

Monopolistic Nonlinear Pricing With Consumer Entry ^{*}

Lixin Ye[†] Chenglin Zhang[‡]

August 2014

Abstract

We consider consumer entry in the canonical monopolistic nonlinear pricing model (Mussa and Rosen, 1978) wherein consumers learn their preference “types” after incurring privately known entry costs. We show that by taking into account consumer entry, the nature of optimal nonlinear pricing contracts changes significantly: compared to the benchmark without costly entry, in our model both quality distortion and market exclusion are reduced, sorting is more likely, and whenever bunching occurs, the bunching interval is necessarily smaller. Additionally, under certain conditions the monopoly solution may even achieve the first best (i.e., production efficiency). We also demonstrate that the optimal monopoly solutions can be ranked according to inverse hazard rate functions of the entry cost, which suggests an interesting dynamic for monopolistic nonlinear pricing with consumer entry.

Keywords: Monopoly, nonlinear pricing, information acquisition, consumer entry, quality distortion, market exclusion.

JEL Classification: D82, D23, L12, L15

1 INTRODUCTION

Since the pioneering work of Mussa and Rosen (1978) and Maskin and Riley (1984), there has been a growing literature on nonlinear pricing. In a typical nonlinear pricing model with vertically differentiated products, the varieties of a product are indexed by *quality*, q , which summarizes the underlying attributes of the product.¹ One central task in this literature is how to construct optimal nonlinear pricing contracts in which different “types” of consumers are induced to sort themselves to different varieties of products.

^{*}We thank Mark Armstrong, Hongbin Cai, Johannes Hörner, Jingfeng Lu, George Mailath, Alessandro Pavan, James Peck, Mike Peters, Huanxing Yang, and seminar audiences at Ohio State University, Peking University, Michigan State University, University of Hong Kong, and the Workshop on Information and Incentives (University of Guelph, Canada) for very helpful comments and suggestions. All remaining errors are our own.

[†]Department of Economics, The Ohio State University, 449A Arps Hall, 1945 North High Street, Columbus, OH 43210. Email: ye.45@osu.edu.

[‡]Department of Economics, The Ohio State University, 410 Arps Hall, 1945 North High Street, Columbus, OH 43210. Email: zhang.4515@osu.edu.

¹The term “nonlinear pricing” is more accurate in settings where q is the quantity as in Maskin and Riley (1984). However, such settings are mathematically equivalent to those in which q is interpreted as quality. We thus follow the literature and use the phrase “nonlinear pricing” throughout this paper.

An implicit assumption in this well-developed literature is that consumers are endowed with their preference “types.” For example, in the canonical model of Mussa and Rosen (1978), a type, θ , is the preference “intensity” that “measures intensity of a consumer’s taste for quality” (Mussa-Rosen, page 303). More precisely, θ is the marginal utility of quality, or the consumer’s marginal rate of substitution, which completely determines a consumer’s preference over q and money. A fundamental assumption in Mussa-Rosen’s analysis, as well as in the overall nonlinear pricing literature, is that consumers know their θ ’s at the outset of the game and make purchase decisions based on their known types.

For highly familiar products or services (e.g., electricity, telephone service, newspaper subscriptions, etc.), it is reasonable to assume that consumers are well aware of their preference intensities. However, for some relatively new products or services, it may be less reasonable to assume that one is endowed with her preference type for free. For example: without actually watching 3D televisions with two different displaying technologies, one may never know about their “incremental value” of watching a model that does not require to wear eyeglasses over watching one that does;² after Smartphones were introduced, many users have been confused over which data plan to subscribe for, reflecting the uncertainty about their preferences over different data capacities needed;³ even when purchasing a standard product like a new car, one may not settle down with a specific model (say, Mercedes-Benz C350 or E350) until after some test driving.

The above examples suggest that consumers often need to make efforts to discover their preferences (e.g., through trying the product or test driving). We believe that many other products or services share this common feature. For these markets, it would be more sensible to assume that it is costly for consumers to participate in the sales and learn about their true preference types, as trying the product or simply spending some time to learn about its different features is demanding in both effort and time. We believe this is particularly true for new products. According to Clay Christensen at Harvard Business School, 30,000 new consumer products are launched each year.⁴ Given this astounding number of new products, it is unrealistic to assume that consumers know their preference types for all of them.

In this paper, we explicitly take into account the opportunity costs in learning one’s preferences in a standard monopolistic nonlinear pricing model that is otherwise identical to the original Mussa-Rosen model. More specifically, we model this costly learning process as an entry/participation decision. Continuing with the examples above, in order to buy a 3D television, a consumer will need to visit a store to find out the specific features of a 3D TV model; in order to buy a new car, a consumer will need to visit a car dealership for test driving; in order to sign up for a data plan of a Smartphone, a consumer will

²The basic requirement for creating 3D perception is to display offset images that are filtered separately to the left and right eyes. Two technologies are currently available: having the viewer wear eyeglasses to filter the separately offset images to each eye, or having the light source split the images directionally into the viewer’s eyes (no glasses are required).

³“A Comprehensive Guide To All Those Confusing Smartphone Data Plans” by Steve Kovach, March 06, 2012: http://articles.businessinsider.com/2012-03-06/tech/31126289_1_unlimited-data-data-plan-unlimited-plans

⁴“Clay Christensen’s Milkshake Marketing,” by Carmen Nobel, Harvard Business School Working Knowledge, February 14, 2011.

need to talk to a sales representative to understand the subtlety of different data plans, etc. We thus add a costly entry/participation stage to the Mussa-Rosen model, so that each consumer needs to incur a privately known entry cost, c_i , in order to participate in the sale and learn her preference type, θ_i .⁵

In the traditional nonlinear pricing setting, where consumers are passively endowed with private information about their preference types, the analysis usually focuses on optimal elicitation of that private information. When costly entry is taken into account, optimal nonlinear pricing is potentially challenging as it has to balance entry and information elicitation, which are interdependent: the nonlinear pricing contract has a direct effect on the set of entrants to be induced (and hence the actual market base for the product), and consumer entry imposes restrictions on the optimal nonlinear pricing contracts to be offered.

Nevertheless, using standard techniques in calculus of variations,⁶ we are able to characterize the optimal monopolistic nonlinear pricing contract in this new setting. The analytical framework we develop is general enough to encompass the Mussa-Rosen benchmark as a special case. As in Mussa-Rosen, the monopolistic optimal quality provision (q^*) is characterized by two fundamental types: segments where $q^{*'} > 0$ (perfect sorting) and segments where $q^{*'} = 0$ (bunching). In the perfect sorting intervals, q^* is chosen so that marginal revenue equals marginal cost of increments in quality; when marginal revenue fails to be monotonically increasing over some interval, however, q^* must involve bunching, in which case the bunching interval and quality can be identified using standard ironing techniques (e.g., Myerson, 1981, and Maskin and Riley, 1984). A key difference in our analysis, however, is that in our model the magnitude of marginal revenue of quality provision is always higher than its counterpart in Mussa-Rosen due to an additional component from consumer entry. So in our model with entry, the monopoly has an incentive to increase quality provision (or to lower the price schedule). As a result, quality distortion and market exclusion are both smaller in our model with costly entry. Moreover, we show that whenever sorting occurs in Mussa-Rosen, it also occurs in our model; whenever bunching occurs in Mussa-Rosen, the bunching interval is smaller or simply disappears in our model. A rough intuition is that as high types enjoy higher informational rents with consumer entry, the incentive compatibility (or sorting) condition is relaxed; consequently, in our case with entry, perfect sorting is more likely and bunching is smaller.

Not only is quality distortion smaller compared to the Mussa-Rosen benchmark, quality distortion may even disappear completely in our model (*production or allocation efficiency*, which is also referred to as *the first best solution* throughout this paper). We identify exact conditions under which the first best occurs. The condition identified can be explained intuitively. In our model, the monopolistic nonlinear

⁵An implicit assumption is that buyers cannot or do not make purchases without incurring entry costs to learn their true preference types. This is the case when the entry cost is interpreted as the shopping cost, i.e., the cost of visiting the store, inspecting the product, and buying it. This will also be the case if there is a small probability that the product is terrible (i.e., gives the consumer $-\infty$ utility), in which case no one makes purchase without learning about her true preference type or the match value of the product.

⁶Given the more complicated structure of the objective function in our new setting, we can no longer use the more conventional point-wise maximization (*a la* Myerson, 1981) or optimal control techniques in our general analysis allowing for bunching.

pricing contract affects the consumers' equilibrium payoffs only through its effect on the entry cutoff. As such, the monopolist has an incentive to provide as efficient as possible quality provision. This point can be made more precise as follows. By the envelope theorem, a buyer's equilibrium payoff (gross of the entry cost) can be expressed as the sum of a common rent provision (i.e., the rent provision to the lowest type covered) and an integration over the quality provision schedule. Note that the common rent provision is nondistortionary, as it is applied to all the types covered. Our finding is that, by fixing the first-best quality provision, whenever optimal entry can be induced by adjusting the common rent provision, then the first-best quality provision is *feasible* and will indeed emerge in the monopoly solution. On the other hand, by fixing the first-best quality provision, if optimal entry cannot be induced by adjusting the common rent provision alone, then the first-best solution is not feasible and quality provision has to be distorted downward.

It turns out that the comparison with the Mussa-Rosen benchmark and the condition for the first-best solution to arise can both be unified in a more general ranking of monopoly solutions across different markets characterized by different inverse hazard rate functions of the entry cost ($\eta(c)$). We demonstrate that this inverse hazard rate reflects a measure of a cost/benefit ratio in raising rent provision to consumers, which is also inversely related to the price-elasticity of entry: the higher η , the smaller the price-elasticity of entry, which implies higher price or larger quality distortion in the monopoly solution. This result has an interesting implication for pricing dynamics in a monopoly market: when a product is newly launched (η is low), the price should start low to encourage consumer entry; when the product becomes more and more established, the price-elasticity of entry becomes increasingly smaller and the monopoly may increase the price gradually; in the limit, as the market base becomes stabilized (no new entry occurs), the Mussa-Rosen solution emerges, which is characterized by the highest pricing schedule (and maximum quality distortion).

We also explore the monopoly solution when the monopolist can charge entry fees, e.g., in the form of club membership fees. We show that even when the condition for the first-best solution mentioned above fails, the *infeasibility* issue can be overcome by the monopolist's ability to charge an entry fee before consumer entry occurs, as charging an entry fee effectively relaxes the *ex post* individual rationality (IR) constraint (imposed after entry occurs) to the *ex ante* IR constraint (imposed before entry occurs).

Lastly, we show that even when the optimal nonlinear pricing involves no quality distortion (production efficiency is achieved), the presence of a monopoly always induces insufficient entry in our model compared to the socially efficient benchmark. In other words, the monopoly in our model is mainly characterized by its distortion in entry, rather than by its distortion in production efficiency. This suggests a subtle implication for anti-trust practices in nonlinear pricing contexts with consumer entry.

Despite its importance, consumer entry in nonlinear pricing has received little attention from the current literature. To our knowledge, the only exception is Rochet and Stole (2002), who introduce a random participation component into the Mussa-Rosen framework. Unlike our model, in their setting consumers

know both their preference types and participation costs before entry occurs. Therefore entry in their model is purely a participation process, while in our model entry is both a participation and information acquisition process. This contrast in modeling leads to some different results. For example, while they also show that quality distortion is reduced (compared to the Mussa-Rosen benchmark), the first best can never be achieved, which is different from our case. Interestingly, our results are somewhat more in line with those obtained from the competitive nonlinear pricing literature.⁷ In particular, Rochet and Stole (2002) also extend their analysis of monopolistic nonlinear pricing with random participation to a duopoly case and show that, under full-market coverage, quality distortions disappear and the equilibrium is characterized by the cost-plus-fee pricing feature. A similar result is also obtained in Armstrong and Vickers (2001). When partial market coverage (along vertical dimension) is allowed, Yang and Ye (2008) show that quality distortion is reduced and market coverage is increased under a duopoly compared to a monopoly benchmark. Our results from this current research share many of these flavors, suggesting that entry has a similar effect as competition on nonlinear pricing schedules. This is perhaps not too surprising, given that in both models, a firm's market share is endogenously determined (either by entry or competition). As such, a firm has an incentive to reduce quality distortion, although the exact workings are quite different between models with competition and entry.

The role of information acquisition has been examined in several papers in the context of principal-agent settings (e.g, Crémer and Khalil, 1992; Crémer, Khalil, and Rochet, 1998a, 1998b).⁸ Crémer and Khalil (1992) incorporate a costly information acquisition stage to a standard adverse selection model similar to Baron and Myerson's (1982) setting of regulating a monopolist with unknown cost. They show that, although the firm does not acquire information in equilibrium, the ability to acquire information decreases the downward distortion at the production stage. Crémer, Khalil, and Rochet (1998a) modify this setting so that all information about the cost structure has to be acquired at some fixed cost. They show that when the cost is not too small, distortion is reduced for low cost types but increased for high cost types in the optimal contract. Crémer, Khalil, and Rochet (1998b) further modify the setting so that the firm's information acquisition decision is taken covertly before the contract is offered. This reversal in timing introduces strategic uncertainty for the principal as the firm may randomize over information acquisition. In all these papers, agents (firms) do not have to acquire information in order to accept a contract, which is different from our setting. Besides, in Crémer and Khalil (1992) and Crémer, Khalil, and Rochet (1998b), the agent can learn its true cost type at zero cost after signing the contract, so information acquisition is socially wasteful, which is another difference from our model.

Our paper is also closely related to a well-developed literature on auctions with costly entry.⁹ As

⁷See, for example, Spulber (1989), Champsaur and Rochet (1989), Wilson (1993), Gilbert and Matutes (1993), Stole (1995), Verboven (1999), Villas-Boas and Schmidt-Mohr (1999), Armstrong and Vickers (2001, 2010), Rochet and Stole (2002), Ellison (2005), and Yang and Ye (2008).

⁸Also see Bergemann and Välimäki (2006) for an excellent survey on information acquisition in the context of mechanism design

⁹See, for example, McAfee and McMillan (1987), Tan (1992), Engelbrecht-Wiggans (1993), and Levin and Smith (1994).

in our approach, this literature also models information acquisition as an entry decision where each bidder has to incur a cost in order to participate in an auction and learn her value of the object for sale. More recently, Lu (2010) and Moreno and Wooders (2011) extend the analysis to auctions with privately known (heterogeneous) entry costs, which is closest to our setting. While Lu’s analysis focuses on entry coordination, Moreno and Wooders focus on the optimal screening value (e.g., the optimal reserve price) depending on whether entry fees are feasible or not. They show that the distortionary reserve price is reduced with costly information acquisition, which is largely consistent with our finding that entry reduces quality distortion. Our paper differs from theirs in the following aspects. First, unlike in auctions, in our setting a consumer’s allocation and payment depend on her own reported type only, thus entry of an individual consumer does not impose an externality on the rest of the entrants. As such, the need for entry coordination does not arise in our analysis, while it is one central issue in their work. Second, there is a single indivisible item for sale in their auction setting, while in our setting, the supply of products is endogenously determined. In fact we work with a continuum of products (and buyers). Third, allocation efficiency is characterized in terms of quality distortion in our setting, while it is characterized by the distortion of reserve prices in their setting. As such, there may not exist exact correspondences between our results and theirs.¹⁰ Finally, we work with general distributions of buyer types, and hence unlike in their work, substantial analysis is devoted to bunching in our paper.

