SCREENING AND SEGMENTING: A CONSUMER SURPLUS PERSPECTIVE

By

Dirk Bergemann, Tibor Heumann and Michael C. Wang

June 2026

COWLES FOUNDATION DISCUSSION PAPER NO. 2498R1

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Screening and Segmenting: A Consumer Surplus Perspective

Dirk Bergemann Tibor Heumann Michael C. Wang*

June 24, 2026

Abstract

We analyze consumer surplus when a monopolist can adjust both prices and product qualities across segments, engaging in second- and third-degree price discrimination simultaneously. We characterize the consumer-optimal segmentation and show that it has a striking structure: consumers with the same value receive the same quality in every segment, though prices differ. Under mild conditions, any segmentation harms consumers if and only if demand is sufficiently more elastic than supply. Hence, potential benefits for consumers depend critically on demand and supply elasticities. These findings have implications for regulatory policy regarding price discrimination and market segmentation.

JEL Classification: D42, D83, L12

Keywords: Price Discrimination, Nonlinear Pricing, Private Information, Second Degree Price Discrimination, Third Degree Price Discrimination, Bayesian Persuasion

---

*Bergemann: Department of Economics, Yale University (email: dirk.bergemann@yale.edu); Heumann: Instituto de Economía, Pontificia Universidad Católica de Chile (email: tibor.heumann@uc.cl); Wang: Department of Economics, Yale University (email: michael.wang.mcw75@yale.edu). The order of authors is alphabetical. An early version of this paper working in a more limited setting appeared under the title "A Unified Approach to Second and Third Degree Price Discrimination." We acknowledge financial support from NSF grants SES-2001208 and SES-2049744 and ANID Fondecyt Regular 1241302. We acknowledge many productive suggestions from Marzena Rostek (Co-Editor) and three anonymous referees. We have benefited from many conversations and related joint work with Ben Brooks and Stephen Morris. We thank Eugenio Miravete and Justin Johnson for many helpful suggestions. We thank David Wambach for excellent research assistance.

1 Introduction

1.1 Motivation and Results

Market segmentation and product screening are two of the most extensively studied forms of price discrimination, yet their interaction remains poorly understood. A firm that segments its market into distinct groups can tailor not only prices but entire product menus to each group, combining third-degree price discrimination—different pricing for different segments—with second-degree price discrimination—screening consumers through quality-differentiated menus within each segment. This combination is pervasive in practice: streaming services offer differentiated subscription tiers to behaviorally segmented audiences, airlines pair yield management across customer classes with fare-class menus within each class, and digital platforms simultaneously personalize prices and adjust product offerings based on consumer data. The central question of this paper is: when does such combined price discrimination benefit consumers?

The welfare implications of market segmentation have long interested economists. Third-degree price discrimination—charging different prices to distinct market segments—may either benefit or harm consumers. A similar ambiguity exists for second-degree price discrimination, where sellers screen consumers through quality-differentiated product menus. However, most research has analyzed these practices in isolation, leaving open the question of how they interact when deployed simultaneously, as is increasingly common in practice.

We provide a complete characterization of the consumer-optimal market segmentation when a monopolist sells goods of varying quality to privately informed buyers. In our model, the market is partitioned into arbitrary segments (or submarkets), each characterized by its own distribution of buyer values, subject only to the requirement that the segments aggregate to the original market. Taking this segmentation as given, the seller offers a profit-maximizing screening menu within each segment. We then study the segmentation that maximizes aggregate consumer surplus.

The consumer-optimal segmentation has a striking structure. Theorem 1 establishes that buyers with the same willingness to pay receive the same quality in every segment even though the monopolist could offer them different qualities in different segments. Prices, however, vary across segments: identical buyers may pay different amounts for the same product across segments. This result is obtained through a dramatic simplification of the market segmentation problem. The original optimization is over segmentations—distributions over distributions of values—an inherently infinite-dimensional and unwieldy object. We can reduce this to maximizing a single functional, the expected *local information rent*, over a single *inverse hazard rate function*, subject to a *majorization constraint*. The local information rent captures the contribution of buyers of value v to aggregate consumer surplus as a function of the inverse hazard rate at that value. The majorization constraint characterizes the precise limits on how segmentation can redistribute buyers across segments. This reduction is the paper's central methodological contribution, and it relies on a novel application of majorization techniques to the space of inverse hazard rates.

A second consequential finding concerns the conditions under which segmentation benefits consumers at all. Under a regularity condition on demand and supply, we identify a sharp threshold determined by demand and supply elasticities (Corollary 3). When aggregate demand is sufficiently elastic relative to the supply, no segmentation can improve consumer surplus. In such markets, the monopolist's screening already serves consumers well enough that any redistribution of buyers across segments only makes consumers worse off. When this condition fails, the consumer-optimal segmentation takes a particularly simple form (Theorem 2): it consists of "convex segments" with nested interval supports that shrink from below. Moreover, the optimal pricing differs only by a fixed fee, while the quality menu remains uniform across segments.

These results delineate precisely how the multi-product setting departs from the unit-demand environment of Bergemann et al. (2015). There, segmentation can always achieve efficiency; here, the scope for beneficial segmentation depends on a quantifiable interaction between the demand and cost primitives. Our results also clarify the relationship to Haghpanah and Siegel (2023), who showed that in generic markets with a finite number of goods, some segmentation always improves consumer surplus. We show that as the quality space becomes richer—approaching a continuum—the gains they identify can vanish entirely, and we provide the precise conditions under which this occurs.

These results have important implications for competition policy and regulation of price discrimination practices. They suggest that blanket restrictions on market segmentation may harm consumers by preventing welfare-enhancing price discrimination, while highlighting specific market conditions where segmentation is more likely to be beneficial. The findings also inform ongoing debates about big data and personalized pricing by showing how consumer heterogeneity and cost structures interact to determine the nature and effects of optimal segmentation strategies.

The paper is organized as follows. Section 2 introduces our model of nonlinear pricing with market segmentation. We consider a seller who can engage simultaneously in second- and third-degree price discrimination. Our model consists of a monopolist that offers goods of varying quality to a continuum of buyers. The willingness-to-pay (value) is private information to each buyer, and the seller only knows the distribution of values. The seller may segment the market into submarkets, each with its own distribution of values, subject only to the condition that the distribution of values across all submarkets must conform to the aggregate market. The monopolist offers an optimal pricing scheme in each submarket.

In Section 3, we consider a simple binary value environment, which illustrates the main concepts we will use throughout the paper. In the screening problem, the consumer surplus is the information rent. The contribution of each value to the consumer surplus is identified by the product of the inverse hazard rate and the allocation. We provide a convenient representation of this contribution as a function that depends only on the inverse hazard rate, we refer to this function as the *local information rent*. In the binary value environment, valuable segmentations have the effect of lowering the inverse hazard rate, by *concentrating* low-value buyers into a single segment, leaving them missing in the remaining segments. The maximum consumer surplus corresponds to the highest local information rent achievable with concentration.

Section 4 then extends this analysis to the general environment. With more than two values, the designer has a second tool alongside concentration: *dilution*, which spreads buyers of a given value across more segments and so *raises* their inverse hazard rate. Concentration and dilution are linked—diluting one value requires concentrating higher values—and our first result (Proposition 1) characterizes exactly what they can jointly achieve. The key condition turns out to be a familiar one: a *majorization* constraint on inverse hazard rates.

Next, we introduce a class of segmentations, called *uniform segmentations*, in which the inverse hazard rate of every buyer of the same value is equalized across segments. While a significant restriction, this class of segmentations is still quite rich. Nonetheless, we show that the optimal segmentation must lie within this class. Our main result, Theorem 1, fully characterizes the consumer surplus attained by the consumer-optimal segmentation. Theorem 1 shows that the value of segmenting is found by maximizing the expected local information rent over all inverse hazard rates satisfying the majorization constraint we found previously. Interestingly, while the original problem consists of a maximization over segmentations, which are distributions over distributions of values, Theorem 1 yields a maximization over a single distribution of values. Theorem 1 thus provides a remarkable simplification of the original problem.

In Section 5, we introduce a symmetric regularity condition—log-concavity—on demand and supply under which the consumer-optimal segmentation (Theorem 2) is easy to characterize. The resulting segmentation has the feature that the demand elasticity of low-value buyers is increased to a fixed level determined by the cost function, while the segments themselves have support shrinking from the bottom. We refer to these as *convex segmentations* because the constituent segments are characterized by their nested convex interval supports, and their construction is a generalization of the direct segmentations of Bergemann et al. (2015). The particularly simple characterization of the optimal segmentation under these conditions provides more insight into the conditions under which zero segmentation is optimal, i.e., any segmentation reduces consumer surplus (Corollary 3). When the model is specialized further to power cost functions, we uncover a tight connection between supply and demand elasticities in the optimal segmentation. Section 6 then discusses how our methodology can be generalized to other settings, including the case of discrete distributions, finding all Pareto-efficient outcomes, and an application to asset trading with adverse selection. Section 7 concludes. The appendix collects proofs omitted in the main text.

1.2 Related Literature

Our results are related to a large literature on price discrimination, beginning with Pigou (1920) and now encompassing a wide range of research on the output and welfare implications of price discrimination, such as Robinson (1933); Schmalensee (1981); Varian (1985); Aguirre et al. (2010); Cowan (2012) and Bergemann et al. (2015). We use the classic model of second-degree price discrimination via quality and quantity differentiated products first presented by Mussa and Rosen (1978) and Maskin and Riley (1984). Johnson and Myatt (2003) consider a model of second degree price discrimination under monopoly and duopoly and provide conditions under which the monopolist will offer a single good.

The problem of extending the results of Bergemann et al. (2015) to a multi-product setting was analyzed earlier by Haghpanah and Siegel (2022, 2023). Haghpanah and Siegel (2022) show that when the optimal menu in the aggregate market consists of a menu of more than one item, thus a screening menu, then the consumer surplus maximizing allocation cannot attain the Pareto frontier and hence the full surplus triangle cannot be obtained. Based on this insight, they provide a sufficient condition when the full surplus triangle will be attained, namely when all value distributions lead to the efficient single item menu. By contrast, we provide necessary and sufficient conditions for there to be a segmentation that improves consumer surplus and provide the value of the consumer surplus maximizing allocation, whether it is efficient or not. Haghpanah and Siegel (2023) show that in generic markets there is always a segmentation relative to the single aggregate market that improves consumer surplus, however they do not provide properties of the consumer-optimal segmentation. In contrast to Haghpanah and Siegel (2022, 2023), who work with a finite number of products, we allow for (but do not require) a continuum of qualities. Therefore, the results in Haghpanah and Siegel (2023) do not generally apply to our setting, and we find a large class of markets in which zero segmentation is optimal for consumers, in addition to completely solving for the consumer-optimal segmentation. We view our work as complementary to

Haghpanah and Siegel (2023) in the sense that, with a continuum of qualities, the improvement they identify sometimes becomes negligible, and we provide precise conditions under which this does (or does not) occur.

Finally, the representation of consumer surplus through the inverse hazard rate, central to our analysis, has not played this role in the earlier nonlinear-pricing literature. The inverse hazard rate is of course not new: it converts values into virtual values in Mussa and Rosen (1978), Maskin and Riley (1984), and Myerson (1981). There, however, it is an exogenous summary of the value distribution, used to characterize the seller's optimal menu. The *joint* segmentation-and-screening problem promotes it from a primitive to an instrument: segmentation chooses the within-segment distribution, and hence the local inverse hazard rate subject to the majorization constraint, while screening makes the allocation respond to it. The inverse hazard rate then carries both the extensive margin (the mass of higher values who earn rent) and the intensive margin (the allocative distortion through the virtual value). This dual role requires a continuum of qualities together with endogenous segmentation, a combination absent from both the single-market screening literature and the finite-good analyses of Haghpanah and Siegel (2022, 2023).

2 Model

Payoffs and Menu Pricing A monopolist produces vertically differentiated products of varying quality $q \in \mathbb{R}_+$ at cost $c(q)$, where $c : \mathbb{R}_+ \to \mathbb{R}_+$ is a thrice differentiable, strictly increasing, and strictly convex function with $c'(0) = 0$. There is a continuum of buyers, each characterized by a value $v \in V = [\underline{v}, \overline{v}] \subset \mathbb{R}_+$ that is privately known. A buyer with value (type) v who purchases quality q at price p obtains net utility $vq - p$.

The monopolist posts a menu of prices $p : \mathbb{R}_+ \to \mathbb{R}_+$, or *tariff*, which specifies a price $p(q)$ for each quality level q. By default, every menu includes the option to buy zero quality at zero price, or $p(0) = 0$. (We use p both for the function (menu) and for a particular item price p. The context will identify the relevant interpretation.) Given a menu p, a buyer chooses the quality that maximizes net utility:

$$U(v, p) \equiv \max_{q \in \mathbb{R}_+} \left[vq - p(q)\right],$$

and the corresponding quality choice is denoted by

$$q(v) \equiv \arg\max_{q \in \mathbb{R}_+} \left[vq - p(q)\right].$$

If multiple qualities maximize the utility of the buyer, ties are broken in favor of the seller. The profit of the seller from a buyer with value v when offering menu p is:

$$\Pi(v, p) \equiv p(q(v)) - c(q(v)).$$

Market A *market* $m \in \Delta V$ is a probability distribution over values V. F_m denotes the cumulative distribution function (cdf) associated with m (right-continuous by convention), $f_m(v)$ the density where F_m is absolutely continuous, and $a_m(v)$ the mass of any atom at v.

In a given market m, the profit-maximizing menu p_m solves:

$$p_m \equiv \arg\max_{p(q)} \int_{\underline{v}}^{\overline{v}} \Pi(v, p)\; dF_m(v).$$

If there are multiple optimal price menus, we select the one that results in the highest consumer surplus. We pose the problem for the seller directly in terms of a menu (or tariff), i.e. as an indirect rather than a direct revelation mechanism. By the taxation principle of Guesnerie and Laffont (1984), these two approaches are equivalent; our arguments are more transparent when stated in the quantity-price space.

We denote by $q_m(v)$ the optimal choice of value v in market m when the menu of prices is p_m. The aggregate consumer surplus in market m is given by:

$$U(m) \equiv \int_{\underline{v}}^{\overline{v}} U(v, p_m)\; dF_m(v). \tag{1}$$

Segmentations We identify the *aggregate market* with superscript "$*$":

$$m^* \in \Delta(V),$$

which is the distribution of values of all buyers present in the economy. And all the associated quantities associated to the aggregate market are identified with a superscript "$*$". For example, in the aggregate market, F^* is the cdf and p^* is the optimal menu of prices. We assume that the aggregate market is absolutely continuous with respect to the Lebesgue measure with strictly positive density on $[\underline{v}, \overline{v}]$ and that F^* is a real-analytic function. Section 6.3 explains how our results change when F^* is discrete.

A *segmentation* σ is a distribution over markets $\sigma \in \Delta(\Delta V)$ such that the composition across market segments equals the aggregate market m^*. Formally, endow ΔV with the weak topology, and let σ be a Borel probability measure on ΔV such that

$$\int_{m \in \Delta V} m \; d\sigma(m) = m^*. \tag{2}$$

We write $\mathrm{MPS}(m^*)$ to denote the set of segmentations which satisfy the aggregation constraint (2), that is, are mean-preserving spreads of m^* in $\Delta(\Delta V)$.

Under a segmentation σ, the monopolist observes each buyer's segment but not their value, and posts an optimal menu independently within each segment or submarket. The seller thus engages simultaneously in second-degree price discrimination (screening within each segment) and third-degree price discrimination (across segments).

Consumer Surplus Maximization Problem Our central question is: which segmentation σ of m^* generates the highest aggregate consumer surplus?

$$\max_{\sigma \in \mathrm{MPS}(m^*)} \left[\int_{m \in \Delta V} U(m) \; d\sigma(m) \right]. \tag{3}$$

We are interested in the value of this maximization problem, as well as the segmentation σ that attains the maximum. Two features of this formulation deserve emphasis. First, the seller is unconstrained in how they price within each segment. Second, segmentations are unrestricted: any partition of the aggregate market is permissible, including those with gaps and atoms in the constituent submarkets, so long as the aggregation condition (2) is satisfied.

