When Are Single-Contract Menus Profit-Maximizing?*

Christopher Sandmann†

January 2023

Abstract

This paper revisits the canonical screening problem due to Mussa and Rosen (1978): A seller proposes a menu of price-quality (or quantity) contracts so that buyers of distinct valuations purchase distinct varieties. While menu pricing is *feasible* in their model, a result by Stokey (1979) implies that sometimes it is *profit-maximizing* to offer a single take-it-or-leave-it contract instead. This article identifies conditions that tell the optimality of separation and single-contract menus apart. I show that a single-contract menu (continuous separation) is profit-maximizing if the surplus function (utility minus cost) is more concave (convex) for higher types. This is akin to saying that higher buyer types' surplus-utility represents more (less) risk-averse preferences over quality. Finally, I prove a (partial) generalization that applies when there are common values and contrast my findings with menu pricing in competitive environments.

Keywords: menu pricing, second-degree price discrimination, nonlinear pricing
JEL classification: D42, D82, D86, L11, L12

*I am grateful to my advisor Thomas Mariotti for his guidance and support. I also wish to thank Alae Baha, Nicolas Bonneton, James Dana, Daniel Garrett, Bertrand Gobillard, Daniel Gottlieb, Marina Halac, Bruno Jullien, Lucas Maestri, Thomas Mariotti, Benny Moldovanu, Patrick Rey, Jean Tirole and Leeat Yariv for their thoughtful comments, as well as various seminar audiences.
†London School of Economics, Department of Economics, London WC2A 3PH, United Kingdom, c.sandmann@lse.ac.uk

1 Introduction

Menu pricing, also known as second-degree price discrimination, is a strategy used by businesses to charge high prices for high-quality goods without forgoing demand from buyers looking for a more affordable option.[1] In this paper, I investigate when it is profit-maximizing for businesses to employ this strategy. As in the seminal article by Mussa and Rosen (1978), a monopolist seller can offer buyers a menu of contracts with different prices and qualities. The menu allows buyers that do not particularly value quality to purchase a lesser quality good at a more affordable price. According to conventional wisdom (e.g., Shapiro and Varian (1998)), offering more quality varieties strictly increases the seller's profit. However, not all businesses use this strategy. For example, Bloomberg does not offer a discounted version of its financial trading terminal. A financial trader interested in one Bloomberg dataset, e.g., fixed income, must acquire all of the company's data. Similarly, Steve Jobs ended software licensing agreements and reduced the number of Apple computer models offered when he returned as CEO. These companies realized that by not offering lower-quality options, they could raise their prices for their unmatched, high-quality offerings.

The literature on menu pricing offers conflicting examples of when single- and multi-contract menus are profit-maximizing. Itoh (1983) shows that offering several qualities strictly increases profit when there are increasing marginal unit costs. In the absence of unit costs, by contrast, Stokey (1979) and Myerson (1981) prove that the opposite is true. Stokey, in the context of intertemporal price-discrimination, shows that sellers never want to drop prices below their initial high level. And Myerson's optimal auction is a posted price when there is a single bidder. Both results are limited in scope in that they assume that agents with different valuations have identical von-Neumann-Morgenstern preferences over qualities. Formally, utility functions are multiplicatively separable in type and quality. Work due to Salant (1989) has since shown that separable preferences are not an isolated example: there exists a broader class of complementarity conditions under which single-contract menus are most profitable. And Anderson and Dana (2009) argue, pending further regularity assumptions, that single-contract menus are profit-maximizing if complementarity takes the form of log submodular surplus. Log submodular surplus however is sufficient only if the buyers' utility of consuming their outside option is normalized to zero. This paper relaxes their sufficiency conditions and offers a novel explanation of the underlying economic forces.

The main economic insight of this paper is to link the optimality of single-contract menus to a ranking of the buyers' risk preferences: we will see that if high-valuation buyers are more risk-averse in the sense of Arrow-Pratt, then the profit-maximizing menu consists of a single contract offering a high quality good at a high price. This result assumes the absence of unit costs but can easily be generalized by relabeling surplus as the buyer's utility. To see why a comparison of risk preferences is key, suppose the seller were to actively sell two goods of distinct qualities, one high and one low. Then, instead of offering the low-quality good, the seller could offer a lottery where the buyer has a chance of receiving the high-quality good or no good at all. Clearly, there exists such a lottery that leaves the buyer of the low-quality good indifferent to the initial purchase. And under the prescribed ranking of risk preferences, the buyer of the high-quality good would prefer the low-quality good over the lottery, allowing the seller to increase the price of the high-quality good without altering the buyers' self-selection incentives.

[1] Menu pricing is also referred to as versioning (Shapiro and Varian (1998)). Following Belleflamme and Peitz (2015), I adopt the more suggestive terminology of menu pricing.

Example To illustrate, consider two travellers, one for business and one for leisure. A flag-carrier airline chooses between three regimes: only offer business class, only offer economy class, or both economy and business class. Using the ranking of risk preferences, we identify circumstances in which offering both business and economy class tickets decreases the seller's profit. Specifically, set the utility of not travelling with the flag-carrier to k_L and k_H. One may think of these as the utility provided by another low-cost carrier airline. Then denote u_L and u_H the willingness-to-pay (or added benefit) for travelling with the flag-carrier for leisure or business via economy class. And let $\triangle u_L$ and $\triangle u_H$ be the leisure and the business traveller's added benefit of travelling via business class. Following Arrow-Pratt, a high-valuation buyer is more risk-averse than a low-valuation buyer if raising quality from low to high relative to none to low leads to a smaller percentage increase in his willingness-to-pay. Expressed algebraically, business travellers are more risk-averse if $\triangle u_H/u_H < \triangle u_L/u_L$. This holds as long as the business traveller's gain from travelling u_H are sufficiently large. To apply the result to profit-maximization, we must instead investigate the implied risk preferences of surplus-utility, i.e., utility minus cost. Denoting c the cost attributed to economy class and $\triangle c$ the cost increase for business class, surplus is given as follows:

Surplus	Business Class	Economy Class	Outside Option
Business Traveller	$u_H + \triangle u_H + k_H - c - \triangle c$	$u_H + k_H - c$	k_H
Leisure Traveller	$u_L + \triangle u_L + k_L - c - \triangle c$	$u_L + k_L - c$	k_L

The same comparison of risk preferences, now applied to surplus-utility, then entails that

$$\frac{\triangle u_H - \triangle c}{u_H - c} < \frac{\triangle u_L - \triangle c}{u_L - c}.$$

If this holds, a single-contract menu is profit-maximizing: the seller only offers economy or business class tickets, but never both. How many business and leisure travellers there are does not change the optimality of single-contract menus. Standard single-crossing or increasing differences may hold: $u_H > u_L$ and $\triangle u_H > \triangle u_L$. And the ranking on risk preferences neither implies nor is implied by surplus being log submodular.[2]

The absence of multi-contract menus coincides with counterintuitive welfare implications: if offering both business and economy class is not profit-maximizing, it is neither surplus-maximizing. In this situation the monopolist seller will inadvertedly maximize social welfare if there are sufficiently many leisure travellers: all travellers will travel via business class. But if the number of business travellers is sufficiently large, the monopolist seller will instead exclude leisure travellers altogether. Only business travellers travel. In the latter case, consumer surplus can be raised by requiring that the seller must actively sell economy class tickets. This reverses the standard intuition: The seller does not capture the greatest profit by offering different versions. Instead, welfare is raised by obliging the seller to do so. □

Mathematically, a single-contract menu is a global expression of what we shall refer to as two-sided bunching. Consider a continuum of types and qualities. Two-sided bunching occurs locally on an interval of types when only one low-quality contract (possibly corresponding to the null contract or exclusion) and one high-quality contract are actively traded; in this case, the quality assignment is necessarily discontinuous at some cutoff type (see the right panel of Figure 1). More generally, discontinuous bunching occurs when the number

[2] Surplus is log submodular if both $\frac{\triangle u_H - \triangle c}{u_H + k_H - c} < \frac{\triangle u_L - \triangle c}{u_L + k_L - c}$ and $\frac{u_H - c}{k_H} < \frac{u_L - c}{k_L}$. Whether these hold depends on the utility of the outside option, k_L and k_H.

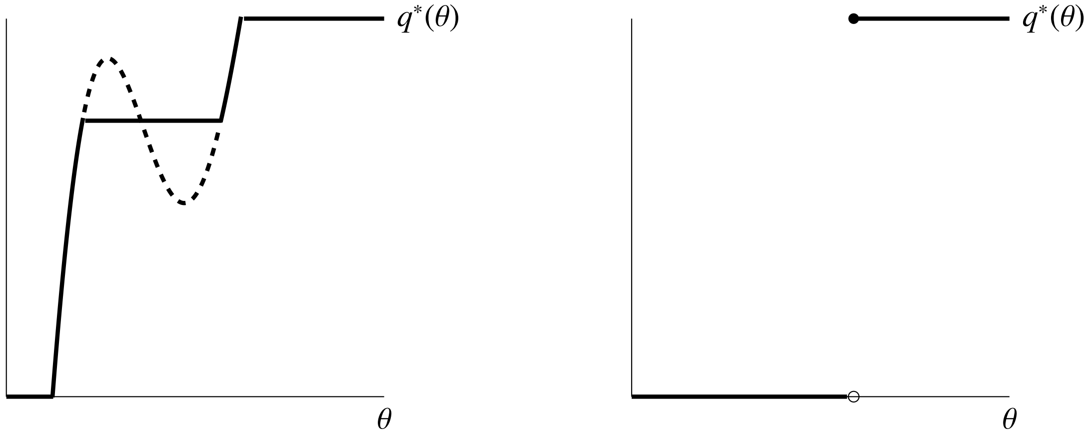

Figure 1: (Left) In Mussa-Rosen, surplus is $S(q;\theta) = \theta q - \frac{q^2}{2}$. The profit-maximizing quality assignment to types $\theta \mapsto q(\theta) \in [0,1]$ (left) is continuous and may involve continuous bunching. The dashed line depicts the virtual surplus maximizing quality. (Right) Stokey considers surplus $S(q;\theta) = \theta q$. A single-contract menu turns out to be profit-maximizing.

of discontinuities is not limited to one. The more familiar notion of bunching, by contrast, is continuous. Here many heterogeneous buyer types purchase the same contract, yet the quality assignment is continuous in types. An example is depicted in the left panel of Figure 1. While seemingly related, continuous and discontinuous bunching give rise to different predictions for the empirically observed price-quality tariff $p(q)$: Under discontinuous bunching, some qualities are simply not offered (or offered at the same price as some good of superior quality); in effect, the tariff $p(q)$ is a discontinuous step function, marginal prices ill-defined. Under continuous bunching the tariff $p(q)$ is continuous instead; marginal prices however exhibit a kink at the bunching quality.

The main contribution of this paper is to identify a condition on surplus-utility $q \mapsto S(q;\theta)$ (consumption utility minus cost) that characterizes which regime is profit-maximizing: two-sided (or discontinuous) bunching or continuous separation? This condition captures the degree of complementarity between the intensity of preferences over higher qualities and buyers' risk preferences. While the analysis generalizes to common value environments, the stronger result is found in the context of private values:[3] two-sided bunching is profit-maximizing (Proposition 1) for any interval of types where higher types' surplus is a concave transformation of lower types' surplus. If we interpret surplus as a utility, this is akin to saying that *single-contract menus are profit-maximizing if higher buyer types are more risk-averse in their surplus-utility in the sense of Arrow (1965)-Pratt (1964).* Then say that surplus is log submodular in differences (or that marginal surplus $D_q S(q;\theta)$ is log submodular). Conversely, in any interval where higher types' surplus is a convex transformation of lower types' surplus (so that higher types' surplus-utility represents less risk-averse preferences), the profit-maximizing quality assignment is continuous (Proposition 2). Following Bonneton and Sandmann (2022), say that surplus is log supermodular in differences (or that $D_q S(q;\theta)$ is log supermodular).[4] Both results hold irrespective of whether the hazard rate is monotone.

[3]Common values describe environments where the seller's costs vary with the buyer's type. Common values can arise due to selection effects or restrictions on pricing, which stipulate that prices must be identical in geographically distinct markets. In a private values environment, by contrast, the seller's cost and hence the profitability of a contract does not depend on the buyer's type.

[4]Log supermodularity in differences also drives sorting results in unrelated areas such as moral hazard and random search and matching (e.g., Chade and Swinkels (2019) and Bonneton and Sandmann (2022)).

The analysis yields two important economic implications. First, in a standard environment (where surplus is concave in quality) surplus is log submodular in differences only if the surplus-maximizing quality is identical for all buyers. In practice, this means that quality provision should be as large as technologically-feasible. From an economic point of view, this tends to be the case for information goods such as data, digital content, software, or medicine where unit production costs are of lesser importance. It tends to not be the case for goods that rely on manual labor or input of expensive raw materials. In consequence, the results in this paper predict that single-contract menus are most pervasive for information goods. Here an investigation of risk preferences is warranted: are high-valuation buyers such as financial institutions, highly-skilled programmers and graphic designers, or affluent medical patients relatively more severely affected (in terms of willingness-to-pay) by a system breakdown, loss of continued access to data, or end of medical treatment? If so, offering a single-contract menu that delivers only the highest-quality good at a high price is profit-maximizing. Consumer surplus is negatively affected by this pricing strategy, however, because high prices lead to the exclusion of some consumers. Here regulatory intervention that mandates offering quality-discounted versions of high-quality products may enhance consumer surplus.

A second insight relates to competitive menu pricing. Even if single-contract menus are profit-maximizing, they need not be entry-proof in the presence of operating costs. To see this consider a seller that offers a single quality contract only. To cover his operating costs, the seller must necessarily exclude some low-valuation buyers who he could otherwise serve at low unit production costs. While exclusion may be profit-maximizing, it also creates an incentive for competitors to enter the market and sell to its lower end. Such entry, in turn, undermines the incumbent's demand from intermediate-valuation buyers. Since no two identical sellers could be simultaneously active in a competitive equilibrium, it follows that the incumbent must offer a menu that covers the entire market. Yet doing so serves the purpose of entry-deterrence, not profit-maximization.

On a technical note, it is tempting to attribute the prevalence of single-contract menus to a convex objective function (on this theme note the important contribution by Kleiner et al. (2021)): two-sided bunching describes an extreme point in the assignment function space, and extreme points are commonly associated with convex objective functions. This is not the case here. The seller's objective is to maximize virtual surplus. But surplus being log submodular in differences does not imply that virtual surplus is convex. To the contrary, it could even be concave.

Finally, we ask: when are single-contract menus profit-maximizing in a common values environment? In a strict sense the answer is negative: Theorem 1 asserts that the profit-maximizing assignment must either coincide with the pointwise maximum of virtual surplus, or satisfy discontinuous bunching but not necessarily two-sided bunching. Yet discontinuous bunching does not imply that a single-contract menu is profit-maximizing. To illustrate, we study an example where a generalized notion of log submodularity in differences holds so that virtual surplus is quasi-convex. Offering two and not one contract is profit-maximizing nonetheless. Identifying the optimal points of discontinuity in this environment is non-trivial. Theorem 2 characterizes the optimal points of discontinuity with a familiar (yet generalized) tool from the literature: the (hybrid) maximum principle (see the textbook by Clarke (2013) and references therein). This result from optimal control theory allows for discontinuities in the controlled state (see Theorem 3) and may be of independent interest in related fields.

Related Literature: Beginning with Salant (1989), many authors have sought to reconcile the seemingly contradictory findings by Mussa and Rosen (1978) and Stokey (1979).[5] The focus on complementarity in the surplus function is a common theme. Anderson and Dana (2009) provide a set of (more restrictive) regularity conditions that imply the ranking of risk preferences identified in this article. They assert that single-contract menus are profit-maximizing if surplus is log submodular and reservation utilities are zero. Said differently, the surplus ratio must be decreasing in types. Haghpanah and Hartline (2021) study the optimality of single-contract menus in the related context of bundling (which corresponds to menu-pricing with a given finite set of qualities). Their analysis presents a sufficient condition for the optimality of the grand bundle in terms of a stochastic order over the ratio of surplus. This paper shows that the key economic condition is a ranking of risk preferences depending on both surplus utility and reservation utilities; if reservation utilities are zero this ranking corresponds to log submodular surplus.[6] As a byproduct, the optimality of single-contract menus is easy to interpret and admits a simple proof.

Bunching has been extensively studied, albeit only when virtual surplus is concave or quasi-concave so that the profit-maximizing assignment $\theta \mapsto q(\theta)$ is continuous (Lollivier and Rochet (1983), Guesnerie and Laffont (1984), and Jullien (2000)). By contrast, when virtual surplus is quasi-convex, discontinuous bunching occurs so that the classical maximum principle (which requires continuity of the quality assignment) does not apply. By drawing on the hybrid maximum principle I show that the well-known characterization of the profit-maximizing assignment generalizes and characterizes the discontinuities.

Corollary 2 on random contracts builds on Strausz (2006). Relatedly, Maskin and Riley (1984) (Assumption 3 condition (21)) show that deterministic contracts are profit-maximizing if higher valuation buyers are less risk-averse. Strausz (2006) (Proposition 2) shows this if the solution to the seller's relaxed optimization problem with deterministic menus satisfies the standard monotonicity constraint on quality. But this need not be the case when surplus is log submodular in differences.

Several directions are left unexplored: unlike in classical insurance (Chade and Schlee (2020)), utility is linear in transfers. Dilmé and Garrett (2022) show that with CARA utility and identical risk preferences across buyers, single-contract menus are optimal if utility is close to being linear. Second, contracts are two-dimensional. Gershkov et al. (2022) study multi-dimensional screening with Yaari dual utility.[7,8]

The analysis of competitive menu pricing complements Johnson and Myatt (2006). They show that firms compete on all qualities if it is profit-maximizing to offer a multi-contract menu (see their Footnote 8). This paper shows that this can also happen as a means of entry-deterrance even if a single-contract menu is profit-maximizing.

[5] Villas-Boas (1998) and Deneckere and McAfee (1996) analyze the two-type case.

[6] Consider, for instance, surplus $S(q;\theta) = \theta + q + \theta q - \alpha q^2$. This is log submodular for $\alpha < \frac{1}{3}$, but does not satisfy Anderson and Dana (2009)'s regularity conditions. $S(q;\theta) = \theta + q + \theta q$ is however log supermodular in differences and satisfies increasing differences, whence Proposition 2 in this paper establishes that the profit-maximizing assignment is continuous. Figure 3 illustrates this point.

[7] Bergemann et al. (2022) embed Mussa and Rosen (1978)'s original screening problem in an information design problem where the seller chooses the buyers' information structure. This changes the properties of the profit-maximizing menu. They provide conditions including convex cost under which single-contract menus are profit-maximizing, whereas in the original model the profit-maximizing assignment is continuous and features separation (because surplus is log supermodular in differences, see Proposition 2).

[8] Doval and Skreta (2022) study menu pricing by monopolist seller without commitment in a repeated trade relationship where buyers' valuations are persistent over time. Two-sided bunching is profit-maximizing because it entails a commitment not to learn.

2 The Model

2.1 Set-up

Consider a single seller that sells a variety of different qualities to a continuum of differentiated buyer types $\theta \in [0,1]$. We denote $M(\theta)$ the distribution of θ and $\mu(\theta)$ its density.[9] As in Mussa and Rosen (1978), the market is impersonal in that all buyers pay the same price for the same quality.

