Theoretical Economics 12 (2017), 141–173 1555-7561/20170141

Monopolistic nonlinear pricing with consumer entry

Lixin Ye
Department of Economics, The Ohio State University

Chenglin Zhang
School of Finance, Southwestern University of Economics and Finance

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.

Lixin Ye: ye.45@osu.edu
Chenglin Zhang: chenglinzhangecon@gmail.com
We thank Mark Armstrong, Alain Bensoussan, Hongbin Cai, Jingfeng Lu, George Mailath, Alessandro Pavan, James Peck, Mike Peters, Xi Shan, Huanxing Yang, and seminar audiences at Ohio State University, Peking University, Michigan State University, University of Hong Kong, the Workshop on Information and Incentives (University of Guelph, Canada), and the Midwest Economic Theory Conference (Spring 2014) for very helpful comments. We also greatly benefited from constructive and insightful suggestions from Johannes Hörner and two anonymous referees. All remaining errors are our own.

[1]The 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.

DOI: 10.3982/TE1944

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, p. 303), a type, θ, is the preference "intensity" that "measures intensity of a consumer's taste for quality." 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 and 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 (three dimensional) televisions with two different displaying technologies, one may never know about the "incremental value" of watching a model that does not require wearing eyeglasses over watching one that does;[2] since Smartphones were introduced, many users have been confused over which data plan to subscribe to, 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, to buy a 3D television, a consumer will need to visit a store to find out the specific features of a 3D television model; to buy a new car, a consumer will need to visit a car dealership

[2]The 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://www.businessinsider.com/smartphone-carrier-data-plans-2012-3.

[4]"Clay Christensen's Milkshake Marketing," by Carmen Nobel, Harvard Business School Working Knowledge, February 14, 2011.

for test driving; to sign up for a data plan for a Smartphone, a consumer will 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, so as 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, 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 and Rosen, the monopolistic optimal quality provision (q^*) is characterized by two fundamental types: segments where $q^{*\prime} > 0$ (perfect sorting) and segments where $q^{*\prime} = 0$ (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 determined using ironing techniques similar to those identified in, e.g., Mussa and Rosen (1978) and Myerson (1981). 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 and 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 and Rosen, it also occurs in our model; whenever bunching occurs in Mussa and Rosen, the bunching interval is smaller or simply disappears in our model.

Not only is quality distortion smaller compared to the Mussa–Rosen benchmark, but 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). This result is somewhat surprising but can be explained intuitively. When the monopolist can charge a fee (entry fee) before consumers learn their preference types, it is easy to demonstrate that the monopoly solution is the first best: the optimal mechanism can be implemented by charging an entry fee and committing to cost-plus-fee pricing (i.e., $p(q) = C(q) + p_0$, where $C(q)$ is the production cost and p_0 is a fixed fee). When the

[5]An 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 a purchase without learning about her true preference type or the match value of the product.

monopolists do not charge entry fees (as commonly observed in business and hence assumed in our main analysis), we demonstrate that the cost-plus-fee pricing is still optimal if and only if some condition holds. Such a condition basically ensures that a fixed fee p_0 can be chosen to (i) induce optimal entry, and (ii) satisfy the post-entry individual rationality constraint (IR) for buyers. We identify such a feasibility condition in our model; should this condition fail, the optimal nonlinear pricing contract must involve quality distortion.

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 the rent provision to consumers, which is also inversely related to the price elasticity of entry: the higher is η, the smaller is 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).

Even when the optimal nonlinear pricing involves no quality distortion (production efficiency is achieved), 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 antitrust 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 an 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.[6] 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

[6]See, 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).

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 an effect similar to that of 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 et al. 1998a, 1998b).[7] 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 et al. (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 et al. (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 et al. (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.[8] As 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. 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

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

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

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, we work with general distributions of buyer types; 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 Hao 2000, Esö and Szentes 2007, and, more recently, Pavan et al. 2014 and Bergemann and Wambach 2015). 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 also no benefit to running 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, and Section 3 characterizes our monopoly solutions and compares it with the solution in Mussa and 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 discusses some assumptions/restrictions in our analysis, and Section 5 offers concluding remarks. All long proofs are relegated to Appendix A. Appendix B provides an analysis on sufficient conditions for optimality.