3 The Binary Value Case

This section develops the key economic ideas in a simple environment with two possible values for the buyers.[1] The binary setting is rich enough to illustrate the central trade-off underlying our results—the tension between the *extensive* and *intensive margins* of the information rents—and to introduce the main analytical tool, the *local information rent function*, which will carry over essentially unchanged to the general analysis. At the same time, the setting is simple enough that all the economic forces are transparent.

Screening with Binary Values Suppose buyer values v_i are drawn from $\{v_1, v_2\}$ with $0 < v_1 < v_2 < \infty$. We fix an aggregate market m^* where the low value v_1 arises with

[1] This two-value example is purely illustrative and outside the continuum model of Section 2, as a discrete distribution necessarily violates the assumption that F^* be absolutely continuous with full support.

probability $a(v_1)$ and the high value v_2 with complementary probability $a(v_2) = 1 - a(v_1)$. We initally just consider the aggregate market and hence can omit the index m^* or *.

Consider a profit-maximizing seller who offers a menu $((q_1, p_1), (q_2, p_2))$ of quality–price pairs. As is standard in the screening literature, the binding constraints are participation for the low value and incentive compatibility for the high value, which pin down optimal prices as a function of qualities:

$$p_1 = v_1 q_1, \quad p_2 = v_1 q_1 + v_2(q_2 - q_1).$$

Hence, the quality upgrade (or increment) $q_2 - q_1$ is priced at v_2 but the baseline quality q_1 is priced at v_1 (we used that incentive compatibility requires that $q_1 < q_2$). Consequently, only high-value buyers get any information rent and the consumer surplus U is given by the expected information rents:

$$U = a(v_2)(v_2 - v_1)q_1.$$

Consumer surplus depends on the quality offered to *low-value* buyers, even though it is *high-value* buyers who capture the rents. This is because the rents arise from the high value's ability to mimic the low value: the better the low value's product, the larger the rent the high value extracts by threatening to "trade down." The qualities supplied are of course chosen by the seller to maximize profits. Hence, the seller chooses qualities (q_1, q_2) that solve:

$$\max_{(q_1,q_2)\in\mathbb{R}_+^2} \left\{a(v_1)(v_1 q_1 - c(q_1)) + a(v_2)(v_2 q_2 - c(q_2)) - a(v_2)(v_2 - v_1)q_1\right\}.$$

The first two terms are the social surplus generated by qualities (q_1, q_2), while the last term is the information rents earned by buyers. The quality q_2 offered to a buyer with high value v_2 solves the corresponding first order condition:

$$v_2 = c'(q_2).$$

As usual, the highest value does not see any distortion in its allocation. The quality q_1 offered to a buyer with low value v_1 solves:

$$\phi(v_1) \equiv v_1 - (v_2 - v_1)\frac{a(v_2)}{a(v_1)} \leq c'(q_1), \tag{4}$$

where the above expression holds with equality if the *virtual value* $\phi(v_1)$ is positive. Note that the virtual value depends on the value distribution. To write this condition explicitly, we denote the inverse of the marginal cost $c'(q)$ by Q:

$$Q(v) \equiv \mathbb{1}[v \geq 0](c')^{-1}(v). \tag{5}$$

We refer to Q as the *supply function*, since $Q(v)$ is the quality that the monopolist would sell to a buyer of value v if the monopolist were to offer a socially efficient pricing scheme: $p'(q) = c'(q)$. The supply function is 0 if the value v (or virtual value) is negative, captured by the indicator function $\mathbb{1}[v \geq 0]$. We can write (4) using the supply Q by replacing the value v_1 with the virtual value $\phi(v_1)$:

$$q(v_1) = Q\left(v_1 - (v_2 - v_1)\frac{a(v_2)}{a(v_1)}\right) = Q(\phi(v_1)). \tag{6}$$

The Extensive vs. Intensive Margin Trade-Off The consumer surplus as a function of the distribution in aggregate market is thus

$$U \equiv a(v_2)(v_2 - v_1)q(v_1), \tag{7}$$

and reveals a fundamental tension in how the market composition affects consumer surplus. Consider what happens as the fraction of low-value buyers $a(v_1)$ varies. Two forces work in opposite directions.

When there are more low-value buyers, the seller has a stronger incentive to serve them well. A larger $a(v_1)$ means a larger customer base whose quality the seller would sacrifice to improve profits. Hence, the quality q_1 offered to the low value rises with $a(v_1)$, and so does the per-buyer rent $(v_2 - v_1)q_1$ enjoyed by each high-value buyer, the *intensive margin* of the rent. But, as $a(v_1)$ rises, the fraction of high-value buyers—the *extensive margin* of the information rent—$a(v_2) = 1 - a(v_1)$ falls. Fewer buyers are collecting the rents. Eventually, this force dominates, and consumer surplus declines.

Figure 1a illustrates this trade-off for a quadratic cost $c(q) = q^2/2$ with $V = \{1, 2\}$. We display the quantity $q(v_1)$ delivered to the low-value buyers and the resulting expected consumer surplus U as a function of the probability $a(v_2)$ of high values. Observe that when $a(v_2)$ is too high, the seller offers nothing to the low-value buyers, and hence the consumer surplus is also zero. As we decrease $a(v_2)$, $q(v_1)$ increases and hence so does U. But, when $a(v_2)$ gets too low, there are too few high values to benefit from the information rents generated by $q(v_1)$, and U eventually goes down again.

Binary Segmentation We now consider segmentations of the aggregate market m^*. We focus on a natural class of binary segmentations $\sigma \in \Delta\{m, m'\}$ in which low-value buyers are absent in market m':

$$a_m(v_1) > 0 \text{ and } a_{m'}(v_1) = 0.$$

Segmentation as Concentration Since low-value buyers appear only in market m, their concentration must satisfy

$$a_m(v_1) \geq a^*(v_1) \text{ and } \sigma(m) = \frac{a^*(v_1)}{a_m(v_1)}, \tag{8}$$

and with equality when all high-value buyers are in market m (no concentration) and $a_m(v_1) \to 1$ representing maximal concentration. The aggregation constraint for the high-value buyers is then

$$\sigma(m)a_m(v_2) + \sigma(m')a_{m'}(v_2) = a^*(v_2).$$

In market m', where only high-value buyers are present, the seller charges the gross utility, so no information rents arise. Hence, consumer surplus will be generated only in market m:

$$\sum_{x\in\{m,m'\}} \sigma(x)U(x) = \sigma(m)(v_2 - v_1)a_m(v_2)q_m(v_1).$$

The effect of concentration is intuitive. By gathering low-value buyers into a single segment, the designer *raises* their prevalence in this segment, which increases the quality the seller offers them and hence the information rents. The cost is that some high-value buyers are stranded in the other segment, where they earn zero rents. The question is whether the deeper rents in m compensate for the narrower base of rent-earning buyers.

The Inverse Hazard Rate To analyze this trade-off more cleanly, observe that the market composition affects consumer surplus entirely through a single statistic: the inverse hazard rate of the low value,

$$h_m(v_1) \equiv (v_2 - v_1)\frac{a_m(v_2)}{a_m(v_1)}.$$

This quantity has a direct economic interpretation: it is the mass of high-value buyers *per unit of low-value buyers*, scaled by the difference in values $v_2 - v_1$. The inverse hazard rate captures both the intensive margin of distortion (through the virtual value):

$$\phi_m(v_1) \equiv v_1 - h_m\left(v_1\right),$$

and the extensive margin of rent-earning buyers (since it is proportional to the high-to-low ratio). Concentration of low-value buyers *lowers* the inverse hazard rate: making the low value more prevalent reduces the high-to-low ratio. This is the only tool available in the binary setting: the segmentation designer can decrease $h_m(v_1)$ relative to $h^*(v_1)$, but never increase it.

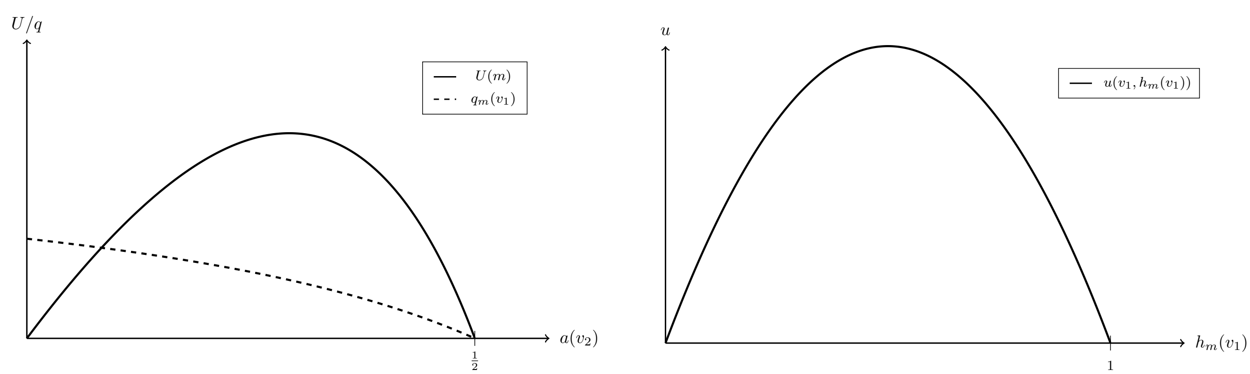
The Local Information Rent We can now express the consumer surplus Ucompactly using the *local information rent* function:

$$u(v, h) \equiv h \cdot Q(v - h). \tag{9}$$

We refer to $u(v, h)$ as the local information rent. Observe that h plays two roles, both captured by the product $h \cdot Q(v-h)$. First, h pins down the allocation to value v: the seller serves v as if its value were the virtual value $v - h$, so it receives quality $Q(v - h)$, decreasing in h. Second, the leading h is proportional to the (scaled) mass of higher-value buyers who collect a rent off that allocation, per unit of value-v buyers. Raising h thus widens the base of rent-earning buyers while shrinking the allocation they earn rent on—the extensive-intensive margin trade-off, compressed into the single scalar h. For the purpose of this example, we will fix $v = v_1$, and so the relevant variable is the inverse hazard rate $h = h(v_1)$. Using the newly introduced notation, we have that total consumer surplus is:

$$\sum_{x \in \{m, m'\}} \sigma(x) U(x) = a^*(v_1) u(v_1, h_m(v_1)).$$

Figure 1b plots the local information rent function in the same example, where costs are quadratic and $V = \{1, 2\}$.



(a) U and $q_m(v_1)$ as a function of $a(v_2)$. (b) $u(v_1, h)$ as a function of $h_m(v_1)$.

Figure 1: Two ways of finding consumer surplus with $V = \{1, 2\}$ and $c(q) = q^2/2$.

Now the only constraint on the segmentation is given by the constraint (8). Namely, the segmentation will increase the concentration of low-value buyers in market m relative to the aggregate market. Equivalently, we have that:

$$h_m(v_1) \leq h^*(v_1). \tag{10}$$

The maximum consumer surplus that can be attained with this binary segmentation is thus:

$$\max_{h \leq h^*(v_1)} \left[a^*(v_1)u(v_1, h)\right]. \tag{11}$$

The local information rent $u(v, h)$ encodes the extensive-intensive margin trade-off in a single function. At $h = 0$, there are no high-value buyers, so no rents: $u(v, 0) = 0$. As h increases, more high-value buyers benefit from the low value's allocation, raising total rents. But simultaneously, the virtual value $v - h$ falls, so the seller distorts quality downward more aggressively. When $h \to v$, the distortion is so severe that the seller excludes the low value entirely: $Q(v - h) = 0$, and rents vanish. Between these extremes, $u(v, h)$ is hump-shaped, attaining its maximum at an interior value

$$\overline{h}(v) \equiv \arg\max_{h} \big[u(v, h)\big].$$

If the information rents are quasi-concave, the solution to the maximization problem (11) is:

$$h = \begin{cases} h^*(v_1), & \text{if } h^*(v_1) \leq \overline{h}(v_1); \\ \overline{h}(v_1), & \text{if } h^*(v_1) > \overline{h}(v_1). \end{cases}$$

In other words, if the mass of high-value consumers is too low, then the optimal segmentation is zero segmentation; if the mass of high-value consumers is too high, then some of them are separated into another market.

Preview of the General Analysis The analysis of this section is incomplete, as we did not show that the restricted binary segmentations are optimal (with binary values, they are). However, even as we move to continuum values and consider arbitrary segmentations, the expressions for the consumer surplus in terms of expected local information rents will remain essentially the same.

The binary setting illustrates the core mechanics, but it is special in two respects. First, the designer can only *concentrate* buyer values (lower h), never *dilute* them (raise h). With a continuum of values, dilution becomes possible by creating gaps in the support of some segments, but it requires concentrating adjacent values first to create the gaps. This interplay between concentration and dilution is governed by a *majorization constraint* that generalizes the pointwise constraint $h_m(v_1) \leq h^*(v_1)$ in (10). Second, with binary values there is only a single inverse hazard rate to optimize. With a continuum of values, the designer must choose an entire function $h(v)$, balancing rents across all values simultaneously. Despite these added dimensions, the structure of the solution is remarkably parallel: consumer surplus is still the expected local information rent, and the consumer-optimal segmentation is still found by maximizing $u(v, h(v))$ over feasible inverse hazard rates.

4 Consumer-Optimal Segmentations

We now turn to the general environment with a continuum of buyer values and unrestricted segmentations. Our goal is to show that the problem of finding a consumer-optimal segmentation—which is, on its face, an optimization over distributions of distributions—can be reduced to a far simpler problem: maximizing a single functional over a single function, subject to a family of linear constraints. We build toward the main result in four steps.

First, we show that the local information rent representation of consumer surplus from Section 3 extends to the general setting, so that consumer surplus in any regular market equals the expected local information rent (Section 4.1). Second, we characterize the limits of what segmentation can achieve by identifying the precise constraints on inverse hazard rates that any segmentation must satisfy (Section 4.2). The key insight is that the feasible inverse hazard rates are governed by a *majorization constraint*—a generalization of the pointwise constraint $h \leq h^*$ that arose in the binary case. Third, we introduce a class of segmentations, called *uniform segmentations*, in which every buyer of the same value faces the same inverse hazard rate in every segment where they appear, and we argue that it is sufficient to restrict attention to this class (Section 4.3). Finally, we combine these ingredients to state and prove our main result (Section 4.4). All the illustrations in this section correspond to the optimal segmentation when a seller has two indivisible goods with different marginal costs; here, we use them only to illustrate the concepts and results, rather than provide a detailed analysis.

4.1 From Binary to Continuum: Consumer Surplus as Expected Local Information Rent

In the binary-value setting, we showed that consumer surplus is represented by the local information rent which depends on the inverse hazard rate. This representation extends naturally to a continuum of values. The key step is defining the inverse hazard rate for general distributions that may contain both continuous and discrete components, density and atoms, and the support of a market may have gaps.[2]

For any market m and any value v in the support of m, define the *increment* at v, denoted by $\Delta_m(v)$, as the distance between v and the next highest value present in m:

$$\Delta_m(v) \equiv \inf \big\{t \in \mathrm{supp}(m) \mid t > v\big\} - v.$$

We define the inverse hazard rate in segment m as:

$$h_m(v) \equiv \begin{cases} \frac{1-F_m(v)}{f_m(v)}, & \text{if } a_m(v) = 0 \text{ and } 0 < f_m(v) < \infty; \\ \Delta_m(v)\frac{1-F_m(v)}{a_m(v)}, & \text{if } a_m(v) > 0; \\ 0, & \text{if } a_m(v) = f_m(v) = 0. \end{cases} \tag{12}$$

The definition of the inverse hazard rate has three components. The first one is when the distribution is continuous at some v, the second one is when there is an atom at v, and the third one is when v is not in support of m, and hence there is a gap around v. We define the inverse hazard rate to be 0 if $v \notin \mathrm{supp}(m)$ because when v is not present in market m, it does not generate information rents.[3] The economic content of this definition is the same as in the binary case: the inverse hazard rate measures the mass of buyers above v per unit of value v, scaled by the local spacing of the support. When m is continuous, it reduces to the familiar ratio of the survival function to the density.