Contracts. Each contract specifies a quality $q \in [0,1]$, a price $p \in \mathbb{R}$, and a participation dummy $x \in \{0,1\}$.[10]

Preferences and cost. The seller's cost of producing a quality of variety q and selling it to type θ is $C(q;\theta)$. Buyer type θ's preferences over a price-quality pair (p,q) are represented by quasi-linear utility $U(q;\theta) - p$. We denote $\hat{u}(\theta)$ the buyer's utility when not purchasing.

We impose three standard assumptions. First, the utility and cost functions must be sufficiently smooth.[11]

Assumption 1. *$(\theta,q) \mapsto U(q;\theta)$, $(\theta,q) \mapsto D_\theta U(q;\theta)$, $\theta \mapsto \hat{u}(\theta)$, $(\theta,q) \mapsto C(q;\theta)$, and $\theta \mapsto \mu(\theta)$ are $\mathscr{C}$.*

Second, utility satisfies increasing differences so that higher buyer types have greater marginal utility.

Assumption 2 (increasing differences)**.** *$D_\theta U(q;\theta)$ is increasing in q for all $q \in [0,1]$ and $\theta \in [0,1]$.*

Third, utility and surplus $S(q;\theta) = U(q;\theta) - C(q;\theta)$ are increasing when below the first-best.

Assumption 3. *$q \mapsto U(q;\theta)$ and $q \mapsto S(q;\theta)$ are increasing on $[0, \underline{q}^{fb}(\theta)]$ where $\underline{q}^{fb}(\theta) = \min \operatorname{argmax}_{q\in[0,1]} S(q;\theta)$.*

The seller's problem. The seller proposes a menu, i.e., a set of contracts $\{p(\theta), q(\theta), x(\theta)\}_{\theta\in[0,1]}$.[12] Profit is

$$\int_0^1 x(\theta)[p(\theta) - C(q(\theta);\theta)]\mu(\theta)d\theta.$$

As usual, any admissible menu satisfies incentive and participation constraints so that each agent prefers the contract tailored for his type θ over any other type θ': If $x(\theta) = 0$, then $\hat{u}(\theta) \geq U(q(\theta');\theta) - p(\theta')$ for all $\theta' : x(\theta') = 1$. If instead $x(\theta) = 1$, then $U(q(\theta);\theta) - p(\theta) \geq U(q(\theta');\theta) - p(\theta')$ for all $\theta' : x(\theta') = 1$ and $U(q(\theta);\theta) - p(\theta) \geq \hat{u}(\theta)$.

[9] It is without loss to assume that θ is uniformly distributed. Given a cdf $M(\theta)$ it suffices to re-parameterize the model: denote $x \in [0,1]$ the agent's type and set $u(q;x) = U(q;M^{-1}(x))$ and $c(q;x) = C(q;M^{-1}(x))$. The underlying distribution of x is uniform, and u and c satisfy single-crossing and the complementarity assumptions CV-LSD or CV-LsubD if and only if $U(q;\theta)$ and $C(q;\theta)$ do.

[10] The selection of admissible qualities precedes pricing decisions and is strategically irrelevant for a monopolist seller.

[11] We write $\mathscr{C}$ if a function is continuous.

[12] Applications of menu pricing are not limited to price-quality contracts as in Mussa and Rosen (1978). Maskin and Riley (1984) give quality q the interpretation of quantity. Stokey (1979) studies intertemporal price discrimination. Quality q is the time of purchase, the agent's type encodes time preferences. In the taxation literature (Mirrlees (1971)) the agent's type is his ability to generate taxable income, and quality is net income. In the regulation literature (Baron and Myerson (1982)) the principal is a regulator, the agent a firm that produces at a privately known cost. Under menu pricing over lotteries absent expected utility (see the generalized local bilinear utility model in DeJarnette et al. (2020)), quality corresponds to the objective probabilities of outcomes x and x' as given by q and $1-q$, $\pi(q;\theta)$ and $1-\pi(q;\theta)$ are subjective probability weights and $u(x)$ and $u(x')$ the utility attached to the certain outcomes x and x'. Then $U(q;\theta) = \pi(q;\theta)u(x) + (1-\pi(q;\theta))u(x')$. The results presented in this paper apply to all applications of the principal-agent model with hidden information and partial exclusion that maintain the standard assumption of quasi-linearity of utility and profit.

2.2 Implementability and Virtual Surplus

As a first step we would like to reduce the dimensionality of the problem. Rather than optimizing over both $\theta \mapsto q(\theta)$ and $\theta \mapsto p(\theta)$, the choice of one should imply the other. To that end say that a quality assignment $\theta \mapsto q(\theta)$ is implementable if there exist transfers $\theta \mapsto p(\theta)$ so that the menu is incentive-compatible. The following is well-known.

Lemma 1. *Posit Assumption 2. An assignment $q(\theta)$ is implementable if and only if it is non-decreasing.*

In a model with a continuum of types, the non-decreasing quality assignment alone fully characterises prices in any interval where participation constraints are slack; the implementing prices are uniquely determined by a differential equation that states that incentive constraints are locally binding. The idea of locally binding incentive constraint can be more easily grasped when there are finitely many types $\theta \in \{\theta_1, ..., \theta_N\}$ only. Transfers are no longer uniquely determined; incentive constraints can either be upward or downward binding. The latter turns out to be profit-maximizing. This means that, for a given quality assignment, the seller chooses prices so that buyer type θ_{j+1} is indifferent between contracts $(p(\theta_j), q(\theta_j))$ and $(p(\theta_{j+1}), q(\theta_{j+1}))$.

Thus equipped, we can recall the key insight from the literature (the unfamiliar sum allows to incorporate non-standard participation constraints as in Jullien (2000)): The seller's optimization problem is to maximize surplus adjusted to account for the agent's informational rent. This objective, given here in the square braces, is often referred to as virtual surplus and henceforth denoted $\Lambda(q;\theta)$.

Lemma 2. *Posit Assumptions 1 and 2. Fix a non-decreasing quality assignment $\theta \mapsto q(\theta)$ and let $\{(\underline{\theta}_k, \overline{\theta}_k)\}$ be a finite or countable collection of maximal disjoint participation intervals in $[0,1]$; i.e., $x(\theta) = 1$ on $(\underline{\theta}_k, \overline{\theta}_k)$ and $\lim_{\theta \uparrow \underline{\theta}_k} x(\theta) = \lim_{\theta \downarrow \overline{\theta}_k} x(\theta) = 0$ for all k. The seller's profit under the unique implementing transfers $\theta \mapsto \pi(\theta)$ is*

$$\Pi(q) = -\sum_k \left\{ \hat{u}(\underline{\theta}_k) + \int_{\underline{\theta}_k}^{\overline{\theta}_k} \left[S(q(\theta);\theta) - D_\theta U(q(\theta);\theta) \frac{\int_\theta^{\overline{\theta}_k} \mu(\vartheta) d\vartheta}{\mu(\theta)} \right] \mu(\theta) d\theta \right\}. \tag{1}$$

3 Private Values

We first explore private value environments where, as in Mussa and Rosen (1978), the seller's cost $C(q;\theta) \equiv C(q)$ do not depend on the buyer's type.

The analysis of private values is appealing because the seller's problem can be reformulated as an optimization problem in which the seller incurs no production cost. To see this, define a transfer-quality-participation menu $\{\pi(\theta), q(\theta), x(\theta)\}_{\theta \in [0,1]}$ where a transfer corresponds to the seller's profit, $\pi(\theta) = p(\theta) - C(q(\theta))$. In keeping with this interpretation the buyer's utility is given by surplus $S(q;\theta) = U(q;\theta) - C(q)$. Since $D_\theta S(q;\theta) = D_\theta U(q;\theta)$, surplus-utility satisfies increasing differences. The seller then chooses a menu that maximizes

$$\int_0^1 x(\theta)\pi(\theta)\mu(\theta)d\theta \quad \text{such that} \quad S(q(\theta);\theta) - \pi(\theta) \geq \max\{S(q(\vartheta);\theta) - \pi(\vartheta); \hat{u}(\theta)\} \quad \forall\, \theta, \vartheta : x(\theta) = x(\vartheta) = 1.^{13}$$

[13] It is without loss of generality to ignore the participation constraint of non-participating types: if some actively traded transfer-quality pair was desirable to an excluded type, including said type would further increase the seller's profit. This is a consequence of private values and Assumption 3.

3.1 A Novel Representation of the Seller's Profit

Unlike the standard approach, we now pursue another representation of the seller's objective that separates surplus-utility, $S(q(\theta);\theta)$, from the distribution over types, $\mu(\theta)$. From a technical point of view, one may think of it as one iteration of integration by parts away from the preceding virtual surplus representation.[14,15]

Lemma 3. *Posit Assumptions 1 and 2. Fix a non-decreasing quality assignment $\theta \mapsto q(\theta)$ and let $\{(\underline{\theta}_k, \overline{\theta}_k)\}$ be a finite or countable collection of maximal disjoint participation intervals in* $[0,1]$. *The seller's profit under the unique implementing transfers $\theta \mapsto \pi(\theta)$ is*

$$\Pi(q) = \sum_k \left\{ \Big[S\big(\lim_{\theta \downarrow \underline{\theta}_k} q(\theta); \underline{\theta}_k\big) - \hat{u}(\underline{\theta}_k) \Big] \Big(M(\overline{\theta}_k) - M(\underline{\theta}_k) \Big) + \lim_{N \to \infty} \sum_{\ell=1}^{N} \Big[S(q(\theta_{\ell,k}); \theta_{\ell,k}) - S(q(\theta_{\ell-1,k}); \theta_{\ell,k}) \Big] \Big(M(\overline{\theta}_k) - M(\theta_{\ell,k}) \Big) \right\} \tag{2}$$

for all k and any $\underline{\theta}_k = \theta_{0,k} < \theta_{1,k} < \ldots < \theta_{N,k} = \overline{\theta}_k$ so that $|\theta_{j+1,k} - \theta_{j,k}| < \delta_N$ where $\delta_N \to 0$ as $N \to \infty$.

The representation (2) is more transparent when participation constraints are slack for all but one type, e.g., $\hat{u}(\theta) = S(0;\theta)$ (see also condition (4)). The finite type case further rids the representation of the seller's objective of its technical dimension.

Lemma 3′. *Posit Assumption 2 and consider finitely many types $\theta \in \{\theta_0, \ldots, \theta_N\}$ so that type θ_0 is excluded and the participation constraint is binding for θ_1 and slack for all $\ell > 1$. Fix a non-decreasing quality assignment $(q(\theta_\ell))_{\ell=1}^N$. The seller's profit under the profit-maximizing transfers $(\pi(\theta_\ell))_{\ell=1}^N$ is*

$$\Big[S(q(\theta_1); \theta_1) - \hat{u}(\theta_1) \Big] \Big(M(\theta_N) - M(\theta_0) \Big) + \sum_{\ell=2}^{N} \Big[S(q(\theta_\ell); \theta_\ell) - S(q(\theta_{\ell-1}); \theta_\ell) \Big] \Big(M(\theta_N) - M(\theta_{\ell-1}) \Big). \tag{2'}$$

From an economic point of view, the representation captures succinctly the trade-offs associated with menu pricing. The central observation is that downward-binding incentive constraints imply that marginal surplus must equal marginal per-type profit:

$$S(q(\theta_j); \theta_j) - S(q(\theta_{j-1}); \theta_j) = \pi(\theta_j) - \pi(\theta_{j-1}).$$

Then consider providing better quality to θ_j. Doing so increases θ_j's marginal surplus, whence marginal profit earned from buyers of type θ_j. At the same time, it reduces θ_{j+1}'s marginal surplus, whence marginal profit earned from buyers of type θ_{j+1} and above. Increasing differences then implies that total profit for types greater than θ_j goes down. But this is the familiar trade-off: an increase in quality provision leads to higher profit from marginal types θ_j, but lower profit from inframarginal types.

[14] If there exists a partition $0 = \theta_0 < \theta_1 < \ldots < \theta_n = 1$ so that $q \in \mathscr{C}^1(\theta_j, \theta_{j+1})$, then $\lim_{N\to\infty} \sum_{\ell=1}^N [S(q(\theta_{\ell,k}); \theta_{\ell,k}) - S(q(\theta_{\ell-1,k}); \theta_{\ell,k})](M(\overline{\theta}_k) - M(\theta_{\ell,k}))$ is equal to $\int_{\underline{\theta}_k}^{\overline{\theta}_k} D_q S(q(\theta);\theta) \dot{q}(\theta)(M(\overline{\theta}_k) - M(\theta)) d\theta + \sum_{j=1}^n (S(q(\theta_j^+);\theta_j) - S(q(\theta_j^-);\theta_j))(M(\overline{\theta}_k) - M(\theta))$ where $q(\theta_j^+) = \lim_{\theta \downarrow \theta_j} q(\theta)$ and $q(\theta_j^-) = \lim_{\theta \uparrow \theta_j} q(\theta)$ (see Iorio and Iorio (2001), Section 3.2, Proposition 3.32). But this pre-supposes a lot of regularity on the profit-maximizing quality assignment $\theta \mapsto q(\theta)$.

[15] Observe that single-contract menus are the extreme points among all non-decreasing functions $q : [0,1] \to [0,1]$. It follows that profit is maximized by a single-contract menu if $\Pi(q)$ is convex. It is then tempting to deduce from Lemma 3 that the convexity of $q \mapsto D_q S(q;\theta)$ is a sufficient condition for single-contract menus to be profit-maximizing. This turns out to be wrong. As footnote 14 points out, single-contract menus lead to discontinuities that have to be accounted for. The surplus function $S(q;\theta) = \alpha q + \frac{1}{3}\theta q^3$ illustrates this point: $D_q S(q;\theta) = \alpha + \theta q^2$ is convex. Yet surplus is log supermodular in differences, so Proposition 2 establishes that the profit-maximizing quality assignment must be continuous.

3.2 Log Submodularity in Differences and Risk Preferences

In Mussa and Rosen (1978), it is taken for granted that the optimal policy "smokes out" consumer preferences and assigns (under appropriate distributional assumptions) different buyer types to distinct quality varieties. But, as Stokey (1979) illustrates, offering a single as opposed to many contracts may be profit-maximizing. We now seek to delineate both cases. We will see that whether continuous screening or discontinuous bunching is profit-maximizing depends on a comparison of the curvature of surplus across buyer types.

Definition 1. *Surplus is weakly log submodular in differences (weak LsubD) on* $[\underline{\theta}, \overline{\theta}] \subseteq [0, 1]$*, if, for all* $\theta_H > \theta_L$ *in* $[\underline{\theta}, \overline{\theta}]$ *and* $q^{fb}(\theta_L) \geq q_1 > q_2 > q_1 \geq 0$*,*

$$\frac{S(q_3;\theta_H) - S(q_2;\theta_H)}{S(q_2;\theta_H) - S(q_1;\theta_H)} \leq \frac{S(q_3;\theta_L) - S(q_2;\theta_L)}{S(q_2;\theta_L) - S(q_1;\theta_L)}.$$

Surplus is log submodular in differences if the preceding inequality is strict for all types and qualities.

This condition is well-known in the economics literature.[16] In particular, weak LsubD admits an interpretation in terms of risk preferences. Interpret $q \mapsto S(q;\theta)$ as agent type θ's utility function. Then, owing to a theorem by Pratt (1964), the following are equivalent to weak LsubD:

1. Marginal surplus $D_q S(q;\theta)$ is weakly log submodular;
2. $q \mapsto S(q;\theta_H)$ is a concave transformation of $q \mapsto S(q;\theta_L)$ for all $\theta_H > \theta_L$;
3. Buyer type $\theta_H > \theta_L$ is weakly more risk-averse than buyer type θ_L; that is, θ_H does not accept a lottery over qualities that is rejected by θ_L in favor of a certain quality q.

If $U(q;\theta) = \theta q$ as assumed by Mussa and Rosen (1978), one can verify that LsubD holds if and only if $C(q)$ is concave.

From an economic point of view it is worth pondering a key implication of surplus satisfying increasing differences and log submodularity in differences. Both conditions jointly imply that social welfare is maximized when giving every consumer the same quality assignment. This is for instance the case when surplus is $S(q;\theta) = \theta q$. It is also a reasonable assumption for information goods such as data, digital content, software, or medicine where unit production cost is of lesser importance as well as many of the examples discussed in Shapiro and Varian (1998) in their Chapter 3 on versioning. (Also refer to the discussion in the introduction.)

Remark 1. *Posit Assumptions 1, 2, 3. Suppose that surplus is weakly log submodular in differences. Then the first-best quality assignment is constant in types.*

Proof. Following Topkis (1998), the first-best $\theta \mapsto q^{fb}(\theta)$ is non-decreasing under increasing differences. Thus suppose by contradiction that $q_2 \equiv q^{fb}(\theta_2) > q^{fb}(\theta_1) \equiv q_1$. Then without loss $0 = D_q S(q_2;\theta_2) \geq D_q(q_1;\theta_1)$. And so $D_q S(q_2;\theta_1) > 0$ and $D_q S(q_1;\theta_2) < 0$ due to increasing differences. In effect it holds that

$$0 = D_q S(q_2;\theta_2) D_q S(q_1;\theta_1) > D_q S(q_2;\theta_1) D_q S(q_1;\theta_2)$$

Whence, due to Pratt (Items 1. and 3.), surplus cannot be weakly log submodular in differences. Absurd. □

[16] Equivalently (see Proposition 7 in the textbook by Gollier (2004)) *surplus is weakly log submodular in differences if and only if* $q \mapsto \frac{D^2_{q\theta} S(q;\theta)}{D_q S(q;\theta)}$ *is non-increasing.*

3.3 A Local Result: Two-Sided Bunching

We first prove a more general result that narrows down the optimality of single-contract menus to a subset of types. Our characterization requires the concept of two-sided bunching. Two-sided bunching means that within a given type interval there is pooling at the top and at the bottom.

Definition 2. *There is two-sided bunching on* $[\underline{\theta}, \overline{\theta}]$ *if the quality assignment* $\theta \mapsto q(\theta)$ *is constant on* $[\underline{\theta}, \theta^*)$ *and* $(\theta^*, \overline{\theta}]$ *for some* $\theta^* \in [\underline{\theta}, \overline{\theta}]$.

Locally, our ranking on the concavity of surplus then implies the following:

Proposition 1. *Posit Assumptions 1, 2 and 3. And suppose that participation constraints are slack on* $(\underline{\theta}, \overline{\theta})$*. If surplus is (weakly) log submodular in differences for all types in* $(\underline{\theta}, \overline{\theta}) \subseteq [0, 1]$*, then (some) any profit-maximizing quality assignment* $\theta \mapsto q^*(\theta)$ *satisfies two-sided bunching in such interval.*

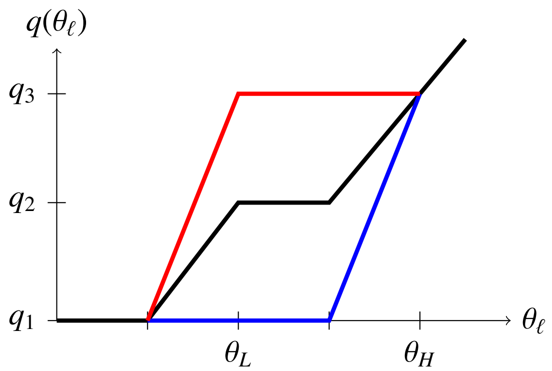

Figure 2: The figure illustrates the proof of Proposition 1. Given a candidate profit-maximizing mechanism (middle), one constructs enveloping deviations that locally sandwich the candidate mechanism.

