. $(c+p)^{-(\theta_L + x + 1)} \geq 0$: This holds because $c, p \geq 0$.

2. $(c - \theta_L p + p)(c+p)^x - c + \theta_L p + px - p \geq 0$.

   Simplify the above expression to get

   $$(c - (\theta_L - 1)p)(c+p)^x - (c - (\theta_L - 1)p) + px = \quad ((c+p)^x - 1)\,(c - (\theta_L - 1)p) + px$$

   As $p(\theta_H) \leq p \leq p(\theta_L)$ and $p(\theta_L) = \frac{c}{\theta_L - 1}$, $c - (\theta_L - 1)p \geq 0$.

   If $((c+p)^x - 1) \geq 0$, we are done. So suppose $((c+p)^x - 1) < 0$. In this case, because $c - (\theta_L - 1)p$ is maximized at $p = p(\theta_H) = p(\theta_L + x) = \frac{c}{\theta_L + x - 1}$, we have

   $$\begin{aligned} &((c+p)^x - 1)\,(c - (\theta_L - 1)p) + px \\ \geq &\frac{cx\,((c+p)^x - 1)}{\theta_L + x - 1} + px. \end{aligned} \tag{38}$$

   The derivative of this expression with respect to p is given by

   $$\frac{x^2 c(c+p)^{x-1}}{\theta_L + x - 1} + x,$$

   which is positive. Thus, Equation (38) is minimized when p takes its minimum value at $p = p(\theta_H) = p(\theta_L + x)$. In this case, Equation (38) simplifies to

   $$\frac{cx\left(c\left(\frac{1}{\theta_L + x - 1} + 1\right)\right)^x}{\theta_L + x - 1} \geq 0.$$

Thus, the minimum value of $((c+p)^x - 1)(c - (\theta_L - 1)p) + px$ is at least zero, which completes this part of the proof.

3.

$$\Big(2c^4 + 8c^3p - c^2p^2(\theta_L(3\theta_L+2) + 3\theta_L x + x - 11) + cp^3(2\theta_L + x - 3)(\theta_L(\theta_L + x) - 2) \\ - (\theta_L - 1)p^4(\theta_L + x - 1)(\theta_L(\theta_L + x) - 1)\Big) \geq 0 \tag{39}$$

We show that the derivative of Equation (39) with respect to x is negative, which means that it takes its lowest value at the highest value of x, $x = \frac{1}{2}$. The derivative is given by

$$p^2\left(\theta_L^2 p(3(c+p) - 2px) + \theta_L(c+p)(2px - 3c) - (c+p)^2 - 2\theta_L^3 p^2\right).$$

Dividing this expression by p^2 does not change its sign. So we want to show that the following expression is negative

$$\theta_L^2 p(3(c+p) - 2px) + \theta_L(c+p)(2px - 3c) - (c+p)^2 - 2\theta_L^3 p^2 \leq 0. \tag{40}$$

For this, let us first evaluate Equation (40) at the minimum, i.e. $p(\theta_L + x)$. At $p(\theta_L + x)$, the expression is $-\frac{c^2(\theta_L+x)(2(\theta_L-1)\theta_L+(\theta_L+1)x)}{(\theta_L+x-1)^2}$, which is negative.

Next, consider the derivative of Equation (40) with respect to p. This derivative is given by

$$\begin{aligned} &c(\theta_L(3\theta_L + 2x - 3) - 2) - 2(\theta_L - 1)p(\theta_L(2\theta_L + 2x - 1) - 1) \\ \leq &c(\theta_L(3\theta_L + 2x - 3) - 2) - 2(\theta_L - 1)p(\theta_L + \frac{1}{2})(\theta_L(2\theta_L + 2x - 1) - 1) \\ = &-\frac{c\left(2\theta_L^3 + \theta_L + \theta_L^2(4x-3) - 6\theta_L x + 2\right)}{2\theta_L - 1}. \end{aligned}$$

where the inequality follows because the multiplier of p in the first expression is negative, so the expression takes its largest value at the lowest possible p, which is $p = p(\theta_L + x) \geq p(\theta_L + \frac{1}{2})$.

Consider the term in numerator in parenthesis and rewrite as

$$\left(2\theta_L^3 + \theta_L + \theta_L^2(4x-3) - 6\theta_L x + 2\right) = \left((2\theta_L^3 - 3\theta_L^2 + \theta_L + 2) + 2\theta_L x(2\theta_L - 3)\right).$$

We want to show that this term is positive when $\theta_L > 1$. Both $(2\theta_L^3 - 3\theta_L^2 + \theta_L + 2)$ and $\theta_L(2\theta_L - 3)$ have a positive derivatives when $\theta_L > 1$, so they both achieve their minimum value at $\theta_L = 1$. In particular, the minimum value of $\theta_L(2\theta_L - 3)$ is -1, thus $2\theta_L x(2\theta_L - 3)$ is minimized when $x = \frac{1}{2}$. Putting this together, the minimum value of the parenthesis in the numerator is $1 > 0$. As such, the lowest value of the derivative is

negative.