For any market m and any $v \in \mathrm{supp}(m)$, we define the *virtual value* as:

$$\phi_m(v) \equiv v - h_m(v). \tag{13}$$

This definition incorporates both the case of continuous and discrete distributions. As in Section 3, we denote by Q the inverse of the marginal cost function (see (5)). A market is regular if for every v, the quality that the buyer with value v consumes $q_m(v)$ under the profit-maximizing menu is the supply evaluated at the corresponding virtual value:

$$q_m(v) = Q(\phi_m(v)).$$

We can restrict attention to segmentations that place positive weight only on regular markets.

Two implications of regularity recur below. (i) In a regular market, the profit-maximizing menu involves no ironing: value v consumes exactly $q_m(v) = Q(\phi_m(v))$, which is what makes

[2] It is without loss to restrict to distributions of this form, as markets with a singular-continuous component are irregular; see proof of Lemma 1.

[3] We could have replaced the value of $h_m(v)$ with any finite value when $v \notin \mathrm{supp}(m)$ because it will be integrated over a null density, but we want to keep it finite to avoid multiplying infinity times zero.

the local information rent representation of Lemma 2 available. When m has full support, regularity is equivalent to the virtual value ϕ_m being non-decreasing. (ii) Regularity requires an atom at the lower endpoint of any gap in the support. Roughly, if instead the density continued up to the gap, the quality allocation would jump downward at the top of that density, violating monotonicity.

Lemma 1 (Consumer Surplus in Irregular Markets)
For every market m', there exists a segmentation σ of m' into regular markets such that

$$U(m') = \int_{\Delta V} U(m)\ d\sigma(m). \tag{14}$$

This lemma shows that any irregular market can be segmented into regular markets while preserving the same profit-maximizing pricing. Our notion of regularity requires that the ironed virtual values agree with the original ones (details in proof of Lemma 1). When m is full support, it reduces to having monotone virtual values. Hence, this statement uses no properties of how consumer surplus is determined, but simply shows that one can segment irregular markets into regular ones in an inconsequential way. Thus, for the remainder of the section, we will only consider segmentations supported on regular markets, which is sufficient to characterize the maximum consumer surplus achievable with segmentations. It will turn out that (14) can be strengthened to a strict inequality, that is, segmentations supported on irregular markets are strictly suboptimal; we relegate the details of this argument to the proof of Theorem 1 in the appendix.

The *local information rent* is defined the same way as (9), which we display again:

$$u(v, h) \equiv h \cdot Q(v - h).$$

As in the binary value case, we can express the consumer surplus in regular markets as the expected local information rent via the envelope theorem as it is standard.

Lemma 2 (Consumer Surplus in Regular Markets)
In every regular market m, the consumer surplus is given by:

$$U(m) = \int_{\underline{v}}^{\overline{v}} u(v, h_m(v))\ dF_m(v). \tag{15}$$

The local information rent $u(v, h)$ captures the total rents generated by value v per unit mass of buyers, so when multiplied by the density of buyers, it gives the total consumer surplus generated by v. The two roles of the inverse hazard rate identified in the binary case carry over unchanged: the quality consumed by value v is $q_m(v) = Q(v - h_m(v))$, which depends on the market only through $h_m(v)$, while $h_m(v)$ is itself the mass of higher values benefiting from that allocation per unit of value v. This is why the entire segmentation problem can be recast in the single function h: fixing the inverse hazard rate at each value determines both the allocation and the mass benefiting from it, and hence consumer surplus. We emphasize that the local information rents do not identify which values are *receiving* the rents, but rather which values are *generating* the rents. Value v generates rent for higher values because the seller offers a quality level at price v, which in turn generates rent for all higher values that can purchase this quality level at a price below their own valuation. Hence, throughout this section, we consider the following problem:

$$\begin{aligned}\max_{\sigma \in \mathrm{MPS}(m^*)} \int_{\Delta V} \int_{\underline{v}}^{\overline{v}} u(v, h_m(v))\ dF_m(v)\ d\sigma(m) \\ \text{subject to: every } m \in \mathrm{supp}(\sigma) \text{ is regular,}\end{aligned} \tag{16}$$

which delivers the same value as (3).

4.2 What Can Segmentation Achieve? Concentration, Dilution and Majorization

Our expression for consumer surplus in regular markets (15) invites us to consider how we might change the inverse hazard rate h to increase local information rents. In general, higher h means more high-value buyers benefit because it raises the mass of buyers above v per unit of value v (as measured by the inverse hazard rate). Simultaneously, however, it means lower quality for value v, so distortions and prices increase with the inverse hazard rate. These opposing forces imply that local information rents are maximized at an interior inverse hazard rate $\overline{h}(v)$, defined by

$$\overline{h}(v) \equiv \arg\max_{h \in \mathbb{R}_+} \big[u(v, h)\big]. \tag{17}$$

If multiple maximums exist, we take the lowest one.

Two Tools: Concentration and Dilution The binary-value analysis identified a single tool available to the segmentation designer: *concentration*, gathering buyers of a given value into one segment, thereby lowering their inverse hazard rate. With more values, a second tool becomes available: *dilution*, raising the inverse hazard rate of a given value by placing that value in segments where immediately higher values are *missing*. When a given value is spread out over more markets than the immediately higher values, a gap opens in the support above value v, which increases the increment $\Delta_m(v)$ and hence the inverse hazard rate (12).

Dilution is economically quite different from concentration. Concentrating a value is straightforward: place more of those buyers in some segments and remove them from others. But diluting a value—raising its inverse hazard rate—requires that some values just above v have been *removed* from the segment. And those values can only be removed if they have been concentrated elsewhere. In other words, dilution at one value requires concentration at adjacent higher values.

The Majorization Constraint For this reason, concentration is valuable both as a tool in its own right to modify the inverse hazard rate, and to enable dilution of lower values. Our key challenge, then, is to characterize the limits of what inverse hazard rates are achievable through segmentation. For this purpose, let $h : V \to \mathbb{R}_+$ be an arbitrary function. We say h^* *majorizes* h and write $h^* \succ h$ if:

$$\int_v^{\overline{v}} h^*(t)dF^*(t) \geq \int_v^{\overline{v}} h(t)dF^*(t), \tag{18}$$

for all $v \in V$. We refer to (18) as the *majorization* constraint, as it is closely related to the familiar notion of majorization (with the distinction that h here need not be monotone in v). We define the slack in the majorization constraint at v as $E_h(v)$:

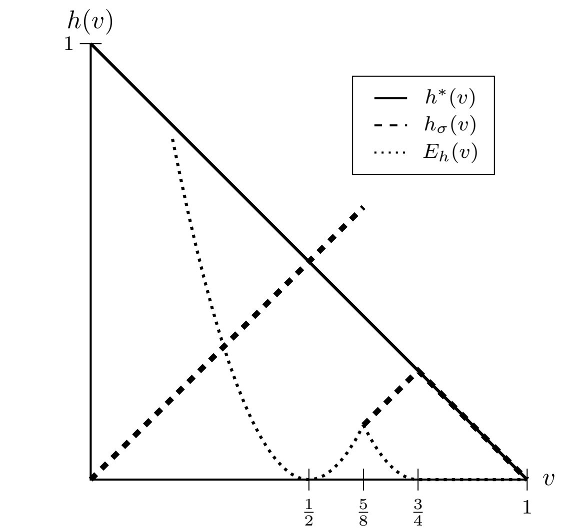
$$E_h(v) \equiv \int_v^{\overline{v}} h^*(t) - h(t)dF^*(t) \geq 0. \tag{19}$$

In Figure 2a we plot the inverse hazard rate of an aggregate market and an inverse hazard rate that satisfies the majorization constraint (labeled as h_σ). We also plot the slack in the majorization constraint E and observe that its slope changes sign depending on whether the hazard rate is above or below the aggregate market.

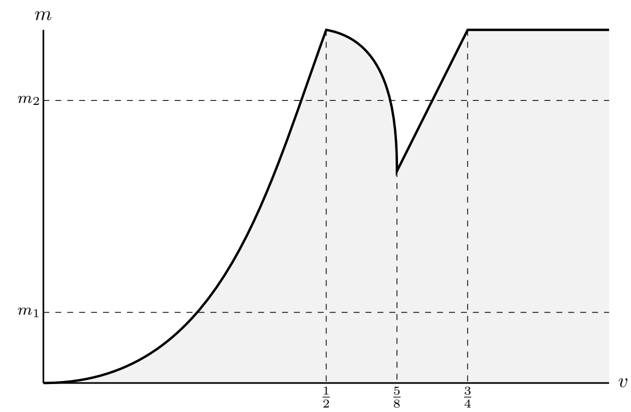
For any segmentation σ, let $\Sigma(v, m)$ be the induced joint distribution over values and markets, and $\Sigma_v(m)$ the marginal distribution of Σ over markets, conditional on value. Define the average inverse hazard rate h_σ as:

$$h_\sigma(v) \equiv \int_{\Delta V} h_m(v) \; d\Sigma_v(m). \tag{20}$$

This is, in general, not the same as the average value of $h_m(v)$ over $\sigma(m)$, unless $h_m(v)$ is constant across all $m \in \mathrm{supp}(\sigma)$. If F_m were continuous for all $m \in \mathrm{supp}(\sigma)$, this would



(a) Inverse hazard rate satisfying majorization constraint, and that of the aggregate market.



(b) Support of a segmentation implementing h_σ shown in Figure 2a.

Figure 2: Inverse hazard rates and a segmentation that implements it.

simplify to the weighted average value of $h_m(v)$,

$$h_\sigma(v) = \int_{\Delta V} h_m(v) \frac{f_m(v)}{f^*(v)} \, d\sigma(m),$$

where more weight is given to markets that have more buyers of value v. The integral expression (20) extends this to the case where σ contains distributions with atoms.

This notion of average hazard rate formalizes our intuition that diluting buyers of value v requires gaps to have been created above v. Specifically, for all v,

$$E_h(v) = \int_{\{m : v \notin \operatorname{supp}(m)\}} \Delta_m(v)(1 - F_m(v)) \, d\sigma(m). \tag{21}$$

The right-hand side (rhs) integrates the "missing demand" in each market where v is missing (that is, $v \notin \operatorname{supp}(m)$), weighted by the distance to the nearest value in the support. Since it is weakly positive by construction, we get that the majorization constraint (19) is satisfied.

Proposition 1 (Inverse Hazard Rates in Segmentations)
For any segmentation σ of m^*, $h_\sigma \prec h^*$.

In the next section, we will use a partial converse of this result: given $h \prec h^*$, we can construct a segmentation σ such that $h_\sigma = h$, provided that h satisfies certain technical conditions. Thus, in a strong sense, the majorization constraint captures the maximal extent to which the redistribution of buyers can affect the inverse hazard rates, which is itself sufficient for capturing the outcomes of segmentation.

The contrast with the binary case is illuminating. There, the only feasible change was $h \leq h^*$ pointwise; the designer could only concentrate (lower h), never dilute (raise h). With a continuum of values, the majorization constraint is *weaker* than the pointwise constraint: it permits the inverse hazard rate to *exceed* h^* at some values, provided it falls sufficiently below h^* at other (higher) values to maintain the cumulative inequality. This additional flexibility is precisely what dilution provides, and it can strictly expand the achievable consumer surplus.

4.3 Uniform Segmentations: Equalizing Returns across Segments

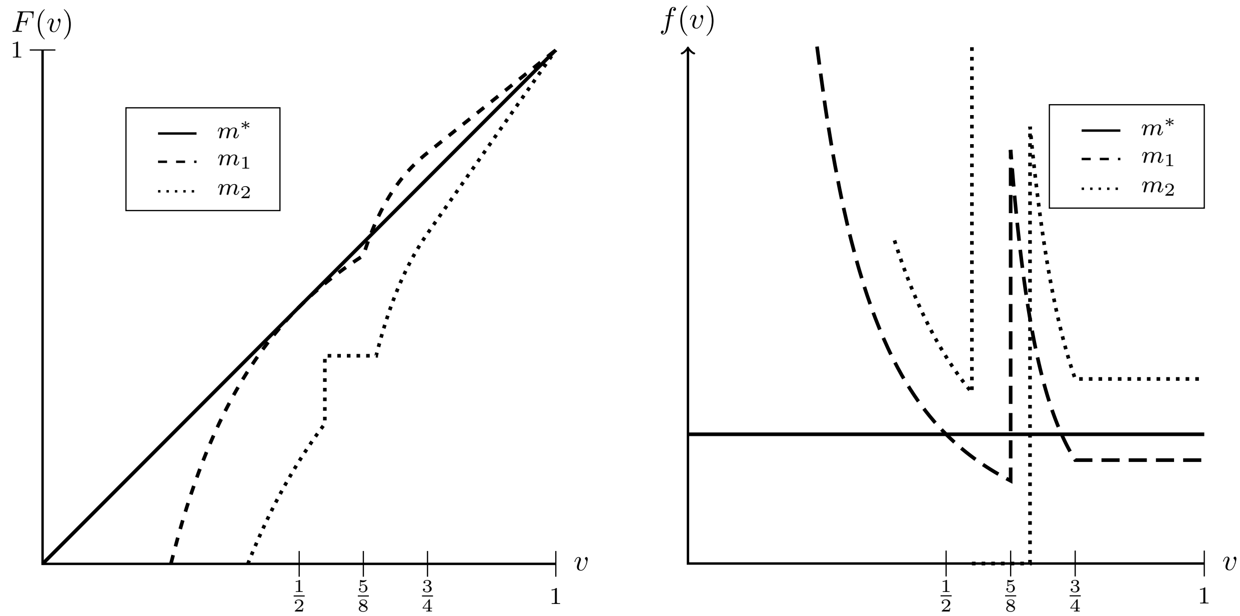
The local information rent at value v describes the contribution of every unit of buyers of value v to the consumer surplus in a given segment. If the same value makes different contributions in different segments, then we might want to redistribute buyers of this value to wherever they can make the largest contribution. Thus, we should expect that the optimal segmentation equalizes the returns of given value v across all segments. But, this suggests that all segments should have a uniform inverse hazard rate for a given value v, meaning every value generates the same inverse hazard rate in every segment that the value is present.

Formally, we say that σ is a *uniform segmentation* if, for almost every $m \in \text{supp}(\sigma)$ and $v \in \text{supp}(m)$,

$$h_m(v) = h_\sigma(v). \tag{22}$$

Segments within a uniform segmentation share a common shape along their shared supports that may differ from the aggregate market. In Figure 3a, we plot two distributions that have the same inverse hazard rates (which is the one in Figure 2a). The aggregate market F^* is drawn only for reference. Segments m_1 and m_2 are two members of the continuum making up the segmentation, so they need not—and in general do not—average to F^*; only the full collection does. Figure 2b shows how m_1 and m_2 sit among all the segments. Market m_1 is a gapless distribution, while market m_2 has a gap (flat segment) and an atom (discontinuous jump) that are in exact proportion so that the discrete inverse hazard rate is the same as market m_1. Even though both distributions have the same inverse hazard rates (over their overlapping support) they do look quite different, illustrating the richness that exists even among uniform segmentations. In Figure 3b, we plot the corresponding densities, where we again observe similar shapes in their supports.

A uniform segmentation σ will consist only of regular markets if and only if $v - h_\sigma(v)$ is non-decreasing. Every uniform segmentation σ that consists only of regular markets



(a) Cdf of aggregate market and segments. (b) Densities of the same markets.

Figure 3: Two segments (out of a continuum) in a uniform segmentation of $\mathcal{U}[0,1]$. The aggregate market is shown for reference only; the two segments alone do not average to it.

generates consumer surplus

$$\int_{\Delta V} \int_{\underline{v}}^{\overline{v}} u(v, h_m(v))\ dF_m(v)\ d\sigma(m) = \int_{\underline{v}}^{\overline{v}} u(v, h_\sigma(v))\ dF^*(v),$$

where here we are using expression (15) to express the consumer surplus as the expected local information rents and (22) to replace the hazard rates. We can maximize this expression over $h_\sigma \prec h^*$ to obtain an upper bound on how much consumer surplus a uniform segmentation supported only on regular markets can generate:

$$\sup_{h \prec h^*} \int_{\underline{v}}^{\overline{v}} u(v, h(v))\ dF^*(v) \tag{23}$$

In order for this upper bound to be tight, there must be an h which achieves the maximum of (23), and a uniform segmentation implementing this h which is supported only on regular markets. All these conditions are satisfied.