Pratt's characterization of comparative risk preferences allows to formulate an intuitive proof of this result: Suppose that buyer types $\theta_H > \theta_L > \theta_{L-1}$ consume qualities $q_3 > q_2 > q_1 \geq 0$. This corresponds to the black quality assignment depicted in Figure 2. Then consider the lottery that with probability P assigns quality q_3 and with probability $1 - P$ assigns quality q_1 to buyer type θ_L. In Figure 2, this corresponds to randomizing over the blue and the red quality assignment. Clearly, P can be chosen so that from type θ_L's point of view the sure quality q_2 is the lottery's certainty equivalent. That is, type θ_L is prepared to pay just as much for the random quality as for the sure quality q_2. If buyer type θ_H is strictly more risk-averse than θ_L, the same cannot be said about θ_H: θ_H is strictly better off under the sure quality q_2. In effect, by replacing the sure quality q_2 with the described lottery, one strengthens the high type's incentive constraint. This, in turn, allows the seller to increase the transfer from θ_H without upsetting incentive constraints.

The proof discussed in the Appendix holds without recourse to random contracts. We will show that if surplus is log submodular in differences, then the candidate quality assignment cannot yield greater profit than both enveloping deterministic deviations depicted in Figure 2. This suggests that the seller could profitably discard the intermediate-quality contract in favor of either the greater or the lesser quality provided to adjacent buyer types.

3.4 When Is a Single-Contract Menu Profit-Maximizing?

Let us now address the main question of this paper: the optimality of single-contract menus. Informed by Proposition 1, we deduce that single-contract menus are profit-maximizing if (i) the ranking of risk preferences holds not just locally but for all types, and (ii) the seller faces no competitors so that $\hat{u}(\theta) = S(0;\theta)$.

Inspection of the probabilistic proof of Proposition 1 reveals that this result can be strengthened, or, equivalently, conditions (i) and (ii) can be relaxed: First (i), consider the subset of enveloping lotteries depicted in Figure 2. Here we compared a sure quality (then denoted q_2) against lotteries that admitted only two possible realizations: upside risk (then denoted q_3) and downside risk (then denoted q_1). And, crucially, in the context of single-contract menus, q_1 is pinpointed as the outcome of not purchasing at all. Whence, rather than considering all lotteries, it suffices to compare a sure contract offering quality q_L and transfer t' with a random contract offering either $q_H > q_L$ with some probability P or no purchase at all with probability $1-P$ at a transfer t''. We then require that if a consumer with surplus utility $q \mapsto S(q;\theta_H)$ and outside option $\hat{u}(\theta_H)$ prefers the lottery over the sure contract, then so does a consumer with surplus utility $q \mapsto S(q;\theta_L)$ and outside option $\hat{u}(\theta_L)$. Following Pratt, an algebraic representation of this ranking of risk preferences is:

$$\frac{S(q_H;\theta_H) - S(q_L;\theta_H)}{S(q_L;\theta_H) - \hat{u}(\theta)} \leq \frac{S(q_H;\theta_L) - S(q_L;\theta_L)}{S(q_L;\theta_L) - \hat{u}(\theta)} \qquad \text{for all } \theta_H > \theta_L \text{ in } [0,1] \text{ and } 1 \geq q_H > q_L > 0. \tag{3}$$

Second (ii), we require that participation constraints bind only once so that $x(\theta)$ is non-decreasing. A sufficient condition is for higher types to exceed their reservation utility when selecting a participating lower type's contract. This extends the notion of increasing differences to reservation utilities:

$$S(q;\theta_H) - \hat{u}(\theta_H) \geq S(q;\theta_L) - \hat{u}(\theta_L) \qquad \text{for all } \theta_L \in [0,1] \text{ and } q \in [0,1] \text{ so that } S(q;\theta_L) \geq \hat{u}(\theta_L). \tag{4}$$

Corollary 1. *Posit Assumptions 1, 2, 3. Suppose that surplus and reservation utilities satisfy conditions* (3) *and* (4)*. Then a single-contract menu is profit-maximizing: buyers either purchase the highest quality or nothing.*[17]

Note that when normalizing $\hat{u}(\theta) = 0$ for all $\theta \in [0,1]$, Condition (3) holds if surplus is log submodular (as in Anderson and Dana (2009)).[18] But this obfuscates the link to risk preferences. What happens when outside options change?

Example. *To illustrate the role of outside options, consider two parametrizations of the earlier example:*

Surplus	*Business Class*	*Economy Class*	*Outside Option*
Business Traveller	8	5	0 *or* 3
Leisure Traveller	5	3	0 *or* 1

In one model (left), there is no low-cost carrier and we normalize the outside option to zero. Surplus is log submodular, whence a single-contract menu is profit-maximizing. In the other model (right) there is a low-cost carrier and outside options have improved. Surplus continus to be log submodular, but not in differences. So in response to the improved outside option and depending on the number of business and economy travellers, the flag-carrier may now want to offer both business and economy class tickets.

[17] If, in addition, (3) is strict, then any profit-maximizing menu is a single contract.

[18] Corollary 1 is a stronger result than Proposition 3 in Anderson and Dana (2009) in that it dispenses with their regularity conditions on the sign of higher-order derivatives of surplus and the distribution function.

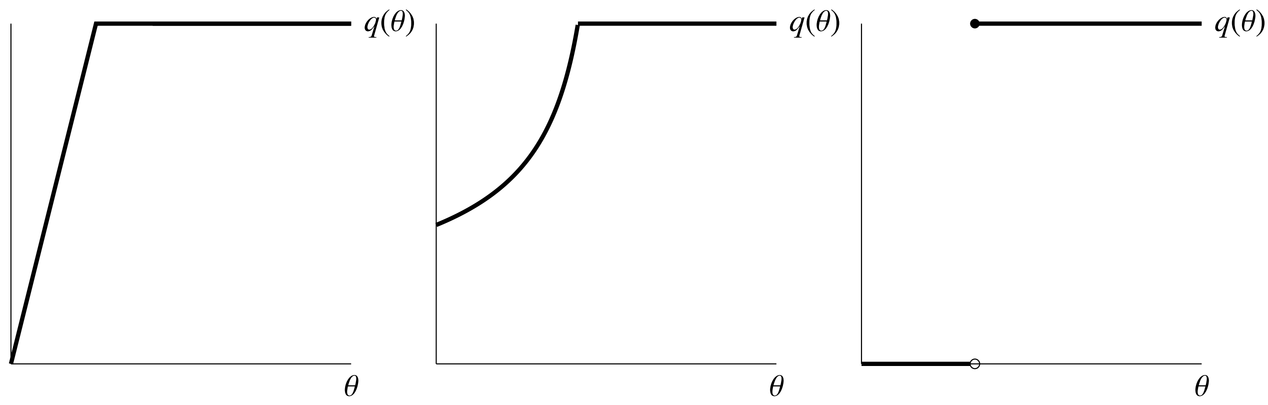

Figure 3: (Left) Consider utility $U(q;\theta) = \theta + q + \theta q$ and cost $C(q) = \frac{1}{4}q^2$. Further assume that types are uniformly distributed throughout. Surplus $S(q;\theta) = \theta + q + \theta q - \frac{1}{4}q^2$ is log submodular, yet the profit-maximizing quality assignment is separating and continuous. (Also note that the seller chooses not to exclude any buyer type. But this is a result, not an assumption.) This shows that it is not log submodularity but its namesake in differences and its implied ranking of risk preferences that determines as to whether single-contract menus are profit-maximizing. (Middle) Consider $S(q;\theta) = \frac{1}{6}q + \frac{1}{3}\theta q^3$. The objective function appears to be convex in q, so we could expect the profit-maximizing solution to be bang-bang. This is not true. Proposition 2 applies because surplus is log supermodular in differences. (Right) Consider concave surplus $S(q;\theta) = \alpha q + 2\theta q - \theta qq$ with $\alpha \in (0, 1]$. In particular, virtual surplus is not convex in quality. Nonetheless surplus is log submodular in differences, whence Proposition 1 implies that the profit-maximizing menu (with $q \in [0, 1]$) consists of a single contract.

Random Contracts Not allowing the seller to screen via random contracts, yet invoking random contracts for intuition is not entirely satisfactory. We finally show that restricting attention to deterministic menus is without loss of generality (assuming that there are constant marginal cost as in Maskin and Riley (1984)).

Corollary 2. *Posit Assumptions 1, 2, 3. Suppose that surplus and reservation utilities satisfy conditions* (3) *and* (4). *Further assume that marginal cost are constant. If surplus is weakly log submodular in differences, then within the larger set of random contracts a deterministic contracts is still profit-maximizing.*

Proof. In light of Lemma 3, we show this for finitely many types only. Fix a candidate (random) menu. Since marginal cost are constant, it is without loss to consider determinstic transfers. Then consider Q_j the certainty equivalent quality of buyer type θ_j's random contract. And observe that there exists another lottery with identical certainty equivalent Q_j with support in $q = 1$ and, depending on which one is less, $S(0;\theta_j)$ or $\hat{u}(\theta_j)$, $q = 0$ or no trade. This lottery is more risky. And due to (3) and LsubD, high valuation buyers weakly prefer the initial lottery over the newly constructed lottery. Then, following Strausz (2006), study the relaxed problem ((14) in his paper) that only considers the local downward binding constraints. Since the relaxed problem is less constrained, profit must be weakly greater than when considering all incentive constraints. In this problem, increasing the riskiness of intermediate qualities allows the seller to increase requested transfers from the more risk-averse high-valuation buyers. In effect, for any random menu there exists another (possibly not incentive-compatible) random menu with deterministic transfers and random qualities solely characterized by the probability $\hat{q}_j$ of assigning quality 1 instead of the lesser of 0 or no trade to type θ_j that guarantees (weakly) more profit to the seller. Then define $\sigma(\hat{q}_j, \theta_j) = \min\{\hat{u}(\theta_j); S(0;\theta_j)\} + \hat{q}_j(S(1;\theta_j) - \min\{\hat{u}(\theta_j); S(0;\theta_j)\})$. σ is weakly log submodular in differences and participation constraints are slack for intermediate types due to (4), whence Proposition 1 asserts that probabilities $\hat{q}_j \in \{0, 1\}$ attain maximal profit in the relaxed problem. □

3.5 When Is Screening Profit-Maximizing?

The optimality of single-contract menus is an appealing proposition because it holds irrespective of distributional assumptions. While presented in the standard context where there is a continuum of buyer types, the proof does not require it. Our ranking on risk preferences guaranteed that the seller would never offer more than one contract, irrespective of whether there are mass points, or finitely many buyer types only. Establishing a counterpoint to two-sided bunching will require more structure on the type space. As is well-known, with finitely many types the prevalence of multi-contract menus also depends on the distribution of buyer types (Anderson and Dana (2009), Proposition 1). We shall now see that with a continuum, a distribution-free result holds if one is willing to assume that the density is continuous (as asserted by Assumption 1).

We analyze the case where surplus is log supermodular in differences (LSD) so that higher types' surplus-utility represents weakly less risk-averse preferences.[19]

Definition 3. *Surplus is log supermodular in differences (LSD) on* $[\underline{\theta}, \overline{\theta}] \subseteq [0, 1]$*, if, for all* $\theta_H > \theta_L$ *in* $[\underline{\theta}, \overline{\theta}]$ *and* $0 \leq q_1 < q_2 < q_3 \leq \underline{q}^{fb}(\theta_L)$*,*

$$\frac{S(q_3; \theta_H) - S(q_2; \theta_H)}{S(q_2; \theta_H) - S(q_1; \theta_H)} > \frac{S(q_3; \theta_L) - S(q_2; \theta_L)}{S(q_2; \theta_L) - S(q_1; \theta_L)}.$$

As before, we may interpret $q \mapsto S(q; \theta)$ as agent type θ's utility function. Then, owing to Pratt (1964), the following are equivalent to weak LSD:

1. Marginal surplus $D_q S(q; \theta)$ is weakly log supermodular;
2. $q \mapsto S(q; \theta_H)$ is a convex transformation of $q \mapsto S(q; \theta_L)$ for all $\theta_H > \theta_L$;
3. Buyer type $\theta_H > \theta_L$ is weakly less risk-averse than buyer type θ_L; that is, θ_H does not reject a lottery over qualities that is accepted by θ_L in favor of a certain quality q.

If $U(q; \theta) = \theta q$ one can verify that LSD holds if and only if $C(q)$ is convex. Fittingly, this is the assumption made by Mussa and Rosen (1978).

I then claim that screening by means of offering several contracts (in fact a continuum) is always profit-maximizing (provided that the density is continuous, as asserted by Assumption 1).

Proposition 2. *Posit Assumptions 1, 2 and 3. And suppose that participation constraints are slack on* $(\underline{\theta}, \overline{\theta})$*. If surplus is log supermodular in differences for all types in* $(\underline{\theta}, \overline{\theta}) \subseteq [0, 1]$*, then any profit-maximizing quality assignment* $\theta \mapsto q^*(\theta)$ *is continuous in* $(\underline{\theta}, \overline{\theta})$*.*

Economic interest derives from the fact that a continuous quality assignment guarantees that the mapping from qualities into prices is also continuous: purchasing slightly better qualities increases the price incrementally only. The result can be proved via a number of perturbations that are commonplace in optimal control theory (see the online appendix).[20,21] However, Proposition 2 is subsumed by a stronger and more intuitive result (Theorem 1, item 1). We shall study it next.

[19] Equivalently, *surplus is log supermodular in differences if and only if* $q \mapsto \frac{D^2_{q\theta} S(q;\theta)}{D_q S(q;\theta)}$ *is increasing.*

[20] Proposition 2 generalizes a finding by Itoh (1983). His Proposition 2 establishes that when $U(q; \theta) = \theta q$ and cost are convex, and the seller is constrained to offer fixed qualities in $\{q_1, ..., q_n\}$ (as under bundling), the addition of a new good $q^* \notin \{q_1, ..., q_n\}$ will always increases the producer's profit if cost are convex. This is in fact true whenever surplus is log supermodular in differences!

[21] Jullien (2000) provides a result that is closely related to Proposition 2. He shows that (see his Lemma 10) that the profit-maximizing assignment $q(\theta)$ is continuous if both $\partial_\theta S(q; \theta)$ and $\partial_q S(q; \theta)$ are log supermodular in differences (see his assumption CVU). In relation to this result, Proposition 2 shows that log supermodularity in differences of $\partial_\theta S(q; \theta)$ is superfluous to establish continuity.

4 A (Partial) Generalization: Common Values

We now extend the analysis to allow for common values. Common values describe instances under which the seller's cost $C(q;\theta)$ depend on the buyer's type. While such dependence does not play a role for most consumer goods, some categories are exempt. A frequent example is adverse selection. Here, a monopolist insurer must take into account a countervailing insurance motive among potential insurees: Those buyers most eager to seek out insurance are also the buyers most costly to insure, e.g., $D_\theta C(q;\theta) > 0$. Another example arises if a seller must charge the same prices in geographically distinct markets.

We will not restrict attention to specific examples. Instead, we seek to characterize complementarity assumptions under which two-sided bunching (but not necessarily a single-contract menu) is profit-maximizing in any common value environment. In keeping with our objective of generalization and mathematical clarity, the discussion will be more technical than in the previous section.

4.1 Maximizing Virtual Surplus

Refer, motivated by our earlier analysis, to the following sorting conditions as CV-LSD and CV-LsubD:[22]

Definition 4. *1. CV-LSD holds on* $[\underline{\theta},\overline{\theta}] \subseteq [0,1]$ *if* $q \mapsto \frac{D^2_{q\theta}U(q;\theta)}{D_q S(q;\theta)}$ *is increasing for all* $\theta \in [\underline{\theta},\overline{\theta}]$ *and* $q \in [0,1]$.

2. *CV-LsubD holds on* $[\underline{\theta},\overline{\theta}] \subseteq [0,1]$ *if* $q \mapsto \frac{D^2_{q\theta}U(q;\theta)}{D_q S(q;\theta)}$ *is decreasing for all* $\theta \in [\underline{\theta},\overline{\theta}]$ *and* $q \in [0,1]$.
   If instead the ratio is non-increasing, weak CV-LsubD holds.

When there are private values, CV-LSD (CV-LsubD) is equivalent to LSD (LsubD), for $D_\theta U(q;\theta) = D_\theta S(q;\theta)$. Unfortunately, the common value conditions CV-LSD and CV-LsubD do not admit intuitive interpretations in terms of risk preferences like their private value counterparts. The least one can say is that cost cannot be too convex for CV-LsubD to hold.

To understand the role CV-LsubD plays, we must recall the seller's objective: maximize expected virtual surplus. Virtual surplus (recall Lemma 2; $\Theta \subseteq [\theta, 1]$ depends on binding participation constraints) is given by

$$\Lambda(q;\theta) = S(q(\theta);\theta) - D_\theta U(q(\theta);\theta)\frac{\int_\Theta \mu(\vartheta)d\vartheta}{\mu(\theta)}.$$

Lemma 2 showed the well-known result that the seller's profit is equal to integrated virtual surplus. In effect, the seller maximizes virtual surplus subject to never assigning greater quality to low-valuation buyers, i.e., a non-decreasing quality assignment. The key observation is that under CV-LSD virtual surplus is increasing-decreasing, i.e., quasi-concave in quality. Analogously, under CV-LsubD virtual surplus is decreasing-increasing, i.e., quasi-convex in quality.

Lemma 4. *Posit Assumptions 1, 2, 3.*

1. *Suppose that* $(\underline{\theta},\overline{\theta}) \subseteq [0,1]$ *is an interval on which CV-LSD holds. Then virtual surplus is strictly quasi-concave for all* $q \in [0, q^{fb}(\theta)]$ *on such interval.*
2. *Suppose that* $(\underline{\theta},\overline{\theta}) \subseteq [0,1]$ *is an interval on which CV-LsubD (weak CV-LsubD) holds. Then virtual surplus is strictly quasi-convex (quasi-convex) for all* $q \in [0, q^{fb}(\theta)]$ *on such interval*

[22]Observe that CV-LSD corresponds to the second part of condition CVU in Jullien (2000).