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). The cost $C(q)$ 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 one 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 $[\underline{\theta}, \overline{\theta}]$. Under these assumptions, Mussa and Rosen show that the monopolistic nonlinear pricing solution exhibits three features: (i) 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; (ii) market exclusion, i.e., the monopolist frequently prices consumers with the lowest types out of the market; (iii) 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 and Rosen described above. Formally, there is 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 $[\underline{c}, \overline{c}]$. After entry, consumers draw θ's from the distribution $F(\cdot)$ on $[\underline{\theta}, \overline{\theta}]$. 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).[9] Define the inverse hazard rate functions as

$$\xi(\theta) = \frac{1 - F(\theta)}{f(\theta)}, \quad \theta \in [\underline{\theta}, \overline{\theta}]$$

$$\eta(c) = \frac{G(c)}{g(c)}, \quad c \in [\underline{c}, \overline{c}].$$

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

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

As demonstrated in Appendix B, Assumption 1 is needed to ensure that the necessary conditions for optimality derived in our analysis below are also sufficient for optimality.

The monopolist's objective is to maximize its expected profit from the sale. It is easily verified that under complete information about θ, the first-best solution is given by $q^{\text{fb}}(\theta) = C'^{-1}(\theta)$ for $\theta \geq \theta^{*\text{fb}} = \max\{\underline{\theta}, C'(0)\}$. Note that $C'^{-1}(\theta)$ is strictly increasing given the strict convexity of $C(\theta)$.

In our main analysis, we focus on the case where the firm commits to a nonlinear pricing scheme and the consumers engage privately in information acquisition. Formally, the time line is as follows:

(i) The monopolist offers the (nonlinear) pricing schedule, $P(q)$ or, equivalently, the menu of quality–price contracts, $\{q(\theta), p(\theta)\}$.

(ii) 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.

(iii) Consumers who entered make purchase decisions, and sales are realized.

3. The analysis

The firm offers the nonlinear pricing schedule $p(q) : \mathbb{R}_+ \to \mathbb{R}_+$, which is equivalent to offering a menu of direct contracts of the form $\{q(\theta), p(\theta)\}$, where $\theta \in [\underline{\theta}, \overline{\theta}]$. Given the

[9] An 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 remains unaltered under this alternative setting.

menu of contracts $\{q(\theta), p(\theta)\}$, the utility obtained by a consumer with type θ, when choosing the offer $\{q(\widehat{\theta}), p(\widehat{\theta})\}$, is given by

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

Throughout, the quality provision schedule q is defined over $[\underline{\theta}, \overline{\theta}]$, so the lowest type served in the market, θ^*, is implicitly defined as the cutoff type below which $q(\theta) = 0$ and above which $q(\theta) > 0$.[10]

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

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

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

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:

$$u(\theta) = u(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} q(\tau)\, d\tau \quad \text{for all } \theta \in [\underline{\theta}, \overline{\theta}]. \tag{1}$$

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

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

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

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 $\underline{\theta}$ (i.e., it is possible that $u(\underline{\theta}) > 0$). This marks the first departure from the standard screening model.

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

$$Eu = \int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\, dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} [\theta q(\theta) - p(\theta)]\, dF(\theta).$$

[10]The types not served can also be treated as those who accept the null contract where $q(\theta) = p(\theta) = 0$.

Using (1), we have

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

In equilibrium, a consumer with entry cost c_i enters the sale if and only if $c_i \le c^* \equiv Eu$. 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

$$\pi(\theta, u(\theta), q(\theta)) \equiv p(\theta) - C(q(\theta)) = \theta q(\theta) - C(q(\theta)) - u(\theta).$$

The firm's problem can be formulated as

$$\begin{aligned} &\max_{u(\cdot)} G\left(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\, dF(\theta) \right) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), q(\theta))\, dF(\theta) \\ &\text{s.t. } u(\underline{\theta}) \ge 0 \\ &\quad q(\theta) = u'(\theta), q'(\theta) \ge 0. \end{aligned} \tag{2}$$

The firm's maximization problem in the Mussa–Rosen benchmark can be regarded as a special case in which the actual market base $G(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\, dF(\theta)) = 1$, i.e., when all potential consumers enter the sale.