Proposition 2 (Optimal Uniform Segmentation)
There exists an h that attains the supremum of (23), *and a uniform segmentation σ implementing it supported only on regular markets.*

The propositions show that the relevant (optimal) inverse hazard rate that satisfies the majorization constraint can be implemented by a uniform segmentation. We prove the existence of such segmentations constructively, where markets are indexed by $m \in [0,1]$ and a function $s(v)$ marks the upper boundary of the support: only market indices below $s(v)$ contain value v. Hence, markets have nested supports. In Figure 2b we illustrate the support of the segmentation that implements inverse hazard illustrated in Figure 2a; the $x-$axis represents values $v \in [0,1]$, y-axis represent a market index $m \in [0,1]$, and the shaded area is the pairs of value-market (v,m) that have positive support. In the y-axis, we show where markets m_1, m_2 are (illustrated in Figure 3a), and see that the gray shaded area marks the support of these distributions. Whenever the upper boundary of the shaded area has a positive (negative) slope, we get concentration (dilution), corresponding to the case where the inverse hazard rate is smaller (larger) than in the aggregate market (as illustrated in Figure 3b). At the upper boundary of the segment with negative slope (between 1/2 and 5/8), a market has an atom, which significantly shrinks the mass of values available to distribute across other markets; hence, even if a value is not distributed across all the markets, the inverse hazard rate is higher than in the aggregate market when the slope of the upper boundary is negative.

The existence of an optimal h is non-trivial to prove, but the arguments are technical, so we relegate them to the appendix. From an economically substantive perspective, the more important aspect of this proposition is that the optimal h always satisfies regularity: $v - h(v)$ is non-decreasing (alternatively, $h'(v) \leq 1$). This means the segments in the optimal uniform segmentation are themselves regular, which has a clear economic logic.

Recall from the binary case that the local information rent $u(v, h)$ is hump-shaped in h: increasing h raises the extensive margin of rents but increases distortions. A necessary condition for h to be optimal is that the marginal local information rent is that

$$\frac{\partial}{\partial h} u(v, h(v)) \tag{24}$$

is non-decreasing in v. If this failed—if the marginal value of raising h were higher for some low value than for a nearby high value—the designer could concentrate the high value (lowering their h) to create slack for diluting the low value (raising their h), generating a net increase in consumer surplus. Now, because u is "not too supermodular" (in a sense formalized in the appendix), this condition can be used to show that higher values cannot be associated with too much higher hazard rates (formally, $h'(v) \leq 1$ in any solution to (23)).

4.4 The Main Result: The Value of Segmentation

Proposition 2 means that the maximum of (23) is a lower bound on the consumer surplus achievable by a segmentation. To show that this bound is tight, we need to rule out that higher values can be achieved by non-uniform segmentations. We do this with a short technical argument, which tracks the intuition provided at the start of Section 4.3: if buyers of the same value are contributing differentially to the local information rent, they should be reallocated to segments where they contribute the most. If u were concave, this would follow from Jensen's inequality; although u is not concave, we show that the second-order condition for the optimal h is strong enough to apply this argument nonetheless.

Theorem 1 (Value of Segmentation)
The maximum consumer surplus attained by segmentation is:

$$\max_{\sigma \in \mathrm{MPS}(m^*)} \left[\int_{\Delta V} U(m)\ d\sigma(m) \right] = \max_{h \prec h^*} \int_{\underline{v}}^{\overline{v}} u(v, h(v))\ dF^*(v). \tag{25}$$

Furthermore, every segmentation achieving this value is a uniform segmentation implementing some inverse hazard rate h which solves the rhs of (25).

Theorem 1 is the central result of the paper. We highlight several aspects of its content and implications.

Reduction of Complexity The original problem, given by the lhs of (25), requires maximizing over segmentations $\sigma \in \Delta(\Delta V)$: distributions over the infinite-dimensional space of all probability distributions on V. The reduced problem, given by the rhs of (25), requires maximizing a single functional over a single function $h : V \to \mathbb{R}_+$, subject to the family of linear majorization constraints (18). This is a standard optimal control problem, amenable to Karush-Kuhn-Tucker (KKT) methods and often yielding closed-form solutions. The reduction is thus both conceptually clarifying and computationally powerful.

Relation to Extreme Points and Majorization Theorem 1 contrasts with the extreme-point approach to majorization problems (Kleiner et al., 2021). The original problem (3) maximizes an objective which is linear in σ over the convex set MPS(m^*), so an optimum is attained at an extreme point (an "extremal" segmentation into regular markets). However, the set MPS(m^*) is taken over $\Delta(\Delta V)$, not ΔV, and hence lacks the usual tractable one-dimensional reduction to majorization. By contrast, in our simplified problem (25), the objective in terms of inverse hazard rates $\int_{\underline{v}}^{\overline{v}} u(v, h(v))\, dF^*(v)$ is *not* linear (nor convex) in

h, and the optimal inverse hazard rate is therefore generically interior. Thus, the standard tools for maximizing a convex functional over a majorized set do not apply.

Uniform Quality across Segments Since $h(v)$ is the same in every segment containing value v, and the quality allocation in a regular market is $q_m(v) = Q(v - h_m(v))$, it follows that every buyer of value v consumes the *same quality* in every segment where they appear. This is a strong and perhaps counterintuitive property: even though the monopolist could offer different products to the same buyer in different segments, the consumer-optimal segmentation ensures uniform quality provision. What varies across segments is only the *price* for the quality.

Comparison with the Binary Case In the binary setting, the consumer surplus maximization reduced to $\max_{h \leq h^*(v_1)} u(v_1, h)$, a single-variable optimization with a pointwise constraint. Theorem 1 reveals that the natural generalization is *not* the pointwise problem $\max_{h \leq h^*} \int u(v, h(v))\, dF^*(v)$, but rather the majorization-constrained problem (25). The majorization constraint is strictly weaker than the pointwise constraint, because it permits the designer to raise the inverse hazard rate at low values by concentrating higher values. This additional flexibility can strictly improve consumer surplus, and it arises from the dilution mechanism that was absent in the binary case. In many environments, however—including those studied in Section 5— dilution is not optimal. When dilution is not used, the solution coincides with the one obtained if, instead of having the majorization constraint, the inverse hazard rate were a pointwise bound on the markets' inverse hazard rate $h(v) \leq h^*(v)$.

Uniqueness The second part of Theorem 1 establishes that *every* optimal segmentation is uniform. We prove this by showing that irregular markets are strictly suboptimal (hence, strengthening Lemma 1), and non-uniform segmentations are also strictly suboptimal by the concavity argument above. Thus, the optimal segmentation is essentially unique in terms of its economic content (the inverse hazard rate function h and the resulting quality allocations), even though multiple uniform segmentations may implement the same inverse hazard rate h.

5 Convex Segmentations

Theorem 1 reduces the segmentation problem to maximizing a single functional over a single function—a remarkable simplification. Yet the solution it delivers can still be intricate: optimal inverse hazard rates may require segments with gaps and atoms, even in simple environments with finitely many goods. In this section, we show that a widely used regularity condition—on the demand as well as on the supply function—yields a solution of striking simplicity.

The optimal segmentation turns out to consist of "convex" segments: nested intervals all sharing the same upper bound $\overline{v}$, differing only in how far down in the support of the value distribution they reach. These are segmentations where, in contrast to Figure 2b, the slope of the upper boundary is always increasing. Furthermore, the shape of all optimal segments can be expressed in terms of a single transparent interaction between demand and supply elasticities.

We develop these results in three steps. First, we introduce the two regularity conditions and interpret them economically. Second, we characterize the optimal inverse hazard rates and the associated pricing structure. Next, we specialize to iso-elastic costs and uncover a tight connection between supply and demand elasticities. Finally, we use these tools to provide a sharp, easily interpretable condition under which no segmentation can help consumers.

5.1 Log-Concave Demand and Supply

The analysis in this section relies on a symmetric assumption for demand and supply. We assume that the aggregate demand and supply function are log-concave. Specifically, the demand in the aggregate market, the survival function $1 - F^*(v)$, is assumed to be log-concave in v:

$$\frac{d^2}{dv^2} \log(1 - F^*(v)) \leq 0. \tag{26}$$

This is equivalent to the monotone (inverse) hazard rate condition: the assumption that $h^*(v)$ is non-increasing in v. It is satisfied by most commonly used distributions, including the uniform, normal, logistic, and exponential families (see Bagnoli and Bergstrom (2005) for a comprehensive catalog).

Similarly, the supply function is also assumed to be log-concave in v:

$$\frac{d^2}{dv^2} \log(Q(v)) \leq 0. \tag{27}$$

We can also provide this condition in terms of the cost function:

$$\frac{c'''(q)q}{c''(q)} \geq -1.$$

This condition requires the marginal cost to be convex or not too concave, and accommodates a wide range of cost structures, including the entire family of power cost function studied in Section 5.3 (which includes the single unit demand model as a limiting case).

What do these conditions have in common? They require that either demand or supply elasticity does not grow faster than proportional with price. The log-concavity condition appears prominently in the pass-through literature in international trade and industrial organization, see Bulow and Pfleiderer (1983), Weyl and Fabinger (2013) and Mrázová and Neary (2017), where it implies that the monopolist (on the demand side) or the monopsonist (on the supply side) absorbs at least half of any cost increase. This shared "distortion" property is what makes the two conditions work together so powerfully. It ensures that low-value buyers—who are already the most distorted in the aggregate market—are also the ones who stand to gain the most from segmentation. Conversely, high-value buyers, whose allocations are already close to efficient, cannot be profitably helped by rearranging the market.

5.2 Optimal Hazard Rates and Pricing

Under these two conditions, the solution to the majorization problem (25) takes a clean, two-regime form. We recall from the earlier analysis that $\overline{h}(v)$ denotes the inverse hazard rate that maximizes local information rents at value v (see (17)). Under log-concavity of supply, this maximizer is uniquely defined and strictly increasing in v: higher-value buyers can tolerate more distortion before their rents peak, because their baseline surplus is larger. Meanwhile, log-concavity of demand makes the aggregate inverse hazard rate $h^*(v)$ decreasing in v: the aggregate inverse hazard rate falls with v, since higher values are progressively rarer relative to the mass just below them. These two monotonicity properties guarantee that h^* and $\overline{h}$ cross exactly once. Define the threshold:

$$\widehat{v} \equiv \min\left\{v \mid h^*(v) \leq \overline{h}(v)\right\}$$

to be the unique threshold value at which the aggregate inverse hazard rate $h^*(v)$ exceeds the rent-maximizing level $\overline{h}(v)$.

Proposition 3 (Convex Segmentations)
With log-concave demand and supply, the consumer-optimal segmentation displays:

$$h(v) = \begin{cases} \overline{h}(v), & \text{if } v < \widehat{v}; \\ h^*(v), & \text{if } v \geq \widehat{v}. \end{cases} \tag{28}$$

The logic is intuitive. For buyers above the threshold $\widehat{v}$, the aggregate inverse hazard rate is already at or below the rent-maximizing level—there are "too few" high-value buyers relative to these values for the designer to improve their rents. Since local information rents are increasing in h at these values, the designer would like to raise h further, but the majorization constraint is binding: no feasible reallocation of buyers can push h higher. So these values are left untouched.

For buyers below $\widehat{v}$, the situation is reversed. The aggregate market has "too many" higher-value buyers relative to these low values—the inverse hazard rate $h^*(v)$ exceeds the rent-maximizing level $\overline{h}(v)$. This means local information rents are decreasing in h, and the consumer surplus can improve by concentrating low-value buyers into fewer segments, driving down their inverse hazard rate to the rent-maximizing level. Crucially, no dilution is ever needed: the designer uses only the simpler of the two tools identified earlier, concentrating values rather than creating gaps in segment supports. This is precisely because log-concave supply makes local information rents concave in h, while the log-concave demand ensures that the marginal local informational rents are monotonically ordered across values.

The two regularity conditions not only simplify finding the optimal inverse hazard rates—they also yield an especially clean segmentation structure.

Theorem 2 (Convex Segmentations and Pricing)
With log-concave demand and supply, there is a consumer-optimal segmentation such that:

1. *each market m is absolutely continuous with support $[v_m, \overline{v}]$ for some $v_m \leq \widehat{v}$, and*

2. *the pricing differs across markets only by the base price:*

$$p_m(q) = p(q) + T_m,$$

where $p(q)$ does not depend on m and T_m increases in v_m but does not depend on q.

Part 1 says that segments are nested convex intervals, all sharing the same upper bound $\overline{v}$ but with lower bounds that vary across segments. As concentration increases, fewer low-value buyers remain, and the lowest values "drop out" first. We refer to these segmentations as *convex segmentations*, as their supports are nested convex intervals; they are the natural generalization of the direct segmentations of Bergemann et al. (2015).

Part 2 follows from the envelope theorem. By Theorem 1, every buyer of value v consumes the same quality in every segment in which it appears. Within each segment, the indirect utility of value v has the same derivative $q(v)$ at every point in its support. Since the supports of different segments are nested convex intervals, the indirect utilities of any two segments differ only by a constant. The payment of each value is therefore the same up to a quality-independent base price: the monopolist does not redesign its product line across segments, but merely adjusts the base price of the lowest quality offered. What varies is only the level of information rents—because T_m increases in v_m, buyers in more concentrated segments (lower v_m) face a lower base price and enjoy higher rents.

Alternatively, we can associate each nested segment $[v_m, \overline{v}]$ with a nested menu as the qualities below $Q(v_m - h(v_m))$ fail to sell at the base price T_m. In this interpretation, the monopolist offers nested menus all sharing the same efficient upper bound quality $Q(\overline{v})$, differing only in how far down the menu is extended and in the price of the lowest offered quality, the base price.

The theorem shows that the optimal segmentation is easy to construct. Each segment m has a unique support, which is of the lower censored form $[v_m, \overline{v}]$. Within each segment, the inverse hazard rate of values below $\widehat{v}$ has been lowered to $\overline{h}$, while the values above $\widehat{v}$ are unchanged. The key simplifying aspect is optimal segmentation increasingly concentrates low values more than high values. So, as low values are concentrated more, we "run out" of consumers sequentially from $\underline{v}$ upwards, creating segments with support shrinking from the bottom. The combination of (26) and (27) ensures that we run out of consumers sequentially.

5.3 The Elasticity Nexus: The Case of Iso-Elastic Cost

To make the interplay between supply and demand elasticities fully transparent, we now specialize to the family of power cost functions:

$$c(q) = q^\gamma/\gamma, \tag{29}$$

where $\gamma > 1$. We also refer to the cost as *iso-elastic* as

$$\frac{dc\,(q)\,/dq}{c\,(q)\,/q} = \gamma.$$

This family of cost functions satisfies log-concavity (27) for all γ and nests several important cases: $\gamma = 2$ gives quadratic costs (as in Mussa and Rosen (1978)), $\gamma \to \infty$ approximates the unit demand model, and intermediate values parameterize the curvature of the cost function. The supply function Q takes the constant elasticity form:

$$Q(v) = \mathbb{1}[v \geq 0]v^{\frac{1}{\gamma-1}}.$$

with supply elasticity:

$$\frac{dQ(v)}{dv} \cdot \frac{v}{Q(v)} = \frac{1}{\gamma - 1} \in (0, \infty).$$

A direct calculation shows that with the power cost function the inverse hazard rate $\overline{h}$ that maximizes the local information rent has the explicit form:

$$\overline{h}(v) = \frac{\gamma - 1}{\gamma} v. \tag{30}$$

Now, the only class of distributions that has a linear inverse hazard rate is the class of Pareto distributions given by

$$F_P(v) \equiv 1 - (\underline{v}/v)^{\alpha}, \tag{31}$$

where $\alpha \in \mathbb{R}_+$ is referred to as the shape parameter of the Pareto distribution. The inverse hazard rate of the Pareto distribution with shape parameter α is:

$$\frac{1 - F_P(v)}{f_P(v)} = \frac{v}{\alpha}.$$

An immediate Corollary of Proposition 3 for the case of iso-elastic cost functions follows.