Proof. To begin with, select some $\overline{q}(\theta) \in \underset{q \in [0, q^{fb}(\theta)]}{\operatorname{argmax}} \Lambda(q;\theta)$. Then note that, irrespective of the selection of the pointwise maximum $\overline{q}(\theta)$, increasing differences $D^2_{q\theta}U(q;\theta)$ (as stipulated by Assumption 2) implies that $\overline{q}(\theta) \leq q^{fb}(\theta)$.[23] Next, taking the derivative with respect to quality, well-defined due to Assumption 1, gives

$$D_q\Lambda(q;\theta) = D_q S(q;\theta) \frac{\int_\Theta \mu(\vartheta)d\vartheta}{\mu(\theta)} \Big[\frac{\mu(\theta)}{\int_\Theta \mu(\vartheta)d\vartheta} - \frac{D^2_{q\theta}U(q;\theta)}{D_q S(q;\theta)} \Big].$$

Then note that $D_q S(q;\theta) > 0$ due to Assumptions 2 and 3. Under CV-LsubD it follows that $D_q\Lambda(q;\theta)$ is weakly up-crossing, i.e., $D_q\Lambda(q_1;\theta) > (\geq) 0 \Rightarrow D_q\Lambda(q_2;\theta) > (\geq) 0$ for all $q_2 > q_1$. Under CV-LSD it follows that $D_q\Lambda(q;\theta)$ is down-crossing, i.e., $D_q\Lambda(q_1;\theta) \leq 0 \Rightarrow D_q\Lambda(q_2;\theta) < 0$ for all $q_2 > q_1$. □

Since the pointwise maximizing and minimizing quality assignment (whose existence is guaranteed by CV-LSD and CV-LsubD respectively) will play a crucial role in the subsequent analysis, it is worthwhile to introduce new notation for both objects. Define, provided that CV-LSD and CV-LsubD hold respectively,

$$\overline{q}(\theta) = \underset{q \in [0, q^{fb}(\theta)]}{\operatorname{argmax}} \Lambda(q;\theta) \qquad \text{and} \qquad \underline{q}(\theta) = \underset{q \in [0, q^{fb}(\theta)]}{\operatorname{argmin}} \Lambda(q;\theta).$$

4.2 When Is Discontinuous Bunching Profit-Maximizing?

To see how quasi-concavity of virtual surplus relates to continuity of the profit-maximizing assignment, fix a candidate solution to the seller's problem. If this assignment were profit-maximizing, it must either be constant or equal to the pointwise maximum of virtual surplus (in which case it must be continuous).[24] The reason is simple: for any candidate solution not of this form there exists another increasing assignment with the same end points that is uniformly closer to $\overline{q}(\theta)$. Guesnerie and Laffont (1984) first made this observation in the first part of their proof of Proposition 2. The reasoning equally applies if virtual surplus is quasi-convex: here, it is advantageous for the seller to remove the quality assignment as far as possible from the pointwise minimum of virtual surplus. It then suffices to note that the non-decreasing quality assignments that are maximally removed from $\underline{q}(\theta)$ are piecewise constant.

Theorem 1. *Posit Assumptions 1, 2, 3. And suppose that participation constraints are slack on* $(\underline{\theta}, \overline{\theta})$.

1. *If* $(\underline{\theta}, \overline{\theta}) \subseteq [0,1]$ *is an interval on which CV-LSD holds, then any profit-maximizing quality assignment* $q^*(\theta)$ *is continuous, and constant or equal to* $\overline{q}(\theta)$ *in* $(\underline{\theta}, \overline{\theta})$.
2. *If* $(\underline{\theta}, \overline{\theta}) \subseteq [0,1]$ *is a maximal interval on which (weak) CV-LsubD holds, then any (some) profit-maximizing assignment is piecewise constant on a partition of* $(\underline{\theta}, \overline{\theta})$.

Theorem 1 asserts that under condition CV-LsubD the profit-maximizing quality assignment takes the form of discontinuous bunching under both common and private values. But this is a weaker notion of bunching than two-sided bunching as established by Proposition 1 in the context of private values. There we had shown

[23] Albeit well-known, a proof may be re-assuring: First, clearly $q^{fb}(1) = \overline{q}(1)$. Then fix $\theta \in [0,1)$ and denote $q_2 = q^{fb}(\theta)$ and $q_1 = \overline{q}(\theta)$. By construction it must hold that $\Lambda(q_2;\theta) \leq \Lambda(q_1;\theta)$ and $S(q_2;\theta) \geq S(q_1;\theta)$. Multiplying surplus by the density and taking differences implies that $D_\theta U(q_1;\theta) \leq D_\theta U(q_2;\theta)$. Since $D^2_{q\theta}U(q;\theta) > 0$ it follows that $q_2 > q_1$, as desired.

[24] Berge's maximum Theorem applies and guarantees that $\underset{q \in [0, \underline{q}^{fb}(\theta)]}{\operatorname{argmax}} \Lambda(q;\theta)$ is upper hemicontinuous (whence continuous since the argmax is uniquely defined).

that *if* $(\underline{\theta}, \overline{\theta}) \subseteq [0, 1]$ *is a maximal interval on which (weak) LsubD holds, then any (some) profit-maximizing assignment satisfies* $q^*(\theta) \in \{\underline{q}, \overline{q}\}$*, where*

$$\underline{q} \begin{cases} \in \{0, 1\} & \textit{if } \underline{\theta} = 0 \\ = \lim\limits_{\theta \uparrow \underline{\theta}} q^*(\theta) & \textit{otherwise} \end{cases} \qquad \textit{and} \qquad \overline{q} \begin{cases} \in \{0, 1\} & \textit{if } \overline{\theta} = 1 \\ = \lim\limits_{\theta \downarrow \overline{\theta}} q^*(\theta) & \textit{otherwise.} \end{cases}$$

In effect, only two qualities were actively traded by types in $[\underline{\theta}, \overline{\theta}]$. If LsubD holds for all $\theta \in [0, 1]$ it followed in particular that $q^*(\theta)$ is binary in $\{0, 1\}$—a single-contract menu. Since no such result is asserted under common values, Theorem 1 suggests that the optimality of two-sided bunching and single-contract menus does not generalize. An example (see Figure 4 and formally developed in Appendix A.7) confirms this: Single-contract menus need not be profit-maximizing despite surplus satisfying CV-LsubD for all types.

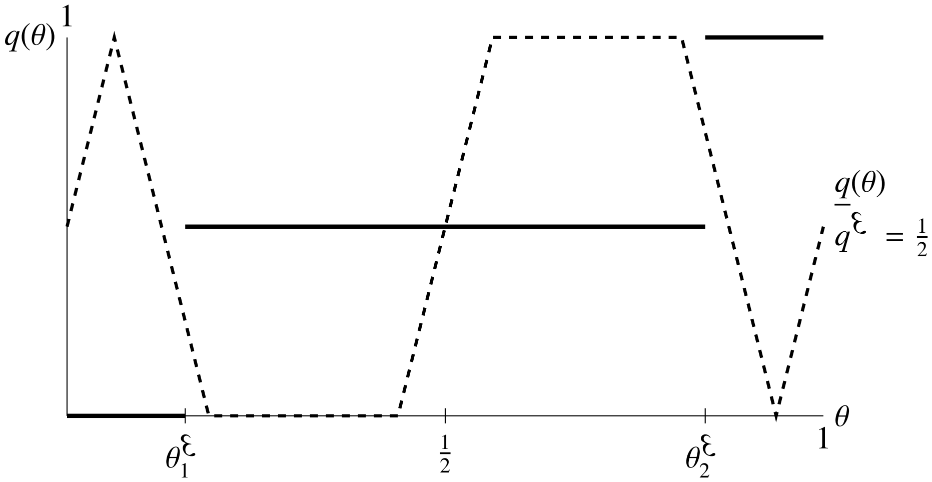

Figure 4: The profit-maximizing quality assignment does not consist of a single contract, despite surplus satisfying CV-LsubD. The thick zick-zack line (dashed) depicts the virtual surplus minimizing quality assignment $\underline{q}(\theta)$. Actively trading qualities in $\{0, \frac{1}{2}, 1\}$ turns out to be profit-maximizing.

But why does this difference between private and common values occur in the first place? One conjecture is that virtual surplus is convex in quality when there are private values and LsbuD holds.[25] This turns out to be wrong: $S(q; \theta) = \alpha q + 2\theta q - \theta qq$ with $\alpha \in (0, 1]$ satisfies Assumptions 1, 2, 3 and LsubD.[26] Yet when types are uniformly distributed, virtual surplus is $\Lambda(q; \theta) = 2\theta q + (1 - 2\theta)qq$. This is concave in quality for types $\theta > \frac{1}{2}$! Short of an explanation, it may be worthwhile to retreat to the proof of Proposition 1 (in Appendix A.2): What sets private values mathematically apart is that for any piecewise constant quality assignment (as occurs when LsubD holds), the integral of virtual surplus over a maximal interval $(\underline{\theta}_k, \overline{\theta}_k)$ where participation constraints are slack reduces to a sum:

$$\Lambda(q; \theta)\mu(\theta) = -\frac{d}{d\theta}\{S(q; \theta)(M(\overline{\theta}_k) - M(\theta))\}.$$

[25] Since convexity is preserved under integration, convexity in quality of virtual surplus would imply that the sellers objective $\Pi(q)$ (introduced in (1)) is convex in the quality assignment. This ensures that any multi-contract menu, being a convex combination of its enveloping deviations (see Figure 2) can impossibly dominate both deviations at the same time. A single-contract menu would be optimal. In particular, this would provide an alternative proof of Proposition 1.

[26] To see that surplus $S(q; \theta) = \alpha q + 2\theta q - \theta qq$ satisfies LsubD, note that $\frac{D^3_{qq\theta} S(q;\theta)}{D^2_{q\theta} S(q;\theta)} = \frac{-2}{2(1-q)} < \frac{-2\theta}{\alpha + 2\theta(1-q)} = \frac{D^2_{qq} S(q;\theta)}{D_q S(q;\theta)}$.

So the fact that quasi-convexity is not preserved under integration seems less critical here.

Remedies to guarantee the optimality of single-contract menus when there are common values do exist but are not as elegant as the previous characterization. If we assume that virtual surplus is not only quasi-convex but convex, then a single-contract menu will likewise be optimal with common values. Convexity holds if $q \mapsto S(q;\theta)$ is convex and $D_\theta U(q;\theta)$ is concave. Alternatively, it suffices to require that the virtual surplus minimizing quality $\underline{q}(\theta)$ is monotone in types.

4.3 Characterization

Having determined the conditions under which the profit-maximizing quality assignment is discontinuous, it remains to characterize it. Such a characterization is commonly achieved via the optimal control approach and is the content of Theorem 2.[27]

But first a strengthening of Assumption 1 is in place:

Assumption 1′. *$(\theta, q) \mapsto U(q;\theta)$, $(\theta, q) \mapsto D_\theta U(q;\theta)$, $(\theta, q) \mapsto C(q;\theta)$ and $\theta \mapsto \mu(\theta)$ are locally Lipschitz.*

Of course, continuous differentiability would imply this.

Theorem 2 (optimal control)**.** *Posit Assumptions 1′, 2. And suppose that the profit-maximizing quality assignment $q^*(\theta)$ admits finitely many discontinuities and is absolutely continuous on any continuous segment. Then the following holds on any maximal interval where participation constraints are slack:*

1. *If $q^*(\theta)$ is increasing then it solves $D_q\Lambda(q^*(\theta);\theta) = 0$, i.e., $q^*(\theta)$ is a critical point of virtual surplus;*
2. *If $(\underline{\theta}, \overline{\theta})$ is a maximal interval on which $q^*(\theta)$ is constant equal to $\hat{q}$ and interior, i.e., $\hat{q} \in (0,1)$ then*

$$\int_{\underline{\theta}}^{\hat{\theta}} D_q\Lambda(\hat{q};\theta)\,\mu(\theta)\,d\theta \;\geq 0 \qquad \textit{holding with equality at } \hat{\theta} = \overline{\theta}.$$

Moreover, $\lim_{\theta\uparrow\hat{\theta}} \Lambda(q(\theta);\theta) = \lim_{\theta\downarrow\hat{\theta}} \Lambda(q(\theta);\theta)$ *if $\hat{\theta}$ is a point of discontinuity.*

The literature exclusively employs the optimal control approach under the restriction that the profit-maximizing quality assignment is absolutely continuous everywhere.[28] In light of Theorem 1 this is too strong a requirement, for discontinuities do arise whenever CV-LsubD holds. The main innovation of Theorem 2 is to accommodate quality assignments with finitely many discontinuities by making use of a hybrid maximum principle (see Clarke (2013) Theorem 22.20 and Theorem 3 in Appendix A.5).

[27]The conclusion of item 2 of Theorem 2 at points of discontinuity can be derived using pointwise optimization over the optimal bunching quality $\hat{q}$ and its boundary points $\underline{\theta}$ and $\overline{\theta}$. Theorem 2 is a stronger result in that item 2 shows that the same conditions must hold regardless of whether $\underline{\theta}$ and $\overline{\theta}$ are points of discontinuity or not, and in that item 1 characterizes the profit-maximizing quality assignment on intervals where it is not constant.

[28]Theorem 2 generalizes a known result (see Guesnerie and Laffont (1984), Theorem 4). It usually appears in the literature under the strong assumption that virtual surplus is concave, as is the case if $q \mapsto S(q;\theta)$ is concave and $q \mapsto D_\theta U(q;\theta)$ is convex. This is to ensure the existence of a profit-maximizing quality assignment that is absolutely continuous *everywhere*. Since Theorem 1 shows that existence of an everywhere absolutely continuous quality assignment is guaranteed by the weaker assumption that CV-LSD holds, it is superfluous to assume concavity in the first place. But so is the focus on everywhere as opposed to piecewise absolutely continuous quality assignments.

5 Competitive Menu Pricing

Motivated by the preceding analysis, one may ask: If the profit-maximizing menu consists of a single contract, will there likewise be only one contract traded in a competitive market? Arguably there is no single model of competition, but exclusive Bertrand competition provides a natural benchmark. Finitely many sellers simultaneously propose a price-quantity menu, and each agent trades with at most one seller and selects her preferred contract. Outside options equal to zero consumption utilities, i.e., $\hat{u}(\theta) = U(0;\theta)$. In a private value environment this model is easy to analyze and gives an unequivocal answer:[29]

Claim 1. *If surplus is log submodular in differences, then the unique competitive equilibrium consists of a single-contract menu only.*

The proof is immediate and rests on the fact that in a competitive equilibrium qualities sold must coincide with the first-best quality when there are private values.

Proof. Formally, recall the identity $\pi(\theta) = S(q(\theta);\theta) - V(\theta)$ where $V(\theta) = \max_{\hat{\theta}} S(q(\hat{\theta}),\theta) - t(\hat{\theta})$ denotes indirect utility. As is well-known, incentive compatibility requires that $\dot{V}(\theta) = D_\theta U(q;\theta)$. In equilibrium profit earned for each type must be zero. Therefore

$$0 = \frac{d}{d\theta}\Big[\underbrace{S(q(\theta);\theta) - V(\theta)}_{\text{per type profit } \pi(\theta)}\Big] = D_q S(q(\theta);\theta)\dot{q}(\theta).$$

Or, the quality either coincides with the first-best or is constant. This concludes the proof, for if surplus is log submodular in differences, Remark 1 asserts that the first-best quality assignment must be constant in types. As to equilibrium uniqueness, it suffices to note that if the single quality sold in equilibrium was different from the first-best, any seller could deviate, offer the first-best and profitably share the gains from trade with the buyers. □

5.1 Entry-Proof Menus with Operating Expenses

The idea that each contract must break even in a competitive equilibrium is often not tenable. From an accounting point of view, this would require that gross profit (sales revenue minus cost of goods sold) is zero. In practice, prices must exceed unit cost so that sellers can recover their operating expenses. While operating expenses are irrelevant to determining the profit-maximizing menu, they alter the characteristics of competitive equilibrium.

We now formalize this intuition. Consider the static game where the sellers must first invest in their capacity to produce, κ, to then be able to sell qualities $q \in [0,\kappa]$. As before, $C(q)$ denotes the seller's unit cost, and, unlike before, $K(\kappa)$ denotes the seller's fixed cost. Clearly, for a given level of capacity investment κ the profit-maximizing menu remains unchanged. In particular, single-contract menus remain profit-maximizing if the surplus function is log submodular in differences. Matters are different when sellers compete with one another.

[29] In a common value environment one may analogously consider the candidate assignment characterized via zero-profit conditions and locally binding incentive constraints. Here, however, it is well known that a symmetric equilibrium in pure strategies typically fails to exist due to possible profitable pooling deviations (see Rothschild and Stiglitz (1976)). In Sandmann (2022), I analyze a model of oligopolistic competition. Seller differentiation imposes limits on the price elasticity of demand. A symmetric equilibrium in pure strategies does exist if sellers are sufficiently differentiated and has the distinctive feature that not all contracts earn the same profit.

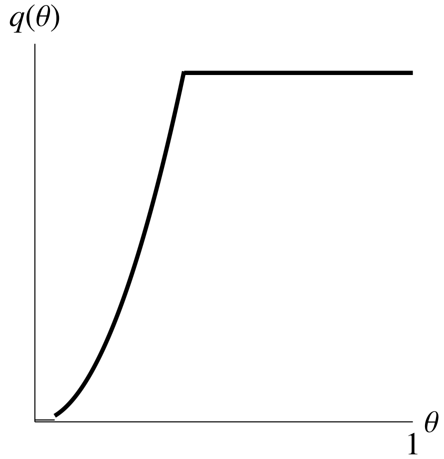

Figure 5: Consider $U(q;\theta) = \theta q$, $C(q) = 0$ and $K(\kappa) = 0.2q^2$ and a uniform type distribution. Since surplus is weakly log submodular in differences, Proposition 1 applies and the profit-maximizing menu consists of a single contract. The entry-proof menu, by contrast, consists of many contracts. The figure depicts the quality assignment when N is large. All types are served, but the lowest types' quality assignment is very small. Notably, there is a discontinuity for some low yet intermediate types. Bunching vanishes for intermediate types, but there is a large bunching interval at the top. Finally, the quality assignment is less than one (due to capacity cost) but uniformly greater than the profit-maximizing quality assignment.

Claim 2. *When generating capacity is costly, the entry-proof menu typically does not consist of a single contract.*

To illustrate, consider buyer utility of the form $U(q;\theta) = \theta q$ and zero marginal unit cost. Then any single contract must exclude the buyers of the lowest type to recover the fixed cost $K(\kappa)$. Exclusion, however, can not be sustained in the presence of possible entrants if the seller's fixed cost is an order of magnitude smaller than buyers' marginal utilities. Convex capacity cost of the form $K(\kappa) = \gamma\kappa^2$ is a case in point. By not serving the lowest types, the incumbent seller creates an incentive for an entrant to enter the market and pick up the unserved demand. If so, the entrant's pricing strategy does not require strategic sophistication: choose an incremental quality capacity whose capacity cost is negligible compared to the price a lower-than-intermediate buyer (hitherto excluded) would be willing to pay.

Construction of an Entry-Proof Menu

Claim 2 is a negative result. It reveals that menu pricing must be prevalent in equilibrium but fails to characterize it. We now set out to define one such possible construction.

An equilibrium of the economy with capacity cost corresponds to an entry-proof menu. A single seller will be active and offer a menu so that (i) no entrant can profitably attract a subset of buyers and recover the initial capacity investment, and (ii) the profit earned is no less than the cost of generating capacity. The construction of this menu divides the continuum of buyers into finitely many groups. Buyers from each group self-select into purchasing the same contracts. In effect, the candidate menu is bunching by construction. The extent of bunching can then be gradually diminished.

Formally, fix $N \in \mathbb{N}$ and consider the partition $0 = \theta_0 < \theta_1 < ... < \theta_N$ of the type space $[0, 1]$ so that

$M(\theta_\ell) - M(\theta_{\ell-1}) = \frac{1}{N}$.