Finally, our model belongs to the general framework of dynamic mechanism design or sequential screening (e.g., Courty and Li, 2000; Eso and Szentes, 2007; and more recently, Bergemann and Wambach, 2013, and Pavan, Segal, and Toikka, 2014). However, note that in our setting there is no benefit for the monopolist to run an additional mechanism to screen consumers at the information acquisition or entry stage. This is due to the following reasons. First, in our setting the entry cost c is independent of the preference type θ and does not contribute to the buyers’ post-entry payoffs. So learning about c does not help in the nonlinear pricing mechanism; second, in our setting entry of an individual consumer does not impose an externality on the rest of the consumers who enter, as the allocation and transfer in a nonlinear pricing mechanism are only functions of one’s report on her own “types.” So there is no benefit either to run a prescreening mechanism at the entry stage to shortlist bidders.

The rest of the paper is organized as follows. Section 2 lays out the model, Section 3 characterizes our monopoly solutions and compares it with the solution in Mussa-Rosen. We also show that the monopoly solutions can be parameterized and ranked by the inverse hazard rate functions of the entry cost. Section 4 offers concluding remarks. All long proofs are relegated to Appendix A, and Appendix B provides an analysis on sufficient conditions for optimality.

¹⁰For example, we show that the optimal solution in our setting may involve the first-best quality provision even when entry fees are infeasible. This result does not have a natural correspondence in their analysis.

2 THE MODEL

We start with a review of the well known Mussa-Rosen model. In their setting, a monopolist offers to sell a commodity at various levels of quality and price, which can be represented by a nonlinear pricing schedule, $P(q)$. Given any quality q , the per unit production cost, $C(q)$, is constant (independent of the number of units produced). $C(q)$ is assumed to be (strictly) increasing and (strictly) convex in q : $C'(q) > 0$, $C''(q) > 0$ for all feasible qualities $q \geq 0$. There is a continuum of consumers with measure 1. Each consumer demands up to 1 unit of the product. The consumer's preference is completely determined by her "type" or the taste parameter, θ , with associated gross utility $\theta q - P(q)$, where θ is the marginal utility of quality, or the marginal rate of substitution of quality for money. The consumer's outside option is normalized to be zero. Ex ante, θ follows distribution $F(\cdot)$ with strictly positive density $F'(\cdot) = f(\cdot)$ over its support $[\underline{\theta}, \bar{\theta}]$. Under these assumptions, Mussa-Rosen show that the monopolistic nonlinear pricing solution exhibits three features: (1) quality distortion, i.e., compared with the competitive setting (the first-best solution), the monopolist reduces the quality sold to any consumer except the highest type; (2) market exclusion, i.e., the monopolist frequently prices consumers with the lowest types out of the market; (3) bunching, i.e., the monopolist may find it optimal to bunch consumers with different types onto the same (quality) product.

We are now ready to describe our model. We introduce costly entry to the monopolistic nonlinear pricing model in Mussa-Rosen described above. Formally, there are a continuum of consumers with measure 1. Consumers are heterogeneous in their entry costs, c_i 's, which are private information to consumers. Ex ante, c_i follows the distribution $G(\cdot)$ with strictly positive density function $G'(\cdot) = g(\cdot)$ on $[\underline{c}, \bar{c}]$. After entry, consumers draw θ 's from the distribution $F(\cdot)$ on $[\underline{\theta}, \bar{\theta}]$. We assume that θ_i and c_i are independent (so consumers are symmetric in terms of preference types even after they learn their c_i 's).¹¹ Define the inverse hazard rate functions as follows:

$$\begin{aligned}\xi(\theta) &= \frac{1 - F(\theta)}{f(\theta)}, \theta \in [\underline{\theta}, \bar{\theta}] \\ \eta(c) &= \frac{G(c)}{g(c)}, c \in [\underline{c}, \bar{c}]\end{aligned}$$

We will maintain the following regularity assumption regarding the distribution of entry cost c :

Assumption 1. $\eta(c)$ is strictly increasing over $c \in [\underline{c}, \bar{c}]$.

As demonstrated in Appendix B, Assumption 1 is needed to ensure that the first variation conditions in our analysis below are both necessary and sufficient for optimization.

The monopolist's objective is to maximize its expected profit from the sale. It is easily verified that,

¹¹An alternative interpretation of our model is that there is only one consumer, whose entry cost, c , follows the distribution $G(\cdot)$. With the size of consumer entry replaced by the probability of (single-consumer) entry, our analysis should remain unaltered under this alternative setting.

under complete information about θ , the first-best solution is given by $q^{fb}(\theta) = C'^{-1}(\theta)$ for $\theta \geq \theta^* = \max\{\underline{\theta}, C'(0)\}$. Note that $C'^{-1}(\theta)$ is strictly increasing given the strict convexity of $C(\theta)$.

Since the preference types of the consumers are uncertain to both parties at the outset, there are a number of ways in which we could formulate the contracting between the consumers and the firm. For example, the firm can contract with the consumers at the stage before learning the preference type. A special form of such contracting is considered in Section 3.3. Another possibility is that the firm cannot commit to a nonlinear pricing contract. In this case, the firm simply sets the Mussa-Rosen prices.¹² As will be demonstrated, the expected payoff to the consumers under Mussa-Rosen contracting may be too low to induce sufficient entry from the monopolist's point of view. In our main analysis, we will focus on the case where the firm commits to a nonlinear pricing scheme and the consumers engage privately in information acquisition. They report to the seller only after they learn their preference types. Individual rationality requires that each type of the buyer make a non-negative payoff from her purchases. This is in contrast to the first case above where contracting occurs at the *ex ante* stage. So formally, the timeline is as follows:

1. The monopolist offers the (nonlinear) pricing schedule, $P(q)$, or equivalently, the menu of quality-price contracts, $\{q(\theta), p(\theta)\}$;
2. The consumers make simultaneous and independent entry decisions. Once a consumer participates, she incurs a cost c_i and learns her preference type θ_i ;
3. Consumers who have learned their θ_i 's make purchase decisions, and the sale is realized.

3 THE ANALYSIS

The firm offers the nonlinear pricing schedule $p(q) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, which is equivalent to offering a menu of direct contracts of the form $\{q(\theta), p(\theta)\}$, where $\theta \in [\underline{\theta}, \bar{\theta}]$. Given the menu of contracts $\{q(\theta), p(\theta)\}$, the utility obtained by a consumer with type θ , when choosing the offer $\{q(\hat{\theta}), p(\hat{\theta})\}$, is given by

$$u(\hat{\theta}, \theta) = \theta q(\hat{\theta}) - p(\hat{\theta}).$$

Let $u(\theta) = u(\theta, \theta)$. Incentive compatibility (IC) implies that

$$u(\theta) = \max_{\hat{\theta}} \theta q(\hat{\theta}) - p(\hat{\theta})$$

By the envelope theorem, we have $u'(\theta) = q(\theta)$. For our setting, the following lemma is standard:

¹²Since we assume that entry costs and the preference types are independent, the distribution of consumers' preference types is not affected by the form of the contract.

Lemma 1. The IC condition is satisfied if and only if (i) $u'(\theta) = q(\theta)$ and (ii) $q(\theta)$ is increasing in θ , i.e., $u''(\theta) \geq 0$.

Condition (i) above is also equivalent to the following integral form of the envelope theorem:

$$u(\theta) = u(\theta^*) + \int_{\theta^*}^{\theta} q(\tau) d\tau \text{ for all } \theta \in (\theta^*, \bar{\theta}], \quad (1)$$

where $\theta^* \in [\underline{\theta}, \bar{\theta}]$ is the lowest type that is served (or covered) in the market.

By (1), the equilibrium rent provision for a type- θ consumer ($u(\theta)$) is completely determined by the rent for the lowest type covered ($u(\theta^*)$) and the quality provision schedule ($q(\cdot)$). Since $u(\theta^*)$ is provided to all consumers who are covered, we also refer to it as *the common rent provision*.

Note that $\{q(\theta), p(\theta)\}$ can be recovered from $u(\theta)$ as follows:

$$q(\theta) = u'(\theta) \text{ and } p(\theta) = \theta u'(\theta) - u(\theta)$$

Thus any menu of IC nonlinear pricing contracts can be characterized by the rent provision schedules $u(\cdot)$. For this reason it suffices for us to identify $u(\cdot)$ in characterizing the optimal monopolistic nonlinear pricing contract.

Given that our objective function contains a term of the demand, we will demonstrate that in our case with entry, the individual rationality constraint (IR) may not bind for type θ^* (i.e., it is possible that $u(\theta^*) > 0$). This marks the first departure from the standard screening model. Given the definition of θ^* , it cannot be the case that $u(\theta^*) > 0$ while $\theta^* \in (\underline{\theta}, \bar{\theta}]$; otherwise a type- θ^* consumer would not be indifferent from accepting a contract and staying out (from being served). Thus we have either (1) $u(\theta^*) = 0$ and $\theta^* \in (\underline{\theta}, \bar{\theta}]$ or (2) $u(\theta^*) \geq 0$ and $\theta^* = \underline{\theta}$.

Given the IC menu of contracts offered in the final sale, the expected utility, gross of entry cost, for a consumer who enters the sale is given by

$$Eu = \int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) = \int_{\theta^*}^{\bar{\theta}} [\theta q(\theta) - p(\theta)] dF(\theta),$$

where $\theta^* \in [\underline{\theta}, \bar{\theta}]$ is the lowest type who purchases the product. Using (1), we have

$$\begin{aligned} Eu &= \int_{\theta^*}^{\bar{\theta}} \left[u(\theta^*) + \int_{\theta^*}^{\theta} q(\tau) d\tau \right] dF(\theta) \\ &= [1 - F(\theta^*)]u(\theta^*) + \int_{\theta^*}^{\bar{\theta}} [1 - F(\theta)]q(\theta) d\theta. \end{aligned} \quad (2)$$

In equilibrium, a consumer with entry cost c_i enters the sale if and only if $c_i \leq c^* \equiv Eu$. In other words, given $\{q(\theta), p(\theta)\}$, a total measure of $G(c^*)$ consumers will enter the sale. Hence $G(c^*) = G(Eu)$

can be interpreted as the *actual* market base of the product.

Define the profit from serving a type- θ consumer as follows:

$$H(\theta, u(\theta), q(\theta)) \equiv p(\theta) - C(q(\theta)) = \theta q(\theta) - C(q(\theta)) - u(\theta) \quad (3)$$

The firm's problem can be formulated as follows:

$$\begin{aligned} \max_{\theta^*, u(\cdot)} \quad & G \left(\int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot \int_{\theta^*}^{\bar{\theta}} H(\theta, u(\theta), q(\theta)) dF(\theta) \\ \text{s.t.} \quad & u(\theta^*) \geq 0, \theta^* \in [\underline{\theta}, \bar{\theta}) \\ & q(\theta) = u'(\theta), q'(\theta) \geq 0 \end{aligned} \quad (4)$$

It is worth noting that the above program cannot be formulated as an optimal control problem. To see that, we rewrite the objective function as follows:

$$\max_{\theta^*, u(\cdot)} \int_{\theta^*}^{\bar{\theta}} G \left(\int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot H(\theta, u(\theta), q(\theta)) dF(\theta)$$

It is clear that the integrand above is a function of not just $u(\theta)$, but also the entire path $u(\cdot)$, exactly due to the additional demand term G . As such, $u(\theta)$ cannot be treated as a state variable in any optimal control problem. In fact, there does not exist an appropriate state variable for us to formulate our program (4) as an optimal control problem. For this reason we turn to the method of calculus variation in our analysis.

3.1 Characterization of The Monopoly Solution

The firm's maximization problem in the Mussa-Rosen benchmark can be regarded as a special case in which the actual market base $G \left(\int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) \right) = 1$, i.e., when all potential consumers enter the sale. Let $u^*(\theta)$ and $q^*(\theta)$, where $\theta^* \leq \theta \leq \bar{\theta}$, be optimal and consider comparison functions $u(\theta)$ and $q(\theta)$, where $\theta^* - \delta\theta^* \leq \theta \leq \bar{\theta}$. The domains of the two functions may differ slightly with $\delta\theta^*$ small in absolute value but of any sign. Since the domains may not be identical, either u^* or u can be extended to have a common domain.¹³

Let $h(\theta) = u(\theta) - u^*(\theta)$ and $k(\theta) = q(\theta) - q^*(\theta)$ be admissible and arbitrary deviation functions, and let $\delta\theta^*$ be a small fixed number. The objective function (the Lagrangian) evaluated at $\{u^*(\theta), q^*(\theta) : \theta \in [\theta^*, \bar{\theta}]\}$ is as follows:

$$J^* = G \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) \right) \cdot \int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta) + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [q^*(\theta) - u^{*'}(\theta)] + \mu(\theta) q^{*'}(\theta) \} dF(\theta) \quad (5)$$

¹³This can be done, e.g., via linear extrapolations as introduced in Kamien and Schwartz (1991, page 57).

The objective function (the Lagrangian) evaluated at $\{u(\theta), q(\theta): \theta \in [\theta^* - \delta\theta^*, \bar{\theta}]\}$ is as follows:

$$J = G \left(\int_{\theta^* - \delta\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot \int_{\theta^* - \delta\theta^*}^{\bar{\theta}} H(\theta, u(\theta), q(\theta)) dF(\theta) + \int_{\theta^* - \delta\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [q(\theta) - u'(\theta)] + \mu(\theta) q'(\theta) \} dF(\theta) \quad (6)$$

Given (5) and (6), using standard techniques in calculus of variation, we can derive the following inequality implied by the optimality of θ^* , $u^*(\theta)$, and $q^*(\theta)$ (i.e., the first variation):¹⁴

$$\begin{aligned} J - J^* &\approx \int_{\theta^*}^{\bar{\theta}} G \cdot \left\{ -(1-b) + \frac{d(\lambda(\theta)f(\theta))}{G \cdot f(\theta)d\theta} \right\} h(\theta) dF(\theta) \\ &\quad + \int_{\theta^*}^{\bar{\theta}} \left\{ G \cdot [\theta - C'(q^*(\theta))] + \lambda(\theta) - \frac{d(\mu(\theta)f(\theta))}{f(\theta)d\theta} \right\} k(\theta) dF(\theta) \\ &\quad - \lambda(\bar{\theta}) f(\bar{\theta}) h(\bar{\theta}) + \mu(\bar{\theta}) f(\bar{\theta}) k(\bar{\theta}) \\ &\quad + [bGu^*(\theta^*) + GH(\theta^*, u^*(\theta^*), q^*(\theta^*)) + \lambda(\theta^*) q^*(\theta^*)] f(\theta^*) \cdot \delta\theta^* \\ &\quad - \mu(\theta^*) f(\theta^*) \cdot \delta q(\theta^*) + \lambda(\theta^*) f(\theta^*) \cdot \delta u(\theta^*) \\ &\leq 0, \end{aligned} \quad (7)$$

where

$$b = \frac{G'}{G} \cdot \int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta) \text{ and } G = G \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) f(\theta) d\theta \right) \quad (8)$$

Since $q^*(\theta) = u^{*\prime}(\theta)$, b defined above is a function of the entire path $u^*(\cdot)$ (and hence a functional). Once the optimal path $u^*(\cdot)$ is given, the value of b is uniquely determined.