Corollary 1 (Segmentation with Pareto Distribution)
With iso-elastic cost and log-concave demand, there is a consumer-optimal segmentation where each segment m is composed of a Pareto distribution and the aggregate demand:

$$h_m(v) = \begin{cases} \frac{\gamma-1}{\gamma} v, & \text{if } v < \widehat{v}; \\ h^*(v), & \text{if } v \geq \widehat{v}; \end{cases} \tag{32}$$

where the shape parameter of the Pareto distribution is $\alpha = \gamma/(\gamma - 1)$.

Thus, we have obtained an explicit shape for every segment. It is the pairing of the Pareto distribution on the lower end and the aggregate distribution itself on the upper end of the distribution. The shape (parameter α) of the Pareto distribution is determined by the convexity (exponent γ) of the cost function.

We can restate the above Corollary in terms of the demand elasticity $\eta_m(v)$ in every segment m. The demand elasticity is given by:

$$\eta_m(v) \equiv -\frac{f_m(v)v}{1 - F_m(v)} = -\frac{v}{h_m(v)}.$$

Corollary 2 (Segmentation and Demand Elasticity)
With iso-elastic cost and log-concave demand, there is a consumer-optimal segmentation where each segment m has constant elasticity below a cutoff and matches the aggregate market

elasticity above it:

$$\eta_m(v) = \begin{cases} \frac{\gamma}{1-\gamma}, & \text{if } v < \widehat{v}; \\ \eta^*(v), & \text{if } v \geq \widehat{v}. \end{cases}$$

The corollary delivers a sharp economic message. The consumer-optimal segmentation makes demand "elastic enough" at the bottom of the distribution, where the monopolist's distortions are most severe. Above the threshold $\widehat{v}$, demand elasticities are left as they are. Below $\widehat{v}$, they are set to a constant that depends only on the cost structure—specifically, $\gamma/(1-\gamma)$, which is one plus the supply elasticity, with a sign reversal.

This connection between supply and demand elasticities has a natural interpretation. More elastic supply (lower γ) means the monopolist can adjust quality more cheaply, amplifying the distortions it imposes on low-value buyers. To counteract these larger distortions, the optimal segmentation must make demand correspondingly more elastic at low values, raising the "cost" to the monopolist of excluding or under-serving these buyers. More rigid supply (higher γ) means the monopolist's quality choices are less responsive to market composition, so less demand-side adjustment is needed.

5.4 When Does Segmentation Always Harm Consumers?

Perhaps the most policy-relevant question our framework can answer is: when should we expect segmentation to hurt consumers regardless of how the market is carved up? An immediate consequence of Proposition 3 provides a sharp answer.

Corollary 3 (Zero Segmentation)
Under log-concave demand and supply, zero segmentation is optimal if and only if

$$h^*(\underline{v}) \leq \overline{h}(\underline{v}). \tag{33}$$

The condition is appealingly simple: segmentation cannot help consumers when the inverse hazard rate at the *lowest* value $\underline{v}$ is already below the rent-maximizing level. Since (26) makes h^* decreasing, if the condition holds at $\underline{v}$, it holds everywhere, so the designer cannot improve rents at any value. The economic interpretation is that when demand is sufficiently elastic throughout the market—when there are relatively few high-value buyers per unit of each lower value—the monopolist's screening already provides consumers with near-optimal informational rents. Any redistribution of values across segments can only reduce them.

When the cost is iso-elastic, the zero-segmentation condition (33) can be written as follows:

$$\eta^*(\underline{v}) \leq \frac{\gamma}{1-\gamma}, \tag{34}$$

which crystallizes the interplay between supply and demand elasticities. In particular, when demand is given by a parametrized class of distributions, we can use the condition (34) to directly verify the optimality of zero segmentation. For example, suppose the demand is given by the above class of Pareto distributions. The demand elasticity is then constant and given by the negative of the shape parameter, $-\alpha$. From Corollary 3 we then have that zero segmentation is optimal when

$$-\alpha \leq \frac{\gamma}{1-\gamma} \iff \alpha \geq \frac{\gamma}{\gamma - 1},$$

and thus demand is sufficiently elastic relative to supply. We can derive similar condition for the class of exponential distributions or the class of uniform distributions, both of which have monotone demand elasticities.

Corollary 3 stands in contrast with Haghpanah and Siegel (2023): they show that for generic markets, there is always a segmentation that improves consumer surplus. The conditions for Corollary 3 to go through are not too restrictive and, in particular, satisfy the notion of genericity in Haghpanah and Siegel (2023). The discrepancy arises because they consider environments with a discrete number of goods, which does not satisfy (27) (in fact, they provide a counterexample showing that their result fails with a continuum of goods). Corollary 3 tells us that as the number of discrete goods grows large, the gains from segmentation they identify sometimes (but not always) disappear, and provides a characterization of when this happens.

Using Theorem 1, we can extend Corollary 3 to the general environment. The general characterization of when zero segmentation is optimal is somewhat technical, so we relegate it to the appendix (see Corollary A.1). Instead, further focusing the cost function to be iso-elastic we can gain further insight into how cost and demand determine when zero segmentation is optimal.

When cost is more elastic (lower γ), the threshold $\gamma/(1-\gamma)$ becomes more negative, making the condition harder to satisfy: the set of markets where zero segmentation is optimal *shrinks*. A more elastic supply means the seller can adjust quality more readily, amplifying the distortions that segmentation can exploit. Conversely, a more inelastic supply (higher γ) makes quality provision rigid, reducing the scope for segmentation to improve consumer welfare. Hence, the condition makes it clear that the set of markets where zero segmentation is optimal grows larger as the exponent γ increases. In fact, this is true even dropping the (26) assumption. Denote by Z_γ the set of markets under which zero segmentation is optimal when the cost function is given by (29).

Proposition 4 (Zero Segmentation—Iso-elastic Cost)
With iso-elastic cost, for any $\gamma' < \gamma$, $Z_{\gamma'} \subset Z_{\gamma}$.

Hence, as the cost becomes more elastic (γ decreases), the potential gains from segmentation increase. This is despite the fact that even in some aggregate markets m^* associated with inefficient allocations, no segmentation can be optimal. In contrast, in the limit $\gamma \to \infty$, zero segmentation is optimal only if the allocation in the aggregate market m^* itself is efficient. One might have thought that this means that it is relatively unlikely to find a market in which zero segmentation is optimal. However, the conclusion is the opposite: in the limit $\gamma \to \infty$, the cost is very inelastic, reducing the potential benefits from segmentation.

6 Generalization and Applications

The analysis of Sections 4 and 5 was developed for a specific objective (consumer surplus maximization) in a specific environment (private values with a convex cost of quality). In this section, we show that the underlying framework—the reduction to an optimization over inverse hazard rates via the local information rent—extends well beyond this setting. The key insight of Theorem 1, reducing the segmentation problem to an optimization over inverse hazard rates via the local information rent, carries over to settings with adverse selection, alternative welfare objectives, and richer preference structures.

The crucial property of the local informational rents that we used in our analysis is:

$$\frac{\partial^2 u(v,h)}{\partial h^2} + \frac{\partial^2 u(v,h)}{\partial h \partial v} < 0, \tag{35}$$

for all (v, h) with $v > h$, thus the marginal local information rent $\partial u/\partial h$ decreases along the diagonal direction in (v, h). We used this property to show that the solution to (23) is regular. One can heuristically verify this by writing the derivative of the optimality condition (24) explicitly; it looks quite similar to (35), the only difference being that the concavity of u is multiplied by $h'(v)$ in the optimality condition. Now, the concavity is negative, so it must be multiplied by a number smaller than one for the optimality condition (24) to have a different sign than (35). Hence, when maximizing over inverse hazard rates (23) we get that a regular inverse hazard rate (ie., $h'(v) \leq 1$) is necessary for optimality.

A generalization of Theorem 1, stated and established as Theorem 3 in the appendix, will apply to any objective function that satisfies (35). We now discuss two extensions: adverse selection environments, where the seller's cost depends on the buyer's type; and Pareto-efficient segmentations, where the designer balances consumer and total surplus. Finally we briefly discuss other extensions where the main elements of Theorem 1 apply such as finite value or weakly convex costs.

6.1 Adverse Selection and Market Fragmentation

Our baseline model assumes that the cost of quality provision is independent of the buyer's type—a private-values environment. Many important applications, however, feature adverse selection: the seller's cost rises with the buyer's willingness to pay. Insurance markets, credit markets, and financial trading all share this feature, and it fundamentally changes the economics of screening. To incorporate adverse selection while maintaining a transparent connection to our baseline, we consider the following specification. A buyer who trades quantity q obtains utility:

$$\tau(v)\,q - p - \frac{q^2}{2}, \tag{36}$$

where $\tau(v)$ is the buyer's gross value and v is the net value known by the buyer, distributed according to $F^*(v)$. Now q is interpreted as a quantity rather than a quality. The quadratic term captures diminishing marginal utility, as in Maskin and Riley (1984), or risk aversion in asset trading models as in Biais et al. (2000). The cost of supplying the good is:

$$c(q) = (\tau(v) - v)q, \tag{37}$$

which is linear in quantity but type-dependent. The derivative of τ measures the severity of adverse selection, and we assume that:

$$\frac{d\tau(v)}{dv} \in [1, \infty), \tag{38}$$

which implies that the cost is (weakly) increasing in the value v. As a special case, if the derivative is equal to 1 at every v, the marginal cost is independent of v, as in a private value environment. Otherwise, we recover the trading model with adverse selection of Biais et al. (2000). We additionally impose that:

$$0 \leq \frac{\tau''(v)v}{\tau'(v)} \leq 1, \tag{39}$$

which requires the degree of adverse selection to be increasing in v but not too quickly.

The local information rents in this environment take the form:

$$w(v, h) \equiv u\left(v, \frac{d\tau(v)}{dv}h\right), \tag{40}$$

where now the supply is $Q(v) = v$ because the utility is quadratic in q. Adverse selection enters by amplifying the inverse hazard rate: the effective inverse hazard rate facing the designer is $\tau'(v)h$ rather than h alone. This amplification has a clear economic interpretation. Under adverse selection, the seller's screening problem is more severe—each unit of information asymmetry (as captured by h) does more damage—because the seller must account for both the downward distortion of quantities and the type-dependent cost wedge.

Proposition 5 (Optimal Segmentation with Adverse Selection)
In the model of adverse selection (36)-(39)*, the maximum consumer surplus attained by segmentation is:*

$$\max_{\sigma \in \text{MPS}(m^*)} \left[\int_{\Delta V} U(m) d\sigma(m) \right] = \max_{h \prec h^*} \int_{\underline{v}}^{\overline{v}} w(v, h(v))\ dF^*(v), \tag{41}$$

where w is given by (40)*. Furthermore, every segmentation achieving this value is a uniform segmentation implementing some h which solves the rhs of* (41)

The result shows that the analytical framework—the reduction to inverse hazard rates, the majorization constraint, and the optimality of uniform segmentations—carries over to this environment. What changes is the shape of the objective function, and with it the quantitative features of the optimal segmentation. In particular, greater adverse selection (higher τ') leads the optimal segmentation to feature more elastic demands within each segment. The intuition is that adverse selection already depresses trade; the designer responds by making segments more elastic, so that more generous terms of trade can be sustained.

It is worth noting why the reduction goes through under (36)-(39), and where it would not. Adverse selection enters this specification only by rescaling the inverse hazard rate, $h \mapsto \tau'(v)h$, leaving an objective of the same form $w(v, h) = u(v, \tau'(v)h)$. Three assumptions do the work. The linear-quadratic payoff in (36) delivers a linear supply $Q(v) = v$, so the local rent retains the product structure $h \cdot Q(\cdot)$ underlying Lemma 2. Condition (38) keeps v an increasing index of willingness to pay, so segment regularity is still governed by monotone virtual values. Most importantly, (39) bounds how fast the severity of adverse selection grows in v, which is exactly what preserves the inequality (35) driving the optimality of regular, uniform segmentations. The reduction does *not* extend automatically beyond these conditions: if (39) fails, so that adverse selection intensifies too quickly in v, then w can violate (35) and the optimality of uniform segmentations may break down.

Market Fragmentation In this context, segmentation has a natural reinterpretation as market fragmentation: buyers trade across different venues, each attracting a different mix of types. A monopolist market maker sets terms in each venue, and our analysis characterizes the fragmentation structure that maximizes buyer welfare. This connection to market microstructure opens a broader reading of our results. The question "when does market segmentation help consumers?" maps onto "when does fragmenting a trading venue improve the terms of trade?" Our characterization—the answer hinges on the interaction of value-dependent costs and the distribution of buyer values— provides a foundation for analyzing fragmentation in financial markets. The assumption that the good is supplied by a single seller (market maker) provides a natural benchmark (see, for example, Glosten (1989)), however, extending the model to allow buyers to submit orders to multiple venues simultaneously, as in Malamud and Rostek (2017), is a natural and promising direction.

6.2 Pareto-Efficient Segmentations

Our main analysis focused on maximizing consumer surplus, but a regulator or a platform may care about broader welfare objectives. We now consider the problem of finding segmentations that are Pareto efficient—those that maximize a weighted sum of consumer surplus $U(m)$ and social surplus $S(m)$:

$$U(m) + \lambda S(m),$$

where $\lambda \geq 0$ represents the relative weight on total welfare (or, equivalently, on the seller's profits, since social surplus includes both consumer surplus and profits). Setting $\lambda = 0$ recovers our baseline problem; the limit $\lambda \to \infty$ corresponds to maximizing social surplus alone. Any objective placing a larger weight on profits than on consumer surplus yields the same solution as this limit, so varying λ traces the entire Pareto frontier. We apply Theorem 3 by setting:

$$w(v,h) \equiv u(v,h) + \lambda(vQ(v-h) - c(Q(v-h))). \tag{42}$$

Here $u(v,h)$ gives the local informational rents while the second term multiplying λ is the local social surplus generated by value v.

Proposition 6 (Linear Combination of Consumer and Social Surplus)
If $\lambda \geq -1$, the maximum linear combination of social surplus and consumer surplus attained by segmentation is:

$$\max_{\sigma \in \mathrm{MPS}(m^*)} \left[\int_{\Delta V} U(m) + \lambda S(m) \; d\sigma(m) \right] = \max_{h \prec h^*} \int_{\underline{v}}^{\overline{v}} w(v, h(v)) \; dF^*(v), \tag{43}$$

where w is given by (42). *Furthermore, every segmentation achieving this value is a uniform segmentation implementing some h which solves the rhs of* (43)

The proposition confirms that the majorization framework applies across the Pareto frontier. What changes as λ increases is the shape of w and, consequently, the optimal inverse hazard rates. Higher λ pushes the optimal segmentation toward more elastic demands within each segment: reducing distortions becomes more valuable when social surplus carries greater weight. In the limit $\lambda \to \infty$, demands become perfectly elastic—the efficient outcome—which corresponds to first-degree price discrimination. This makes economic sense: when the objective is purely allocative efficiency, the designer's ideal is to eliminate all screening distortions, which requires each segment to contain a single value. Under iso-elastic costs, as introduced in Section 5, the demand elasticity in each segment below the threshold takes the closed-form value $(\gamma+\lambda)/(1-\gamma)$. As λ increases, this elasticity rises monotonically, confirming the intuition that moving along the Pareto frontier toward greater total surplus involves progressively reducing market power within segments. Beyond the consumer-optimal level, further increases in demand elasticity no longer benefit consumers: they translate directly into gains for the seller.

Figure 4 illustrates the Pareto frontier for a uniform aggregate market $F^* = \mathcal{U}[0,1]$ and quadratic cost $c(q) = q^2/2$. As λ ranges over $[-1,\infty)$, the induced pairs (U,Π) trace the rightward boundary of the set attainable by segmentation. Two features stand out. First, the consumer-optimal segmentation ($\lambda = 0$) delivers strictly more consumer surplus *and* profit than no segmentation: here, some segmentation benefits buyers and seller alike. Second, as λ rises beyond the consumer-optimal level, the optimum moves along the frontier, trading consumer surplus for profit, and converges as $\lambda \to \infty$ to the efficient outcome implemented by first-degree price discrimination. The outcome is socially efficient only at that point; elsewhere total surplus falls short because the within-segment screening distortion cannot be removed while leaving rents to consumers. Unlike the unit-demand benchmark of Bergemann et al. (2015), the full surplus triangle is not attainable under screening within segments (cf. Haghpanah and Siegel (2022)).