An implicit assumption from the time line of our model is that the monopoly cannot charge entry fees (before entry occurs). When the firm can charge entry fees, the optimal mechanism is simple and the solution is always the first best. The reason is that the fee is charged before θ is learned, so informational rents arising from the private information about θ can be extracted ex ante. This means that the seller does not need to resort to quality distortion for rent extraction (hence the quality provision is the first best). The optimal mechanism can be interpreted as an optimal procurement in which the expected value of each consumer's entry is given by $V = E_\theta[\max_q \theta q - C(q)] = E_\theta[\theta q^{\text{fb}}(\theta) - C(q^{\text{fb}}(\theta))]$. Given Assumption 1, the optimal procurement takes the form of a posted price $P^* = \arg\max_P (V - P) \cdot G(P)$. The following lemma follows straightforwardly.

Lemma 2. *When the monopolist can charge entry fees, the optimal quality provision is the first best, which can be implemented by setting the entry fee $e^* = V - P^* - p_0$ and committing to the price schedule $p(q) = C(q) + p_0$, where $p_0 = 0$ if $\underline{\theta} < C'(0)$ and $p_0 \in (-\infty, \underline{\theta} \cdot C'^{-1}(\underline{\theta}) - C(C'^{-1}(\underline{\theta}))]$ if $\underline{\theta} \ge C'(0)$.*

Basically the first-best nonlinear pricing mechanism must be a sellout contract with cost-plus-fee pricing: $p(q) = C(q) + p_0$, where p_0 is a fixed fee.[11] When $\underline{\theta} < C'(0)$,

[11] Since $p_0 = p(q) - C(q)$, p_0 is the firm's profit from selling each unit of the product (this per-unit profit is the same for all products with $q > 0$).

$p_0 = 0$ is the only fixed fee that induces the first-best market exclusion ($\theta^{*\mathrm{fb}} = C'(0)$); when $\underline{\theta} \geq C'(0)$, any fixed fee works as long as it does not violate the post-entry IR constraint for the lowest type $\underline{\theta}$, which means that $p_0 \leq \underline{\theta} \cdot q^{\mathrm{fb}}(\underline{\theta}) - C(q^{\mathrm{fb}}(\underline{\theta})) = \underline{\theta} \cdot C'^{-1}(\underline{\theta}) - C(C'^{-1}(\underline{\theta}))$.

Given that charging entry fees is uncommon in practice, our analysis will focus on the case where charging entry fees is not feasible, which we now turn to.

3.1 *Characterization of the monopoly solution*

Let $q^*(\theta)$ and $u^*(\theta)$, $\underline{\theta} \leq \theta \leq \overline{\theta}$, be the optimal quality provision and rent provision, respectively. If the solution only involves perfect sorting (i.e., when the monotonicity constraint, $q'(\theta) \geq 0$, is not binding), the optimal solution can be derived 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

$$\begin{aligned}\Pi &= G\left(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\, dF(\theta)\right) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), q(\theta))\, dF(\theta) \\ &= G\left(\int_{\underline{\theta}}^{\overline{\theta}} (u(\underline{\theta}) + q(\theta)\xi(\theta))\, dF(\theta)\right) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta q(\theta) - C(q(\theta)) - q(\theta)\xi(\theta) - u(\underline{\theta})\right] dF(\theta).\end{aligned} \tag{3}$$

Differentiating (3) with respect to $q(\theta)$ and simplifying, we obtain the optimality condition

$$\theta - (1-b)\xi(\theta) = C'(q^*(\theta)), \tag{4}$$

where

$$b = \frac{G'}{G} \cdot \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u^*(\theta), q^*(\theta))\, dF(\theta) \quad \text{and} \quad G = G\left(\int_{\underline{\theta}}^{\overline{\theta}} u^*(\theta) f(\theta)\, d\theta\right). \tag{5}$$

Equality (4) basically equates marginal revenue ($\theta - (1-b)\xi(\theta) \equiv \mathrm{MR}(\theta)$) with the marginal cost ($C'(q^*(\theta))$) of raising the quality provision $q(\theta)$.