We then consider the set of menus that only sell to the n first buyer groups. Define $\mathscr{V}(\{\theta_1, ..., \theta_n\})$ with $n : 1 \leq n < N$ the set of indirect utility vectors $V(\theta_1), ..., V(\theta_n)$ so that there exists a menu $\{(q(\theta_\ell), t(\theta_\ell)\}_{\ell=0}^n$ with $q(\theta_0) = t(\theta_0) = 0$ that is (i) incentive compatible for all types $\theta \in [0, \theta_n]$:

$$S(q(\theta_\ell); \theta) - t(\theta_\ell) \geq S(q(\theta_k); \theta) - t(\theta_k) \qquad \text{for all } \theta \in [\theta_{\ell-1}, \theta_\ell) \text{ and for all } \ell, k,$$

and (ii) budget balanced (where capacity is $\kappa = q(\theta_n)$):

$$\frac{1}{N} \sum_{\ell=1}^{n} t(\theta_\ell) \geq K(q(\theta_n)).$$

The candidate equilibrium is constructed recursively by gradually increasing the number of buyer groups to whom the seller can sell. At stage 1 define

$$\underline{V}^1(\theta_1) = \max_{V(\theta_1) \in \mathscr{V}(\{\theta_1\})} V(\theta_1)$$

where the superindex indicates the stage number. $\underline{V}^1(\theta_1)$ corresponds to the maximal (indirect) utility any seller could provide to the lowest types without incurring a loss if no higher type were to make a purchase. Crucially, the seller seeks to provide greater utility to all buyers than an entrant seller could by selling to only a subset of buyer groups. Inductively we thus derive reservation utilities $\underline{V}_{n-1}(\theta_1), ..., \underline{V}_{n-1}(\theta_{n-1})$. Then, at stage n, define

$$\underline{V}^n(\theta_n) = \max_{(V(\theta_1), ..., V(\theta_n)) \in \mathscr{V}(\{\theta_1, ..., \theta_n\})} V(\theta_n) \qquad \text{such that} \qquad V(\theta_\ell) \geq \underline{V}^{n-1}(\theta_\ell) \quad \text{for all } \ell : 1 \leq \ell \leq n-1.$$

The recursive construction ensures that, as N converges to infinity, no entrant could profitably serve low valuation buyers. Similarly, no seller could profitably fish for buyers at the top. Doing so would deprive high valuation buyers of benefiting from the presence of low valuation buyers. This benefit is twofold: First, each buyer (including those of low valuation) contributes to recovering the cost of the initial capacity investment, paving the way for uniformly reduced prices. Second, providing better quality to low valuation buyers decreases prices on intermediate quality varieties and increases high valuation buyers' outside options, thereby decreasing prices at the top.

We must contend that the inclusion of operating expenses in the model could lead to the existence of multiple competitive equilibria. Many entry-proof menus are conceivable. Some will cater to the tastes of low valuation buyers, others to high valuation buyers. The multiplicity of entry-proof menus (especially so in models with finitely many buyer types) is not an extravagant finding; it is reminiscent of the multiplicity of core allocations found in many cooperative games. Future work could characterize the entire set of entry-proof menus and refine it by introducing diminishing seller market power.

6 Conclusion

This paper studied the canonical monopoly screening problem with quasi-linear utility. I showed that the profit-maximizing quality assignment takes one of two forms: It is either continuous and (as is well-known) any non-constant segment pointwise maximizes virtual surplus. Or, it is piecewise constant (bunching) and plausibly discontinuous in types. This paper identified conditions that tell continuous quality assignments and discontinuous bunching apart. As a main result, I characterized when single-contract menus are profit-maximizing in a private value environment. The result is as follows: Maintain that surplus is non-decreasing in quality below the first-best and satisfies increasing differences (also referred to as single-crossing). Then, if a higher type's surplus function is more concave in quality (i.e., higher types' surplus-utility encodes more risk-averse preferences), the monopolist seller can not do better than selling only a single quality variety. Equivalently, this holds if marginal surplus $D_qS(q;\theta)$ is log submodular. While the characterization is disarmingly simple, it contributes to a literature that had not settled the question conclusively.

A Appendix

A.1 Preliminaries: Representation of the Seller's Profit

Proof of Lemma 2. The derivation of the virtual surplus representation of the seller's profit is standard: Denote $(\underline{\theta}_k, \overline{\theta}_k)$ a maximal interval of participating types. Define $V(\theta) = U(q(\theta);\theta) - p(\theta)$. The envelope theorem establishes that a menu is incentive-compatible if and only if $V(\theta) = \hat{u}(\underline{\theta}_k) + \int_{\underline{\theta}_k}^{\theta} D_\theta U(q(\vartheta);\vartheta)d\vartheta$ for all $\theta \in (\underline{\theta}_k, \overline{\theta}_k)$. If so, $V(\theta)$ corresponds to the buyer's indirect utility. Then, following the standard dual approach, the seller's profit over types $(\underline{\theta}_k, \overline{\theta}_k)$ is

$$\int_{\underline{\theta}_k}^{\overline{\theta}_k} (p(\theta) - C(q;\theta))\mu(\theta)d\theta = \int_{\underline{\theta}_k}^{\overline{\theta}_k} (S(q(\theta);\theta) - V(\theta))\mu(\theta)d\theta$$

$$= -\hat{u}(\underline{\theta}_k)\Big(M(\overline{\theta}_k) - M(\underline{\theta}_k)\Big) + \int_{\underline{\theta}_k}^{\overline{\theta}_k} \Big(S(q(\theta);\theta) - \int_0^{\theta} D_\theta U(q(\vartheta);\vartheta)d\vartheta\Big)\mu(\theta)d\theta.$$

Finally, due to integration by parts, $\int_{\underline{\theta}_k}^{\overline{\theta}_k}\Big[\int_{\underline{\theta}_k}^{\theta} D_\theta U(q(\vartheta);\vartheta)d\vartheta\Big]\mu(\theta)d\theta = \int_{\underline{\theta}_k}^{\overline{\theta}_k} D_\theta U(q(\theta);\theta)(M(\overline{\theta}_k) - M(\theta))d\theta.$ □

Proof of Lemma 3. Observe that for any maximal interval of participating types $(\underline{\theta}_k, \overline{\theta}_k)$ the following holds:

$$\int_{\underline{\theta}_k}^{\overline{\theta}_k} \Big\{S(q(\theta);\theta) - D_\theta S(q(\theta);\theta)\frac{M(\overline{\theta}_k) - M(\theta)}{\mu(\theta)}\Big\}\mu(\theta)d\theta$$

$$= \lim_{N\to\infty}\sum_{\ell=1}^{N}\Big[S(q(\theta_\ell);\theta_\ell)\mu(\theta_\ell) - D_\theta S(q(\theta_\ell);\theta_\ell)(M(\overline{\theta}_k) - M(\theta_\ell))\Big]\frac{1}{N}$$

$$= \lim_{N\to\infty}\sum_{\ell=1}^{N} S(q(\theta_\ell);\theta_\ell)(M(\theta_\ell) - M(\theta_{\ell-1})) - (S(q(\theta_\ell);\theta_{\ell+1}) - S(q(\theta_\ell);\theta_\ell))(M(\overline{\theta}_k) - M(\theta_\ell))$$

$$= \lim_{N\to\infty}\sum_{\ell=1}^{N} S(q(\theta_\ell);\theta_\ell)(M(\overline{\theta}_k) - M(\theta_{\ell-1})) - S(q(\theta_\ell);\theta_{\ell+1})(M(\overline{\theta}_k) - M(\theta_\ell))$$

$$= S(\lim_{\theta\downarrow\underline{\theta}_k} q(\theta);\underline{\theta}_k)\Big(M(\overline{\theta}_k) - M(\underline{\theta}_k)\Big) + \lim_{N\to\infty}\sum_{\ell=1}^{N}\Big[S(q(\theta_\ell);\theta_\ell) - S(q(\theta_{\ell-1});\theta_\ell)\Big]\Big(M(\overline{\theta}_k) - M(\theta_\ell)\Big)$$

where we have used the Lebesgue criterion: a bounded function is Riemann integrable if and only if the set of all discontinuities has measure zero. To see that the criterion applies to $\theta \mapsto S(q(\theta);\theta)\mu(\theta - D_\theta S(q(\theta);\theta)(M(\overline{\theta}_k - M(\theta))$, note that due to Assumption 1, for any non-decreasing $\theta \mapsto q(\theta)$ the functions $\theta \mapsto S(q(\theta);\theta)$ and $\theta \mapsto D_\theta S(q(\theta);\theta)$ are continuous almost everywhere.[30] □

[30] In greater detail: Let $x \mapsto g(x)$ be a non-decreasing and $y \mapsto f(y)$ be a continuous function. Then $x \mapsto f \circ g(x)$ is continuous almost everywhere. To see this, fix x' a point where g is continuous and denote $y' = g(x')$. Then show that $x \mapsto f \circ g(x)$ is continuous at x'. This proves the claim, since $x \mapsto g(x)$, being non-decreasing, is continuous almost everywhere. Or, fix $\epsilon > 0$ and let $\zeta > 0$ so that $f(y) \in B_\epsilon(f(y'))$ for all $y \in B_\zeta(y')$; and fix $\delta > 0$ so that $g(x) \in B_\zeta(g(x'))$ for all $x \in B_\delta(x')$. Then $f \circ g(x) \in B_\epsilon(f \circ g(x'))$ for all $x \in B_\delta(x')$.

Proof of Lemma 3′. Following standard arguments, define indirect utility as follows:

$$V(\theta_1) = \hat{u}(\theta_1) \quad \text{and} \quad V(\theta_j) = \max_{\ell \in \{1,\dots,N\}} S(q(\theta_\ell);\theta_j) - \pi(\theta_\ell).$$

Provided that the menu satisfies incentive and participation constraints, $V(\theta_j) = S(q(\theta_j);\theta_j) - \pi(\theta_j)$. Under the profit-maximizing mechanism incentive constraints are downward binding (see Rochet (1987)) so that $V(\theta_{j+1}) - V(\theta_j) = S(q(\theta_j);\theta_{j+1}) - S(q(\theta_j);\theta_j)$. Then indirect utility reads as

$$V(\theta_j) = \hat{u}(\theta_1) + \sum_{\ell=1}^{j-1} S(q(\theta_\ell);\theta_{\ell+1}) - S(q(\theta_\ell);\theta_\ell).$$

Taking the weighted sum of utilities and denoting $m(\theta_j) = M(\theta_j) - M(\theta_{j-1})$ gives

$$\begin{aligned}\sum_{j=1}^{N} V(\theta_j)m(\theta_j) &= \hat{u}(\theta_1) + \sum_{j=1}^{N}\sum_{\ell=1}^{j-1} [S(q(\theta_\ell);\theta_{\ell+1}) - S(q(\theta_\ell);\theta_\ell)]m(\theta_j) \\ &= \hat{u}(\theta_1) + \sum_{j=1}^{N-1} [S(q(\theta_j);\theta_{j+1}) - S(q(\theta_j);\theta_j)](M(\theta_N) - M(\theta_j)).\end{aligned}$$

This allows to express the seller's profit earned from types $\{\theta_1,\dots,\theta_N\}$ as

$$\begin{aligned}\sum_{j=1}^{N} \pi(\theta_j)m(\theta_j) &= \sum_{j=1}^{N} [S(q(\theta_j);\theta_j) - V(\theta_j)]m(\theta_j) \\ &= \sum_{j=1}^{N} S(q(\theta_j);\theta_j)m(\theta_j) - \hat{u}(\theta_1)(M(\theta_N) - M(\theta_{j-1})) + \sum_{j=1}^{N-1} [S(q(\theta_j);\theta_j) - S(q(\theta_j);\theta_{j+1})](M(\theta_N) - M(\theta_j)) \\ &= \sum_{j=1}^{N} S(q(\theta_j);\theta_j)(M(\theta_N) - M(\theta_{j-1})) - \hat{u}(\theta_1)(M(\theta_N) - M(\theta_{j-1})) - \sum_{j=2}^{N} S(q(\theta_{j-1});\theta_j)(M(\theta_N) - M(\theta_{j-1})) \\ &= (S(q(\theta_0);\theta_1) - \hat{u}(\theta_1))(M(\theta_N) - M(\theta_{j-1})) + \sum_{j=1}^{N} [S(q(\theta_j);\theta_j) - S(q(\theta_{j-1});\theta_j)](M(\theta_N) - M(\theta_{j-1})).\end{aligned}$$

□

A.2 Proof of Proposition 1: Weak Optimality of Single-Contract Menus

Proof. We show that two-sided bunching is profit-maximizing. In light of Lemma 3′ and Lemma 3 it suffices to show this for finitely many types. Among all profit-maximizing quality assignments select one so that no other assignment entails bunching for a greater number of types. If the selected assignment does not give rise to two-sided bunching, there exist types θ_L, θ_H with $\theta_H > \theta_L$ such that $q(\theta_H) > q(\theta_{H-1}) = q(\theta_L) > q(\theta_{L-1}) \geq 0$. Then consider the two deviations where the seller either provides less quality or more quality to intermediate types: $\tilde{q}(\theta_{L-1}) = q(\theta_L) - \epsilon \geq q(\theta_{L-1})$ (with $\epsilon > 0$ small as to preserve participation constraints) or $\tilde{q}(\theta_H) \equiv \min\{q^{fb}(\theta_L); q(\theta_H)\}$ (where raising quality relaxes participation constraints) to all types $\{\theta_L,\dots,\theta_{H-1}\}$:

$$\underline{q}(\theta_\ell) = \begin{cases} \tilde{q}(\theta_{L-1}) & \text{if } \theta_\ell \in \{\theta_L,\dots,\theta_{H-1}\} \\ q(\theta_\ell) & \text{otherwise} \end{cases} \quad \text{and} \quad \overline{q}(\theta_\ell) = \begin{cases} \tilde{q}(\theta_H) & \text{if } \theta_\ell \in \{\theta_L,\dots,\theta_{H-1}\} \\ q(\theta_\ell) & \text{otherwise} \end{cases}$$

Transfers for both deviations are implicitly defined via downward binding incentive constraints. Lemma 3′ characterizes the deviating profits.

Maintaining that among profit-maximizing mechanisms there cannot be more bunching than under $\{q(\theta_\ell)\}_{\ell=1}^N$, the initial profit must be strictly greater than both deviating profits. This yields

$$\begin{aligned}
&[S(q(\theta_L);\theta_L) - S(q(\theta_{L-1});\theta_L)](M(\theta_N) - M(\theta_{L-1})) + [S(q(\theta_H);\theta_H) - S(q(\theta_{H-1});\theta_H)](M(\theta_N) - M(\theta_{H-1})) \\
&> \max\Big\{[S(q(\theta_H);\theta_H) - S(\tilde{q}(\theta_H);\theta_H)](M(\theta_N) - M(\theta_{H-1})) + [S(\tilde{q}(\theta_H);\theta_L) - S(q(\theta_{L-1});\theta_L)](M(\theta_N) - M(\theta_{L-1})); \\
&\qquad [S(q(\theta_H;\theta_H) - S(\tilde{q}(\theta_{L-1});\theta_H)](M(\theta_N) - M(\theta_{H-1})) + [S(\tilde{q}(\theta_{L-1});\theta_L) - S(q(\theta_{L-1});\theta_L)](M(\theta_N) - M(\theta_{L-1}))\Big\}.
\end{aligned}$$

Re-arranging, this is equivalent to

$$\begin{aligned}
&[S(\tilde{q}(\theta_H) - S(q(\theta_L);\theta_L)](M(\theta_N) - M(\theta_{L-1})) < (S(\tilde{q}(\theta_H);\theta_H) - S(q(\theta_{H-1});\theta_H)](M(\theta_N) - M(\theta_{H-1})) \\
&[S(q(\theta_{H-1});\theta_H) - S(\tilde{q}(\theta_{L-1});\theta_H)](M(\theta_N) - M(\theta_{H-1})) < [S(q(\theta_L);\theta_L) - S(\tilde{q}(\theta_{L-1});\theta_L)](M(\theta_N) - M(\theta_{L-1})).
\end{aligned}$$

To facilitate the notation, denote $q_3 = \tilde{q}(\theta_H)$, $q_2 = q(\theta_{H-1}) = q(\theta_L)$, and $q_1 = \tilde{q}(\theta_{L-1})$. It follows that

$$\begin{aligned}
&[S(q_3;\theta_L) - S(q_2;\theta_L)](1 - M(\theta_{L-1})) < [S(q_3;\theta_H) - S(q_2;\theta_H)](1 - M(\theta_{H-1})) \\
&[S(q_2;\theta_H) - S(q_1;\theta_H)](1 - M(\theta_{H-1})) < [S(q_2;\theta_L) - S(q_1;\theta_L)](1 - M(\theta_{L-1})).
\end{aligned}$$

Then dividing the first inequality by the second gives

$$\frac{S(q_3;\theta_L) - S(q_2;\theta_L)}{S(q_2;\theta_L) - S(q_1;\theta_L)} < \frac{S(q_3;\theta_H) - S(q_2;\theta_H)}{S(q_2;\theta_H) - S(q_1;\theta_H)},$$

in spite of weak log submodularity in differences (Definition 1). □

A.3 Proof of Theorem 1: Continuity Versus Discontinuous Bunching

Proof. Consider item 1 first. Fix any implementable assignment $\hat{q}$ that is not of the described form. First, consider a maximal interval $(\underline{\theta}, \overline{\theta})$ where $\hat{q}(\theta) < \overline{q}(\theta)$ almost everywhere. This means that $\hat{q}(\underline{\theta}) = \overline{q}(\underline{\theta})$ or $\underline{\theta} = 0$ and $\hat{q}(\overline{\theta}) = \overline{q}(\overline{\theta})$ or $\overline{\theta} = 1$. Then construct a deviation and set $q^*(\theta) = \min\{\overline{q}(\theta); \min_{\tilde{\theta}\in[\theta,\overline{\theta}]} \overline{q}(\tilde{\theta})\}$ for all $\theta \in [\underline{\theta}, \overline{\theta})$. $q^*(\theta)$ is weakly greater than $\hat{q}(\theta)$ and so participation constraints remain slack and clearly non-decreasing on any such interval. Moreover, $\overline{q}(\theta) \geq \min\{\overline{q}(\theta); \min_{\tilde{\theta}\in[\theta,\overline{\theta}]} \overline{q}(\tilde{\theta})\} \geq \min_{\tilde{\theta}\in[\theta,\overline{\theta}]} \hat{q}(\tilde{\theta}) = \hat{q}(\theta)$ so that $\overline{q}(\theta) \geq q^*(\theta) \geq \hat{q}(\theta)$. Similarly, consider a maximal interval $(\underline{\theta}, \overline{\theta})$ where the candidate assignment is strictly greater than the pointwise maximum of virtual surplus, i.e., $\hat{q}(\theta) > \overline{q}(\theta)$ almost everywhere. This means that $\hat{q}(\underline{\theta}) = \overline{q}(\underline{\theta})$ or $\underline{\theta} = 0$ and $\hat{q}(\overline{\theta}) = \overline{q}(\overline{\theta})$ or $\overline{\theta} = 1$. Then construct a deviation and set $q^*(\theta) = \max\{\hat{q}(\underline{\theta}) - \epsilon; \max_{\tilde{\theta}\in[\underline{\theta},\theta]} \overline{q}(\tilde{\theta})\}$ for all $\theta \in [\underline{\theta}, \overline{\theta})\}$ where $\epsilon > 0$ is small as to preserve participation constraints. $q^*(\theta)$ is clearly non-decreasing on any such interval. Moreover, $\overline{q}(\theta) \leq \max_{\tilde{\theta}\in[\underline{\theta},\theta]} \overline{q}(\tilde{\theta}) \leq \max\{\hat{q}(\underline{\theta}) - \epsilon; \max_{\tilde{\theta}\in[\underline{\theta},\theta]} \overline{q}(\tilde{\theta})\} \leq \max_{\tilde{\theta}\in[\underline{\theta},\theta]} \hat{q}(\tilde{\theta}) = \hat{q}(\theta)$ so that $\overline{q}(\theta) \leq q^*(\theta) \leq \hat{q}(\theta)$.