Note that in the optimal solution, $\int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta)$ cannot be strictly negative, as the monopolist can guarantee zero profit by not offering any contract. Thus $b \geq 0$. In Mussa-Rosen, $G'/G = 0$ so $b = 0$. The economic interpretation of b will become clear below.

Since $h(\theta)$ and $k(\theta)$ are arbitrary admissible deviations and $f(\bar{\theta}) > 0$, we must have the transversality conditions $\lambda(\bar{\theta}) = 0$ and $\mu(\bar{\theta}) = 0$. Moreover, except at points of discontinuity of $u^*(\theta)$ and $q^*(\theta)$, we must have

$$-(1-b) + \frac{d[\lambda(\theta)f(\theta)]}{G \cdot f(\theta)d\theta} = 0 \quad (9)$$

$$G \cdot [\theta - C'(q^*(\theta))] + \lambda(\theta) - \frac{d[\mu(\theta)f(\theta)]}{f(\theta)d\theta} = 0 \quad (10)$$

(9) combined with $\lambda(\bar{\theta}) = 0$ implies that

$$\lambda(\theta) = -(1-b)\xi(\theta)G \quad (11)$$

¹⁴The detailed derivation is provided in the proof of Proposition 1 in the appendix.

Plugging (11) into (10), and using the transversality condition $\mu(\bar{\theta}) = 0$, we can obtain

$$\mu(\theta) = -\frac{G}{f(\theta)} \int_{\theta}^{\bar{\theta}} \{[\theta - C'(q^*(\theta))] - (1-b)\xi(\theta)\} dF(\theta) \quad (12)$$

By the Kuhn-Tucker theorem (pp. 249, Luenberger, 1969), $\mu(\theta) \geq 0$; whenever $\mu(\theta) > 0$, we must have $q^{*'}(\theta) = 0$. In other words, the following complementary slackness condition holds:

$$q^{*'}(\theta) \cdot \left\{ - \int_{\theta}^{\bar{\theta}} \{[\theta - C'(q^*(\theta))] - (1-b)\xi(\theta)\} f(\theta) d\theta \right\} = 0 \text{ for all } \theta \in [\theta^*, \bar{\theta}]$$

Now define a hypothetical quality provision schedule

$$q^s(\theta) \equiv C'^{-1}(\theta - (1-b)\xi(\theta)), \quad (13)$$

where b is given by (8).

Over an interval where $q^{*'}(\theta) > 0$, we have $\mu(\theta) = 0$. This in turn implies, from (12), that

$$q^*(\theta) = C'^{-1}(\theta - (1-b)\xi(\theta)) = q^s(\theta) \text{ when } q^{*'}(\theta) > 0. \quad (14)$$

Next we demonstrate that $\theta - (1-b)\xi(\theta)$ is the marginal revenue from raising the quality provision to type- θ consumers. Consider selling an additional increment of quality to the existing entrants with type θ (with measure $G \cdot f(\theta)$):

1. For the existing entrants with type θ (with measure $G \cdot f(\theta)$), each has incremental value θ . So total revenue increases by $Gf(\theta) \cdot \theta$;
2. For the existing entrants with types above θ (with measure $G \cdot (1 - F(\theta))$), the price that can be charged falls by the increment sold to type θ . So the additional rent provided is given by $G \cdot (1 - F(\theta))$;
3. Before entry, the (*ex ante*) expected rent to consumers is increased by $(1 - F(\theta))$. So the measure of new entrants will increase by $G' \cdot (1 - F(\theta))$. Thus the increased revenue from new entrants is given by

$$G' \cdot (1 - F(\theta)) \cdot \int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta).$$

Taking the above all together, the marginal revenue from an additional increment of quality to type θ is given by:

$$MR(\theta) \equiv \theta - \left[1 - \frac{G'}{G} \cdot \int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta) \right] \frac{1 - F(\theta)}{f(\theta)} = \theta - (1-b)\xi(\theta).$$

Note that it is the effect on the new entrants that makes the expression of marginal revenue in our model differ from that in the Mussa-Rosen benchmark. In Mussa-Rosen, raising quality provision can only affect the existing customers, so $MR(\theta) = \theta - \xi(\theta)$. With consumer entry, there is an additional component in marginal revenue, which is equal to $b\xi(\theta) (\geq 0)$. Thus b can be interpreted as a measure of this additional marginal revenue due to consumer entry.

The above analysis thus suggests that whenever $q^*(\theta)$ is strictly increasing (perfect sorting) over some interval, it must be chosen so that marginal revenue is equated to marginal cost of quality provision (i.e., $MR(\theta) = C'^{-1}(q^*(\theta))$, which implies that $q^*(\theta) = q^s(\theta)$). However, $MR(\theta)$ may not be a monotonically increasing function of θ . Thus the monopoly solution may not be necessarily characterized by equating marginal revenue with marginal cost, in which case the solution involves bunching ($q'^*(\theta) = 0$). We are now ready to state the following characterization proposition, leaving the rest of the proof to the appendix.

Proposition 1. The monopoly solution q^* has the following properties:

1. Bunching does not occur in the neighborhood of $\bar{\theta}$, and $q^*(\bar{\theta}) = C'^{-1}(\bar{\theta})$ (efficiency at the top);
2. Whenever q^* is perfect sorting, it is determined by (14);
3. Whenever bunching occurs over interval $[\theta_1, \theta_2] \subset [\theta^*, \bar{\theta})$, it is determined by (15)-(16) for interior bunching and (16) with θ_1 being replaced by $\underline{\theta}$ for bottom bunching:

$$\theta_1 - (1-b)\xi(\theta_1) = \theta_2 - (1-b)\xi(\theta_2) \quad (15)$$

$$\int_{\theta_1}^{\theta_2} [\theta - C'(q^*(\theta)) - (1-b)\xi(\theta)] dF(\theta) = 0 \quad (16)$$

The complete proof is quite tedious, which is relegated to the appendix. We should point out that the variation method is used in the proof mainly due to the need for bunching analysis. If the solution only involves perfect sorting (i.e., when the monotonicity constraint is not binding), the optimal solution can be derived rather straightforwardly. To see this, when $q'(\theta) > 0$, the monotonicity constraint is dropped from the Lagrangian. Substituting $u'(\theta) = q(\theta)$ into the objective function and using (1), we can verify that the firm's expected profit is given by

$$\begin{aligned} \Pi &= G \left(\int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot \int_{\theta^*}^{\bar{\theta}} H(\theta, u(\theta), q(\theta)) dF(\theta) \\ &= G \left(\int_{\theta^*}^{\bar{\theta}} \left(u(\theta^*) + q(\theta) \frac{1-F(\theta)}{f(\theta)} \right) dF(\theta) \right) \cdot \int_{\theta^*}^{\bar{\theta}} \left[\theta q(\theta) - C(q(\theta)) - q(\theta) \frac{1-F(\theta)}{f(\theta)} - u(\theta^*) \right] dF(\theta) \quad (17) \end{aligned}$$

Differentiating (17) with respect to $q(\theta)$ and simplifying, we obtain

$$\theta - (1-b)\xi(\theta) = C'(q(\theta)),$$

which is exactly the condition (14).

To sum up, our solution (q^*) is characterized by two fundamental types: segments where $q^{*'} > 0$ (perfect sorting) and segments where $q^{*'} = 0$ (bunching). In the perfect sorting intervals, q^* is chosen so that marginal revenue equals marginal cost of increments in quality; when marginal revenue fails to be monotonically increasing over some interval, however, q^* must involve bunching. Intuitively, since $MR'(\theta) = 1 - (1 - b)\xi'(\theta)$, $MR(\theta)$ will be a decreasing function of θ over any interval where $\xi'(\theta) > 1/(1 - b)$. In such an interval the monopolist cannot equate marginal revenue and marginal cost, nor can he exclude the consumers in such an interval, unless it is profitable for him to exclude all the consumers with types lower than this interval. The procedure in identifying the bunching intervals (and bunching qualities), as introduced in the proof, is known as the “ironing” technique (Myerson, 1981, and Masking and Riley, 1984), which is illustrated in Figure 1 below.

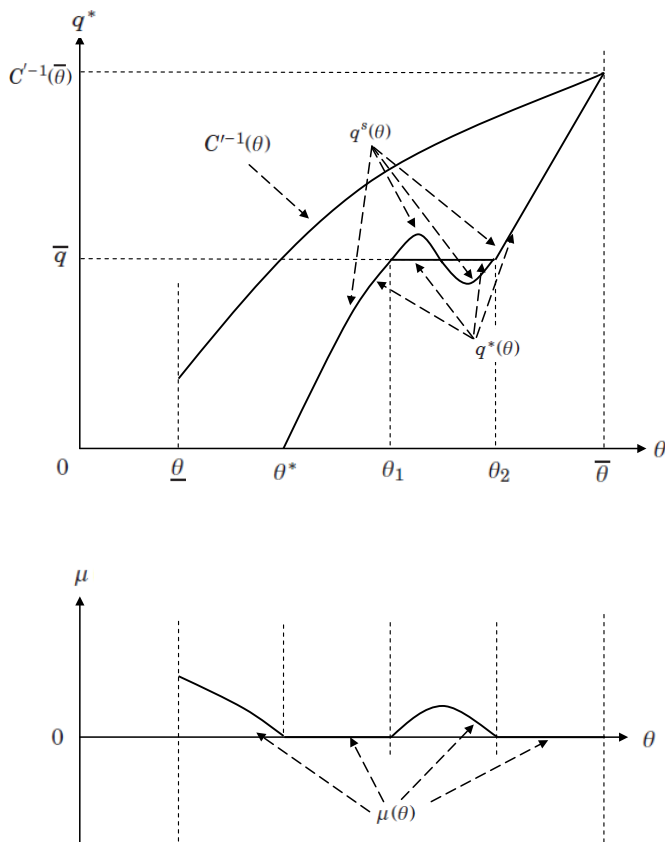


Figure 1: Monopoly Solution with Interior Bunching

It is worth noting that the specific construction of the optimal path of quality provision in our model can be much more involved than in Mussa-Rosen. In Mussa-Rosen, the optimal path q_{MR}^* can be derived straightforwardly, as it can be constructed backwards starting from the top ($\theta = \bar{\theta}$) based on the conditions characterizing sorting and bunching segments. This is no longer true in our model, as the sorting and bunching conditions both involve b , which is a function of the entire path q^* . So, computationally, the construction in our case is a process to identify a pair of fixed points (q^*, b) : given an initial value of $b_1 \in (0, 1]$, we can construct a candidate path q_1^* (backward, starting from $\bar{\theta}$) using the sorting and bunching conditions, and, using the derived q_1^* , we compute the induced value of b from (8). When this induced value (denoted as b_2) coincides with b_1 , we find the optimal solution $q^* = q_1^*$; otherwise we repeat the process by setting $b = b_2$. This process continues until we find a pair of fixed points (q^*, b^*) such that q^* is derived from b^* and b^* is justified by q^* .

The proof of Proposition 1 also implies the following corollary:

Corollary 1. q^* is either the first best ($b = 1$), or involves downward quality distortion ($0 \leq b < 1$).

Proof. When $b = 1$, it is clear that bunching does not occur. So $q^*(\theta) = C'^{-1}(\theta)$ for $\theta \geq \theta^* = \max\{\underline{\theta}, C'(0)\}$, which is the first-best quality provision. When $b \in [0, 1)$, we have $q^*(\theta) = C'^{-1}(\theta - (1-b)\xi(\theta)) \leq C'^{-1}(\theta)$ (with equality at $\theta = \bar{\theta}$ only) in intervals where q^* is perfect sorting; when q^* involves bunching, say, over $[\theta_1, \theta_2]$, the bunching quality $\bar{q} = q^s(\theta_1) < C'^{-1}(\theta_1) \leq \theta$ for all $\theta \in [\theta_1, \theta_2]$. So for all $\theta \in [\theta^*, \bar{\theta}]$, we have $q(\theta) \leq C'^{-1}(\theta)$ (with equality at $\theta = \bar{\theta}$ only). \square

3.2 Comparison of Monopoly Solutions

Let q^* and θ^* denote the quality provision schedule and lowest type served, respectively, in the monopoly solution in our model with consumer entry. Similarly, let q_{MR}^* and θ_{MR}^* denote the quality provision schedule and lowest type served, respectively, in the monopoly solution in the Mussa-Rosen benchmark. We can first establish the following lemma regarding the bunching intervals:

Lemma 2. Suppose bunching occurs over $[\theta_1, \theta_2] \subseteq [\underline{\theta}, \bar{\theta})$ in our model, then either $\theta_{MR}^* > \theta_2$, or bunching occurs over $[\theta_1^{MR}, \theta_2^{MR}] \subseteq [\underline{\theta}, \bar{\theta})$ in Mussa-Rosen, where $\theta_1^{MR} \leq \theta_1 < \theta_2 < \theta_2^{MR}$.

Proof. See Appendix A. \square

If market exclusion is regarded as a special form of bunching (at $\bar{q} = 0$), then Lemma 2 simply says that any bunching interval in our model is contained in a bunching interval in Mussa-Rosen.

We are now ready to compare the monopoly solutions in our model and that in Mussa-Rosen:

Proposition 2. Compared to the Mussa-Rosen benchmark, both quality distortion and market exclusion are smaller with consumer entry, i.e., $q^*(\theta) \geq q_{MR}^*(\theta)$ (with equality only at $\bar{\theta}$) and $\theta^* \leq \theta_{MR}^*$ (with equality only when $\theta^* = \theta_{MR}^* = \underline{\theta}$).

Proof. See Appendix A. □

Intuitively, taking costly information acquisition into account, the monopolist has to balance entry (the actual market base) and profit conditional on consumer entry. By reducing quality distortion and increasing market coverage (conditional on entry), the monopolist makes the product more attractive and induces an optimal set of entrants to maximize expected profit. This can be seen more clearly in the decomposition of $MR(\theta)$ in our setting. With consumer entry, there is an additional benefit from raising the quality provision to type- θ consumers, which is given by $b\xi(\theta)$ (> 0). So compared to the Mussa-Rosen benchmark, the incentive to raise quality provision must be higher in our model with consumer entry.

Proposition 2 implies that whenever a type is covered in Mussa-Rosen, she is also covered in our model. Given this, the comparison stated in Lemma 2 can be made stronger as follows:

Proposition 3. Over a given interval, whenever perfect sorting occurs in Mussa-Rosen, perfect sorting must also occur in our model; whenever bunching occurs in Mussa-Rosen, the bunching range must be smaller or absent in our model.

Proposition 3 reflects the impact of entry on the IC condition (1): as higher informational rents are provided to consumers due to costly entry, the sorting condition (1) is relaxed; as a result, perfect sorting is more likely and bunching is smaller with consumer entry.

Corollary 1 suggests that in our setting the optimal monopolistic solution may even be first best. In order to identify conditions under which the first best arises as the monopolistic optimal solution, we start with the expression of *virtual surplus* (Myerson, 1981) from the sale, which is given by

$$v(q, \theta) = \theta q(\theta) - C(q(\theta)) - \xi(\theta) q(\theta).$$

$v(q, \theta)$ is the total profit contributed from each type- θ buyer.

When the first-best quality provision is offered (coupled with $u(\theta^*) = 0$), we let $Ev(q^{fb})$ and c^{fb} denote the expected virtual surplus and the induced entry cutoff, respectively.