6.3 Other Extensions

Finite Values When the aggregate value distribution is discrete, the analysis carries through with natural modifications. The objective must be replaced by its concave envelope in the inverse hazard rate, and the notion of uniform segmentation is relaxed to an average implementation condition. Buyers of the same value may receive different allocations across segments, but allocations remain monotone in value. As the discrete model approximates a continuum, these differences vanish, and outcomes converge to the continuous benchmark.

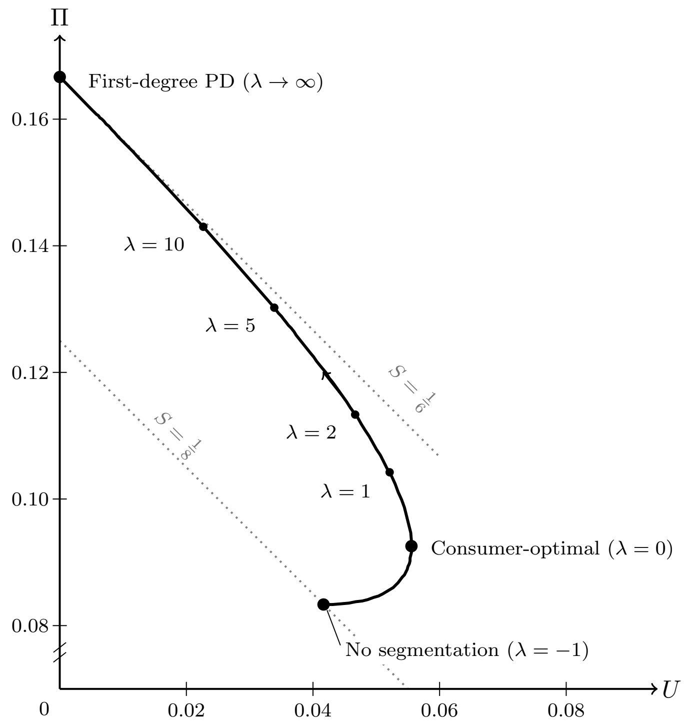


Figure 4: Pareto frontier of (U, Π) achievable with segmentation, where $c(q) = q^2/2$ and $F^* = \mathcal{U}[0,1]$. Dotted lines are iso-surplus loci $U + \Pi = S$.

Weakly Convex Cost The uniqueness of the optimal solution (see discussion after Theorem 1) crucially relies on the strict convexity of cost; if cost is only weakly convex, the solution we characterize continues to be optimal, but non-uniform segmentations can also be optimal. For example, in the unit-demand environment, there are optimal segmentations with varying inverse hazard rates across markets and also with irregular markets. We thus obtain much stronger predictions by assuming the cost function is strictly convex rather than weakly convex.

7 Conclusion

This paper characterizes how market segmentation affects consumer welfare when monopolists can engage in both second- and third-degree price discrimination. Our analysis yields several key insights. First, the consumer-optimal segmentation maintains consistent quality provision across segments while allowing price variation. Second, the benefits of segmentation depend critically on demand and cost elasticities, with no segmentation being optimal when aggregate demand is sufficiently elastic.

Our central methodological contribution is to show that the consumer-optimal segmentation, despite being defined over an infinite-dimensional space of distributions of distributions, reduces to maximizing the expected local information rent over a single inverse hazard rate function subject to a majorization constraint. This reduction enables a complete characterization: consumers of the same value receive the same quality in every segment; the optimal segmentation can be implemented through a common quality menu with segment-specific fixed fees; and whether segmentation helps consumers at all is determined by a sharp threshold on the interaction of demand elasticities and cost convexity.

These theoretical results have direct practical implications. For competition authorities, they suggest that market segmentation should be evaluated based on observable market characteristics. When consumer demand is sufficiently elastic across the whole market, segmentation is uniformly bad for consumers. The monopolist's screening already provides information rents close to what any information structure could achieve, and segmentation can only dilute these rents by redistributing buyer values inefficiently. This contrasts sharply with the unit-demand benchmark, where segmentation is always potentially beneficial, and it provides a concrete, testable criterion for when segmentation raises consumer welfare. For firms, our characterization of optimal segmentation strategies offers guidance for designing market segmentation policies that balance profit maximization with consumer welfare. Our analysis also connects to broader debates about big data and personalized pricing in digital markets. While enhanced ability to segment markets could enable more sophisticated price discrimination, our results suggest this may benefit consumers when properly structured.

Several important directions remain for future research. First, extending the analysis to competitive markets as in the recent analysis of Bergemann et al. (2025) for single unit demand would provide insight into how market structure affects optimal segmentation. Second, empirical work testing our theoretical predictions about the relationship between demand elasticities and optimal segmentation would be valuable.

A Appendix

Proof of Lemma 1. We begin by describing the profit-maximizing menu. Define

$$J_m(t) \equiv F_m^{-1}(t)(t-1).$$

We denote by t_m the largest minimizer of J_m and by vex$[J_m]$ the convexification (i.e. the pointwise largest convex function smaller than $J_m(t)$). Then, a buyer with value v buys quality q satisfying:

$$c'(q) = \frac{\partial}{\partial t} \text{vex}\left[J_m\right](F_m(v)),$$

if $F_m(v) \geq t_m$, and $q = 0$ otherwise. Note that $J_m'(F_m(v)) = \phi_m(v)$, so the expression coincides with the one in the main text when m is regular.

Consider the set of markets X_{p_m} where menu p_m is seller-optimal:

$$X_{p_m} \equiv \left\{m' \in \Delta V : \Pi(m', p_{m'}) = \Pi(m', p_m)\right\}.$$

The seller's profit $\Pi(m', p_{m'})$ is a pointwise supremum of the weak*-continuous affine maps $m' \mapsto \Pi(m', p)$, and hence convex and weak*-lower semicontinuous (lsc); since $\Pi(m', p_{m'}) \geq \Pi(m', p_m)$ and $m' \mapsto \Pi(m', p_m)$ is affine and weak*-continuous, X_{p_m} is a sublevel set of a convex weak*-lsc function, and therefore convex and weak*-closed (Theorem 7.6 and Lemma 2.42, Aliprantis and Border (2006)). Because V is a compact metric space, ΔV is itself a convex, weak*-compact, metrizable set (Theorem 15.11, Aliprantis and Border (2006)), and the closed, convex subset X_{p_m} inherits all three properties. Therefore, by Choquet's theorem (Phelps (2001)), $m \in X_{p_m}$ can be written as a segmentation supported on its extreme points.

Now, observe that $m' \in X_{p_m}$ if and only if

$$t_m = t_{m'} \text{ and } \text{vex}\left[J_m\right](t) = \text{vex}\left[J_{m'}\right](t) \text{ for all } t \geq t_m.$$

On the other hand, m is irregular if and only if $J_m(t) > \text{vex}[J_m](t)$ on some interval $[t_1, t_2]$ where $F_m^{-1}(t)$ is not constant. For such m, sufficiently small perturbations of $F_m(v)$ over $v \in [F_m^{-1}(t_1), F_m^{-1}(t_2)]$ leave vex$[J_m]$ unchanged (and hence remain in X_{p_m}), meaning irregular markets are not extreme points. Thus, irregular markets can be segmented into regular markets without changing the profit-maximizing menu.

The preceding reduction also allows us to restrict to markets without singular-continuous components. Let $A_m \subseteq (0, 1)$ denote the union of the interiors of all maximal intervals on which F_m^{-1} is constant; these correspond exactly to the atoms of F_m. Define

$$H(t) \equiv \frac{\text{vex}[J_m](t)}{t-1} \text{ for } t \in (0,1).$$

Since $\text{vex}[J_m]$ is convex, H is locally absolutely continuous. Moreover, regularity implies $F_m^{-1}(t) = H(t)$ for Lebesgue-a.e. $t \in A_m^c$. Take any $t \in A_m^c$ where $H'(t)$ exists; except at the endpoint of a terminal constant interval, there exists $s > t$ such that $F_m^{-1}(s) > F_m^{-1}(t)$. Convexity and the fact that $\text{vex}[J_m] \leq J_m$ implies

$$\text{vex}[J_m]'(t) \leq \frac{\text{vex}[J_m](s) - \text{vex}[J_m](t)}{s-t} \leq \frac{J_m(s) - J_m(t)}{s-t}.$$

Thus,

$$H'(t) \geq \frac{1-s}{1-t} \cdot \frac{F_m^{-1}(s) - F_m^{-1}(t)}{s-t} > 0 \text{ for Lebesgue-a.e. } t \in A_m^c. \tag{A.1}$$

Let $B \subseteq V$ be any set with Lebesgue measure zero containing no values carrying atoms, meaning that $F_m^{-1}(t) \in B \implies t \in A_m^c$. Since H is locally absolutely continuous, H' exists a.e. on $(0,1)$, and the Serrin-Varberg theorem (Serrin and Varberg (1969)) gives

$$\lambda(\{t : H'(t) \neq 0 \text{ and } H(t) \in B\}) = 0,$$

where λ denotes the Lebesgue measure. Combining this with (A.1) implies that

$$\lambda(\{t \in A_m^c : H(t) \in B\}) = 0.$$

Finally, since $F_m^{-1}(t) = H(t)$ for Lebesgue-a.e. $t \in A_m^c$,

$$m(B) = \lambda(\{t : F_m^{-1}(t) \in B\}) \leq \lambda(\{t \in A_m^c : H(t) \in B\}) = 0,$$

meaning m has no singular-continuous component. □

To make the notation more compact, we denote the survival function by:

$$D_m \equiv 1 - F_m.$$

Proof of Proposition 1. To begin, we claim that for any regular market m,

$$\int_v^{\overline{v}} D_m(t)\ dt = \begin{cases} \int_v^{\overline{v}} h_m(t)\ dF_m(t) & \text{if } \Delta_m(v) = 0; \\ \int_v^{\overline{v}} h_m(t)\ dF_m(t) + \Delta_m(v) D_m(v) & \text{if } \Delta_m(v) > 0. \end{cases} \tag{A.2}$$

In any regular market, $\Delta_m(v) = 0$ if and only if F_m is absolutely continuous at v in the local sense that the continuous non-atomic component admits a density there; by the observation in Lemma 1, regularity rules out a singular-continuous component. If F_m is absolutely continuous at v with density $f_m(v)$, then by definition $D_m(v) = f_m(v)h_m(v)$. Also, in any regular market, at the lower boundary of a gap there is always an atom. That is, at any interval (v_1, v_2) satisfying

$$(v_1, v_2) \cap \text{supp}(m) = \{\emptyset\} \text{ and } v_1, v_2 \in \text{supp}(m),$$

we must have that $a(v_1) > 0$. In any such interval,

$$\int_v^{v_2} D_m(t)\ dt = \Delta_m(v)D_m(v) = \Delta_m(v)D_m(v) + \int_v^{v_2} h_m(t)\ dF_m(t).$$

where in the second equality we are integrating zero because $dF(v)$ is zero when $v \notin \text{supp}(m)$. At the limit $v = v_1$ we obtain

$$\int_{v_1}^{v_2} D_m(t)\ dt = a_m(v_1)h_m(v_1) = \lim_{w \uparrow v_1} \int_w^{v_2} h_m(t)\ dF_m(t)$$

where we take the left limit of the lower boundary of the integral to include the atom at v_1. Thus, combining the intervals, we get (A.2).

We now apply this to the aggregate market:

$$\begin{aligned}\int_v^{\overline{v}} h^*(t)\ dF^*(t) &= \int_v^{\overline{v}} \int_{\Delta V} D_m(t)\ d\sigma(m)\ dt = \int_{\Delta V} \int_v^{\overline{v}} D_m(t)\ dt\ d\sigma(m) \\ &= \int_{\Delta V} \int_v^{\overline{v}} h_m(t)\ dF_m(t)\ d\sigma(m) + \int_{\Delta V} \mathbb{1}\{v \notin \text{supp}(m)\}\Delta_m(v)D_m(v)\ d\sigma(m), \quad \text{(A.3)}\end{aligned}$$

where the change in the order of integration uses Fubini's theorem (Theorem 11.27, Aliprantis and Border (2006)) and the second line is applying (A.2). The first term can be written as:

$$\int_{\Delta V} \int_v^{\overline{v}} h_m(t)\ dF_m(t)\ d\sigma(m) = \int_v^{\overline{v}} \int_{\Delta V} h_m(t)\ d\Sigma_t(m)\ dF^*(t) = \int_v^{\overline{v}} h_\sigma(t)\ dF^*(t).$$

This again uses Fubini's theorem and the disintegration/conditional measure notation embedded in the definition of Σ_t. Subtracting this from the lhs of (A.3) yields the result. $\square$

Proof of Proposition 2. Endow ΔV and $\Delta(\Delta V)$ with the weak topology on measures, so that ΔV and $\Delta(\Delta V)$ are closed and compact. The maximization problem (23) is taken over $h \in L_+^p[\underline{v}, \overline{v}]$, where $1 < p < \infty$. Boundedness guarantees Q (and u) lie in L_+^p, and our tie-breaking assumptions ensure that Q and u are upper semicontinuous (usc). The assumption that $f^*(v) > 0$ and F^* is real-analytic means that h^* is real-analytic. A crucial tool for our analysis will be the concave upper envelope of $u(v, h)$ with respect to h (i.e. the pointwise smallest concave function larger than $u(v, h)$), which we denote by $\overline{u}(v, h)$. Since $u(v, h)$ is bounded and zero when $h \geq v$, $\overline{u}(v, h)$ is constant for $h \geq \overline{h}(v)$.

The proof is divided into two main parts. Proposition A.1 proves that there exists an h which attains the supremum of (23), and characterizes the solution using a KKT system, which allows us to derive certain properties of the solution (Lemma A.1). Proposition A.2 then provides a particular construction method for a uniform segmentation implementing any h which satisfies the technical properties of Lemma A.1.

Proposition A.1 (Supremum is Attained)
The supremum of (23) *is attained. Furthermore, h solves* (23) *if and only if both* (i) $\frac{\partial}{\partial h} u(v, h(v))$ *is non-decreasing in v and constant in v when $E_h(v) > 0$, and* (ii) $u(v, h(v)) = \overline{u}(v, h(v))$ *hold at almost all v.*

Proof. For any $h(v)$ we have that

$$\widehat{h}(v) = \min\{h(v), \overline{h}(v)\}$$

generates higher local informational rents and relaxes the majorization constraint. Hence, it is without loss to make the feasible set

$$\mathcal{H} \equiv \Big\{ h \in L^p_+[\underline{v}, \overline{v}] \mid h \prec h^* \text{ and } 0 \leq h(v) \leq \overline{h}(v) \text{ a.e.} \Big\}.$$

We then study the relaxed problem:

$$\sup_{\mathcal{H}} \int_{\underline{v}}^{\overline{v}} \overline{u}(v, h(v)) \; dF^*(v). \tag{A.4}$$

Endow this space with the weak topology. By construction, $\overline{u}(v, h)$ is concave (and hence also continuous) in h, and additionally bounded because u is bounded. By Rockafeller (1968)'s theorem on integral functionals, the rhs is concave and weakly usc over $\mathcal{H}$. Furthermore, since $\mathcal{H}$ is a closed subset of a weakly compact space, it is itself weakly compact (Theorem 6.25, Aliprantis and Border (2006)). Hence, by Weierstrass Theorem, the supremum is attained.