In both cases $q^*(\theta)$ is uniformly closer to the pointwise maximum of virtual surplus $\overline{q}(\theta)$ than the candidate $\hat{q}(\theta)$. Since the seller's profit is the integral over virtual surplus (Lemma 3) and virtual surplus is quasi-concave (Lemma 4), it follows that $q^*(\theta)$ yields strictly greater profit than the candidate solution $\hat{q}(\theta)$.

Then consider item 2 and recall that under CV-LsubD virtual surplus is quasi-convex so that virtual surplus admits a unique pointwise minimum $\underline{q}(\theta)$. Then consider a candidate solution $q^*(\theta)$ to the monopolist's problem and any maximal interval $(\underline{\theta}, \overline{\theta})$ on which $q^*(\theta)$ is not piecewise constant. Since $\underline{q}(\theta)$ is continuous and $\hat{q}(\theta)$ non-decreasing (actually non-constant, therefore increasing), whence continuous almost everywhere, one can partition $(\underline{\theta}, \overline{\theta})$ into countably many intervals so that on each interval (θ', θ'') either $\hat{q}(\theta) \leq \underline{q}(\theta)$ or $\hat{q}(\theta) \geq \underline{q}(\theta)$ for all $\theta \in (\theta', \theta'')$. By setting $q^*(\theta)$ equal to $\max\{\hat{q}(\theta'); \hat{q}(\theta) - \epsilon\}$ (with $\epsilon > 0$ small as to preserve participation constraints) or equal to $\hat{q}(\theta'')$ respectively, we have constructed a deviation that is uniformly further removed from the pointwise minimum, whence more profitable (weakly more profitable if weak CV-LsubD holds), and without upsetting the constraint that the quality assignment is non-decreasing, whence implementable. □

A.4 Proof of Proposition 1: Unique Optimality of Single-Contract Menus

We now consider the second part of Proposition 1 where surplus is log submodular in differences.

Proof. Due to Theorem 1, $[\underline{\theta}, \overline{\theta}]$ can be partitioned into a countable collection of intervals so that the profit-maximizing assignment is constant on each partition element. Then fix a candidate assignment, equal to q_1, q_2 and q_3 on intervals $(\theta_0, \theta_1]$, $(\theta_1, \theta_2]$ and $(\theta_2, \theta_3]$ respectively. And recall that with private values $D_\theta U_\theta(q;\theta) = D_\theta S(q;\theta)$. In effect, profit over $[\theta_0, \theta_3]$ is (equally implied by Lemma 3)

$$\sum_{j=1,2,3} \int_{\theta_{j-1}}^{\theta^j} \Lambda(q_j;\theta)\mu(\theta)d\theta = \sum_{j=1,2,3} \int_{\theta_{j-1}}^{\theta^j} -\frac{d}{d\theta}\Big\{S(q_j;\theta)(M(\overline{\theta}_k) - M(\theta))\Big\}d\theta = \sum_{j=1,2,3} -\Big[S(q_j;\theta)(M(\overline{\theta}_k) - M(\theta))\Big]_{\theta_{j-1}}^{\theta_j}.$$

(Here $(\underline{\theta}_k, \overline{\theta}_k) \supseteq (\underline{\theta}, \overline{\theta})$ is the maximal interval so that participation constraints are slack.) If $q_2 \in (q_1, q_3)$ is profit-maximizing, it must weakly dominate both the lower deviation where the seller offers $\tilde{q}_1 = \max\{q_1; q_2 - \epsilon\}$ to types (θ_1, θ_2) (with $\epsilon > 0$ small as to preserve participation constraints) and the upper deviation where the seller offers q_3 to types (θ_1, θ_2):

$$-S(q_2;\theta_2)(M(\overline{\theta}_k) - M(\theta_2)) + S(q_2;\theta_1)(M(\overline{\theta}_k) - M(\theta_1)) \geq \max\left\{\begin{array}{l} -S(\tilde{q}_1;\theta_2)(M(\overline{\theta}_k) - M(\theta_2)) + S(\tilde{q}_1;\theta_1)(M(\overline{\theta}_k) - M(\theta_1)); \\ -S(q_3;\theta_2)(M(\overline{\theta}_k) - M(\theta_2)) + S(q_3;\theta_1)(M(\overline{\theta}_k) - M(\theta_1)) \end{array}\right\}.$$

Then re-arranging and dividing one by the other implies that

$$\frac{S(q_3;\theta_1) - S(q_2;\theta_1)}{S(q_2;\theta_1) - S(\tilde{q}_1;\theta_1)} \leq \frac{S(q_3;\theta_2) - S(q_2;\theta_2)}{S(q_2;\theta_2) - S(\tilde{q}_1;\theta_2)}.$$

This gives the desired contradiction to surplus being log submodular in differences. □

A.5 Proof of Theorem 2: The Hybrid Maximum Principle

To allow for discontinuities in the quality assignment across types we apply a hybrid maximum principle for multi processes (see Theorem 22.20 in the textbook by Clarke (2013)). The underlying mathematical idea is that each discontinuous segment corresponds to a different controlled process.

Begin with the problem description: provided that the seller optimizes over quality assignment with at most L discontinuities (switching times corresponding to switching to the next process) and absolutely continuous segments, the seller's problem can be written as an optimal control problem. By mere accounting, type θ profit is $S(q;\theta) - V(\theta) = p(\theta) - C(q;\theta))$. Assumption 2 ensures that the quality assignment is implementable if and only if $q(\theta)$ is non-decreasing. If we then choose as a control its derivative, the relevant control space are the positive reals.

$$\begin{cases} \text{Maximize} & \sum_{\ell=1}^{L} \int_{\theta_{\ell-1}}^{\theta_\ell} \left(S(q_\ell(\theta);\vartheta_\ell(\theta)) - V_\ell(\theta)\right)\mu(\vartheta_\ell(\theta))\,d\theta \\ \text{subject to} & \text{switching times } 0 = \theta_0 \le \theta_1 \le ... \le \theta_L = 1, \\ & \text{state trajectories } \dot{q}_\ell(\theta) = z_\ell(\theta),\ \dot{V}_\ell(\theta) = D_\theta U(q_\ell(\theta);\vartheta_\ell(\theta)),\ \dot{\vartheta}_\ell(\theta) = 1, \\ & \text{control } z_\ell(\theta) \in U \equiv \mathbb{R}_+ \text{ if } 1 < \ell < L, \text{ zero otherwise}, \\ & \text{endpoint constraints } q_1(0) = 0,\ q_L(1) = 1,\ V_1(0) = U(0;0),\ V_L(1) \text{ free}, \vartheta_1(0) = 0,\ \vartheta_L(1) \text{ free}, \\ & \text{switching condition } q_{\ell-1}(\theta_\ell) \le q_\ell(\theta_\ell),\ V_{\ell-1}(\theta_{\ell-1}) = V_\ell(\theta_{\ell-1}),\ \vartheta_{\ell-1}(\theta_\ell) = \vartheta_\ell(\theta_\ell). \end{cases} \tag{5}$$

We define the Hamiltonian for each segment: $\mathscr{H}_\ell^\eta(q_\ell(\theta), V_\ell(\theta), \vartheta_\ell(\theta), \lambda_\ell^q(\theta), \lambda_\ell^V(\theta), \lambda_\ell^\vartheta(\theta), z_\ell(\theta))$ is given by

$$\eta\left(S(q_\ell(\theta);\vartheta_\ell(\theta)) - V_\ell(\theta)\right)\mu(\vartheta_\ell(\theta)) + \lambda_\ell^q(\theta)z_\ell(\theta) + \lambda_\ell^V(\theta)D_\theta U(q_\ell(\theta);\vartheta_\ell(\theta)) + \lambda_\ell^\vartheta(\theta).$$

Further define the maximized Hamiltonian

$$M_\ell^\eta(q_\ell(\theta), V_\ell(\theta), \vartheta_\ell(\theta), \lambda_\ell^q(\theta), \lambda_\ell^V(\theta), \lambda_\ell^\vartheta(\theta)) = \sup_{\zeta\in U} \mathscr{H}_\ell^\eta(q_\ell(\theta), V_\ell(\theta), \vartheta_\ell(\theta), \lambda_\ell^q(\theta), \lambda_\ell^V(\theta), \lambda_\ell^\vartheta(\theta), \zeta).$$

The introduction of ϑ_ℓ transforms the seller's problem into an autonomous control problem: the seller's objective does not depend on 'time' θ, but state $\vartheta_\ell(\theta)$.

Theorem 3 (adapted from Clarke (2013), Theorem 22.20)**.** *Posit Assumption 1′. Let the multiprocess* $(q_\ell, V_\ell, \vartheta_\ell, z_\ell)_{\ell=1}^L$ *be a global maximizer for problem* (5) *with switching times* $(\theta_\ell)_{\ell=1}^L$*. Then there exist arcs* $(\lambda_\ell^q, \lambda_\ell^V, \lambda_\theta^\vartheta)_{\ell=1}^L$ *and a scalar* $\eta \in \{0, 1\}$ *satisfying the nontriviality condition* $(\eta, ((\lambda_\ell^q, \lambda_\ell^V, \lambda_\theta^\vartheta)_{\ell=1}^L) \neq 0$*, the transversality condition* $\lambda_L^V(1) = 0$ *and* $\lambda_L^\vartheta(1) = 0$*, the adjoint equations*

$$\begin{aligned} -\dot{\lambda}_\ell^q(\theta) &= D_q\mathscr{H}_\ell^\eta(q_\ell, V_\ell, \vartheta_\ell, \lambda_\ell^q, \lambda_\ell^V, \lambda_\ell^\vartheta, z_\ell) = \eta\, D_q S(q_\ell(\theta);\vartheta_\ell(\theta))\mu(\vartheta_\ell(\theta)) + \lambda_\ell^V(\theta)D_{q\theta}^2 U(q_\ell(\theta);\vartheta_\ell(\theta)) \\ -\dot{\lambda}_\ell^V(\theta) &= D_V\mathscr{H}_\ell^\eta(q_\ell, V_\ell, \vartheta_\ell, \lambda_\ell^q, \lambda_\ell^V, \lambda_\ell^\vartheta, z_\ell) = -\eta\,\mu(\vartheta_\ell(\theta)) \\ -\dot{\lambda}_\ell^\vartheta(\theta) &= D_\vartheta\mathscr{H}_\ell^\eta(q_\ell, V_\ell, \vartheta_\ell, \lambda_\ell^q, \lambda_\ell^V, \lambda_\ell^\vartheta, z_\ell), \end{aligned}$$

the maximum condition

$$z_\ell(\theta) \begin{cases} = \infty & \text{if } \lambda_\ell^q(\theta) > 0 \\ \in \mathbb{R}_+ & \text{if } \lambda_\ell^q(\theta) = 0 \\ = 0 & \text{if } \lambda_\ell^q(\theta) < 0, \end{cases}$$

the constancy condition (the maximized Hamiltonian is constant almost everywhere on $[0, 1]$*) and the switching condition* $\lambda_{\ell-1}^V(\theta_\ell) = \lambda_\ell^V(\theta_\ell)$ *and* $\lambda_{\ell-1}^\vartheta(\theta_\ell) = \lambda_\ell^\vartheta(\theta_\ell)$ *(as implied by Exercise 22.21 in Clarke (2013)),* $\lambda_{\ell-1}^q(\theta_\ell) =$

$\lambda^q_\ell(\theta_\ell) = 0$ *if* $q_{\ell-1}(\theta_\ell) < q_\ell(\theta_\ell)$ *and*

$M^\eta_{\ell-1}(q_{\ell-1}(\theta_\ell), V_{\ell-1}(\theta_\ell), \vartheta_{\ell-1}(\theta_\ell), \lambda^q_{\ell-1}(\theta_\ell), \lambda^V_{\ell-1}(\theta_\ell), \lambda^\vartheta_{\ell-1}(\theta_\ell)) = M^\eta_\ell(q_\ell(\theta_\ell), V_\ell(\theta_\ell), \vartheta_\ell(\theta_\ell), \lambda^q_\ell(\theta_\ell), \lambda^V_\ell(\theta_\ell), \lambda^\vartheta_\ell(\theta_\ell))$.

Theorem 3 adapts Clarke (2013)'s Theorem 22.20 for our purposes and allows for finitely many multi processes (it suffices to prove the Theorem for just two processes, i.e., one discontinuity). We require Assumption $1'$ to encompass his regularity assumptions.

Theorem 3 differs from Theorem 22.20 in one important regard. For a process to be *locally* optimal, Theorem 22.20 requires the control set U to be compact. This is not a property of the extended maximum principle from which Theorem 22.20 is derived and it is not a premise of Theorem 3. Careful inspection of Clarke's proof (see 'Derivation of the hybrid maximum principle') reveals that all arguments go through with unbounded and constant control sets, namely $U = \mathbb{R}_+$, if we seek to provide necessary conditions for *globally* optimal processes instead. (Compactness is relied upon only once in Clarke's derivation, in 'Let us see how to arrange this' and following which is immaterial to globally optimal processes. It is then immediate to verify that Assumption 1 implies Hypothesis 22.25 of the extended maximum principle which replaces compactness of U.)

Proof of Theorem 3. Fix a profit-maximizing quality assignment $\theta \mapsto q(\theta)$ that admits $L' \leq L$ discontinuities $\theta_1 < ... < \theta_{L'-1}$ and is absolutely continuous on each continuous segment. By abuse of notation, set $L' = L$. Define $q_\ell(\theta) = q(\theta)$, $\vartheta_\ell(\theta) = \theta$ and $z_\ell(\theta) = \dot{q}(\theta)$ for all $\ell \in \{1, ..., L\}$ and $\theta \in (\theta_\ell, \theta_{\ell+1})$ and extend it to $[\theta_\ell, \theta_{\ell+1}]$ by taking left and right limits. Then $(q_\ell, V_\ell, \vartheta_\ell, z_\ell)_{\ell=1}^L$ is a global maximizer for problem (5), and so the necessary conditions from Theorem 3 apply.

First, it is standard to prove that $\eta = 1$. If not, the second adjoint equation stipulates that $\dot{\lambda}^V_\ell(\theta) = 0$ and therefore $\lambda^V_\ell(\theta) = \lambda^V_L(1) = 0$ for all ℓ and θ due to the switching condition and the transversality condition. It then follows from the first adjoint equation that $\dot{\lambda}^q_\ell(\theta) = 0$ for all ℓ and θ, and therefore that $\lambda^q_\ell(\theta) = 0$ due to the switching condition. It then must be that $\dot{\lambda}^\vartheta_\ell(\theta) = 0$ due to the third adjoint equation, whence $\lambda^\vartheta_\ell(\theta) = 0$ due to switching condition and the transversality condition. This however upsets the nontriviality condition. Second, the state trajectories and endpoint constraints clearly imply that $\vartheta_\ell(\theta) = \theta$. Third, the second adjoint equation implies that $\lambda^V_\ell(\theta) = const + \mu(\theta)$. And the switching condition and the transversality condition imply that $const = -1$, i.e., that $-\lambda^V_\ell(\theta) = 1 - M(\theta)$. Fourth, plugging $-\lambda^V_\ell(\theta) = 1 - M(\theta)$ into the first adjoint equation gives

$$-\dot{\lambda}^q_\ell(\theta) = D_q S(q_\ell(\theta); \theta)\mu(\theta) - (1 - M(\theta))D^2_{\theta q} S(q_\ell(\theta); \theta).$$

Or, equivalently, $-\dot{\lambda}^q_\ell(\theta) = D_q \Lambda(q(\theta); \theta)\mu(\theta)$. Due to the maximum condition it follows that if the quality assignment is separating ($\theta \mapsto q(\theta)$ is increasing or $z_\ell(\theta) > 0$) it must hold that $D_q \Lambda(q(\theta); \theta) = 0$, which corresponds to the first claim. As to the second claim, let $[\underline{\theta}, \overline{\theta}] \subseteq [\theta_{\ell-1}, \theta_\ell]$ be a maximal interval on which $q(\theta)$ is constant. Then either $q(\theta)$ is discontinuous at $\underline{\theta}$, so that $\lambda^q_\ell(\underline{\theta}) = 0$ due to the switching condition. Or there exists $\theta' : \theta_{\ell-1} \leq \theta' < \underline{\theta}$ so that $q(\theta)$ is separating on $(\theta', \underline{\theta})$. Then due to the maximum condition $\lambda^q_\ell(\underline{\theta}) = 0$. A symmetric reasoning ensures that $\lambda^q_\ell(\overline{\theta}) = 0$. Finally note that $\lambda^q_\ell(\theta) \leq 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ due to the maximum condition. Then applying the fundamental theorem of calculus and integrating $\dot{\lambda}^q_\ell(\theta)$ from $\underline{\theta}$ gives the desired result.

It remains to prove that $\lim_{\theta\uparrow\hat{\theta}} \Lambda(q(\theta);\theta) = \lim_{\theta\downarrow\hat{\theta}} \Lambda(q(\theta);\theta)$ at a point of discontinuity $\hat{\theta}$. This is a consequence of the switching condition for the maximized Hamiltonian. $\square$

A.6 Example: Continuous and Discontinuous Bunching

Figure 1 (left) illustrates continuous bunching: the profit-maximizing quality assignment is continuous, yet involves bunching. This occurs whenever virtual surplus does not satisfy increasing differences, i.e., if $D^2_{\theta q}\Lambda(q;\theta) \ngeq 0$. While commonly attributed to non-standard distribution functions, discontinuous bunching can in fact arise for any distribution function (see footnote 9).