Next, evaluate Equation (39) at $x = \frac{1}{2}$ to get

$$\frac{1}{4}(c-(\theta_L-1)p)\left(8c^3+8c^2(\theta_L+3)p+2c\left(-2\theta_L^2+\theta_L+9\right)p^2+\left(4\theta_L^3-5\theta_L+2\right)p^3\right)$$

To show that the above expression is positive, it is sufficient to show that each of its two constituents are positive

(a) $(c-(\theta_L-1)p) \geq 0$: as we are only interested in $p(\theta_L+\frac{1}{2}) \leq p \leq p(\theta_L)$ and $p(\theta_L) = \frac{c}{\theta_L-1}$, this is true.

(b) $(8c^3+8c^2(\theta_L+3)p+2c\left(-2\theta_L^2+\theta_L+9\right)p^2+\left(4\theta_L^3-5\theta_L+2\right)p^3) \geq 0$
We will show that this expression is positive at minimum value of p, $p = p(\theta_L+\frac{1}{2})$, and that it is increasing in p.
First, the value of this expression at $p = p(\theta_H) = p(\theta_L+\frac{1}{2}) = \frac{2c}{2\theta_L-1}$ is given by

$$\frac{16c^3(2\theta_L+1)^2}{(2\theta_L-1)^2} \geq 0.$$

Next, take the derivative with respect to p and simplify to get

$$\begin{aligned}&8c^2(\theta_L+3)+4c\left(-2\theta_L^2+\theta_L+9\right)p+3\left(4\theta_L^3-5\theta_L+2\right)p^2\\ =&8c^2(\theta_L+3)+36cp+(2\theta_L-1)(3p^2\theta_L(2\theta_L+1)-4c\theta_Lp-6p^2).\end{aligned}$$

The first two terms are positive. So in order to show that the derivative is positive, we show

$$3(2\theta_L+1)\theta_Lp^2-4c\theta_Lp-6p^2 \geq 0.$$

Dividing by p

$$3(2\theta_L+1)\theta_Lp-4c\theta_L-6p = 6\left(\theta_L^2-1\right)p+\theta_L(3p-4c)$$

In order to show that the latter expression is positive, it is sufficient to show that it is positive at minimum value of p, $p = p(\theta_H) = p(\theta_L+\frac{1}{2}) = \frac{2c}{2\theta_L-1}$:

$$\begin{aligned}6\left(\theta_L^2-1\right)p+\theta_L(3p-4c) >&6\left(\theta_L^2-1\right)p(\theta_L+\frac{1}{2})+\theta_L(3p(\theta_L+\frac{1}{2})-4c)\\ =&2c\left(\theta_L+3(1-\frac{1}{2\theta_L-1})\right) \geq 0\end{aligned}$$

The last line is positive as $\theta_L > 1$, which implies $\frac{1}{2\theta_L-1} < 1$.

Necessity of $\theta_H \leq \theta_L + \frac{1}{2}$. For any pair of (θ_H, θ_L), derivative of the expression in Equation (1) at $p(\theta_H)$ is given by

$$-\frac{(2\theta_H - 2\theta_L + 1)(\theta_H - \theta_L)\left(\frac{c\theta_H}{\theta_H - 1}\right)^{1-\theta_H}}{2c\theta_H^2}$$

As $\theta_H > \theta_L > 1$, this expression is always negative. Furthermore, for (θ_H, θ_L), derivative of the expression in Equation (1) at $p(\theta_L)$ is given by

$$\frac{(2\theta_H - 2\theta_L - 1)(\theta_H - \theta_L)\left(\frac{c\theta_L}{\theta_L - 1}\right)^{1-\theta_H}}{2c\theta_L^2}$$

As $\theta_L > 1$, this expression is strictly increasing in θ_H and is zero if $\theta_H = \theta_L + \frac{1}{2}$, which completes the proof.

The concavity comparison of Aguirre et al. (2010). Aguirre et al. (2010) use a condition that is violated in our example. In particular, the condition requires that the demand curve with a lower monopoly price, θ_1, is less convex in the sense that

$$\frac{D_{pp}(p, \theta_1)}{D_p(p, \theta_1)} \geq \frac{D_{pp}(p, \theta_2)}{D_p(p, \theta_2)}.$$

In our example,

$$\begin{aligned} D_p(p, \theta_i) &= -\theta_i (c + p)^{-\theta_i - 1}, \\ D_{pp}(p, \theta_i) &= \theta_i(\theta_i + 1)(c + p)^{-\theta_i - 2}, \\ \frac{D_{pp}(p, \theta_1)}{D_p(p, \theta_1)} &= -\frac{1 + \theta_i}{c + p}. \end{aligned}$$

Therefore, their ranking requires that

$$\frac{1 + \theta_1}{c + p} \leq \frac{1 + \theta_2}{c + p},$$

which is violated because $\theta_1 > \theta_2$.