Next, we wish to characterize the solution of (A.4) using a KKT system. The classic Slater constraint qualification fails, due to the fact that both the box and majorization constraints have no topological interior in their corresponding function spaces. However, by applying an appropriate generalization of this constraint qualification to infinite dimensional settings

(Jeyakumar and Wolkowicz (1992); Boţ et al. (2008)), we get that the KKT conditions are both necessary and sufficient.[4] Thus, for any optimal $h \prec h^*$, there exists a function $\lambda(v) \in L^q_+[\underline{v}, \overline{v}]$, where $\frac{1}{p} + \frac{1}{q} = 1$, and a finite non-negative measure μ such that (Theorem 13.18, Aliprantis and Border (2006))

$$\frac{\partial}{\partial h}\overline{u}(v, h(v)) = \mu[\underline{v}, v] - \lambda(v) \text{ and } \lambda(v)h(v) = E_h(v)d\mu(v) = 0 \text{ a.e.}$$

Here λ is the multiplier on the non-negativity constraint, and μ the multiplier on the majorization constraint (we know the upper bound never binds).

In fact, the non-negativity constraint is never binding (i.e. $\lambda(v) = 0$ a.e.). Suppose there is an interval (v_1, v_2) such that the constraint binds in this interval (that is, $\lambda(v) > 0$ and, consequently, also $h(v) = 0$ on this interval). We must clearly have that the constraint slacks in this interval, so $\mu[\underline{v}, v]$ is constant for all $v \in [v_1, v_2]$. Furthermore, if $v_1 > \underline{v}$, then the constraint must slack in $v \in [v_1 - \epsilon, v_2]$ for an ϵ small enough (which follows from the fact that $E_h(v)$ is strictly decreasing in $[v_1, v_2]$). If $v_1 = \underline{v}$, then $\mu[\underline{v}, v]$ must be equal to zero in $v \in [v_1, v_2]$, but this contradicts the first-order condition. If $v_1 > \underline{v}$, then we must have that:

$$\frac{\partial}{\partial h}\overline{u}(v_1 - \epsilon, h(v_1 - \epsilon)) = \mu[\underline{v}, v_1] > \mu[\underline{v}, v_1] - \lambda(v_1 + \epsilon),$$

where we simply used that the constraint slacks so $\mu[\underline{v}, v]$ stays constant in $[v_1 - \epsilon, v_1 + \epsilon]$ (for a small enough ϵ). However, we also have that:

$$\frac{\partial}{\partial h}\overline{u}(v_1 - \epsilon, h(v_1 - \epsilon)) \leq \frac{\partial}{\partial h}\overline{u}(v_1 - \epsilon, 0) \leq \frac{\partial}{\partial h}\overline{u}(v_1 + \epsilon, 0),$$

where in the first inequality we used that $\overline{u}$ is concave and in the second one we used that $\partial\overline{u}(v, 0)/\partial h$ is increasing in v. We thus reach a contradiction with the fact that the first-order condition is satisfied at $v = v_1 + \epsilon$. Thus, the KKT conditions simplify to

$$\frac{\partial}{\partial h}\overline{u}(v, h(v)) = \mu[\underline{v}, v] \text{ and } E_h(v)d\mu(v) = 0 \text{ a.e.} \tag{A.5}$$

Our final step is to show that (A.5) implies that there is no relaxation gap by considering the concavification in (A.4). We prove that for any h satisfying (A.5), $u(v, h(v)) = \overline{u}(v, h(v))$

[4] In particular, we need a feasible point in the *quasi-relative interior* of $\mathcal{H}$, i.e. an $H \in \mathcal{H}$ such that

$$0 < h(v) < \overline{h}(v) \text{ a.e. and } \int_v^{\overline{v}} h(t)\, dF^*(t) < \int_v^{\overline{v}} h^*(t)\, dF^*(t) \text{ for all } v < \overline{v},$$

which is satisfied by truncating h^* by $\overline{h}$ and then scaling down by $0 < \alpha < 1$.

a.e. An implication of (A.5) is that

$$\frac{d}{dv}\frac{\partial}{\partial h}\overline{u}(v,h(v)) = \frac{\partial^2}{\partial v \partial h}\overline{u}(v,h(v)) + \frac{dh}{dv}\cdot\frac{\partial^2}{\partial h^2}\overline{u}(v,h(v)) \geq 0 \text{ a.e.}$$

since μ is a non-negative measure. Take any v where $u(v,h(v)) < \overline{u}(v,h(v))$. The second term of the above expression is zero ($\overline{u}$ is affine). Let (h_B, h_T) be the bottom and top of the support point of $\overline{u}(v,h(v))$, respectively, so that

$$\frac{\partial \overline{u}(v,h(v))}{\partial h} = \frac{\partial u(v,h_B(v))}{\partial h} = \frac{\partial u(v,h_T(v))}{\partial h} = \frac{u(v,h_T(v)) - u(v,h_B(v))}{h_T(v) - h_B(v)}. \tag{A.6}$$

Here we are using that $\overline{u}(v,h) = u(v,h)$ for all $h < \epsilon$ (for a small enough ϵ), so the concavification differs from u only at interior hazard rates. We thus get that

$$\frac{d}{dv}\frac{\partial \overline{u}(v,h(v))}{\partial h} = \frac{1}{h_T(v) - h_B(v)}\left(\frac{\partial u(v,h_T(v))}{\partial v} - \frac{\partial u(v,h_B(v))}{\partial v}\right).$$

Here we took the derivative of the rhs of (A.6), and then used (A.6) to cancel out all the terms with derivatives of $h_T(v)$ and $h_B(v)$, so only the partial derivative with respect to v is left. Finally, using the middle equality of (A.6), we get that:

$$\frac{d}{dv}\frac{\partial \overline{u}(v,h(v))}{\partial h} = \frac{1}{h_T(v) - h_B(v)}\int_{h_B(v)}^{h_T(v)}\left(\frac{\partial^2 u(v,h)}{\partial h^2} + \frac{\partial^2 u(v,h)}{\partial h \partial v}\right) dh < 0$$

where the inequality follows from

$$\frac{\partial^2 u(v,h)}{\partial h^2} + \frac{\partial^2 u(v,h)}{\partial h \partial v} = -Q'(v-h) < 0. \tag{A.7}$$

Thus, there can be no positive measure of points where $u(v,h) < \overline{u}(v,h)$ for any h that satisfies (A.5). Since m^* is absolutely continuous, this means any solution of the relaxed problem (A.4) achieves the same value in the original problem:

$$\max_{\mathcal{H}}\int_{\underline{v}}^{\overline{v}} \overline{u}(v,h(v))\ dF^*(v) = \max_{\mathcal{H}}\int_{\underline{v}}^{\overline{v}} u(v,h(v))\ dF^*(v),$$

and the solution to (A.4) is also a maximizer of (23). □

Before moving on to the construction of the uniform segmentation implementing the solution, we prove that the solution satisfies regularity and the constraint binds near $\overline{v}$.

Lemma A.1 (Regularity of Optimal h)
For any $h \prec h^*$ which satisfies (A.5), (i) $h'(v) < 1$ *a.e., and* (ii) *there exists a $\delta > 0$ such that $h(v) = h^*(v)$ for almost all $v \in [\overline{v} - \delta, \overline{v}]$.*

It is then without loss to assume that (i) and (ii) hold exactly everywhere, since measure 0 sets do not matter when m^* is absolutely continuous. Then, (i) both implies that $v - h(v)$ is monotone, i.e. the canonical segmentation implementing h is supported on (strictly) regular markets only, and that h has bounded total variation.

Proof. Take the total derivative of the marginal local information rent with respect to v:

$$\frac{d}{dv}\frac{\partial u(v,h(v))}{\partial h} = h'(v)\frac{\partial^2 u(v,h(v))}{\partial h^2} + \frac{\partial^2 u(v,h(v))}{\partial h \partial v} \geq 0.$$

Observe that $u(v,h(v)) = \overline{u}(v,h(v))$ implies u is concave in h at $h(v)$ (Theorem 7.23, Aliprantis and Border (2006)); (i) then follows from (A.7). For (ii), suppose by way of contradiction that there is an interval $(\overline{v} - \delta, \overline{v})$ such that $E_h(v) > 0$ on this interval (E_h is continuous even if h is not continuous). By (A.5), $\frac{\partial}{\partial h}\overline{u}(v,h(v))$ must be constant on this interval, and at most $Q(\overline{v} - \delta)$. But, $h^*(v) \to 0$, meaning $\frac{\partial}{\partial h}\overline{u}(v,h^*(v)) \to Q(\overline{v})$. This means for all v sufficiently close to $\overline{v}$, $h(v) > h^*(v)$, which means the majorization constraint is not satisfied. We thus reach a contradiction. □

Proposition A.2 (Implementability of Optimal h)
There exists a uniform segmentation σ implementing the solution to (23).

Proof. Consider a continuum of markets indexed by $m \in [0,1]$. The support is determined by a function $s : V \to [0,1]$ such that $v \in \mathrm{supp}(m)$ if and only if $m \leq s(v)$. The distribution is absolutely continuous in market m at all v such that $m < s(v)$ with density $f_m(v)$:

$$f_m(v) = \frac{D_m(v)}{h(v)}, \tag{A.8}$$

and there is possibly an atom of size $a_m(v)$ at v at market $m = s(v)$:

$$a_m(v) = \frac{\Delta_m(v) D_m(v)}{h(v)}. \tag{A.9}$$

Finally, $s(v)$ satisfies (21), which in this case can be written more simply as follows:

$$\int_{s(v)}^{1} \Delta_m(v) D_m(v)\, dm = E_h(v). \tag{A.10}$$

The proof proceeds in two steps. First, we show that given a solution to (A.10), the associated segmentation given by (A.8)-(A.9) aggregates to m^*. That is, for all v,

$$D^*(v) = \int_0^1 D_m(v)\ dm \iff \int_v^{\overline{v}} D^*(t)\ dt = \int_v^{\overline{v}} \int_0^1 D_m(t)\ dm\ dt. \tag{A.11}$$

Here, the equivalence simply states that two functions are the same if and only if their integrals are the same. As before, we have that for any m,

$$\int_v^{\overline{v}} D_m(t)\ dt = \int_v^{\overline{v}} h(t)\ dF_m(t) + \Delta_m(v) D_m(v).$$

The second term is only nonzero when $m > s(v)$ (or, equivalently, $v \notin \text{supp}(m)$), so:

$$\int_v^{\overline{v}} \int_0^1 D_m(t)\ dm\ dt = \int_0^1 \int_v^{\overline{v}} h(t)\ dF_m(t)\ dm + \int_{s(v)}^1 \Delta_m(t) D_m(t)\ dm,$$

where we apply Fubini's theorem to exchange the integration order. Now, using (A.10) to replace the second term on the rhs, we get

$$\int_v^{\overline{v}} \int_0^1 D_m(t)\ dm\ dt = \int_0^1 \int_v^{\overline{v}} h(t)\ dF_m(t)\ dm + \int_v^{\overline{v}} D^*(t)\ dt - \int_v^{\overline{v}} h(t)\ dF^*(t).$$

Let $\widehat{v}$ be the largest value such that the lhs of (A.11) fails; note that the rhs of (A.11) still holds for $v = \widehat{v}$. Plugging this into the above equation, we get

$$\int_{\widehat{v}}^{\overline{v}} D^*(t)\ dt = \int_0^1 \int_{\widehat{v}}^{\overline{v}} h(t)\ dF_m(t)\ dm + \int_{\widehat{v}}^{\overline{v}} D^*(t)\ dt - \int_{\widehat{v}}^{\overline{v}} h(t)\ dF^*(t)$$

which contradicts that the lhs of (A.11) fails at $\widehat{v}$.

The second part of the proof is to show that (A.10) has a solution. We do this by taking a sequence $\{h_n\}$ approximating h, finding a solution s_n for each h_n, and argue that the sequence $\{s_n\}$ converges. We take h_n to be a function with the following properties:

$$h_n \text{ is real-analytic, } \epsilon_n \leq h_n(v) \leq \overline{h}(v), \text{ and (19) holds for all } v \in [\underline{v}, \overline{v} - \delta],$$

where $\epsilon_n \to 0$ and $h_n(v) = h^*(v)$ for $v \geq \overline{v} - \delta$. Since the space of real-analytic functions is dense in the space of functions with bounded total variation, we can construct a sequence $\{h_n\}$ converging pointwise a.e. to h.[5]

[5] Technically, the sequence converges to h in the L^p-norm, and by the Riesz-Fisher theorem we can then extract a pointwise convergent subsequence.

We construct $s(v)$ inductively. Consider v_2 such that $s(v)$ satisfies (A.10) for all $v \in [v_2, \overline{v}]$ and $s'(v_2) = 0$. We construct $s(v)$ in a new interval $[v_1, v_2]$ satisfying feasibility and such that $s'(v) = 0$. We consider two cases, and show that we move between cases only when E has a strict inflection point. Since h_n and h^* are real-analytic, E_{h_n} has a finite number of strict inflection points. Hence, the process finishes after a finite number of steps. Finally, note that $s(v) = 1$ for $v \geq \overline{v} - \delta$, so we start the induction with Case 1.

(Case 1: $E''_h(v) \geq 0$ for all v in some neighborhood $[v_2 - \epsilon, v_2]$) Define, for each m,

$$\widehat{D}_m(v) = D_m(v_2) \exp\left(\int_v^{v_2} \frac{1}{h_n(t)}\, dt\right).$$

Since h_n is strictly bounded away from 0, $\frac{1}{h_n}$ is integrable and this is well-defined. These survival functions satisfy the constant hazard rate requirement. Now, define $s_n(v)$ by the following differential equation:

$$\frac{ds_n}{dv} = \frac{1}{\widehat{D}_{s_n(v)}(v)} \cdot \frac{d^2 E_{h_n}}{dv^2}. \tag{A.12}$$

This is an ordinary differential equation where the rhs is continuous in s (recall that h_n is real-analytic), so by Carathéodory's existence theorem, a solution exists. It is clear that in this interval, s is increasing. Furthermore, (A.12) is derived from (A.10) by taking the second derivative and setting $\Delta_m(v) = 0$ (since s_n is increasing), and hence (A.10) is satisfied.

(Case 2: $E''_h(v) < 0$ for all v in some neighborhood $[v_2 - \epsilon, v_2]$) For each m, let $\widehat{v}(m)$ be the lowest inverse of $s_n(v)$ in $[v_2, \overline{v}]$:

$$\widehat{v}(m) \equiv \inf\big\{v \in [v_2, \overline{v}] \mid s_n(v) \geq m\big\}.$$

For all $m \geq s_n(v_2)$, this value exists, since $s_n(\overline{v}) = 1$. We define $s_n(v)$ so that

$$\int_{s_n(v)}^{1} \big(\widehat{v}(m) - v\big) D_m(\widehat{v}(m))\, dm = E_{h_n}(v). \tag{A.13}$$

Note that this is just (A.10) under the assumption that $s_n(v)$ is decreasing in $[v, v_2]$. Let v_1 be the largest value less than v_2 such that the solution to (A.13) is decreasing in $[v_1, v_2]$. Taking the second derivative of (A.10), and evaluating at v_1 (we have $s'_n(v_1) = 0$) we get that

$$E''_{h_n}(v_1) = -s''_n(v_1)\big(\widehat{v}(s_n(v_1)) - v_1\big) D_m(\widehat{v}(s_n(v_1))) \geq 0$$

meaning that $[v_1, v_2]$ is at least as large as the gap between the strict inflection points of

E_{h_n}. Thus, we move back to Case 1, and after a finite number of steps, the process ends.

Finally, we note that since $h'(v) < 1$ everywhere, when s_n is increasing,

$$\frac{ds_n}{dv} = \frac{1}{\widehat{D}_{s(v)}(v)} E''_{h_n}(v) < \frac{1}{D^*(\overline{v} - \delta)} \left(\max_v \left[2f^*(v) + \frac{df^*(v)}{dv} \overline{h}(v) \right] \right).$$

Hence, all s_n are bounded in $[0, 1]$ with uniform upper bound on $s'_n(v)$, which in turn means they have uniformly upper bounded total variation. Thus, by Helly's Selection Theorem, there is a subsequence of s_n converging pointwise. □

Together, these results establish Proposition 2. □

As a corollary of Proposition A.1, we get necessary and sufficient conditions for h^* to solve (23), extending Corollary 3 to the general setting.