When the optimal solution involves bunching, however, the Lagrangian contains $q'(\theta)$ and the analysis involves optimal control techniques (e.g., Myerson 1981 and Maskin and Riley 1984). So as to formulate our program (2) into a standard optimal control problem (e.g., Pontryagin et al. 1962, Kamien and Schwartz 2012), we define the new variables

$$\begin{aligned} w(\theta) &= \int_{\underline{\theta}}^{\theta} u(t)\, dF(t) \\ v(\theta) &= \int_{\underline{\theta}}^{\theta} \pi(t, u(t), q(t))\, dF(t) \\ z(\theta) &= G(w(\theta)) \cdot v(\theta). \end{aligned}$$

We thus have

$$\begin{aligned} w'(\theta) &= u(\theta)f(\theta) \\ v'(\theta) &= \pi(\theta, u(\theta), q(\theta))f(\theta) \\ z'(\theta) &= G'(w(\theta))w'(\theta)\cdot v(\theta) + G(w(\theta))\cdot v'(\theta) \\ &= G'(w(\theta))u(\theta)f(\theta)\cdot v(\theta) + G(w(\theta))\cdot \pi(\theta, u(\theta), q(\theta))f(\theta). \end{aligned}$$

The firm's objective function can be rewritten as

$$G\left(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta)\right)\cdot \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), q(\theta))\,dF(\theta) = z(\overline{\theta}) = z(\overline{\theta}) - z(\underline{\theta}) = \int_{\underline{\theta}}^{\overline{\theta}} z'(\theta)\,d\theta.$$

Treating $a(\theta)\equiv q'(\theta)$ as the control variable, and $q(\theta)$, $u(\theta)$, $w(\theta)$, and $v(\theta)$ as the state variables, our program (2) can now be formulated as a standard optimal control problem:

$$\begin{aligned} &\max_{q(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} \left[G'(w(\theta))u(\theta)\cdot v(\theta) + G(w(\theta))\cdot \pi(\theta, u(\theta), q(\theta))\right] f(\theta)\,d\theta \\ &\quad \text{s.t. } u(\underline{\theta}) \geq 0,\, u'(\theta) = q(\theta),\, q'(\theta) = a(\theta) \geq 0 \\ &\qquad w'(\theta) = u(\theta)f(\theta),\, v'(\theta) = \pi(\theta, u(\theta), q(\theta))f(\theta). \end{aligned}$$

The Hamiltonian for this problem is given by

$$\begin{aligned} \mathcal{H} = {} & \left[G'(w(\theta))u(\theta)\cdot v(\theta) + G(w(\theta))\cdot \pi(\theta, u(\theta), q(\theta))\right] f(\theta) + \mu(\theta)a(\theta) \\ & + \lambda_1(\theta)q(\theta) + \lambda_2(\theta)u(\theta)f(\theta) + \lambda_3(\theta)\pi(\theta, u(\theta), q(\theta))f(\theta). \end{aligned}$$

The monopoly solution is characterized by the following proposition.

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

(i) Whenever q^* is perfect sorting, it is determined by

$$q^*(\theta) = C'^{-1}(\theta - (1-b)\xi(\theta)) \equiv q^s(\theta), \tag{6}$$

where the expression of b is given by (5).

(ii) Bunching does not occur in the neighborhood of $\overline{\theta}$, and $q^*(\overline{\theta}) = C'^{-1}(\overline{\theta})$ (efficiency at the top).