Proposition 4. The monopolistic nonlinear pricing achieves the first best if and only if the following condition holds:

$$Ev(q^{fb}) \geq \eta(c^{fb}). \tag{18}$$

If condition (18) fails, the monopolistic nonlinear pricing contract involves downward quality distortion for all but the highest type.

Proof. See Appendix A. □

While the proof in the appendix provides the formal arguments, Proposition 4 can be understood intuitively. In our setting, the rent provision schedule $u(\cdot)$ determines the entry cutoff type c^* (or the

actual market base, $G(c^*)$). Note that once c^* is determined, the expected payoff to each buyer is also determined. The reason is as follows. The cutoff type c^* makes zero expected payoff in equilibrium. Before entry, an entrant with a type $c_i < c^*$ differs from the cutoff type only in her entry cost. Thus, the equilibrium expected payoff to a buyer with type c_i is given by $c^* - c_i$. So in our setting, the monopolist extracts the rents from the buyers via the induced entry cutoff c^* , regardless of the specific nonlinear pricing contract offered in the sale – in this sense consumer rents are extracted by endogenous entry in our model.

It is thus clear that, given any targeted c^* , the monopolist has an incentive to make the quality provision as efficient as possible (since the expected rent to the buyers is fixed given c^*). An implication is, whenever feasible, the monopolist would offer the first-best quality provision, and only if the first-best allocation is achieved, should the monopolist give the consumers “cash” in terms of $u(\theta^*)$.

To check such “feasibility”, substituting $q(\theta) = q^{fb}(\theta)$ into (17), we have

$$\Pi = G \left(\int_{\theta^*}^{\bar{\theta}} \left(u(\theta^*) + q^{fb}(\theta) \frac{1-F(\theta)}{f(\theta)} \right) dF(\theta) \right) \cdot \int_{\theta^*}^{\bar{\theta}} \left[\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - q^{fb}(\theta) \frac{1-F(\theta)}{f(\theta)} - u(\theta^*) \right] dF(\theta).$$

Differentiating by $u(\theta^*)$ and then evaluating at $u(\theta^*) = 0$ yields

$$\begin{aligned} \frac{d\Pi}{du(\theta^*)} \Big|_{u(\theta^*)=0} &= G' \cdot \int_{\theta^*}^{\bar{\theta}} \left[\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - q^{fb}(\theta) \frac{1-F(\theta)}{f(\theta)} \right] dF(\theta) - G \\ &= G' \cdot \left[Ev(q^{fb}) - \eta(c^{fb}) \right]. \end{aligned} \quad (19)$$

The marginal profit expression derived above is fairly intuitive. By raising the expected (common) rent provision by $du(\theta^*)$, an additional measure $G'(c^{fb})du(\theta^*)$ of consumers will enter the sale, which brings additional expected gain of $G'(c^{fb}) \cdot \int_{\theta^*}^{\bar{\theta}} [\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - u^{fb}(\theta)] dF(\theta) \cdot du(\theta^*)$; on the other hand, the additional rent provision $du(\theta^*)$ applies to all the consumers who enter, leading to a total additional “cost” of $G(c^*)du(\theta^*)$. Therefore marginal expected profit is given by (19). When (18) fails, $d\Pi/du(\theta^*)|_{u(\theta^*)=0} < 0$. This suggests that the first-best solution is not feasible, as the “optimal” common rent provision $u(\theta^*)$ would have to be strictly negative, which violates the individual rationality constraint (IR) for consumers in the neighborhood of type θ^* after entry. Conversely, when (18) holds, the first-best solution is achieved ($q(\cdot) = q^{fb}(\cdot)$), and the optimal $u(\theta^*)$ (and hence c^*) is chosen such that

$$\int_{\theta^*}^{\bar{\theta}} \left[\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - u^{fb}(\theta) \right] dF(\theta) = \eta(c^*),$$

where $c^* = u(\theta^*) + \int_{\theta^*}^{\bar{\theta}} \int_{\theta^*}^{\theta} q^{fb}(\tau) d\tau dF(\theta)$. This can be regarded as the “interior” solution for $u(\theta^*)$ but “corner” solution for $q^*(\cdot)$, in which the monopoly solution achieves the first best ($b = 1$).

To further understand condition (18), we provide the following corollaries/examples.

Corollary 2. The monopolist never achieves the first best if $\underline{\theta} = 0$.

Proof. See Appendix A. □

As demonstrated in the proof, $Ev(q^{fb}) = 0$ when $\underline{\theta} = 0$, so the first best (18) can only be achieved when $\underline{\theta} > 0$.

We further assume that θ is distributed uniformly and that the firm's production cost is given by the quadratic form: $C(q) = q^2/2$. Let $\Delta = (\bar{\theta} - \underline{\theta})$ be the range of the support, then η^{-1} is well-defined given Assumption 1.

Corollary 3. When $C(q) = q^2/2$ and θ is distributed uniformly over $[\underline{\theta}, \bar{\theta}]$, the monopoly solution achieves the first best if $0 < \Delta \leq \Delta^*$, and involves downward quality distortion if $\Delta > \Delta^*$, where $\Delta^* = \left(\sqrt{(3\underline{\theta})^2 + 24\eta^{-1}(\underline{\theta}^2/2)} - 3\underline{\theta} \right) / 2$.

Proof. See Appendix A. □

Corollary 3 can be interpreted rather intuitively. If the range of support Δ is not too big, sorting via the first-best quality provision is optimal. However, if the range of support Δ is sufficiently large, sorting via the first-best quality provision is too costly for the monopolist: recall that by (1), the higher Δ , the larger required rent for the consumers.

Suppose the entry cost c is also distributed uniformly over, say, $[0, \bar{c}]$. Define the relative measure of consumers' vertical type heterogeneity $\gamma = \bar{\theta}/\underline{\theta}$, and let $\gamma^* = (\sqrt{21} - 1)/2$. It can be verified that:

- the solution is the first best with full-market coverage (in the consumers' vertical type dimension) if $\gamma \in (1, \gamma^*]$;
- the solution involves downward quality distortion and full coverage if $\gamma \in (\gamma^*, 4]$;
- the solution involves downward quality distortion and partial coverage if $\gamma > 4$.

So the smaller the relative measure of consumers' vertical type heterogeneity, the more likely that the first-best quality provision will be offered or the more likely that the market will be fully covered.

It turns out that Propositions 2 and 4 can be unified in a more general ranking of the monopoly solutions. Given any two (monopolistic) markets characterized by different inverse hazard rate functions $\eta_i(c) = G_i(c)/G'_i(c)$, $i = 1, 2$, we can establish the following ranking of the monopoly solutions:

Proposition 5. If $\eta_1 \leq \eta_2$, then $q^{fb} \geq q_1^* \geq q_2^* \geq q_{MR}^*$ and $\theta_{MR}^* \geq \theta_2^* \geq \theta_1^* \geq \theta^{fb}$.

Proof. See Appendix A. □

As the discussion following Proposition 4 indicates, $\eta = G/G'$ reflects the relative cost/benefit ratio when raising the expected rent provision. When η is lower, the relative benefit to raise the rent provision is higher, or b is higher, so the incentive for the monopolist to offer a higher q^* is also higher.

Note that condition (18) provides an upper bound of η for the first-best solution to emerge. On the other extreme, in Mussa-Rosen, $G = 1$ hence $\eta = +\infty$, and condition (18) never holds. This explains why the first best is never optimal in the Mussa-Rosen benchmark. For the case in between, the higher η , the bigger the quality provision distortion from the efficient provision level. In a sense, η also measures the price-elasticity of entry: the higher η , the lower the price-elasticity of entry, and hence the monopolist may charge a higher price (or provide a lower q^*).

Proposition 5 thus has an implication for the monopolistic pricing dynamic: when a product is relatively new, the price-elasticity of entry is large, and the monopoly should charge a lower price (or provide a higher quality with less distortion); when the product becomes well established, the price-elasticity of entry is low, and the monopoly should charge a higher price (or provide a lower quality with more distortion). This is a potentially testable implication.

3.3 The Monopoly Solution with Entry Fees

An implicit assumption in our previous analysis is that the monopoly cannot charge entry fees (before entry occurs), which is a commonly observed business practice. When given the ability to charge entry fees (i.e., club membership fees), it turns out that even if condition (18) fails, the monopolist can still achieve the first-best solution .

Proposition 6. When the monopolist can charge entry fees, the optimal solution is always the first best.

Proof. See Appendix A. □

As the discussion following Proposition 4 reveals, when condition (18) fails, optimal entry under the first-best quality provision cannot be induced by any positive $u(\theta^*)$. This is the sense in which the first-best solution is not *feasible*: if $u(\theta^*)$ could be made negative, the efficient quality schedule would still be provided. However, a negative $u(\theta^*)$ would be a direct violation of the post-entry IR constraints for consumers in the neighborhood of type θ^* . This problem can be overcome with entry fee payments, which are collected before entry occurs. In a sense, by charging entry fees, the *ex post* IR constraint (imposed after buyers learn their θ_i 's) is replaced by the *ex ante* IR constraint (imposed before buyers learn their θ_i 's). It is clear that the requirement for *ex ante* IR is less stringent than that for the *ex post* IR, making the first-best solution easier to emerge. In fact, as Proposition 6 shows, the first best is always achieved when the monopolist has the option to charge entry fees.

So when condition (18) fails, charging entry fees (e.g., club membership fees) proves to be a superior

instrument for the monopolist to maximize profit, as it does not distort production efficiency and is less constrained compared to the option of charging a higher price. With the ability to charge entry fees, the monopolist simply maximizes the production efficiency in the final sale by completely removing quality distortion, leaving the market base to be entirely adjusted by entry fees. It also appears that the monopolistic profit maximization can be done sequentially: $q^{fb}(\cdot)$ is chosen to maximize expected surplus in the final sale, followed by an optimal entry fee s^* to induce an “optimal” set of entrants. Note that if (18) holds, the first-best quality provision (along with an appropriate positive $u(\theta^*)$) is optimal even when charging entry fees is not feasible.

3.4 Socially Optimum Entry

As demonstrated above, even in the presence of a monopoly, the first-best solution can be achieved, either due to condition (18) or the ability to charge entry fees. So production inefficiency does not have to be associated with the existence of a monopoly, which may appear inconsistent with the basic wisdom from a microeconomics textbook. However, as the following proposition illustrates, even when production inefficiency is absent in our setting, distortion will arise in the form of inefficient entry.

Proposition 7. The monopoly induces insufficient entry compared to socially optimal entry.

Proof. See Appendix A. □

In the socially efficient benchmark, the social planner maximizes the expected total social surplus, which is equal to the expected total surplus generated from the sale less the expected total entry cost. We show that the socially optimal solution is characterized by first-best quality provision coupled with full entry subsidization: all the rents are returned to the consumers in the form of entry subsidies. While formal arguments are provided in the proof, this result can be explained intuitively. In equilibrium, a potential buyer will enter the sale if and only if her expected profit from entry is larger than her entry cost. Such entry is socially efficient as (1) there is no production inefficiency due to the first-best quality provision, and (2) a potential buyer enters if and only if the expected profit, and hence her contribution to the social surplus, is greater than zero.

In general, monopolistic inefficiency takes the form of both production distortion and entry distortion in our setting. Even when production distortion is absent, entry distortion persists. Our model thus suggests a subtle implication for anti-trust experts in nonlinear pricing settings with consumer entry.

4 CONCLUDING REMARKS

Our paper contributes to the literature by incorporating an analysis of nonlinear pricing with costly consumer entry wherein the cost mainly stems from information acquisition. Compared to the Mussa-Rosen benchmark, the optimal solution in our model involves less quality distortion, more market coverage, and

less bunching. We also show that under certain conditions the first-best quality provision can emerge as the monopoly solution in our model, which never occurs in Mussa-Rosen. More generally, we can establish similar rankings of monopoly solutions across different markets characterized by different inverse hazard rate functions of the entry cost, which can be interpreted as a measure of the price-elasticity of entry. Our result suggests an interesting dynamic for monopolistic pricing, which is potentially testable. With the large number of new products introduced every year, consumer entry is becoming increasingly critical for the market viability of new products. Our result may shed some new light on how a firm will adjust its nonlinear pricing schedule in response to consumer entry.

Our analysis relies on some simplifying assumptions. First of all, we assume that the two-dimensional private information, c and θ , are independent. We adopt this assumption mainly because there does not seem to be a consensus in the literature over whether they should be positively or negatively correlated. Nevertheless, a more general analysis allowing for correlation between c and θ may potentially produce more insights. Second of all, our current analysis is restricted to a monopoly regime. A more general analysis should extend costly information acquisition to a competitive setting with more than one firm. Note that the original Mussa-Rosen framework is inappropriate for such an extension, as the post-entry competition between firms would result in an unrealistic Bertrand outcome. Therefore it is not trivial to incorporate information acquisition in a competitive nonlinear pricing framework. Finally, we assume that buyers do not make purchases without incurring entry costs to learn their true preference types. This is reasonable in some situations (such as the examples mentioned in the introduction), but may be less reasonable in other situations. If the buyers can make purchase decisions based on prior beliefs of their preference types (e.g., behave like the mean type), we will need to worry that the informed and uninformed consumers may mimic each other, which makes our analysis more involved. Besides, if all consumers, either informed or uninformed, will enter the final sale, then there will be no effect of nonlinear pricing on entry, which eliminates a major motivation of this current research. Given the challenges in extending our current analysis along all these directions, they are left for future research.