I here consider surplus of the form $S(q;\theta) = (1+\theta)q - \frac{q^2}{2}$ where $\theta \in [0,1]$ and $\hat{u}(\theta) = S(0;\theta)$. (This corresponds to surplus $S(q;\theta) = \theta q - \frac{q^2}{2}$ as studied by Mussa and Rosen (1978) where $\theta \in [1,2]$.) The density is taken to be bi-modal (the second term $\frac{3}{2}\theta(1-\theta)$ merely ensures that $\mu(\theta) > 0$ for all $\theta \in (0,1)$):

$$\mu(\theta) = \frac{3}{4}(1-\cos(4\pi\theta)) + \frac{3}{2}\theta(1-\theta) \quad \text{and} \quad M(\theta) = \frac{1}{4}\theta(-2\theta^2+3\theta+3) - \frac{3}{4}\frac{\sin(4\pi\theta)}{4\pi}.$$

Since surplus is log supermodular in differences, it admits a unique virtual surplus (1) maximizing quality

$$\overline{q}(\theta) = 1+\theta - \frac{1-M(\theta)}{\mu(\theta)} = 1+\theta - \frac{1-\frac{1}{4}\theta(-2\theta^2+3\theta+3)+\frac{3}{4}\frac{\sin(4\pi\theta)}{4\pi}}{\frac{3}{4}+\frac{3}{2}\theta(1-\theta)-\frac{3}{4}\cos(4\pi\theta)}.$$

This is not non-decreasing in θ, and attains a local minimum at $\theta_3 \approx 0.495106$. So, some bunching must prevail. Following Proposition 2 (or Theorem 1) the profit-maximizing quality assignment is continuous. And Theorem 2 gives the familiar characterization of optimal bunching: If $(\underline{\theta}, \overline{\theta})$ is an optimal bunching interval, i.e., a maximal interval on which the profit-maximizing assignment $q^*(\theta)$ is constant equal to q, then $q = q^*(\underline{\theta}) = \overline{q}(\underline{\theta})$, $q = q^*(\overline{\theta}) = \overline{q}(\overline{\theta})$ and $\int_{\underline{\theta}}^{\hat{\theta}} D_q\Lambda(q;\theta)\mu(\theta)d\theta \geq 0$ holding with equality for $\hat{\theta} = \overline{\theta}$. Since one easily observes that $\overline{q}(\theta)$ is increasing-decreasing-increasing (as depicted by the dashed line in Figure 1) and virtual surplus is quasi-convex, the latter condition implies that there exists exactly one pooling interval. Then let $\theta_1 < \theta_2 < \theta_3 < \theta_4$ be the unique parameter values so that $\overline{q}(\theta_1) = \overline{q}(\theta_2)$, $\overline{q}(\theta_2) = \overline{q}(\theta_4)$ and $\theta \mapsto \overline{q}(\theta)$ is increasing on (θ_1, θ_2) and (θ_3, θ_4) and decreasing on (θ_2, θ_3). The profit-maximizing bunching interval must be such that $\underline{\theta} \in (\theta_1, \theta_2)$ and $\overline{\theta} \in (\theta_3, \theta_4)$. To solve for the profit-maximizing bunching interval numerically, perform two steps. First, identify for fixed $\underline{\theta} \in (\theta_1, \theta_2)$ the unique $\overline{\theta} \in (\theta_3, \theta_4)$ so that $\overline{q}(\underline{\theta}) = \overline{q}(\overline{\theta})$. Second, identify the (set of) $\underline{\theta} \in (\theta_1, \theta_2)$ so that $\int_{\underline{\theta}}^{\overline{\theta}} D_q\Lambda(q;\theta)\mu(\theta)d\theta$ where $q = \overline{q}(\underline{\theta})$ and $\overline{\theta}$ is identified by step 1. Numerical analysis reveals that the profit-maximizing bunching interval thus constructed is unique.

As to Figure 1 (right): When surplus is of the form $S(q;\theta) = (1+\theta)q$ with $\theta \in [0,1]$ (or $S(q;\theta) = \theta q - \frac{q^2}{2}$ with $\theta \in [1,2]$), it is weakly log submodular in differences. A single-contract menu is profit-maximizing due to Proposition 1. (Also observe that unless $D_q\Lambda(q;\theta) = 0$ for all q, or, $\theta = \frac{1-M(\theta)}{\mu(\theta)}$, any profit-maximizing menu consists of a single contract.) The profit-maximizing contract serves quality $q = 1$. It remains to determine the type where the quality assignment is discontinuous. Due to Lemma 3 such type is in $\operatorname{argmax}_{\theta\in[0,1]} [S(1;\theta) - S(0;\theta)](1-M(\theta))$. This set is a singleton, the unique maximizer is given by $\theta \approx 0.587646$.

A.7 Example: Single-Contract Menus Are Not Profit-Maximizing

In the main text (see Figure 4) it was claimed that in a common value environment single-contract menus need not be profit-maximizing even when CV-LsubD holds. But the construction is not straightforward. Suppose that CV-LsubD holds so that virtual surplus is quasi-convex in quality. For single-contract menus not to be profit-maximizing, it must be that the virtual surplus minimizing quality assignment is non-monotone in types. If so, the requirement that the quality assignment be non-decreasing (due to incentive compatibility) conflicts with the seller's objective to remove quality as far as possible from the pointwise minimum of virtual surplus. To then create an incentive for the seller to actively sell interior qualities in addition to $q \in \{0, 1\}$, it must further be that virtual surplus is not convex in quality. A minimal example satisfying both properties is the following:

Fix some distribution over types and let surplus and virtual surplus be given by[31]

$$S(q;\theta) = \psi(\theta) + \begin{cases} 1 - \alpha\sqrt{\underline{q}(\theta) - q} & \text{if } q < \underline{q}(\theta) \\ 1 + \alpha\sqrt{q - \underline{q}(\theta)} & \text{if } q \geq \underline{q}(\theta) \end{cases} \qquad \text{and} \qquad \Lambda(q;\theta) = \begin{cases} 1 - \beta(\gamma + q - \underline{q}(\theta))^2 & \text{if } q < \underline{q}(\theta) \\ 1 - \beta(-\gamma + q - \underline{q}(\theta))^2 & \text{if } q \geq \underline{q}(\theta) \end{cases}$$

where $\underline{q}(\theta) \in [0, 1]$. Observe that surplus satisfies Assumption 3. Furthermore note that CV-LsubD holds for all $\alpha > 0, \beta > 0, \gamma \geq 3$, for CV-LsubD is equivalent to

$$0 < D^2_{qq}\Lambda(q;\theta)D_qS(q;\theta) - D_q\Lambda(q;\theta)D^2_{qq}S(q;\theta) = \begin{cases} \frac{\alpha\beta(\gamma - 3(\underline{q}(\theta) - q))}{2(\underline{q}(\theta) - q)^{\frac{3}{2}}} & \text{if } q < \underline{q}(\theta) \\ \frac{\alpha\beta(\gamma - 3(q - \underline{q}(\theta)))}{2(q - \underline{q}(\theta))^{\frac{3}{2}}} & \text{if } q \geq \underline{q}(\theta). \end{cases}$$

The key observation is that virtual surplus is quasi-convex but not convex and $\underline{q}(\theta)$ corresponds to the pointwise minimum of virtual surplus.

Next, verify Assumption 2. The definition of virtual surplus entails that

$$D_\theta U(q;\theta)(1 - M(\theta)) = -\Lambda(q;\theta) + S(q;\theta) = \begin{cases} \psi(\theta) + \beta(\gamma + q - \underline{q}(\theta))^2 - \alpha\sqrt{\underline{q}(\theta) - q} & \text{if } q < \underline{q}(\theta) \\ \psi(\theta) + \beta(-\gamma + q - \underline{q}(\theta))^2 + \alpha\sqrt{q - \underline{q}(\theta)} & \text{if } q \geq \underline{q}(\theta). \end{cases}$$

Therefore

$$D^2_{q\theta}U(q;\theta)(1 - M(\theta)) = \begin{cases} 2\beta(\gamma + q - \underline{q}(\theta)) + \frac{\alpha}{2\sqrt{\underline{q}(\theta) - q}} & \text{if } q < \underline{q}(\theta) \\ 2\beta(-\gamma + q - \underline{q}(\theta)) + \frac{\alpha}{2\sqrt{q - \underline{q}(\theta)}} & \text{if } q \geq \underline{q}(\theta). \end{cases}$$

Assumption 2 is satisfied if this expression is positive. This holds if $\alpha > \beta(\frac{4}{3}\gamma)^{\frac{3}{2}}$. In effect, $\alpha = 1, 0 < \beta \leq \frac{1}{8}, \gamma = 3$ are admissible parameter values for which Assumptions 2,3 and condition CV-LsubD hold.[32] Then consider a non-monotone virtual surplus minimizing quality and let $\underline{q}(\theta)$ be piecewise affine with slope in

[31] Virtual surplus is not differentiable in quality when $q = \underline{q}(\theta)$. This is for ease of exposition only. It could be arbitrarily well approximated by a smooth function by means of convolution using mollifiers (effectively guaranteeing that Assumption 1 holds). This would convexify virtual surplus around this point and thereby strengthen the CV-LsubD characterizing inequality stated below.

[32] The definition of surplus and virtual surplus is consistent with any utility and cost function that can be expressed as

$$U(q';\theta') = \int_0^{\theta'} D_\theta U(q';\theta)d\theta + \varphi(q') \qquad \text{and} \qquad C(q';\theta') = S(q';\theta') - \int_0^{\theta'} D_\theta U(q';\theta)d\theta - \varphi(q').$$

$\{-m, 0, m\}$ as depicted in Figure 4. To conclude the description of the example, let θ be uniformly distributed.

It is easy to verify that when m is sufficiently large there exists a two-contract menu that guarantees strictly greater profit than any single-contract menu: consider the two-contract menu depicted in Figure 4. $q^* \in \{0, \frac{1}{2}, 1\}$ with points of discontinuity given by $\frac{5}{4m}$ and $1 - \frac{5}{4m}$. The profit-maximizing single-contract menu, by contrast, is characterized by a discontinuity at either $\frac{1}{m}$ or $1 - \frac{1}{m}$ as those points are the only solutions to $\Lambda(0; \theta) = \Lambda(1; \theta)$. Symmetry of virtual surplus implies that the choice of either discontinuity yields the same seller profit. Thus suppose that $q^*(\theta) = 0$ for $\theta < \frac{1}{m}$ and $q^*(\theta) = 1$ otherwise. Pointwise comparing the seller's single-contract menu with the two-contract menu reveals that the single-contract menu does strictly better for types $\theta \in [\frac{1}{m}, \frac{3}{2m})$, yet strictly worse for types $\theta \in [\frac{3}{2m}, 1 - \frac{3}{2m})$ because virtual surplus is concave in quality for $q > \underline{q}(\theta)$ and $q < \underline{q}(\theta)$. As m grows large the former interval converges to a point, and the latter to the unit interval. Since both differences are uniformly bounded from below and above for all m, it then must be that the two-contract menu dominates for m sufficiently large.

In what follows we instead formally ascertain that selling qualities $q^* \in \{0, \frac{1}{2}, 1\}$ is profit-maximizing by drawing on the characterizations in Theorems 1 and 2. First observe that the proof of Theorem 1 item 2 applies, so that the profit-maximizing quality assignment is piecewise constant. In light of the seller's objective (1) one immediately notices that setting $q^*(\theta) = 0$ for all θ on the first increasing segment $[0, \frac{1}{2m}]$ and $q^*(\theta) = 1$ for all θ on the last increasing segment $[1 - \frac{1}{2m}, 1]$ must be profit-maximizing, for those points are maximally removed from the pointwise minimum of virtual surplus and do not reduce the set of implementable quality assignments for interior θ. Optimizing over points of discontinuity further implies that $\lim_{\theta \uparrow \hat{\theta}} \Lambda(q(\theta); \theta) = \lim_{\theta \downarrow \hat{\theta}} \Lambda(q(\theta); \theta)$ at any point of discontinuity $\hat{\theta}$. This implies that at a point of discontinuity $\hat{\theta}$ the quality assignment must cross $\underline{q}(\hat{\theta})$ from below. It then follows from the sign changes of the slope of $\underline{q}(\theta)$ that the profit-maximizing assignment includes at least one and at most two points of discontinuity. We have already seen that two-contract menus must dominate, so focus on the the latter case. Denote $q^{\xi} \in (0, 1)$ the interior quality and $\theta_1^{\xi} < \theta_2^{\xi}$ the points of discontinuity. Those must satisfy $\theta_1^{\xi} \in (\frac{1}{2m}, \frac{3}{2m})$ and $\theta_2^{\xi} \in (1 - \frac{3}{2m}, 1 - \frac{1}{2m})$. Moreover, due to symmetry of $q \mapsto \Lambda(q; \theta)$ around $\underline{q}(\theta)$, $\Lambda(0; \theta_1^{\xi}) = \Lambda(q^{\xi}; \theta_1^{\xi})$ and $\Lambda(q^{\xi}; \theta_2^{\xi}) = \Lambda(1; \theta_2^{\xi})$ imply that $q^{\xi} = 2\underline{q}(\theta_1^{\xi})$ and $q^{\xi} = 2\underline{q}(\theta_2^{\xi}) - 1$. Then note that $\underline{q}(\theta)$ is piecewise affine so that $\underline{q}(\theta_1^{\xi}) = 1 - (\theta_1^{\xi} - \frac{1}{2m})m = \frac{3}{2} - \theta_1^{\xi} m$ and $\underline{q}(\theta_2^{\xi}) = 1 - (\theta_2^{\xi} - (1 - \frac{3}{2m}))m = m - \frac{1}{2} - \theta_2^{\xi} m$. It follows that

$$\theta_1^{\xi} = \frac{3}{2m} - \frac{q^{\xi}}{2m} \qquad \text{and} \qquad \theta_2^{\xi} = \frac{m-1}{m} - \frac{q^{\xi}}{2m}.$$

Finally, optimizing over q^{ξ} implies that $0 = \int_{\theta_1^{\xi}}^{\theta_2^{\xi}} D_q \Lambda(q^{\xi}; \theta) d\theta$. (Both optimality conditions are asserted by Theorem 2 and equally follow from pointwise optimization.) Setting $q^{\xi} = \frac{1}{2}$ clearly is a solution. Are there

Here $\varphi(q)$ is arbitrary. One can moreover choose $\psi(\theta)$ so that $C(0; \theta) = 0$ for all $\theta \in [0, 1]$. Furthermore, one easily notes that Assumption 2 implies that utility is increasing in q (provided that $\varphi(q)$ is non-decreasing). The same can not be said about the cost function: Since in the construction above $\Lambda(q; 1) \neq S(q; 1)$, it follows from the definition of virtual surplus that $\lim_{\theta \uparrow 1} D_{q\theta} U(q; \theta) = \infty$. In effect, the example entails an extreme form of advantageous selection (meaning that the most eager to trade are also the least costly to serve): marginal cost $D_q C(q; \theta)$ is positive for low types and negative for sufficiently high types. It should be feasible to construct another example that does not share this extreme property, e.g., virtual surplus given by $(M(\theta))^{\alpha} S(q; \theta) + (1 - M(\theta))^{\alpha} \Lambda(q; \theta)$ with $S(q; \theta)$ and $\Lambda(q; \theta)$ as given above and $\alpha \in (0, 1)$ small. But this would be more tedious to analyze.

others? No! Since $\underline{q}(\theta)$ is piecewise affine, the latter condition can be rewritten as

$$
\begin{aligned}
0 =& \int_{\frac{3}{2m}-\frac{q^{\varepsilon}}{2m}}^{\frac{3}{2m}} -2\beta(-\gamma + q^{\varepsilon} - (\frac{3}{2} - \theta m))d\theta + \int_{\frac{3}{2m}}^{\frac{1}{2}-\frac{1}{2m}} -2\beta(-\gamma + q^{\varepsilon})d\theta + \int_{\frac{1}{2}-\frac{1}{2m}}^{\frac{1}{2}-\frac{1}{2m}+\frac{q^{\varepsilon}}{m}} -2\beta(-\gamma + q^{\varepsilon} - (\theta - (\frac{1}{2} - \frac{1}{2m}))m)d\theta \\
&+ \int_{\frac{1}{2}-\frac{1}{2m}+\frac{q^{\varepsilon}}{m}}^{\frac{1}{2}+\frac{1}{2m}} -2\beta(\gamma + q^{\varepsilon} - (\theta - (\frac{1}{2} - \frac{1}{2m}))m)d\theta + \int_{\frac{1}{2}+\frac{1}{2m}}^{1-\frac{3}{2m}} -2\beta(\gamma + q^{\varepsilon} - 1)d\theta + \int_{1-\frac{3}{2m}}^{1-\frac{3}{2m}+\frac{1-q^{\varepsilon}}{2m}} -2\beta(\gamma + q^{\varepsilon} - (m - \frac{1}{2} - \theta m))d\theta \\
=& 2\beta\left\{\left((\gamma - \frac{3q^{\varepsilon}}{4})\frac{q^{\varepsilon}}{2m} + (\gamma - q^{\varepsilon})(\frac{1}{2} - 2m) + (\gamma - \frac{q^{\varepsilon}}{2})\frac{q^{\varepsilon}}{m}\right) - \left((\gamma - \frac{1-q^{\varepsilon}}{2})\frac{1-q^{\varepsilon}}{m} + (\gamma + q^{\varepsilon} - 1)(\frac{1}{2} - 2m) + (\gamma - \frac{3}{4}(1-q^{\varepsilon}))\frac{1-q^{\varepsilon}}{2m}\right)\right\}.
\end{aligned}
$$

Then note that this expression is affine in q^{ε}. It follows that there exists a unique q^{ε} for which the integral is zero, whence $q^{\varepsilon} = \frac{1}{2}$ as claimed.

Online: proof of Proposition 2

Suppose by contradiction that $q^*(\theta)$ is discontinuous in $(\underline{\theta}, \overline{\theta})$. Then there exists $\theta^* \in (\underline{\theta}, \overline{\theta})$ and an implementable and profit-maximizing $q^* : [0,1] \to \mathbb{R}_+$ that exhibits a discontinuity at θ^*:

$$\triangle q \equiv \lim_{\theta \downarrow \theta^*} q^*(\theta) - \lim_{\theta \uparrow \theta^*} q^*(\theta) > 0.$$

For ease of notation, denote $\overline{q} \equiv \lim_{\theta \downarrow \theta^*} q^*(\theta)$ and $\underline{q} \equiv \lim_{\theta \uparrow \theta^*} q^*(\theta)$ and assume that $q^*(\theta^*) = \overline{q}$.

Claim A: If surplus is log supermodular in differences, then for all $\theta_L, \theta_H : \theta_L < \theta^* < \theta_H$ and $\hat{q} \in [\underline{q}, \overline{q}]$

$$[S(\overline{q}; \theta^*) - S(\hat{q}; \theta^*)](1 - M(\theta^*)) \geq [S(\overline{q}; \theta_H) - S(\hat{q}; \theta_H)](1 - M(\theta_H))$$
$$[S(\hat{q}; \theta^*) - S(\underline{q}; \theta^*)](1 - M(\theta^*)) \geq [S(\hat{q}; \theta_L) - S(\underline{q}; \theta_L)](1 - M(\theta_L)).$$

Proof. Fix arbitrary $\theta_L < \theta^* < \theta_H$ and $\underline{q} \leq q_L < q_H \leq \overline{q}$. Let $(\delta_N)_{N \in \mathbb{N}}$ be a sequence so that $\delta_N \to 0$. Then consider, for fixed N, a partition $\theta_0 < \theta_L = \theta_1 < ... < \theta_N = \theta_H$ so that there exists $\theta_M \in \{\theta_0, ..., \theta_N\}$: $\theta_M = \theta^*$ and $|\theta_{j+1} - \theta_j| < \delta_N$ for all $j : 0 \leq j < N$.