(iii) Whenever bunching occurs over the interval $[\theta_1, \theta_2] \subset (\underline{\theta}, \overline{\theta})$, it is determined by (7) and (8) for interior bunching and by (8) with θ_1 being replaced by $\underline{\theta}$ for bottom bunching:

$$\theta_1 - (1-b)\xi(\theta_1) = \theta_2 - (1-b)\xi(\theta_2) \tag{7}$$

$$\int_{\theta_1}^{\theta_2} \left[(\theta - (1-b)\xi(\theta)) - C'(q^*(\theta))\right] dF(\theta) = 0. \tag{8}$$

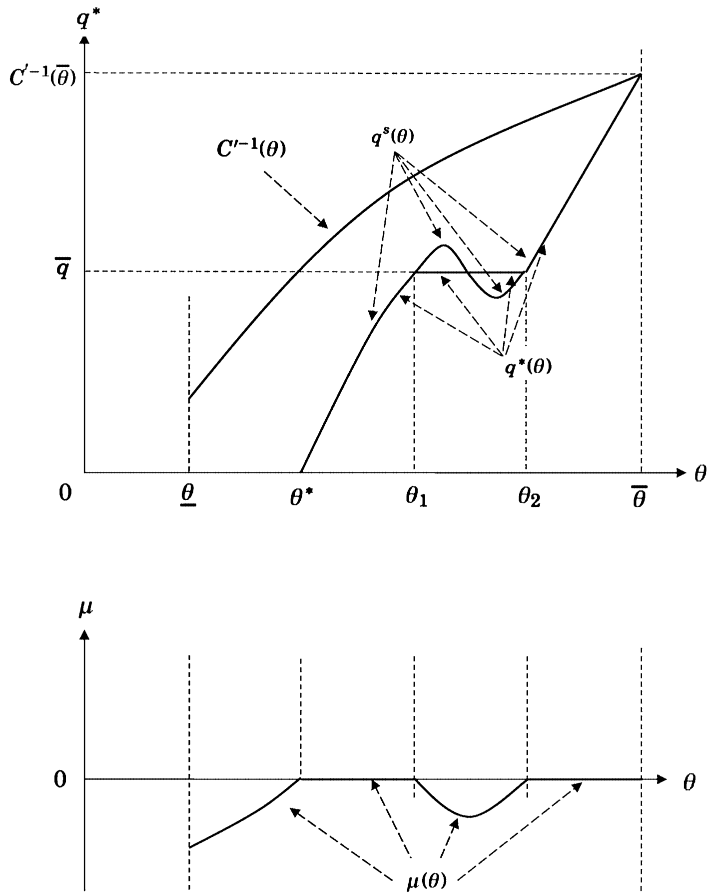

FIGURE 1. Monopoly solution with interior bunching.

Recall that most long proofs are given in Appendix A.

The general intuition for Proposition 1 is clear. Our solution (q^*) is characterized by two fundamental types: segments where $q^{*\prime} > 0$ (perfect sorting) and segments where $q^{*\prime} = 0$ (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 $\mathrm{MR}'(\theta) = 1 - (1 - b)\xi'(\theta)$, $\mathrm{MR}(\theta)$ 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; neither 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. What he can do is to equate the expected marginal revenue with the expected marginal cost over the bunching range, which gives rise to condition (8). The procedure to identify the bunching intervals (and bunching qualities), as spelled out in the proof in the Appendix, is known as the *ironing* technique (e.g., Myerson 1981 and Maskin and Riley 1984), which is illustrated in Figure 1.

To further understand that $\theta - (1-b)\xi(\theta)$ is the marginal revenue from raising the quality provision to type-θ consumers in our setting, consider selling an additional increment of quality to the existing entrants with type θ (with measure $G \cdot f(\theta)$):

(i) For the existing entrants with type θ (with measure $G \cdot f(\theta)$), each has incremental value θ. So total revenue increases by $Gf(\theta) \cdot \theta$.

(ii) 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 $G \cdot (1 - F(\theta))$.

(iii) 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

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

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

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

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 and Rosen, raising quality provision can only affect the existing customers, so $\mathrm{MR}(\theta) = \theta - \xi(\theta)$. With consumer entry, there is an additional component in marginal revenue, which is equal to $b\xi(\theta)$ (≥ 0). Thus b can be interpreted as a measure of this additional marginal revenue due to consumer entry.