APPENDIX A: PROOFS

Proof of Proposition 1: We first verify (7). Using Taylor expansion, we have

$$\begin{aligned}
J &= G \left(\int_{\theta^* - \delta\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot \int_{\theta^* - \delta\theta^*}^{\bar{\theta}} H(\theta, u(\theta), q(\theta)) dF(\theta) + \int_{\theta^* - \delta\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [q(\theta) - u'(\theta)] + \mu(\theta) q'(\theta) \} dF(\theta) \\
&= G \left(\int_{\theta^* - \delta\theta^*}^{\theta^*} u(\theta) dF(\theta) + \int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot \left\{ \int_{\theta^* - \delta\theta^*}^{\theta^*} H(\theta, u(\theta), q(\theta)) dF(\theta) + \int_{\theta^*}^{\bar{\theta}} H(\theta, u(\theta), q(\theta)) dF(\theta) \right\} \\
&\quad + \int_{\theta^* - \delta\theta^*}^{\theta^*} \{ \lambda(\theta) [q(\theta) - u'(\theta)] + \mu(\theta) q'(\theta) \} dF(\theta) + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [q(\theta) - u'(\theta)] + \mu(\theta) q'(\theta) \} dF(\theta) \\
&\approx G \left(u(\theta^*) f(\theta^*) \delta\theta^* + \int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) + \int_{\theta^*}^{\bar{\theta}} h(\theta) dF(\theta) \right) \\
&\quad \cdot \left\{ \begin{aligned} &H(\theta^*, u^*(\theta^*), q^*(\theta^*)) f(\theta^*) \delta\theta^* \\ &+ \int_{\theta^*}^{\bar{\theta}} [H(\theta, u^*(\theta), q^*(\theta)) + H_u(\theta, u^*(\theta), q^*(\theta)) h(\theta) + H_q(\theta, u^*(\theta), q^*(\theta)) k(\theta)] dF(\theta) \end{aligned} \right\} \\
&\quad + \int_{\theta^* - \delta\theta^*}^{\theta^*} \{ \lambda(\theta) [q(\theta) - u'(\theta)] + \mu(\theta) q'(\theta) \} dF(\theta) + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [q^*(\theta) - u^{*'}(\theta)] + \mu(\theta) q^{*'}(\theta) \} dF(\theta) \\
&\quad + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [k(\theta) - h'(\theta)] + \mu(\theta) k'(\theta) \} dF(\theta) \\
&\approx \left[G \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) \right) + G' \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) \right) \cdot \left(\int_{\theta^*}^{\bar{\theta}} h(\theta) dF(\theta) + u(\theta^*) f(\theta^*) \delta\theta^* \right) \right] \\
&\quad \cdot \left[\begin{aligned} &\int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta) + H(\theta^*, u^*(\theta^*), q^*(\theta^*)) f(\theta^*) \delta\theta^* \\ &+ \int_{\theta^*}^{\bar{\theta}} \{ H_u(\theta, u^*(\theta), q^*(\theta)) h(\theta) + H_q(\theta, u^*(\theta), q^*(\theta)) k(\theta) \} dF(\theta) \end{aligned} \right] \\
&\quad + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [q^*(\theta) - u^{*'}(\theta)] + \mu(\theta) q^{*'}(\theta) \} dF(\theta) + \{ \lambda(\theta^*) [q^*(\theta^*) - u^{*'}(\theta^*)] + \mu(\theta) q^{*'}(\theta^*) \} f(\theta^*) \delta\theta^* \\
&\quad + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [k(\theta) - h'(\theta)] + \mu(\theta) k'(\theta) \} dF(\theta). \tag{20}
\end{aligned}$$

We thus have

$$\begin{aligned}
J - J^* &\approx \left[G \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) \right) + G' \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) \right) \cdot \left(\int_{\theta^*}^{\bar{\theta}} h(\theta) dF(\theta) + u(\theta^*) f(\theta^*) \delta\theta^* \right) \right] \\
&\quad \cdot \left[\begin{aligned} &\int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta) + H(\theta^*, u^*(\theta^*), q^*(\theta^*)) f(\theta^*) \delta\theta^* \\ &+ \int_{\theta^*}^{\bar{\theta}} \{ H_u(\theta, u^*(\theta), q^*(\theta)) h(\theta) + H_q(\theta, u^*(\theta), q^*(\theta)) k(\theta) \} dF(\theta) \end{aligned} \right] \\
&\quad + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [q^*(\theta) - u^{*'}(\theta)] + \mu(\theta) q^{*'}(\theta) \} dF(\theta) + \{ \lambda(\theta^*) [q^*(\theta^*) - u^{*'}(\theta^*)] + \mu(\theta) q^{*'}(\theta^*) \} f(\theta^*) \delta\theta^* \\
&\quad + \int_{\theta^*}^{\bar{\theta}} \{ \lambda(\theta) [k(\theta) - h'(\theta)] + \mu(\theta) k'(\theta) \} dF(\theta) - G \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) \right) \int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta)
\end{aligned}$$

$$-\int_{\theta^*}^{\bar{\theta}} \{\lambda(\theta)[q^*(\theta) - u^{*\prime}(\theta)] + \mu(\theta)q^{*\prime}(\theta)\} dF(\theta).$$

Hence we have

$$\begin{aligned} J - J^* &\approx \int_{\theta^*}^{\bar{\theta}} G \left\{ b + \left[H_u(\theta, u^*(\theta), q^*(\theta)) + \frac{d(\lambda(\theta)f(\theta))}{Gf(\theta)d\theta} \right] \right\} h(\theta) dF(\theta) \\ &\quad + \int_{\theta^*}^{\bar{\theta}} \left[GH_q(\theta, u^*(\theta), q^*(\theta)) + \lambda(\theta) - \frac{d(\mu(\theta)f(\theta))}{f(\theta)d\theta} \right] k(\theta) dF(\theta) \\ &\quad - \lambda(\bar{\theta})f(\bar{\theta})h(\bar{\theta}) + \mu(\bar{\theta})f(\bar{\theta})k(\bar{\theta}) + \lambda(\theta^*)f(\theta^*)h(\theta^*) - \mu(\theta^*)f(\theta^*)k(\theta^*) \\ &\quad + [Gbu^*(\theta^*) + GH(\theta^*, u^*(\theta^*), q^*(\theta^*)) + \lambda(\theta^*)q^*(\theta^*)] f(\theta^*) \cdot \delta\theta^* \\ &\quad + \mu(\theta^*)f(\theta^*) \cdot q^{*\prime}(\theta^*)\delta\theta^* - \lambda(\theta^*)f(\theta^*) \cdot u^{*\prime}(\theta^*)\delta\theta^*. \end{aligned}$$

Note that

$$\begin{aligned} \delta u(\theta^*) &= u(\theta^* - \delta\theta^*) - u^*(\theta^*) \approx u(\theta^*) - u^{*\prime}(\theta^*)\delta\theta^* - u^*(\theta^*) = h(\theta^*) - u^{*\prime}(\theta^*)\delta\theta^* \\ \delta q(\theta^*) &= q(\theta^* - \delta\theta^*) - q^*(\theta^*) \approx q(\theta^*) - q^{*\prime}(\theta^*)\delta\theta^* - q^*(\theta^*) = k(\theta^*) - q^{*\prime}(\theta^*)\delta\theta^* \end{aligned}$$

Which imply

$$h(\theta^*) \approx \delta u(\theta^*) + u^{*\prime}(\theta^*)\delta\theta^* \quad (21)$$

$$k(\theta^*) \approx \delta q(\theta^*) + q^{*\prime}(\theta^*)\delta\theta^* \quad (22)$$

Substituting (21) and (22) into the expression of $(J - J^*)$ above, and simplifying, we obtain:

$$\begin{aligned} J - J^* &\approx \int_{\theta^*}^{\bar{\theta}} G \left\{ b + \left[H_u(\theta, u^*(\theta), q^*(\theta)) + \frac{d(\lambda(\theta)f(\theta))}{Gf(\theta)d\theta} \right] \right\} h(\theta) dF(\theta) \\ &\quad + \int_{\theta^*}^{\bar{\theta}} \left[GH_q(\theta, u^*(\theta), q^*(\theta)) + \lambda(\theta) - \frac{d(\mu(\theta)f(\theta))}{f(\theta)d\theta} \right] k(\theta) dF(\theta) \\ &\quad - \lambda(\bar{\theta})f(\bar{\theta})h(\bar{\theta}) + \mu(\bar{\theta})f(\bar{\theta})k(\bar{\theta}) \\ &\quad + [bGu^*(\theta^*) + GH(\theta^*, u^*(\theta^*), q^*(\theta^*)) + \lambda(\theta^*)q^*(\theta^*)] f(\theta^*) \cdot \delta\theta^* \\ &\quad - \mu(\theta^*)f(\theta^*) \cdot \delta q(\theta^*) + \lambda(\theta^*)f(\theta^*) \cdot \delta u(\theta^*) \\ &\leq 0, \end{aligned} \quad (23)$$

where b is given by (8).

Substituting $H_u(\theta, u^*(\theta), q^*(\theta)) = -1$ and $H_q(\theta, u^*(\theta), q^*(\theta)) = \theta - C'(q^*(\theta))$ into (23), we have (7).

Next, we will complete the proof for the proposition. Substituting the transversality conditions and (9)-(11) into (7), we have

$$\begin{aligned}
\frac{J - J^*}{Gf(\theta^*)} &\approx \{bu^*(\theta^*) + [\theta^* q^*(\theta^*) - C(q^*(\theta^*)) - u^*(\theta^*)] - (1-b)\xi(\theta^*)q^*(\theta^*)\} \delta\theta^* \\
&\quad - (1-b)\xi(\theta^*)\delta u(\theta^*) - \frac{\mu(\theta^*)}{G} \delta q(\theta^*) \\
&= [q^*(\theta^*)C'(q^s(\theta^*)) - C(q^*(\theta^*)) - (1-b)u^*(\theta^*)] \delta\theta^* - (1-b)\xi(\theta^*)\delta u(\theta^*) - \frac{\mu(\theta^*)}{G} \delta q(\theta^*) \\
&\leq 0
\end{aligned} \tag{24}$$

By (12),

$$\mu(\theta^*) = -\frac{G}{f(\theta^*)} \int_{\theta^*}^{\bar{\theta}} [C'(q^s(\theta)) - C'(q^*(\theta))] f(\theta) d\theta$$

If $\mu(\theta^*) > 0$, then $-\mu(\theta^*)/G < 0$, which implies that $q^*(\theta^*) = 0$ (suppose not, then $q^*(\theta^*) > 0$ and $\delta q^*(\theta^*)$ can be both positive and negative, which implies $\mu(\theta^*) = 0$, a contradiction). By the continuity of $\mu(\theta)$, there is an interval $[\theta^*, \tilde{\theta}]$ such that $\mu(\theta) > 0$ for all $\theta \in [\theta^*, \tilde{\theta}]$. So $q^*(\theta) = 0$ for all $\theta \in [\theta^*, \tilde{\theta}]$. But this contradicts the assumption that θ^* is the lowest type covered.

So we must have $\mu(\theta^*) = 0$. When $\theta^* > \underline{\theta}$, we have $q^*(\theta^*) = 0$, where θ^* is determined by $MR(\theta^*) = C'(0)$, and bunching does not occur in a (right) neighborhood of θ^* ; when $\theta^* = \underline{\theta}$, bunching may occur at the bottom (in which case $q^*(\underline{\theta})$ may not equal $q^s(\underline{\theta})$).

Substituting $\mu(\theta^*) = 0$ into (24), we have

$$\frac{J - J^*}{Gf(\theta^*)} \approx [q^*(\theta^*)C'(q^s(\theta^*)) - C(q^*(\theta^*)) - (1-b)u^*(\theta^*)] \delta\theta^* - (1-b)\xi(\theta^*)\delta u(\theta^*) \leq 0 \tag{25}$$

If $u(\theta^*) > 0$, $\delta u(\theta^*)$ can be both positive and negative, in which case we have $b = 1$; if $u(\theta^*) = 0$, $\delta u(\theta^*)$ can only be positive, in which case we have $b \leq 1$. In both cases, we have $b \leq 1$. Since we have previously argued that $b \geq 0$, we conclude that $0 \leq b \leq 1$. Note that when $0 \leq b < 1$, we must have $u(\theta^*) = 0$.

It is now clear that $0 < b \leq 1$ corresponds to our model with costly entry, while $b = 0$ corresponds to the Mussa-Rosen benchmark without costly entry.

We now turn to the bunching analysis. First, we will demonstrate that bunching never occurs at the top (in the left neighborhood of $\bar{\theta}$). Suppose in negation, bunching occurs at $q^*(\theta) = \bar{q} > 0$ over some interval $[\theta_1, \bar{\theta}]$. First suppose $\bar{q} < q^s(\bar{\theta})$. Then for θ to be in the neighborhood of $\bar{\theta}$, we have $C'(\bar{q}) < C'(q^s(\bar{\theta})) = MR(\bar{\theta})$. So in this neighborhood, we must have $\mu(\theta) < 0$, a contradiction. Now suppose $\bar{q} \geq q^s(\bar{\theta}) = C'^{-1}(\bar{\theta})$. Then $\theta_1 = \underline{\theta}$, as there exists no $\theta < \bar{\theta}$ for which $q^s(\theta) \geq q^s(\bar{\theta})$. But then we have $C'(\bar{q}) \geq C'(q^s(\bar{\theta})) = \bar{\theta} > MR(\bar{\theta})$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. This implies $\mu(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$, and in particular, $\mu(\underline{\theta}) > 0$, contradicting $\mu(\theta^*) = 0$ established earlier. So the solution must exhibit perfect sorting at the top, hence $q^*(\bar{\theta}) = C'^{-1}(\bar{\theta})$. Thus efficiency at the top is a robust finding beyond the Mussa-Rosen benchmark.

So in our model, bunching can only occur over subintervals in $(\theta^*, \bar{\theta})$ (interior bunching), where $\theta^* > \underline{\theta}$,

or in a neighborhood of $\underline{\theta}$ (bottom bunching).

We first consider an interior bunching interval $[\theta_1, \theta_2] \subset (\theta^*, \bar{\theta})$. By the continuity of $q^*(\theta)$, we have $q^s(\theta_1) = q^s(\theta_2)$, or (15). By the continuity of $\mu(\theta)$, we have $\mu(\theta_1) = \mu(\theta_2) = 0$, which implies (16). So the bunching interval endpoints θ_1 and θ_2 are determined by solving equations (15) and (16). Once θ_1 and θ_2 are determined, the bunching quality is determined by $\bar{q} = q^s(\theta_1) = q^s(\theta_2)$.

For a bottom bunching interval, say, $[\underline{\theta}, \theta_2] \subseteq [\underline{\theta}, \bar{\theta})$. The right bunching endpoint θ_2 is determined by (16) with θ_1 being replaced by $\underline{\theta}$, and the bunching quality is determined by $\bar{q} = q^s(\theta_2)$.

Proof of Lemma 2: When $b = 1$, our previous analysis reveals that the optimal solution is first best, where bunching does not occur. Thus, in the rest of the proof, we focus on the case where $b \in [0, 1)$.

Suppose bunching occurs over $[\theta_1, \theta_2] \subseteq [\underline{\theta}, \bar{\theta})$ in our model. Then there exists an interval $[\theta', \theta''] \subseteq [\theta_1, \theta_2]$ such that $q^{s'}(\theta) < 0$ for $\theta \in [\theta', \theta'']$, which in turn implies that $q_{MR}^{s'}(\theta) < 0$ for $\theta \in [\theta', \theta'']$. So either $[\theta', \theta'']$ is not served, or bunching also occurs in Mussa-Rosen over some interval, say, $[\theta_1^{MR}, \theta_2^{MR}] \subseteq [\underline{\theta}, \bar{\theta})$.

Next we consider a generic bunching interval associated with any $b \in [0, 1)$. We first consider an interior bunching such that $[\theta_1, \theta_2] \subset (\underline{\theta}, \bar{\theta})$.