Then construct a sequence of deviations[33] $(q_N)_{N \in \mathbb{N}}$, one for each N, so that (i) q_N coincides with q^* for $\theta \notin [\theta_L, \theta_H]$ and (ii) q_N is constant on each interval (θ_j, θ_{j+1}):

- $q_N(\theta_0) = q^*(\theta_0)$, $q_N(\theta_N) = q^*(\theta_N)$, $q_N(\theta_M) = q_H$ and $q_N(\theta_{M-1})$ is such that

$$S(q_L; \theta_M) - S(q_N(\theta_{M-1}); \theta_M) = S(\underline{q}; \theta_M) - S(q^*(\theta_{M-1}); \theta_M).$$

  This implies that $\lim_{N \to \infty} q_N(\theta_{M-1}) = q_L$;

- For $\ell \in \{1, ..., M-2, M+1, ..., N-1\}$, $q_N(\theta_\ell)$ is implicitly defined by

$$S(q_N(\theta_\ell); \theta_\ell) - S(q_N(\theta_{\ell-1}); \theta_\ell) = S(q^*(\theta_\ell); \theta_\ell) - S(q^*(\theta_{\ell-1}); \theta_\ell).$$

Observe that by construction $q^*(\theta_\ell) \leq q_N(\theta_\ell)$ for $\ell \in \{1, ..., M-1\}$ and $q^*(\theta_\ell) \geq q_N(\theta_\ell)$ for $\ell \in \{M, ..., N-1\}$ and the inequalities are strict if and only if $\underline{q} < q_L$ and $\overline{q} > q_H$ respectively.

Since q^* is profit-maximizing, lemma 3 implies that for all fixed $N' \in \mathbb{N}$

$$\lim_{N \to \infty} \sum_{\ell=1}^{N} [(S(q^*(\theta_\ell); \theta_\ell) - S(q^*(\theta_{\ell-1}); \theta_\ell)) - (S(q_{N'}(\theta_\ell); \theta_\ell) - S(q_{N'}(\theta_{\ell-1}); \theta_\ell))](1 - M(\theta_\ell)) \geq 0.$$

Here $N \to \infty$ concerns the partition $\{\theta_0, ..., \theta_N\}$. Then take the limit $N' \to \infty$ and notice that the limits $\lim_{N' \to \infty}$ and $\lim_{N \to \infty}$ are interchangeable (use the virtal surplus representation, e.g., Lemma 2, then apply the dominated convergence theorem). In particular, both coincide with the limit $N = N' \to \infty$. Whence

$$\lim_{N \to \infty} \Big\{ \Big[S(q^*(\theta_L); \theta_L) - S(q_N(\theta_L); \theta_L)\Big](1 - M(\theta_L)) - \Big[S(q^*(\theta_{N-1}); \theta_H) - S(q_N(\theta_{N-1}); \theta_H)\Big](1 - M(\theta_H))$$
$$+ \Big[(S(q^*(\theta_M); \theta_M) - S(q^*(\theta_{M-1}); \theta_M)) - (S(q_N(\theta_M); \theta_M) - S(q_N(\theta_{M-1}); \theta_M))\Big](1 - M(\theta_M)) \Big\} \geq 0. \quad (6)$$

[33] We ignore participation constraints that can be accomodated by bounding the deviation by a sufficiently small ϵ as in the main text.

Then distinguish between two cases:

Case 1: $q_H = \overline{q}$ and $q_L > \underline{q}$. Inequality (6) simplifies as follows:

$$\lim_{N\to\infty}\left\{\Big[S(q_N(\theta_L);\theta_L) - S(q^*(\theta_L);\theta_L)\Big](1-M(\theta_L)) - \Big[S(q_N(\theta_{M-1});\theta_M) - S(q^*(\theta_{M-1});\theta_M)\Big](1-M(\theta_{M-1}))\right\} \le 0$$

$$\Leftrightarrow \lim_{N\to\infty}\left\{\left(\sum_{\ell=2}^{M-1}\Big[S(q^*(\theta_\ell);\theta_L) - S(q^*(\theta_{\ell-1});\theta_L)\Big] - \Big[S(q_N(\theta_\ell);\theta_L) - S(q_N(\theta_{\ell-1});\theta_L)\Big]\right)(1-M(\theta_L))\right.$$
$$\left. + \Big[S(q_N(\theta_{M-1});\theta_L) - S(q^*(\theta_{M-1});\theta_L)\Big](1-M(\theta_L)) - \Big[S(q_N(\theta_{M-1});\theta_M) - S(q^*(\theta_{M-1});\theta_M)\Big](1-M(\theta_{M-1}))\right\} \le 0.$$

Then note that $q_N(\theta_\ell) > q^*(\theta_\ell)$ and consider the preceding summand:

$$\Big[S(q^*(\theta_\ell);\theta_L) - S(q^*(\theta_{\ell-1});\theta_L)\Big] - \Big[S(q_N(\theta_\ell);\theta_L) - S(q_N(\theta_{\ell-1});\theta_L)\Big]$$
$$= \Big(S(q^*(\theta_\ell);\theta_\ell) - S(q^*(\theta_{\ell-1});\theta_\ell)\Big)\left[\frac{S(q^*(\theta_\ell);\theta_L) - S(q^*(\theta_{\ell-1});\theta_L)}{S(q^*(\theta_\ell);\theta_\ell) - S(q^*(\theta_{\ell-1});\theta_\ell)} - \frac{S(q_N(\theta_\ell);\theta_L) - S(q_N(\theta_{\ell-1});\theta_L)}{S(q_N(\theta_\ell);\theta_\ell) - S(q_N(\theta_{\ell-1});\theta_\ell)}\right].$$

Provided that surplus is log supermodular in differences this term is positive. In effect,

$$\lim_{N\to\infty}\left\{\Big[S(q_N(\theta_{M-1});\theta_L) - S(q^*(\theta_{M-1});\theta_L)\Big](1-M(\theta_L)) - \Big[S(q_N(\theta_{M-1});\theta_M) - S(q^*(\theta_{M-1});\theta_M)\Big](1-M(\theta_{M-1}))\right\} \le 0$$
$$\Leftrightarrow \quad \Big[S(q_L;\theta_L) - S(\underline{q};\theta_L)\Big](1-M(\theta_L)) - \Big[S(q_L;\theta_M) - S(\underline{q};\theta_M)\Big](1-M(\theta^*)) \le 0.$$

Case 2: $q_H < \overline{q}$ and $q_L = \underline{q}$. Inequality (6) simplifies as follows:

$$\lim_{N\to\infty}\left\{\Big[S(q^*(\theta_{N-1});\theta_H) - S(q_N(\theta_{N-1});\theta_H)\Big](1-M(\theta_H)) - \Big[S(q^*(\theta_M);\theta_M) - S(q_N(\theta_M);\theta_M)\Big](1-M(\theta_M))\right\} \le 0$$

$$\Leftrightarrow \lim_{N\to\infty}\left\{\left(\sum_{\ell=M+1}^{N-1}\Big[S(q^*(\theta_\ell);\theta_H) - S(q^*(\theta_{\ell-1});\theta_H)\Big] - \Big[S(q_N(\theta_\ell);\theta_H) - S(q_N(\theta_{\ell-1});\theta_H)\Big]\right)(1-M(\theta_H))\right.$$
$$\left. + \Big[S(q^*(\theta_M);\theta_H) - S(q_N(\theta_M);\theta_H)\Big](1-M(\theta_H)) - \Big[S(q^*(\theta_M);\theta_M) - S(q_N(\theta_M);\theta_M)\Big](1-M(\theta_M))\right\} \le 0.$$

Then note that $q^*(\theta_\ell) > q_N(\theta_\ell)$ and consider the preceding summand:

$$\Big[S(q^*(\theta_\ell);\theta_H) - S(q^*(\theta_{\ell-1});\theta_H)\Big] - \Big[S(q_N(\theta_\ell);\theta_H) - S(q_N(\theta_{\ell-1});\theta_H)\Big]$$
$$= \Big(S(q^*(\theta_\ell);\theta_\ell) - S(q^*(\theta_{\ell-1});\theta_\ell)\Big)\left[\frac{S(q^*(\theta_\ell);\theta_L) - S(q^*(\theta_{\ell-1});\theta_H)}{S(q^*(\theta_\ell);\theta_\ell) - S(q^*(\theta_{\ell-1});\theta_\ell)} - \frac{S(q_N(\theta_\ell);\theta_H) - S(q_N(\theta_{\ell-1});\theta_L)}{S(q_N(\theta_\ell);\theta_\ell) - S(q_N(\theta_{\ell-1});\theta_\ell)}\right].$$

Provided that surplus is log supermodular in differences this term is positive. In effect,

$$\lim_{N\to\infty}\left\{\Big[S(q^*(\theta_M);\theta_H) - S(q_N(\theta_M);\theta_H)\Big](1-M(\theta_H)) - \Big[S(q^*(\theta_M);\theta_M) - S(q_N(\theta_M);\theta_M)\Big](1-M(\theta_M))\right\} \le 0$$
$$\Leftrightarrow \quad \Big[S(\overline{q};\theta_H) - S(q_H;\theta_H)\Big](1-M(\theta_H)) - \Big[S(\overline{q};\theta_M) - S(q_H;\theta_M)\Big](1-M(\theta^*)) \le 0.$$

□

For ease of notation define, for $q \in (\underline{q}, \overline{q})$,

$$\phi(q;\theta) = [S(q;\theta) - S(\underline{q};\theta)](1 - M(\theta)) \quad \text{and} \quad \varphi(q;\theta) = [S(\overline{q};\theta) - S(q;\theta)](1 - M(\theta)).$$

Claim B: $D_\theta\phi(\overline{q};\theta^*) = 0$ and $D_\theta\varphi(\underline{q};\theta^*) = 0$.

Proof of claim B. Both claims are equivalent. Consider the following deviation:

$$\hat{q}_\epsilon^- = \begin{cases} q^*(\theta) & \text{if } \theta \notin (\theta^* - \epsilon, \theta^*) \\ q^*(\theta) + \triangle q & \text{if } \theta \in (\theta^* - \epsilon, \theta^*) \end{cases}$$

Since q^* is profit-maximizing, $\lim\limits_{\epsilon \downarrow 0} \frac{\Pi(\hat{q}_\epsilon^-) - \Pi(q^*)}{\epsilon} \leq 0$. And Lemma 2 implies that

$$\begin{aligned}\frac{\Pi(\hat{q}_\epsilon^-) - \Pi(q^*)}{\epsilon} &= \frac{1}{\epsilon}\int_{\theta^*-\epsilon}^{\theta^*} \Big\{[S(q^*(\theta) + \triangle q;\theta) - S(q^*(\theta);\theta)]\mu(\theta) - [D_\theta S(q^*(\theta) + \triangle q;\theta) - D_\theta S(q^*(\theta);\theta)](1 - M(\theta))\Big\} d\theta \\ &\longrightarrow [S(\overline{q};\theta^*) - S(\underline{q};\theta^*)]\mu(\theta^*) - [D_\theta S(\overline{q};\theta^*) - D_\theta S(\underline{q};\theta^*)](1 - M(\theta^*)) = D_\theta\phi(\overline{q};\theta^*) \leq 0.\end{aligned}$$

□

Claim C: If surplus is log supermodular in differences, then $\exists\, \theta_L, \theta_H : \theta_L < \theta^* < \theta_H$ and $q_L, q_H \in (\underline{q}, \overline{q})$ so that

$$\begin{aligned}&[S(\overline{q};\theta_L) - S(q_L;\theta_L)](1 - M(\theta_L)) > [S(\overline{q};\theta^*) - S(q_L;\theta^*)](1 - M(\theta^*)) \\ &[S(q_H;\theta_H) - S(\underline{q};\theta_H)](1 - M(\theta_H)) > [S(q_H;\theta^*) - S(\underline{q};\theta^*)](1 - M(\theta^*)).\end{aligned}$$

Proof. First consider $\phi(q;\theta)$. Distinguish between two cases:

Case 1: $D^2_{q\theta}\phi(\overline{q};\theta^*) > 0$. Then there exists $\theta_L < \theta^*$ and $q_H < \overline{q}$ so that $\phi(q_H;\theta_L) > \phi(q_H;\theta^*)$. Proof: Since $D^2_{q\theta}\phi(\overline{q};\theta^*)$ is continuous (by assumption 1), there exists some $q_H < \overline{q}$ so that $D^2_{q\theta}\phi(q;\theta^*) < 0$ for all $q \in [q_H, \overline{q})$. And due to claim B and the fundamental theorem of calculus, $0 = D_\theta\phi(\overline{q};\theta^*) = D_\theta\phi(q_H;\theta^*) + \int_{q_H}^{\overline{q}} D^2_{q\theta}\phi(q;\theta^*)dq > D_\theta\phi(q_H;\theta^*)$. Whence there exists $\theta_L < \theta^*$ so that $\phi(q_H;\theta_H) > \phi(q_H;\theta^*)$. Since this contradicts claim A, deduce that $D^2_{q\theta}\phi(\overline{q};\theta^*) \not> 0$.

Case 2: $D^2_{q\theta}\phi(\overline{q};\theta^*) \leq 0$. Then there exists $\theta_H > \theta^*$ and $q_H < \overline{q}$ so that $\phi(q_H;\theta_H) > \phi(q_H;\theta^*)$. Proof: Since surplus is log supermodular in differences, or equivalently, $q \mapsto \frac{D^2_{q\theta}S(q;\theta)}{D_q S(q;\theta)}$ is increasing in q, it follows that $q \mapsto D^2_{q\theta}\phi(q;\theta)$ is increasing. Then clearly there exists some $q_H < \overline{q}$ so that $D^2_{q\theta}\phi(q;\theta^*) < 0$ for all $q \in [q_H, \overline{q})$. Then, as before, due to claim B and the fundamental theorem of calculus, $0 = D_\theta\phi(\overline{q};\theta^*) = D_\theta\phi(q_H;\theta^*) + \int_{q_H}^{\overline{q}} D^2_{q\theta}\phi(q;\theta^*)dq < D_\theta\phi(q_H;\theta^*)$. Whence there exists $\theta_H > \theta^*$ so that $\phi(q_H;\theta_H) > \phi(q_H;\theta^*)$.

As to $\varphi(q;\theta)$, the existence of $q_L > \underline{q}$ and $\theta_L < \theta^*$ so that $\varphi(q_L;\theta_L) \geq \varphi(q_L;\theta^*)$ follows analogously. □

To conclude the proof, observe that due to claims A and C there exist $\theta_L < \theta_H$ and $q_L, q_H \in (\underline{q}, \overline{q})$ so that

$$\begin{aligned}&[S(\overline{q};\theta_L) - S(q_L;\theta_L)](1 - M(\theta_L)) > [S(\overline{q};\theta^*) - S(q_L;\theta^*)](1 - M(\theta^*)) \geq [S(\overline{q};\theta_H) - S(q_L;\theta_H)](1 - M(\theta_H)) \\ &[S(q_H;\theta_H) - S(\underline{q};\theta_H)](1 - M(\theta_H)) > [S(q_H;\theta^*) - S(\underline{q};\theta^*)](1 - M(\theta^*)) \geq [S(q_H;\theta_L) - S(\underline{q};\theta_L)](1 - M(\theta_L)).\end{aligned}$$

Dividing one by the other yields $\frac{S(\overline{q};\theta_L) - S(q_L;\theta_L)}{S(q_H;\theta_L) - S(\underline{q};\theta_L)} > \frac{S(\overline{q};\theta_H) - S(q_L;\theta_H)}{S(q_H;\theta_H) - S(\underline{q};\theta_H)}$, in spite of log supermodularity in differences.

References

Anderson, E. T. and Dana, J. (2009). When is price discrimination profitable? *Management Science*, 55(6):980–989. 2, 6, 12, 14

Arrow, K. J. (1965). *Aspects of the theory of risk-bearing*. Yrjö Jahnssonin Säätiö. 4

Baron, D. P. and Myerson, R. (1982). Regulating a monopolist with unknown costs. *Econometrica*, 50(4):911–30. 7

Belleflamme, P. and Peitz, M. (2015). *Industrial Organization: Markets and Strategies*. Cambridge University Press. 2

Bergemann, D., Heumann, T., and Morris, S. (2022). Screening with persuasion. 6

Bonneton, N. and Sandmann, C. (2022). Non-stationary search and assortative matching. 4

Chade, H. and Schlee, E. E. (2020). Insurance as a lemons market: Coverage denials and pooling. *Journal of Economic Theory*, 189:105085. 6

Chade, H. and Swinkels, J. M. (2019). The no-upward-crossing condition, comparative statics, and the moral-hazard problem. *Theoretical Economics*. 4

Clarke, F. (2013). *Functional Analysis, Calculus of Variations and Optimal Control*. Springer. 5, 18, 26, 27, 28

DeJarnette, P., Dillenberger, D., Gottlieb, D., and Ortoleva, P. (2020). Time lotteries and stochastic impatience. *Econometrica*, 88(2):619–656. 7

Deneckere, R. J. and McAfee, R. P. (1996). Damaged goods. *Journal of Economics and Management Strategy*, 5(2):149–174. 6

Dilmé, F. and Garrett, D. (2022). A dynamic theory of random price discounts. 6

Doval, L. and Skreta, V. (2022). Purchase history and product personalization. 6

Gershkov, A., Moldovanu, B., Strack, P., and Zhang, M. (2022). Optimal insurance: Dual utility, random losses and adverse selection. 6

Gollier, C. (2004). *The Economics of Risk and Time*, volume 1. The MIT Press, 1 edition. 10

Guesnerie, R. and Laffont, J.-J. (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. 6, 16, 18

Haghpanah, N. and Hartline, J. (2021). When is pure bundling optimal. *Review of Economic Studies*, 88:1127–1156. 6

Iorio, R. J. J. and Iorio, V. d. M. (2001). *Fourier Analysis and Partial Differential Equations*. Cambridge University Press. 9

Itoh, M. (1983). Monopoly, product differentiation and economic welfare. *Journal of Economic Theory*, 31:88–104. 2, 14

Johnson, J. P. and Myatt, D. P. (2006). Mutli-product cournot oligopoly. *RAND Journal of Economics*, 37(3):583–601. 6

Jullien, B. (2000). Participation constraints in adverse selection models. *Journal of Economic Theory*, 93(1):1–47. 6, 8, 14, 15

Kleiner, A., Moldovanu, B., and Strack, P. (2021). Extreme points and majorization: Economic applications. *Econometrica*, 89(4):1557–1593. 5

Lollivier, S. and Rochet, J.-C. (1983). Bunching and second-order conditions: A note on optimal tax theory. *Journal of Economic Theory*, 31(2):392–400. 6

Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *RAND Journal of Economics*, 15(2):171–196. 6, 7, 13

Mirrlees, J. A. (1971). An exploration in the theory of optimal income taxation. *Review of Economic Studies*, 38:175–208. 7

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18(2):301–317. 1, 2, 6, 7, 8, 10, 14, 29

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, pages 58–73. 2

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1-2):122–136. 4, 10, 14

Rochet, J. (1987). A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191–200. 24

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, 90(4):629–649. 19

Salant, S. W. (1989). When is inducing self-selection suboptimal for a monopolist? *The Quarterly Journal of Economics*, 104(2):391–397. 2, 6

Sandmann, C. (2022). Oligopolistic nonlinear pricing: a random search model. 19

Shapiro, C. and Varian, H. R. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press. 2, 10

Stokey, N. L. (1979). Intertemporal price discrimination. *The Quarterly Journal of Economics*, 93(3):355–371. 1, 2, 6, 7, 10

Strausz, R. (2006). Deterministic versus stochastic mechanisms in principal–agent models. *Journal of Economic Theory*, 128(1):306 – 314. 6, 13

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton University Press. 10

Villas-Boas, J. M. (1998). Product line design for a distribution channel. *Marketing Science*, 17(2):156–169.
6