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 and Rosen. In Mussa and Rosen, the optimal path q^*_{MR} can be derived straightforwardly, as it can be constructed backward starting from the top ($\theta = \overline{\theta}$) 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^*(hence a functional). 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 $\overline{\theta}$) using the sorting and bunching conditions, and, using the derived q^*_1, we compute the induced value of b from (5). 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 establishes that $0 \le b \le 1$, which implies the following corollary.

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

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

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

3.2 *Comparison of monopoly solutions*

Let q^*_{MR} and q^* denote the optimal quality provision schedules, and θ^*_{MR} and θ^* the lowest types served in the Mussa–Rosen benchmark and our model with consumer entry, respectively. We can first establish the following lemma regarding the bunching intervals:

Lemma 3. *Suppose bunching occurs over $[\theta_1, \theta_2] \subseteq [\underline{\theta}, \overline{\theta})$ in our model. Then either $\theta^*_{\text{MR}} > \theta_2$, or bunching occurs over $[\theta_1^{\text{MR}}, \theta_2^{\text{MR}}] \subseteq [\underline{\theta}, \overline{\theta})$ in Mussa and Rosen, where $\theta_1^{\text{MR}} \le \theta_1 < \theta_2 < \theta_2^{\text{MR}}$.*

If market exclusion is regarded as a special form of bunching (at $\overline{q} = 0$), then Lemma 3 simply says that any bunching interval in our model is contained in a bunching interval in Mussa and Rosen. This should make sense. If bunching occurs over an interval, say, $[\theta_1, \theta_2]$ in our model, $\theta - (1-b)\xi(\theta)$ must be decreasing over some subinterval, say, $[\theta_1', \theta_2'] \subseteq [\theta_1, \theta_2]$. This implies that $\theta - \xi(\theta)$ must be decreasing over some interval $[\theta_1^{\text{MR}\prime}, \theta_2^{\text{MR}\prime}] \supset [\theta_1', \theta_2']$. This in turn suggests that the bunching interval in Mussa and Rosen should be larger than $[\theta_1, \theta_2]$. This intuition is made precise in the proof in the Appendix.

We are now ready to compare the monopoly solutions in our model with those in Mussa and Rosen.

Proposition 2. *Compared to the Mussa–Rosen benchmark, both quality distortion and market exclusion are smaller with consumer entry, i.e., $q^*(\theta) \ge q^*_{\text{MR}}(\theta)$ (with equality only at $\overline{\theta}$) and $\theta^* \le \theta^*_{\text{MR}}$ (with equality only when $\theta^* = \theta^*_{\text{MR}} = \underline{\theta}$).*

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 $\mathrm{MR}(\theta)$ 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 and Rosen, she is also covered in our model. Given this, the comparison stated in Lemma 3 can be strengthened as follows.

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

Corollary 1 suggests that in our setting the optimal monopolistic solution may even be first best. So as 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

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

Let $E\pi(q)$ be the firm's expected profit from selling one unit of product of quality q. Then

$$E\pi(q)=E\big[\theta q(\theta)-C(q(\theta))-u(\theta)\big]=Ev(q,\theta)-u(\underline{\theta}).$$

When the first-best quality provision is offered (coupled with $u(\underline{\theta})=0$), we let $Ev(q^{\mathrm{fb}})$ and $c^{*\mathrm{fb}}$ 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:*

$$Ev(q^{\mathrm{fb}})\geq\eta(c^{*\mathrm{fb}}). \tag{9}$$

If condition (9) fails, the monopolistic nonlinear pricing contract involves downward quality distortion for all but the highest type.

In light of Lemma 2, that the monopolist cannot charge an entry fee should be regarded as a constraint in the monopolist's maximization program. Proposition 4 suggests that this constraint is nonbinding if and only if condition (9) holds. It also suggests that the first best can be achieved as long as, under efficient pricing, a constant fixed fee p_0^{fb} exists, which induces optimal entry while maintaining the post-entry IR for all types of consumers served.