Define

$$n(\theta, b) = \theta - (1 - b)\xi(\theta),$$

Equations (15) and (16) can be rewritten as

$$n(\theta_1, b) = n(\theta_2, b) \tag{26}$$

$$\int_{\theta_1}^{\theta_2} F(\theta) \cdot n_{\theta}(\theta, b) d\theta = 0 \tag{27}$$

Differentiating equations (26) and (27) with respect to b , we have

$$\begin{aligned} n_{\theta}(\theta_1, b) \frac{d\theta_1}{db} + n_b(\theta_1, b) &= n_{\theta}(\theta_2, b) \frac{d\theta_2}{db} + n_b(\theta_2, b) \\ F(\theta_2) \cdot n_{\theta}(\theta_2, b) \frac{d\theta_2}{db} - F(\theta_1) \cdot n_{\theta}(\theta_1, b) \frac{d\theta_1}{db} + \int_{\theta_1}^{\theta_2} F(\theta) \cdot n_{\theta b}(\theta, b) d\theta &= 0 \end{aligned}$$

which can be written as

$$\begin{pmatrix} n_{\theta}(\theta_1, b) & -n_{\theta}(\theta_2, b) \\ F(\theta_1) \cdot n_{\theta}(\theta_1, b) & -F(\theta_2) \cdot n_{\theta}(\theta_2, b) \end{pmatrix} \begin{pmatrix} \frac{d\theta_1}{db} \\ \frac{d\theta_2}{db} \end{pmatrix} = \begin{pmatrix} n_b(\theta_2, b) - n_b(\theta_1, b) \\ \int_{\theta_1}^{\theta_2} F(\theta) \cdot n_{\theta b}(\theta, b) d\theta \end{pmatrix}.$$

$$|A| = \begin{vmatrix} n_{\theta}(\theta_1, b) & -n_{\theta}(\theta_2, b) \\ F(\theta_1) \cdot n_{\theta}(\theta_1, b) & -F(\theta_2) \cdot n_{\theta}(\theta_2, b) \end{vmatrix}$$

$$= -[1 - (1 - b)\xi'(\theta_1)] \cdot [1 - (1 - b)\xi'(\theta_2)] \cdot (F(\theta_2) - F(\theta_1))$$

$$\begin{aligned} |B| &= \begin{vmatrix} n_b(\theta_2, b) - n_b(\theta_1, b) & -n_\theta(\theta_2, b) \\ \int_{\theta_1}^{\theta_2} F(\theta) \cdot n_{\theta b}(\theta, b) d\theta & -F(\theta_2) \cdot n_\theta(\theta_2, b) \end{vmatrix} \\ &= [1 - (1 - b)\xi'(\theta_2)] \cdot \int_{\theta_1}^{\theta_2} [\xi(\theta_1) - \xi(\theta)] f(\theta) d\theta \end{aligned}$$

$$\begin{aligned} |C| &= \begin{vmatrix} n_\theta(\theta_1, b) & n_b(\theta_2, b) - n_b(\theta_1, b) \\ F(\theta_1) \cdot n_\theta(\theta_1, b) & \int_{\theta_1}^{\theta_2} F(\theta) \cdot n_{\theta b}(\theta, b) d\theta \end{vmatrix} \\ &= [1 - (1 - b)\xi'(\theta_1)] \cdot \int_{\theta_1}^{\theta_2} [\xi(\theta_2) - \xi(\theta)] f(\theta) d\theta \end{aligned}$$

Note that

$$\begin{aligned} (1 - b) \cdot \int_{\theta_1}^{\theta_2} [\xi(\theta_1) - \xi(\theta)] f(\theta) d\theta &= \int_{\theta_1}^{\theta_2} [\theta_1 - C'(q^*(\theta))] f(\theta) d\theta - \int_{\theta_1}^{\theta_2} [\theta - C'(q^s(\theta))] f(\theta) d\theta \\ &= \int_{\theta_1}^{\theta_2} [\theta_1 - \theta] f(\theta) d\theta + \int_{\theta_1}^{\theta_2} [C'(q^s(\theta)) - C'(q^*(\theta))] f(\theta) d\theta \\ &= \int_{\theta_1}^{\theta_2} [\theta_1 - \theta] f(\theta) d\theta \\ (1 - b) \int_{\theta_1}^{\theta_2} [\xi(\theta_2) - \xi(\theta)] f(\theta) d\theta &= \int_{\theta_1}^{\theta_2} [\theta_2 - \theta] f(\theta) d\theta \end{aligned}$$

Moreover, we have $n_\theta(\theta, b) = d[\theta - (1 - b)\xi(\theta)]/d\theta = q^{s'}(\theta)$, and $q^{s'}(\theta_1) > 0$, $q^{s'}(\theta_2) > 0$.

By Cramer's rule, we have

$$\frac{d\theta_1}{db} = \frac{|B|}{|A|} = -\frac{\int_{\theta_1}^{\theta_2} [\xi(\theta_1) - \xi(\theta)] f(\theta) d\theta}{[1 - (1 - b)\xi'(\theta_1)] \cdot (F(\theta_2) - F(\theta_1))} = -\frac{\int_{\theta_1}^{\theta_2} [\theta_1 - \theta] f(\theta) d\theta}{(1 - b)[1 - (1 - b)\xi'(\theta_1)] \cdot (F(\theta_2) - F(\theta_1))} > 0 \quad (28)$$

$$\frac{d\theta_2}{db} = \frac{|C|}{|A|} = -\frac{\int_{\theta_1}^{\theta_2} [\xi(\theta_2) - \xi(\theta)] f(\theta) d\theta}{[1 - (1 - b)\xi'(\theta_2)] \cdot (F(\theta_2) - F(\theta_1))} = -\frac{\int_{\theta_1}^{\theta_2} [\theta_2 - \theta] f(\theta) d\theta}{(1 - b)[1 - (1 - b)\xi'(\theta_2)] \cdot (F(\theta_2) - F(\theta_1))} < 0 \quad (29)$$

Since $b = 0$ in Mussa-Rosen and $b \in (0, 1]$ in our model, we thus have $\theta_1 > \theta_1^{MR}$ and $\theta_2 < \theta_2^{MR}$.

Next we consider bunching at the bottom, say, over the interval $[\underline{\theta}, \theta_2]$. Substituting $\theta_1 = \underline{\theta}$ into (16) and manipulating, we have

$$\int_{\underline{\theta}}^{\theta_2} F(\theta) \cdot n_\theta(\theta, b) d\theta = 0. \quad (30)$$

Differentiating (30) with respect to b , we have

$$\frac{d\theta_2}{db} = -\frac{\int_{\underline{\theta}}^{\theta_2} [\xi(\theta_2) - \xi(\theta)] dF(\theta)}{[1 - (1 - b)\xi'(\theta_2)] F(\theta_2)} = -\frac{\int_{\underline{\theta}}^{\theta_2} [\theta_2 - \theta] dF(\theta)}{(1 - b)[1 - (1 - b)\xi'(\theta_2)] F(\theta_2)} < 0. \quad (31)$$

We thus have $\theta_2^{MR} > \theta_2$.

The above analysis assumes that $q_{MR}^s(\theta_2^{MR}) > 0$ (so bunching occurs at a positive quality level). If $q_{MR}^s(\theta_2^{MR}) \leq 0$, there exists $\theta_{MR}^* \geq \theta_2^{MR} > \theta_2$, such that $q_{MR}^s(\theta_{MR}^*) = 0 = q_{MR}^*(\theta_{MR}^*)$, i.e., all types below $\theta_{MR}^* (> \theta_2)$ are excluded from the market.

In summary, whenever bunching occurs over $[\theta_1, \theta_2] \subseteq [\underline{\theta}, \bar{\theta})$ in our model, either $\theta_{MR}^* > \theta_2$, or bunching also occurs over $[\theta_1^{MR}, \theta_2^{MR}] \subseteq [\underline{\theta}, \bar{\theta})$, where $\theta_1^{MR} \leq \theta_1 < \theta_2 < \theta_2^{MR}$.

Proof of Proposition 2: The idea is to trace the optimal quality schedules backward, starting from the top. We have demonstrated that bunching cannot occur at the top. So in a sufficiently small neighborhood of $\bar{\theta}$, q^* and q_{MR}^* must be perfect sorting. Let $[\theta_1, \bar{\theta}]$ be the longest interval in this neighborhood over which q_{MR}^* is sorting. q^* must also be sorting over $[\theta_1, \bar{\theta}]$ (otherwise contradicting Lemma 2). So by (14), we have $q^*(\theta) > q_{MR}^*(\theta)$ over $[\theta_1, \bar{\theta})$. If $\theta_{MR}^* = \theta_1$, we must have $\theta^* \leq \theta_{MR}^*$ (with equality only at $\theta^* = \theta_{MR}^* = \underline{\theta}$), and we are done with the proof.

If $\theta_{MR}^* < \theta_1$, q_{MR}^* must be bunching in a neighborhood to the left of θ_1 . Let $[\theta_2, \theta_1]$ be the longest interval in such a neighborhood. By Lemma 2, q^* is either sorting over $[\theta_2, \theta_1]$, or bunching over some interval contained in $[\theta_2, \theta_1]$. In either case, we have $q^*(\theta) \geq q^*(\theta_2) > q_{MR}^s(\theta_2) = q_{MR}^*(\theta)$ for $\theta \in [\theta_2, \theta_1]$. If $\theta_{MR}^* = \theta_2$, we must have $\theta^* \leq \theta_{MR}^*$ (with equality only at $\theta^* = \theta_{MR}^* = \underline{\theta}$), and we are done with the proof.

Otherwise this process proceeds and will eventually get to some $\theta_n = \theta_{MR}^*$, in which case we establish that $q^*(\theta) > q_{MR}^*(\theta)$ for all $\theta \in [\theta_{MR}^*, \bar{\theta})$. This also implies that $\theta^* \leq \theta_{MR}^*$ (with equality only when $\theta^* = \theta_{MR}^* = \underline{\theta}$).

Proof of Proposition 4: To show that the first best is achieved if and only if (18) holds, we start with sufficiency. By Assumption 1 and continuity, (18) implies that there exists a unique $\underline{u} \in [0, Ev(q^{fb})]$, such that $Ev(q^{fb}) = \underline{u} + \eta(c^{fb} + \underline{u})$. Thus we can choose $u^*(\theta^*) = \underline{u} \geq 0$, such that

$$\int_{\theta^*}^{\bar{\theta}} \left[\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - \xi(\theta) q^{fb}(\theta) \right] dF(\theta) - u^*(\theta^*) = Ev(q^{fb}) - u^*(\theta^*) = \eta \left(c^{fb} + u^*(\theta^*) \right).$$

This implies

$$b = \frac{1}{\eta} \left\{ \int_{\theta^*}^{\bar{\theta}} \left[\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - \xi(\theta) q^{fb}(\theta) \right] dF(\theta) - u^*(\theta^*) \right\} = 1,$$

where $\eta = \eta(c^{fb} + u^*(\theta^*)) = \eta \left(\int_{\theta^*}^{\bar{\theta}} \xi(\theta) q^{fb}(\theta) dF(\theta) + u^*(\theta^*) \right)$.

As such, the schedule $u^*(\cdot)$ defined by $u^*(\theta) = \int_{\theta^*}^{\theta} q^{fb}(t) dt + u^*(\theta^*)$, where $u^*(\theta^*)$ is chosen above, verifies the variational conditions (14) and (25):

$$q^*(\theta) = C'^{-1}(\theta - (1-b)\xi(\theta)) = C'^{-1}(\theta)$$

$$\begin{aligned}
\frac{J - J^*}{Gf(\theta^*)} &\approx [q^*(\theta^*)C'(q^*(\theta^*)) - C(q^*(\theta^*)) - (1-b)u^*(\theta^*)] \delta\theta^* - (1-b)\xi(\theta^*)\delta u(\theta^*) \\
&= [q^*(\theta^*)C'(q^*(\theta^*)) - C(q^*(\theta^*))] \delta\theta^* \leq 0
\end{aligned}$$

The above inequality holds when either $\theta^* > \underline{\theta}$ or $\theta^* = \underline{\theta}$: when $\theta^* > \underline{\theta}$, $q^{fb}(\theta^*) = 0$; when $\theta^* = \underline{\theta}$, $\delta\theta^* \leq 0$, but $q^{fb}(\theta^*)C'(q^{fb}(\theta^*)) - C(q^{fb}(\theta^*)) \geq 0$ given the convexity of $C(\cdot)$.

Given Assumption 1, the above variational conditions are also sufficient for optimality (Appendix B). Therefore, we have shown that under condition (18), the monopolistic solution is the first best.

Next we show necessity. Suppose that the monopolistic solution achieves the first best, $q^*(\theta) = q^{fb}(\theta)$, then $u^{fb}(\theta) = \int_{\theta^*}^{\theta} q^{fb}(t)dt + u^*(\theta^*)$, $\theta \in [\theta^*, \bar{\theta}]$, and $u^*(\theta^*) \geq 0$. Substituting this into $b = 1$ in (8), we have

$$\int_{\theta^*}^{\bar{\theta}} [\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - \xi(\theta)q^{fb}(\theta)] dF(\theta) - u^*(\theta^*) = \eta \left(\int_{\theta^*}^{\bar{\theta}} \xi(\theta)q^{fb}(\theta) dF(\theta) + u^*(\theta^*) \right),$$

which implies (18), given Assumption 1 and continuity.

Proof of Corollary 2: We first derive the expression of $Ev(q^{fb})$ and show that $Ev(q^{fb}) \geq 0$. Note that

$$\begin{aligned}
\int_{\theta^*}^{\bar{\theta}} \theta q^{fb}(\theta) dF(\theta) &= [\theta q^{fb}(\theta)F(\theta)]|_{\theta^*}^{\bar{\theta}} - \int_{\theta^*}^{\bar{\theta}} F(\theta)\theta dq^{fb}(\theta) - \int_{\theta^*}^{\bar{\theta}} F(\theta)q^{fb}(\theta)d\theta \\
\int_{\theta^*}^{\bar{\theta}} C(q^{fb}(\theta))dF(\theta) &= \int_{\theta^*}^{\bar{\theta}} C((C')^{-1}(\theta))dF(\theta) \\
&= [C((C')^{-1}(\theta)) \cdot F(\theta)]|_{\theta^*}^{\bar{\theta}} - \int_{\theta^*}^{\bar{\theta}} F(\theta) \cdot C'((C')^{-1}(\theta))d[(C')^{-1}(\theta)] \\
&= [C(q^{fb}(\theta)) \cdot F(\theta)]|_{\theta^*}^{\bar{\theta}} - \int_{\theta^*}^{\bar{\theta}} F(\theta)\theta dq^{fb}(\theta) \\
\int_{\theta^*}^{\bar{\theta}} dC(q^{fb}(\theta)) &= \int_{\theta^*}^{\bar{\theta}} \theta dq^{fb}(\theta) = [\theta q^{fb}(\theta)]|_{\theta^*}^{\bar{\theta}} - \int_{\theta^*}^{\bar{\theta}} q^{fb}(\theta)d\theta
\end{aligned}$$

We thus have

$$\begin{aligned}
Ev(q^{fb}) &= \int_{\theta^*}^{\bar{\theta}} \left[\theta q^{fb}(\theta) - C(q^{fb}(\theta)) - \frac{1-F(\theta)}{f(\theta)} q^{fb}(\theta) \right] dF(\theta) \\
&= [(\theta q^{fb}(\theta) - C(q^{fb}(\theta))) \cdot F(\theta)]|_{\theta^*}^{\bar{\theta}} - \int_{\theta^*}^{\bar{\theta}} F(\theta)q^{fb}(\theta)d\theta - \int_{\theta^*}^{\bar{\theta}} [1-F(\theta)]q^{fb}(\theta)d\theta \\
&= [\bar{\theta}q^{fb}(\bar{\theta}) - C(q^{fb}(\bar{\theta}))] - \int_{\theta^*}^{\bar{\theta}} q^{fb}(\theta)d\theta \\
&= [\bar{\theta}q^{fb}(\bar{\theta}) - C(q^{fb}(\bar{\theta}))] + [C(q^{fb}(\theta)) - \theta q^{fb}(\theta)]|_{\theta^*}^{\bar{\theta}} \\
&= \theta^* q^{fb}(\theta^*) - C(q^{fb}(\theta^*)) \\
&= C'(q^{fb}(\theta^*))q^{fb}(\theta^*) - C(q^{fb}(\theta^*))
\end{aligned} \tag{32}$$

$$\begin{aligned}
&= q^{fb}(\theta^*) \left[C'(q^{fb}(\theta^*)) - \frac{C(q^{fb}(\theta^*))}{q^{fb}(\theta^*)} \right] \\
&\geq 0,
\end{aligned}$$

The third equality above is due to $(\theta^* q^{fb}(\theta^*) - C(q^{fb}(\theta^*))) \cdot F(\theta^*) = 0$, as either $F(\theta^*) = 0$ or $q^{fb}(\theta^*) = 0$. The inequality above is due to the convexity of $C(\cdot)$.