Corollary A.1 (Zero Segmentation)
Zero segmentation is optimal if and only if both (i) $\frac{\partial}{\partial h} u(v, h^*(v))$ *is non-decreasing and* (ii) $u(v, h^*(v)) = \overline{u}(v, h^*(v))$ *hold a.e.*

Proof of Theorem 1. Theorem 1 consists of two parts. The first, that the value of segmentation is characterized by (25), follows from Proposition 2 after we rule out non-uniform segmentations. Any segmentation that is supported only on regular markets generates consumer surplus

$$\int_{\Delta V} \int_{\underline{v}}^{\overline{v}} u(v, h_m(v)) \; d\Sigma_v(m) \; dF^*(v) \leq \int_{\underline{v}}^{\overline{v}} \int_{\Delta V} \overline{u}(v, h_m(v)) \; d\Sigma_v(m) \; dF^*(v).$$

But, by Jensen's inequality,

$$\int_{\underline{v}}^{\overline{v}} \int_{\Delta V} \overline{u}(v, h_m(v)) \; d\Sigma_v(m) \; dF^*(v) \leq \int_{\underline{v}}^{\overline{v}} \overline{u}(v, h_\sigma(v)) \; dF^*(v),$$

with strict inequality whenever $h_m \neq h_\sigma$ a.e. (indeed, equality in Jensen would require $h_\sigma(v)$ to lie on an affine face of $\overline{u}(v, \cdot)$, which is ruled out at an optimum by the no-relaxation-gap argument). But, by Propositions 1-2, the maximum possible value of the rhs is achievable with a uniform segmentation, so any non-uniform segmentation achieves lower consumer surplus than the optimal uniform segmentation. This proves that the optimal segmentation attains value (25).

The second part of the theorem requires us to rule out segmentations supported on irregular markets that nonetheless achieve the same value. To do this, take any irregular market, and decompose it via Lemma 1 to weakly regular markets (of which there must be a positive measure). Our first step is to show that

$$\max_{\sigma \in \mathrm{MPS}(m)} \int_{x \in \Delta V} U(x)\ d\sigma(x) \geq \sup_{h \prec h_m} \int_{\underline{v}}^{\overline{v}} \overline{u}(v, h(v))\ dF_m(v). \tag{A.14}$$

If m is absolutely continuous, this is immediate because it follows from (25). To accommodate markets with atoms, we first prove that the rhs of (A.14) is lsc while the lhs is usc in m.

For the rhs, we assume all $h \in \mathcal{H}$ are lsc (this convention obviously does not matter when m is continuous). Since u is increasing in h, the objective function is lsc in F_m (Theorem 15.3, Aliprantis and Border (2006)). On the other hand, the set

$$\{h \in \mathcal{H} : h \prec h_m\}$$

is lower hemicontinuous in m. Lower hemicontinuity is straightforward because any $\widehat{h}(v)$ that is pointwise smaller than $h(v)$, with $h \prec h_m$, satisfies $\widehat{h} \prec h_{m'}$ for all m' close enough to m. Hence, following the maximum theorem, the value is lsc in m.

For the lhs of (A.14), note that, under our tie-breaking rules, $U(m)$ is bounded and usc (see Yang (2023)), meaning that (Theorem 15.3, Aliprantis and Border (2006))

$$\int_{x \in \Delta V} U(x)\ d\sigma(x)$$

is itself usc in σ. Furthermore, by standard results, the mapping $m \mapsto \mathrm{MPS}(m)$ is nonempty, compact-valued, and continuous. Hence, by Berge's theorem, the lhs is usc.

Now, if m is only weakly regular, there exists a positive measure subinterval $[v_1, v_2]$ (possibly with atoms) such that $\frac{\partial}{\partial h}\overline{u}(v, h_m(v))$ is strictly decreasing (if $u(v, h_m(v)) = \overline{u}(v, h_m(v))$, this follows from (A.7); else, see (A.6)). Hence, h_m does not satisfy the first-order condition, meaning it does not solve (A.4).[6] We thus have that:

$$\int_{\underline{v}}^{\overline{v}} \overline{u}(v, h_m(v))\ dF_m(v) < \sup_{h \prec h_m} \int_{\underline{v}}^{\overline{v}} \overline{u}(v, h(v))\ dF_m(v) \leq \max_{\sigma \in \mathrm{MPS}(m)} \int_{x \in \Delta V} U(x)\ d\sigma(x).$$

But, the lhs is weakly larger than the consumer surplus in m. So, there is a further segmentation of m which strictly improves consumer surplus. □

[6] When m has atoms, the KKT system may not characterize the solution, but the FOC is still necessary; see (24) and the discussion that follows.

Proof of Proposition 3. We first note that for any h,

$$\frac{\partial u(v,h)}{\partial h} \geq 0 \implies \frac{\partial^2 u(v,h)}{\partial h^2} < 0. \tag{A.15}$$

This follows from the fact that

$$\frac{\partial u(v,h)}{\partial h} \geq 0 \iff \frac{1}{h} \geq \frac{Q'(v-h)}{Q(v-h)} \tag{A.16}$$

and

$$h\left(Q''(v-h) - \frac{2Q'(v-h)}{h}\right) \leq h\left(Q''(v-h) - 2\frac{(Q'(v-h))^2}{Q(v-h)}\right) \leq -h\frac{(Q'(v-h))^2}{Q(v-h)},$$

where the first inequality uses (A.16) while the second one follows from (27). We thus obtain (A.15). This implies that $u(v,h)$ is strictly quasi-concave in h. Furthermore, we have that:

$$\frac{\partial^2 u(v,h)}{\partial h \partial v} = h\left(\frac{Q'(v-h)}{h} - Q''(v-h)\right) \geq 0,$$

where the inequality follows as before from (A.16) and (27). This implies that $\overline{h}(v)$ is increasing, which, in turn, implies that $\overline{h}(v)$ and $h^*(v)$ cross at most once (where we use that h^* is decreasing for this last implication). Hence, we have that $u(v,h(v)) = \overline{u}(v,h(v))$ for $h(v)$ defined as in (28). Finally, we show that $\partial u(v,h(v))/\partial h$ is increasing in v. For all $v \leq \widehat{v}$ we have that the derivative is zero; for all $v \geq \widehat{v}$ we have that the derivative is positive. Finally, we have that the derivative is also increasing for all $v \geq \widehat{v}$:

$$\begin{aligned}\frac{d}{dv}\left(\frac{\partial u(v,h^*(v))}{\partial h}\right) &= \left(1 - \frac{dh^*(v)}{dv}\right)\left(\frac{\left(1 - 2\frac{dh^*(v)}{dv}\right)}{\left(1 - \frac{dh^*(v)}{dv}\right)} Q'(v-h^*(v)) - h^*(v)Q''(v-h^*(v))\right) \\ &> \left(1 - \frac{dh^*(v)}{dv}\right)\left(Q'(v-h^*(v)) - h^*(v)Q''(v-h^*(v))\right) > 0,\end{aligned}$$

where the first inequality follows from the fact that $h^*(v)$ is decreasing, so the fraction is larger than 1, while the second inequality follows from (A.16) and (27). Thus, $\partial u(v,h(v))/\partial h$ is increasing in v. Following Proposition A.1, this is the solution to (25). $\square$

Proof of Theorem 2. The only thing left is to show that the canonical segmentation is gapless and atomless. By the proof of Proposition A.2, this is equivalent to s being increasing, or equivalently, E_h being convex. By (A.10),

$$\frac{d}{dv}E_h(v) = \big[h(v) - h^*(v)\big] f^*(v).$$

For $v \geq \widehat{v}$, $h(v) = h^*(v)$; otherwise, $h(v) < h^*(v)$ with h increasing and h^* decreasing. So, it is immediate that E_h is convex. $\square$

Proof of Proposition 4. We use the conditions in Corollary A.1 to know when zero segmentation is optimal. First, we note that $u(v, h^*(v)) = \overline{u}(v, h^*(v))$ if and only if

$$h^*(v) \leq \overline{h}(v) = \frac{\gamma - 1}{\gamma} v.$$

Since $h(v)$ is increasing in γ, the condition becomes less restrictive as γ increases. Second, we note that $\partial u(v, h^*(v))/\partial h$ is increasing in v if and only if:

$$(\gamma - 1)\left(v - \frac{\partial h^*(v)}{\partial v}(2v - h^*(v))\right) + \left(\frac{\partial h^*(v)}{\partial v} - 1\right) h^*(v) \geq 0. \tag{A.17}$$

Finally, we have that $\partial h^*(v)/\partial v \leq 1$ so the second term is negative (otherwise, this is an irregular market, so zero segmentation is not optimal due to Theorem 1). Thus, if at some v, (A.17) is satisfied with equality, then (A.17) will not be satisfied also for all lower values of γ. Hence, among regular markets, (A.17) also becomes less restrictive as γ increases. $\square$

We now provide the proofs of the results in Section 6. We consider the following problem:

$$\begin{aligned} \overline{W} \equiv \max_{\sigma \in \mathrm{MPS}(m^*)} \int_{\Delta V} \int_{\underline{v}}^{\overline{v}} w(v, h_m(v))\ dF_m(v)\ d\sigma(m) \\ \text{subject to: every } m \in \mathrm{supp}(\sigma) \text{ is regular.} \end{aligned} \tag{A.18}$$

This is the same as (16), but replaced $u(v, h) \to w(v, h)$. Lemma 1 will be satisfied in all the extensions we consider, so considering this formulation will indeed be justified. We assume that w is not too supermodular relative to the concavity:

$$\frac{\partial^2 w(v, h)}{\partial h^2} + \frac{\partial^2 w(v, h)}{\partial h \partial v} \leq 0. \tag{A.19}$$

This is an economically substantive assumption—it implies that higher values are associated with only moderately higher inverse hazard rates (or lower inverse hazard rates)—and is used to guarantee the solution to (23) is regular. Second, we assume that there exists ϵ such that for all $h \in (0, \epsilon)$:

$$\frac{\partial w(v,h)}{\partial h} \text{ is non-decreasing in } v \text{ and is greater than } \frac{w(v,h')}{h'} \text{ for all } h' \in [\epsilon, \infty). \tag{A.20}$$

This assumption rules out corner cases at $h = 0$ in the relaxed problem. In particular, it ensures that whenever the concavification differs from the original objective, this happens away from the boundary. This is used immediately after (A.6) in our main argument.

Theorem 3 (Generalized Upper Bound)
Under (A.19)-(A.20)*, the maximum of* (A.18) *is given by:*

$$\overline{W} = \max_{h \prec h^*} \int_{\underline{v}}^{\overline{v}} w(v, h(v))\ dF^*(v). \tag{A.21}$$

Furthermore, every segmentation achieving this value is a uniform segmentation implementing some h which solves the rhs of (A.21)*.*

The theorem generalizes our main result to objective functions that do not consist of maximizing the local informational rents. The theorem simply identifies the conditions of u we used to prove Theorem 1, so it does not involve any new analysis. We now provide the proofs of the results in the main text.

Proof of Proposition 5. Following standard techniques, the seller's supply is:

$$Q(\phi_m(v)) = v - \frac{d\tau(v)}{dv}\frac{1 - F_m(v)}{f_m(v)}.$$

Hence, the expected consumer surplus,

$$\int_{\underline{v}}^{\overline{v}} \frac{d\tau(v)}{dv}\frac{1 - F_m(v)}{f_m(v)} \left(v - \frac{d\tau(v)}{dv}\frac{1 - F_m(v)}{f_m(v)} \right) dF_m(v).$$

We thus obtain that the consumer surplus is of the form (A.18) with $w(v,h)$ given by (40). Lemma 1 is clearly satisfied as we can apply the same arguments. Furthermore,

$$\frac{\partial^2 w(v,h)}{\partial h^2} + \frac{\partial^2 w(v,h)}{\partial h \partial v} = \frac{d\tau(v)}{dv}\left(1 + \frac{\frac{v d^2\tau(v)}{dv^2}}{\frac{d\tau(v)}{dv}} - 2\frac{d\tau(v)}{dv}\right) - 4h(v)\frac{d\tau(v)}{dv}\frac{d^2\tau(v)}{dv^2} < 0$$

and

$$\frac{\partial w(v,h)}{\partial h}\ |_{h=0} = v\frac{d\tau(v)}{dv}.$$

Thus, (A.19)-(A.20) are satisfied. □

Proof of Proposition 6. We can verify that for all $\lambda \geq -1$, Lemma 1 holds and $w(v,h)$ satisfies:

$$\frac{\partial^2 w(v,h)}{\partial h^2} + \frac{\partial^2 w(v,h)}{\partial h \partial v} = -(1+\lambda)Q'(v-h) \text{ and } \frac{\partial w(v,h)}{\partial h} \mid_{h=0} = Q(v).$$

Thus, (A.19)-(A.20) are satisfied. $\square$

References

Aguirre, I., S. Cowan, and J. Vickers (2010): "Monopoly Price Discrimination and Demand Curvature," *American Economic Review*, 100, 1601–1615.

Aliprantis, C., and K. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer.

Bagnoli, M., and T. Bergstrom (2005): "Log-concave Probability and its Applications," *Economic Theory*, 26, 445–469.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–957.

——— (2025): "On the Alignment of Consumer Surplus and Total Surplus under Competitive Price Discrimination," *American Economic Journal: Microeconomics*, 17, 234–239.

Biais, B., D. Martimort, and J.-C. Rochet (2000): "Competing Mechanisms in a Common Value Environment," *Econometrica*, 68, 799–837.

Boţ, R. I., E. R. Csetnek, and G. Wanka (2008): "Regularity Conditions via Quasi-relative Interior in Convex Programming," *SIAM Journal on Optimization*, 19, 217–233.

Bulow, J. I., and P. Pfleiderer (1983): "A note on the effect of cost changes on prices," *Journal of Political Economy*, 91, 182–185.

Cowan, S. (2012): "Third-Degree Price Discrimination and Consumer Surplus," *Journal of Industrial Economics*, 60, 333–345.

Glosten, L. R. (1989): "Insider Trading, Liquidity, and the Role of the Monopolist Specialist," *The Journal of Business*, 62, 211–235.

Guesnerie, R., and J.-J. Laffont (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, 25, 329–369.

Haghpanah, N., and R. Siegel (2022): "The Limits of Multi-Product Price Discrimination," *American Economic Review: Insights*, 4, 443–458.

——— (2023): "Pareto Improving Segmentation of Multi-Product Markets," *Journal of Political Economy*, 131, 1546–1575.

Jeyakumar, V., and H. Wolkowicz (1992): "Generalizations of Slater's Constraint Qualification for Infinite Convex Programs," *Mathematical Programming*, 57, 85–101.

Johnson, J. P., and D. P. Myatt (2003): "Multiproduct Quality Competition: Fighting Brands and Product Line Pruning," *American Economic Review*, 93, 748–774.

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593.

Malamud, S., and M. Rostek (2017): "Decentralized Exchange," *American Economic Review*, 107, 3320–3362.

Maskin, E., and J. Riley (1984): "Monopoly with Incomplete Information," *RAND Journal of Economics*, 15, 171–196.

Mrázová, M., and J. P. Neary (2017): "Not so demanding: Demand structure and firm behavior," *American Economic Review*, 107, 3835–3874.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–317.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Phelps, R. (2001): *Lectures on Choquet's Theorem*, Springer.

Pigou, A. (1920): *The Economics of Welfare*, London: Macmillan.

Robinson, J. (1933): *The Economics of Imperfect Competition*, London: Macmillan.

Rockafeller, R. T. (1968): "Integrals which are Convex Functionals," *Pacific Journal of Mathematics*, 24, 525–539.

Schmalensee, R. (1981): "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *American Economic Review*, 71, 242–247.

Serrin, J., and D. E. Varberg (1969): "A General Chain Rule for Derivatives and the Change of Variables Formula for the Lebesgue Integral," *The American Mathematical Monthly*, 76, 514–520.

Varian, H. R. (1985): "Price Discrimination and Social Welfare," *American Economic Review*, 75, 870–875.

Weyl, E. G., and M. Fabinger (2013): "Pass-through as an economic tool: Principles of incidence under imperfect competition," *Journal of political economy*, 121, 528–583.

Yang, K. H. (2023): "On the Continuity of Outcomes in a Monopoly Market," *Journal of Mathematical Economics*, 108, 102887.