Note that the existence of such a fixed fee p_0^{fb} is equivalent to finding a common rent provision $u(\underline{\theta})$ in our direct mechanism analysis. More specifically, whether the first best can be achieved depends on whether we can adjust $u(\underline{\theta})$ alone to induce optimal entry under efficient nonlinear pricing, subject to the only constraint that $u(\underline{\theta})$ cannot be made negative (the post-entry IR constraint). To demonstrate such a condition, substituting $q(\theta) = q^{\text{fb}}(\theta)$ into (3), we have

$$\Pi = G\biggl(\int_{\underline{\theta}}^{\overline{\theta}} (u(\underline{\theta}) + q^{\text{fb}}(\theta)\xi(\theta))\,dF(\theta)\biggr) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \bigl[\theta q^{\text{fb}}(\theta) - C(q^{\text{fb}}(\theta)) - q^{\text{fb}}(\theta)\xi(\theta) - u(\underline{\theta})\bigr]\,dF(\theta).$$

Differentiating by $u(\underline{\theta})$ and then evaluating at $u(\underline{\theta}) = 0$ yields

$$\begin{aligned} \frac{d\Pi}{du(\underline{\theta})}\bigg|_{u(\underline{\theta})=0} &= G' \cdot \int_{\underline{\theta}}^{\overline{\theta}} \bigl[\theta q^{\text{fb}}(\theta) - C(q^{\text{fb}}(\theta)) - q^{\text{fb}}(\theta)\xi(\theta)\bigr]\,dF(\theta) - G \\ &= G' \cdot [Ev(q^{\text{fb}}) - \eta(c^{*\text{fb}})]. \end{aligned}$$

When (9) fails, $d\Pi/du(\underline{\theta})|_{u(\underline{\theta})=0} < 0$. This suggests that the first-best solution is not feasible, as the "optimal" common rent provision $u(\underline{\theta})$ would have to be strictly negative, which violates the IR constraint for consumers in the neighborhood of type $\underline{\theta}$ after entry. In this case the optimal nonlinear pricing contract has to involve quality distortion (along with $u(\underline{\theta}) = 0$). When (9) holds, however, the first-best solution is achieved ($q(\cdot) = q^{\text{fb}}(\cdot)$), and the optimal $u(\underline{\theta})$ (and hence c^*) is chosen such that

$$\int_{\underline{\theta}}^{\overline{\theta}} \bigl[\theta q^{\text{fb}}(\theta) - C(q^{\text{fb}}(\theta)) - u^{\text{fb}}(\theta)\bigr]\,dF(\theta) = \eta(c^*),$$

where $c^* = u(\underline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\theta} q^{\text{fb}}(\tau)\,d\tau\,dF(\theta)$.

It is also shown in the proof of Proposition 4 that $\underline{\theta} \geq C'(0)$ (full market coverage) is a necessary condition for the first-best solution. This is intuitive: only when the market is fully covered can the common rent provision $u(\underline{\theta})$ be strictly greater than zero, which can then possibly be adjusted to induce optimal entry (while fixing first-best quality provision and maintaining post-entry IR).

To further understand condition (9), we consider the example where θ is distributed uniformly and the firm's production cost is given by the quadratic form $C(q) = q^2/2$. Let $\Delta = (\overline{\theta} - \underline{\theta})$ be the range of the support. Then η^{-1} is well defined given Assumption 1.

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

Corollary 2 can be interpreted rather intuitively. If the range of support Δ is not too large, 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 isΔ, the larger is the rent required for the consumers.

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

- 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 is the relative measure of consumers' vertical type heterogeneity, the more likely it is that the first-best quality provision will be offered or the more likely it is 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* $q^{\text{fb}} \geq q_1^* \geq q_2^* \geq q_{\text{MR}}^*$ *and* $\theta_{\text{MR}}^* \geq \theta_2^* \geq \theta_1^* \geq \theta^{*\text{fb}}$.

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 (9) provides an upper bound of η for the first-best solution to emerge. On the other extreme, in Mussa and Rosen, $G = 1$; hence $\eta = +\infty$, and condition (9) never holds. This explains why the first best is never optimal in the Mussa–Rosen benchmark. For the case in between, the higher is η, the greater is the quality provision distortion from the efficient provision level. In a sense, η also measures the price elasticity of entry: the higher is η, the lower is 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.