When $\underline{\theta} = 0$, we argue that $Ev(q^{fb}) = 0$. To see this, if $q^{fb}(\theta^*) = 0$, then $Ev(q^{fb}) = \theta^* q^{fb}(\theta^*) - C(q^{fb}(\theta^*)) = 0$ (by (32)); if $q^{fb}(\theta^*) > 0$, we must have $\theta^* = \underline{\theta} = 0$, which implies $Ev(q^{fb}) = \underline{\theta} q^{fb}(\underline{\theta}) - C(q^{fb}(\underline{\theta})) = -C(q^{fb}(\underline{\theta})) \leq 0$. Thus we have $Ev(q^{fb}) = 0$ as $Ev(q^{fb}) \geq 0$.

So $Ev(q^{fb}) = 0$, and hence condition (18) fails, which implies that the first best can never be achieved.

Proof of Corollary 3: It can be verified that with $C(q) = \frac{1}{2}q^2$, $Ev(q^{fb}) = \frac{\theta^2}{2}$ (by (32)). Based on this and $F(\theta) = (\theta - \underline{\theta})/(\bar{\theta} - \underline{\theta})$, (18) becomes

$$\begin{aligned}
\frac{\theta^2}{2} &\geq \eta \left(\frac{\bar{\theta} + 2\underline{\theta}}{3} (\bar{\theta} - \underline{\theta}) \right), \text{ or} \\
\eta^{-1} \left(\frac{\theta^2}{2} \right) &\geq \frac{\bar{\theta} + 2\underline{\theta}}{3} (\bar{\theta} - \underline{\theta}),
\end{aligned}$$

which gives rise to $0 < \Delta \leq \left(\sqrt{(3\underline{\theta})^2 + 24\eta^{-1}(\underline{\theta}^2/2)} - 3\underline{\theta} \right)/2$. When $\Delta > \left(\sqrt{(3\underline{\theta})^2 + 24\eta^{-1}(\underline{\theta}^2/2)} - 3\underline{\theta} \right)/2$, downward distortion follows.

Proof of Proposition 5: Note that when $b \in (0, 1]$, there is a one-to-one correspondence between the value of b and the optimal quality schedule q^* : given q^* , b is uniquely determined by (8);¹⁵ given b , the schedule of q^* is also uniquely determined by (14) or (15)-(16) whenever bunching is involved. As such, the value of b can be regarded as the index of the optimal quality schedule q^* , with $b = 0$ corresponding to the solution in the Mussa-Rosen benchmark and $b = 1$ corresponding to the first-best solution. Let (q_i^*, θ_i^*) be the quality schedule and the lowest type served, respectively, in the monopoly solution associated with $b_i \in (0, 1)$, $i = 1, 2$, where $b_1 > b_2$. Based on (28)-(31) established in the proof of Lemma 2, we can conclude that whenever bunching occurs over $[\theta_{11}, \theta_{12}] \subseteq [\underline{\theta}, \bar{\theta}]$ in q_1^* , then either $\theta_2^* > \theta_{12}$, or bunching occurs over $[\theta_{21}, \theta_{22}] \subseteq [\underline{\theta}, \bar{\theta}]$ in q_2^* , where $\theta_{21} \leq \theta_{11} < \theta_{12} < \theta_{22}$. That is, any bunching interval associated with q_1^* is contained in a bunching interval associated with q_2^* (with market exclusion being regarded as a special bunching). Given this, the proof of Proposition 2 can be adapted straightforwardly to show that $q^{fb} \geq q_1^* \geq q_2^* \geq q_{MR}^*$ (with equality only at $\theta = \bar{\theta}$) and $\theta_{MR}^* \geq \theta_2^* \geq \theta_1^* \geq \theta^{*fb}$ (with equality only at $\underline{\theta}$). More generally, we have $\partial q^*(\theta)/\partial b \geq 0$ (with equality only at $\theta = \bar{\theta}$) for $b \in [0, 1)$.

Proposition 4 has shown that when $\eta(c^{*fb}) \leq Ev(q^{fb})$, $b = 1$ and $q^* = q^{fb}$. We thus focus on the

¹⁵When $b \in (0, 1]$, $x^*(\theta^*) = 0$ so $x^*(\theta)$ is uniquely determined by q^* through (1).

case when $\eta(c^{*fb}) > Ev(q^{fb})$ (and hence $b \in [0, 1)$). In this case $u^*(\theta^*) = 0$ and hence $u^*(\theta)$ is uniquely determined by $q^*(\theta)$. We thus have

$$\begin{aligned}\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) &= \int_{\theta^*}^{\bar{\theta}} [1 - F(\theta)] q^*(\theta) d\theta, \\ \int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta) &= \int_{\theta^*}^{\bar{\theta}} [\theta q^*(\theta) - C(q^*(\theta)) - \xi(\theta) q^*(\theta)] dF(\theta).\end{aligned}$$

Note that

$$\frac{\partial}{\partial b} \int_{\theta^*}^{\bar{\theta}} [1 - F(\theta)] q^*(\theta) d\theta = \int_{\theta^*}^{\bar{\theta}} [1 - F(\theta)] \frac{\partial q^*(\theta)}{\partial b} d\theta > 0, \quad (33)$$

and

$$\frac{\partial}{\partial b} \int_{\theta^*}^{\bar{\theta}} [\theta q^*(\theta) - C(q^*(\theta)) - \xi(\theta) q^*(\theta)] dF(\theta) = \int_{\theta^*}^{\bar{\theta}} [\theta - \xi(\theta) - C'(q^*(\theta))] \frac{\partial q^*(\theta)}{\partial b} dF(\theta).$$

We consider the following cases:

1. When $q^*(\theta)$ is perfect sorting, say, over $(\theta', \theta'') \subseteq [\underline{\theta}, \bar{\theta}]$, we have

$$\begin{aligned}& \int_{\theta'}^{\theta''} [\theta - \xi(\theta) - C'(q^*(\theta))] \frac{\partial q^*(\theta)}{\partial b} dF(\theta) \\ &= \int_{\theta'}^{\theta''} [\theta - \xi(\theta) - (\theta - (1-b)\xi(\theta))] \frac{\partial q^*(\theta)}{\partial b} dF(\theta) \\ &= -b \int_{\theta'}^{\theta''} [1 - F(\theta)] \frac{\partial q^*(\theta)}{\partial b} d\theta;\end{aligned}$$

2. When $q^*(\theta)$ is bunching, say, over some interval $[\theta_1, \theta_2] \subseteq [\underline{\theta}, \bar{\theta})$, we have

$$\begin{aligned}& \int_{\theta_1}^{\theta_2} [\theta - \xi(\theta) - C'(q^*(\theta))] \frac{\partial q^*(\theta)}{\partial b} dF(\theta) \\ &= \frac{\partial q^*(\theta)}{\partial b} \int_{\theta_1}^{\theta_2} [\theta - \xi(\theta) - C'(q^*(\theta))] dF(\theta) \\ &= \frac{\partial q^*(\theta)}{\partial b} \int_{\theta_1}^{\theta_2} [\theta - \xi(\theta) - (\theta - (1-b)\xi(\theta))] dF(\theta) \text{ (by (16))} \\ &= -b \int_{\theta_1}^{\theta_2} [1 - F(\theta)] \frac{\partial q^*(\theta)}{\partial b} d\theta.\end{aligned}$$

Combining the two cases above, we have

$$\frac{\partial}{\partial b} \int_{\theta^*}^{\bar{\theta}} H(\theta, u^*(\theta), q^*(\theta)) dF(\theta) = -b \int_{\theta^*}^{\bar{\theta}} [1 - F(\theta)] \frac{\partial q^*(\theta)}{\partial b} d\theta < 0. \quad (34)$$

We are now ready to show that if $\eta_1 \leq \eta_2$, then $0 \leq b_2 \leq b_1 \leq 1$.

Suppose in negation, we have $b_1 < b_2$. By (33), (34), and that $q_1^*(\theta) = u_1^*(\theta) = 0$ for $\theta \in [\theta_2^*, \theta_1^*]$, we have

$$\int_{\theta_1^*}^{\bar{\theta}} H(\theta, u_1^*(\theta), q_1^*(\theta)) dF(\theta) = \int_{\theta_2^*}^{\bar{\theta}} H(\theta, u_1^*(\theta), q_1^*(\theta)) dF(\theta) > \int_{\theta_2^*}^{\bar{\theta}} H(\theta, u_2^*(\theta), q_2^*(\theta)) dF(\theta),$$

and

$$\int_{\theta_1^*}^{\bar{\theta}} u_1^*(\theta) dF(\theta) = \int_{\theta_2^*}^{\bar{\theta}} u_1^*(\theta) dF(\theta) < \int_{\theta_2^*}^{\bar{\theta}} u_2^*(\theta) dF(\theta).$$

We thus have

$$\eta_1 \left(\int_{\theta_1^*}^{\bar{\theta}} u_1^*(\theta) dF(\theta) \right) < \eta_1 \left(\int_{\theta_2^*}^{\bar{\theta}} u_2^*(\theta) dF(\theta) \right) \leq \eta_2 \left(\int_{\theta_2^*}^{\bar{\theta}} u_2^*(\theta) dF(\theta) \right).$$

But then we have

$$1 > \frac{b_1}{b_2} = \frac{\eta_2 \left(\int_{\theta_2^*}^{\bar{\theta}} u_2^*(\theta) dF(\theta) \right) \int_{\theta_1^*}^{\bar{\theta}} H(\theta, u_1^*(\theta), q_1^*(\theta)) dF(\theta)}{\eta_1 \left(\int_{\theta_1^*}^{\bar{\theta}} u_1^*(\theta) dF(\theta) \right) \int_{\theta_2^*}^{\bar{\theta}} H(\theta, u_2^*(\theta), q_2^*(\theta)) dF(\theta)} > 1,$$

which is a contradiction. Thus we have established that if $\eta_1 \leq \eta_2$, we have $0 \leq b_2 \leq b_1 \leq 1$, which further implies $q^{fb} \geq q_1^* \geq q_2^* \geq q_{MR}^*$ and $\theta_{MR}^* \geq \theta_2^* \geq \theta_1^* \geq \theta^{fb}$.

Proof of Proposition 6: We consider a more general setting where the monopolist can provide an entry subsidy s to every consumer who enters, where s can be either positive or negative (a negative s denotes an *entry fee*). The firm's problem can be formulated as follows:

$$\begin{aligned} \max_{s, u(\theta)} \quad & G \left(\int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) + s \right) \cdot \left\{ \int_{\theta^*}^{\bar{\theta}} H(\theta, u(\theta), u'(\theta)) dF(\theta) - s \right\} \\ \text{s.t.} \quad & u(\theta^*) \geq 0, \theta^* \geq \underline{\theta}, u''(\theta) \geq 0 \end{aligned}$$

where $H(\theta, u(\theta), u'(\theta)) = \theta u'(\theta) - C(u'(\theta)) - u(\theta)$.

For now, we will work with the relaxed program, dropping the monotonicity constraint $u''(\theta) \geq 0$ (we will do the consistency check after obtaining the solution). Suppose that the optimal solution to the maximization problem above exists and is given by $\{s^*, u^*(\theta) : \theta \in [\theta^*, \bar{\theta}]\}$. Then the first-order condition with respect to s is given by

$$G' \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) + s^* \right) \left\{ \int_{\theta^*}^{\bar{\theta}} (\theta u^{*'}(\theta) - C(u^{*'}(\theta))) dF(\theta) - \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) + s^* \right) \right\} = G \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) + s^* \right)$$

which leads to

$$\eta \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) + s^* \right) + \left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) + s^* \right) = \int_{\theta^*}^{\bar{\theta}} (\theta u^{*'}(\theta) - C(u^{*'}(\theta))) dF(\theta) \quad (35)$$

From the above equation, we can write

$$\left(\int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) + s^* \right) = \psi \left(\int_{\theta^*}^{\bar{\theta}} (\theta u^{*'}(\theta) - C(u^{*'}(\theta))) dF(\theta) \right) \equiv \psi^* \quad (36)$$

for some strictly increasing function ψ (by Assumption 1). Substituting (36) into the objective function and using (35), we can obtain

$$\begin{aligned} J^* &= G(\psi^*) \cdot \left\{ \int_{\theta^*}^{\bar{\theta}} [\theta u^{*'}(\theta) - C(u^{*'}(\theta)) - u^*(\theta)] dF(\theta) - s^* \right\} \\ &= G(\psi^*) \cdot \eta(\psi^*) \end{aligned}$$

Thus maximizing J^* is equivalent to maximizing ψ^* or the expected trade surplus $\left(\int_{\theta^*}^{\bar{\theta}} (\theta u^{*'}(\theta) - C(u^{*'}(\theta))) dF(\theta) \right)$ which gives rise to the first-best quality provision:

$$q(\theta) = C'^{-1}(\theta) \text{ for all } \theta \in [\theta^*, \bar{\theta}], \text{ where } \theta^* = \max\{\underline{\theta}, C'(0)\} \quad (37)$$

Since $q(\theta) = C'^{-1}(\theta)$ is strictly increasing, the monotonicity constraint is satisfied, which justifies our approach to work with the relaxed program.

Substituting the first-best solution (i.e., $u(\theta) = u(\theta^*) + \int_{\underline{\theta}}^{\theta} q^{fb}(\tau) d\tau$) into (35), we can derive the following condition that determines the optimal entry subsidy, s^* :

$$z = Ev(q^{fb}) - \eta \left(\int_{\theta^*}^{\bar{\theta}} \xi(\theta) q^{fb}(\theta) dF(\theta) + z \right),$$

where $z = u(\theta^*) + s^*$, which can be regarded as a *compound subsidy*. By Assumption 1 and continuity, we have

$$\begin{cases} z \geq 0 & \text{if } Ev(q^{fb}) \geq \eta \left(\int_{\theta^*}^{\bar{\theta}} \xi(\theta) q^{fb}(\theta) dF(\theta) \right) \\ z < 0 & \text{if } Ev(q^{fb}) < \eta \left(\int_{\theta^*}^{\bar{\theta}} \xi(\theta) q^{fb}(\theta) dF(\theta) \right) \end{cases} \quad (38)$$

Apparently, the optimal (net) subsidy $s^* = z - u(\theta^*)$ can be either negative or positive: when condition (18) holds, $z \geq 0$ and the optimal subsidy can be positive (or simply zero); when condition (18) is violated, $z < 0$ and the optimal subsidy has to be negative (which implies entry fees). It is the latter case that the ability of charging entry fees is required for the first-best solution to emerge in the optimal nonlinear pricing contract.

Thus when the monopolist can charge entry fees, it is always optimal to offer the first-best quality provision.

Proof of Proposition 7: The social planner's objective is to maximize the expected total social surplus, which is the expected total surplus generated from the sale less the expected total entry cost, by choosing the nonlinear pricing contract represented by $u(\cdot)$. Write entry cutoff as \hat{c} :

$$\begin{aligned}\hat{c} &= \int_{\theta^*}^{\bar{\theta}} u(\theta) dF(\theta) = \int_{\theta^*}^{\bar{\theta}} \left[u(\theta^*) + \int_{\theta^*}^{\theta} u'(\theta) d\theta \right] dF(\theta) \\ &= \int_{\theta^*}^{\bar{\theta}} \left[\int_{\theta^*}^{\theta} u'(\theta) d\theta \right] dF(\theta) + (1 - F(\theta^*)) u(\theta^*)\end{aligned}$$

Then the social planner's maximization program can be written as follows:

$$\begin{aligned}\max_{u(\theta)} \quad & G(\hat{c}) \cdot \int_{\theta^*}^{\bar{\theta}} [\theta u'(\theta) - C(u'(\theta))] dF(\theta) - \int_{\underline{c}}^{\hat{c}} c dG(c) \\ \text{s.t.} \quad & u(\theta^*) \geq 0, \theta^* \geq \underline{\theta}, u''(\theta) \geq 0, \int_{\theta^*}^{\bar{\theta}} [\theta u'(\theta) - C(u'(\theta)) - u(\theta)] dF(\theta) \geq 0\end{aligned}$$

where the last constraint is the “resource” constraint: the total expected profit cannot be negative.

We will solve the above maximization problem by ignoring all the constraints first. We will perform the consistency check later. Suppose that the optimal solution to the maximization problem above exists and is given by $(u^*(\theta^*), u^{*\prime}(\theta))$. Then the first-order condition with respect to $u(\theta^*)$ is given by

$$G'(\hat{c}^*) (1 - F(\theta^*)) \cdot \int_{\theta^*}^{\bar{\theta}} [\theta u^{*\prime}(\theta) - C(u^{*\prime}(\theta))] dF(\theta) = \hat{c}^* G'(\hat{c}^*) (1 - F(\theta^*)),$$

which implies

$$\hat{c}^* = \int_{\theta^*}^{\bar{\theta}} [\theta u^{*\prime}(\theta) - C(u^{*\prime}(\theta))] dF(\theta). \quad (39)$$

This means that the “resource” constraint is binding and all the rents are given to the consumers (by adjusting the common rent provision $u(\theta^*)$). Substituting (39) into the objective function leads to

$$J^* = \int_{\underline{c}}^{\hat{c}^*} G(c) dc$$

Thus maximizing J^* is equivalent to maximizing $\hat{c}^* = \int_{\theta^*}^{\bar{\theta}} [\theta u^{*\prime}(\theta) - C(u^{*\prime}(\theta))] dF(\theta)$, which gives rise to the first-best quality provision (37). Since $q^{fb}(\theta)$ is strictly increasing, the monotonicity constraint is satisfied. In addition, solving from

$$\hat{c}^* = \int_{\theta^*}^{\bar{\theta}} \left[u(\theta^*) + \int_{\theta^*}^{\theta} u'(\theta) d\theta \right] dF(\theta) = \int_{\theta^*}^{\bar{\theta}} [\theta u^{*\prime}(\theta) - C(u^{*\prime}(\theta))] dF(\theta),$$

we have

$$\begin{aligned} u^*(\theta^*) &= \frac{1}{1-F(\theta^*)} \int_{\theta^*}^{\bar{\theta}} \left[\theta u^{*'}(\theta) - C(u^{*'}(\theta)) - \frac{1-F(\theta)}{f(\theta)} u^{*'}(\theta) \right] dF(\theta) \\ &= \frac{Ev(q^{fb})}{1-F(\theta^*)} \geq 0 \end{aligned}$$

by a result demonstrated in the proof of Corollary 2. So all the constraints are satisfied, which justifies our approach to work with the relaxed program.

We are now ready to compare the entry levels induced by the monopolist and the social planner. Let c^* and c^{**} be entry cutoffs under monopolistic and socially optimum solutions, respectively, and $u^*(\theta)$ and $u^{**}(\theta)$ be the rent provisions under monopolistic and socially optimum solutions, respectively. First we consider the case when condition (18) is satisfied. By (38), the first best is achieved by the monopolist without charging entry fees. In this case, $u^*(\theta^*) = Ev(q^{fb}) - \eta \left(\int_{\theta^*}^{\bar{\theta}} \xi(\theta) C'^{-1}(\theta) dF(\theta) \right) < Ev(q^{fb}) = u^{**}(\theta^*)$, which implies that $u^*(\theta) < u^{**}(\theta)$. Thus $c^* = \int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) < \int_{\theta^*}^{\bar{\theta}} u^{**}(\theta) dF(\theta) = c^{**}$.

Next we consider the case when condition (18) is violated. By (38) again, the monopolistic solution involves downward quality distortion when the monopolist cannot charge entry fees, in which case it is obvious that $c^* < c^{**}$. When the monopolist can charge entry fees, the first best is achieved, in which case $u^*(\theta) = s^* + u^*(\theta^*) + \int_{\theta^*}^{\theta} q^{fb}(\tau) d\tau < \int_{\theta^*}^{\theta} q^{fb}(\tau) d\tau < u^{**}(\theta)$, as $s^* + u^*(\theta) < 0$. Thus again, $c^* = \int_{\theta^*}^{\bar{\theta}} u^*(\theta) dF(\theta) < \int_{\theta^*}^{\bar{\theta}} u^{**}(\theta) dF(\theta) = c^{**}$.

In summary, the monopoly induces insufficient entry compared to socially efficient entry in all cases.

APPENDIX B: SUFFICIENT CONDITIONS FOR OPTIMALITY

We will show that the (strict) log-concavity of G (Assumption 1) is sufficient to guarantee that the first variation conditions derived in the main text are both necessary and sufficient for optimization.

Since taking a (positive) monotone transformation preserves the solution to a maximization problem, the firm's problem can be reformulated as follows:

$$\begin{aligned} \max_{u(\cdot)} \quad & \log \left(G \left(\int_{\underline{\theta}}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot \int_{\underline{\theta}}^{\bar{\theta}} H(\theta, u(\theta), u'(\theta)) dF(\theta) \right) \\ = \quad & \log G \left(\int_{\underline{\theta}}^{\bar{\theta}} u(\theta) dF(\theta) \right) + \log \int_{\underline{\theta}}^{\bar{\theta}} H(\theta, u(\theta), u'(\theta)) dF(\theta) \\ \text{s.t.} \quad & u(\theta) \geq 0, u'(\theta) \geq 0, u''(\theta) \geq 0 \end{aligned}$$

or

$$\min_u Q(u) \text{ s.t. } u \in P,$$

where $Q(u) = -\log \left(G \left(\int_{\underline{\theta}}^{\bar{\theta}} u(\theta) dF(\theta) \right) \cdot \int_{\underline{\theta}}^{\bar{\theta}} H(\theta, u(\theta), u'(\theta)) dF(\theta) \right)$ and $P = \left\{ u \in C^2 \left[\underline{\theta}, \bar{\theta} \right], u \geq 0, u' \geq 0, u'' \geq 0 \right\}$.

First we show that P is a convex cone: given any $u_1, u_2 \in P$, and $\alpha \in (0, 1)$, define $\hat{u} = \alpha u_1 + (1 - \alpha)u_2$. It is easily verified that $\hat{u} \in P$, and βu_1 (and βu_2) $\in P$, where $\beta \geq 0$, so P is a convex cone.

Next we show that Q is a convex functional. By the strict concavity of $\log(G(\cdot))$, we have

$$\begin{aligned} \log\left(G\left(\int_{\underline{\theta}}^{\bar{\theta}} \hat{u}(\theta) dF(\theta)\right)\right) &= \log\left(G\left(\alpha \int_{\underline{\theta}}^{\bar{\theta}} u_1(\theta) dF(\theta) + (1 - \alpha) \int_{\underline{\theta}}^{\bar{\theta}} u_2(\theta) dF(\theta)\right)\right) \\ &> \alpha \log\left(G\left(\int_{\underline{\theta}}^{\bar{\theta}} u_1(\theta) dF(\theta)\right)\right) + (1 - \alpha) \log\left(G\left(\int_{\underline{\theta}}^{\bar{\theta}} u_2(\theta) dF(\theta)\right)\right). \end{aligned}$$

Moreover,

$$\begin{aligned} &\log\left(\int_{\underline{\theta}}^{\bar{\theta}} H(\theta, \hat{u}(\theta), \hat{u}'(\theta)) dF(\theta)\right) \\ &= \log\left(\int_{\underline{\theta}}^{\bar{\theta}} [\theta(\hat{u}'(\theta)) - (\hat{u}(\theta)) - C(\hat{u}'(\theta))] dF(\theta)\right) \\ &> \log\left(\int_{\underline{\theta}}^{\bar{\theta}} [\alpha(\theta u_1'(\theta) - u_1(\theta) - C(u_1'(\theta))) + (1 - \alpha)(\theta u_2'(\theta) - u_2(\theta) - C(u_2'(\theta)))] dF(\theta)\right) \\ &= \log\left(\alpha \int_{\underline{\theta}}^{\bar{\theta}} H(\theta, u_1(\theta), u_1'(\theta)) dF(\theta) + (1 - \alpha) \int_{\underline{\theta}}^{\bar{\theta}} H(\theta, u_2(\theta), u_2'(\theta)) dF(\theta)\right) \\ &> \alpha \log\left(\int_{\underline{\theta}}^{\bar{\theta}} H(\theta, u_1(\theta), u_1'(\theta)) dF(\theta)\right) + (1 - \alpha) \log\left(\int_{\underline{\theta}}^{\bar{\theta}} H(\theta, u_2(\theta), u_2'(\theta)) dF(\theta)\right). \end{aligned}$$

The first inequality above is due to the strict convexity of cost function $C(\cdot)$. We have thus shown that $Q(u)$ is convex in u .

Since $G(u)$ and $C(u)$ have continuous derivatives with respect to u , it can be easily verified that Q is Fréchet differentiable at u .

In sum, Q is a Fréchet differentiable convex functional on a real normed space $C^2[\underline{\theta}, \bar{\theta}]$ and P is a convex cone in $C^2[\underline{\theta}, \bar{\theta}]$. By Lemma 1 in Luenberger (1969, pp. 227), the first variation conditions derived in the main text are both necessary and sufficient for optimization.

REFERENCES

- [1] Armstrong, Mark and John Vickers. "Competitive Price Discrimination," *Rand Journal of Economics* 32(4), 2001, pp. 579-605.
- [2] Armstrong, Mark and John Vickers. "Competitive Nonlinear Pricing and Bundling," *Review of Economic Studies* 77, 2010, pp. 30-60.
- [3] Baron, D., and R. Myerson. "Regulating a Monopolist with Unknown Costs," *Econometrica* 50, 1982, pp. 911-930.
- [4] Bergemann, D. and J. Välimäki. "Information in Mechanism Design," in *Advances in Economics and Econometrics*, ed. by R. Blundell, W. Newey, and T. Persson, 2006, pp. 186-221. Cambridge University Press, Cambridge.
- [5] Bergemann, Dirk and Achim Wambach. "Sequential Information Disclosure in Auctions," Working paper, 2013.
- [6] Borenstein, Severin and Nancy L. Rose. "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy* 102, 1994, pp. 653-83.
- [7] Champsaur, Paul and Jean-Charles Rochet. "Multiproduct Duopolists," *Econometrica* 57(3), 1989, pp. 533-57
- [8] Courty, Pascal and Li, Hao (2000), Sequential Screening, *Review of Economic Studies* 67, pp. 697-717.
- [9] Crémer, J., and F. Khalil. "Gathering Information Before Signing a Contract," *American Economic Review* 82, 1992, pp. 566-578.
- [10] Crémer, J., F. Khalil, and J.-C. Rochet. "Contracts and Productive Information Gathering," *Games and Economic Behavior* 25, 1998a, pp. 174-193.
- [11] Crémer, J., F. Khalil, and J.-C. Rochet. "Strategic Information Gathering Before a Contract is Offered," *Journal of Economic Theory* 81, 1998b, pp. 163-200.
- [12] Ellison, Glenn. "A Model of Add-on Pricing," *Quarterly Journal of Economics* 120, 2005, pp. 1349-72.
- [13] Engelbrecht-Wiggans, Richard. "Optimal Auctions Revisited," *Games and Economic Behavior* 5, 1993, 227-239.
- [14] Esö, Péter and Balázs Szentes (2007), Optimal Information Disclosure in Auctions and the Handicap Auction, *Review of Economic Studies* 74, pp. 705-731.

- [15] Gal-Or, Esther. "Quality and Quantity Competition," *Bell Journal of Economics* 14(2), 1983, pp. 590-600.
- [16] Gilbert, Richard J. and Carmen Matutes. "Product Line Rivalry with Brand Differentiation," *Journal of Industrial Economics* 41(3), 1993, pp. 223-40.
- [17] Kamien, Morton and Nancy Schwartz. "Dynamic Optimization: the Calculus of Variations and Optimal Control in Economics and Management," 2nd edition, 1992, Elsevier, North Holland.
- [18] Levin, Dan and James Smith. "Equilibrium in Auctions with Entry," *American Economic Review* 84, 1994, pp. 585-599.
- [19] Lu, Jingfeng. "Entry coordination and auction design with private costs of information acquisition," *Economic Inquiry* 48(2), 2010, pp. 274-289.
- [20] Luenberger, David G. *Optimization by Vector Space Methods*, New York: John Wiley & Sons, 1969.
- [21] Maskin, Eric and John Riley. "Monopoly with Incomplete Information," *Rand Journal of Economics* 15, 1984, pp. 171-96.
- [22] P. McAfee, J. McMillan, "Auctions with Entry," *Economic Letters* 23, 1987, pp. 343-347.
- [23] Moreno, Diego and John Wooders. "Auctions with heterogeneous entry costs," *Rand Journal of Economics* 42(2), 2011, pp. 313-336.
- [24] Mussa, Michael and Sherwin Rosen. "Monopoly and Product Quality," *Journal of Economic Theory* 18, 1978, pp. 301-17.
- [25] Myerson, Roger. "Optimal Auction Design," *Mathematics of Operations Research* 6, 1981, pp. 58-73.
- [26] Pavan, Alessandro, Ilya Segal, and Juuso Toikka (2014), Dynamic Mechanism Design: A Myersonian Approach, *Econometrica* 82 (2), 601-653.
- [27] Rochet, Jean-Charles and Lars Stole. "Nonlinear Pricing with Random Participation," *Review of Economic Studies* 69(1), 2002, pp. 277-311.
- [28] Spulber, Daniel. "Product Variety and Competitive Discounts," *Journal of Economic Theory* 48, 1989, pp. 510-525.
- [29] Stole, Lars. "Nonlinear Pricing and Oligopoly," *Journal of Economics and Management Strategy* 4(4), 1995, pp. 529-62.
- [30] Tan, Guofu. "Entry and R&D in Procurement Contracting," *Journal of Economic Theory* 58, 1992, pp. 41-60.

- [31] Verboven, Frank. "Product Line Rivalry and Market Segmentation," *Journal of Industrial Economics* 47(4), 1999, pp. 399-425.
- [32] Villas-Boas, Miguel and Udo Schmidt-Mohr. "Oligopoly with Asymmetric Information: Differentiation in Credit Markets," *Rand Journal of Economics* 30(3), 1999, pp. 375-96.
- [33] Wilson, Robert. *Nonlinear Pricing*, Oxford, Oxford University Press, 1993.
- [34] Yang, Huanxing and Lixin Ye. "Nonlinear Pricing, Market Coverage, and Competition," *Theoretical Economics* 3, 2008, pp. 123-153.