4. Discussion

Our preceding analysis reveals that the first-best solution arises when either condition (9) holds or when the firm can charge entry fees. This suggests that production inefficiency does not have to be associated with the existence of a monopoly, which may appear to be inconsistent with the basic wisdom from a microeconomics textbook.

However, note that even when production inefficiency is absent in our setting, distortion will arise in the form of inefficient entry, as we can show that the monopoly always induces insufficient entry compared to socially optimal entry, in which the social planner maximizes the expected total social surplus (the expected total surplus generated from the sale less the expected total entry cost). To see this, note that the socially optimal outcome can be achieved by allowing consumers to produce the goods themselves at cost $C(q)$. As a result, all the surplus goes to the consumers and the quality must be provided at the first-best level. 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 thus socially efficient as (i) there is no production inefficiency due to the first-best quality provision, and (ii) a potential buyer enters if and only if the expected profit, and hence her contribution to the social surplus, is greater than zero. It is then straightforward to see that the monopoly induces insufficient entry compared to socially optimal entry.

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 antitrust experts in nonlinear pricing settings with consumer entry.

Another key assumption made in our analysis is 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 not be reasonable in others. If the buyers can make purchase decisions based on prior beliefs of their preference types (e.g., behave like the mean type), the analysis is more complicated. For ease of exposition we relegate such an analysis to the end of Appendix A. A main finding is that quality distortion can now be smaller or larger than the Mussa–Rosen benchmark. More specifically, we show that under environments where learning is "beneficial" to the firm (in the sense that an individual consumer contributes more expected profit to the firm when she learns than when she does not learn in equilibrium), then the firm will reduce quality distortion for high types (the types above the mean type) but *increase* quality distortion for low types (the types below the mean type). This should make sense, as by doing so, the equilibrium rent provision $u(\theta)$ becomes more convex (recall that $u'(\theta) = q(\theta)$), and a more convex rent provision implies stronger (ex ante) incentives for the consumers to learn (due to Jensen's inequality). In environments where learning is not beneficial to the firm (in the sense that an individual consumer contributes less expected profit to the firm when she learns than when she does not learn in equilibrium), however, the firm should discourage learning. In this case the firm makes quality distortion *larger*, rather than *smaller*, for the high types; for types below the mean type, however, the implication for quality distortion is ambiguous. We are unable to obtain more precise analytical results,[12] but the main message should be clear. That is, by allowing those who do not learn their preference types to make purchases, the optimal nonlinear pricing contract becomes more subtle: whenever consumer learning is more desirable to the firm, the firm should make the equilibrium rent provision more

[12]Given the challenge of the new analysis, we can only focus on the analysis of perfect sorting solutions; moreover, we are unable to identify the exact conditions (environments) under which the optimal solution is perfect sorting.

convex (so as to induce more learning), and vice versa. This is different from the finding in Section 3, as we demonstrate that quality distortion is reduced for *all* (but the highest) types. Nevertheless, our findings can be reconciled: learning is modeled as entry in our preceding analysis, so learning has to be beneficial for the firm (in the sense described above), and hence quality distortion should be reduced for all (but the highest) types.[13]

Our results may not be entirely robust to allowing consumers to buy without learning their preference types. Nevertheless we model learning/information acquisition as entry in our preceding analysis mainly for tractability of analysis. Besides, if all consumers, whether informed or not, will enter the final sale, then there will be no effect of nonlinear pricing on consumer entry, which eliminates a major motivation of this current research.

5. Conclusion

Our paper contributes to the literature by incorporating costly consumer entry into a canonical model of monopolistic nonlinear pricing. 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 emerges as the monopoly solution in our model. 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 pattern 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.

Besides those discussed in the previous section, our analysis also relies on other assumptions. First, we assume that the entry cost c and the preference type 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 lead to new insights. Second, our current analysis is also 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. Given the challenges in extending our current analysis in these directions, they are left for future research.

[13] Our analysis in Section 3 can be treated as the case where the consumers who do not learn behave like the lowest type, $\underline{\theta}$, instead of the mean type, $\widehat{\theta} = E\theta$.

Appendix A: Proofs

Proof of Proposition 1. Given the Hamiltonian

$$\mathcal{H} = \big[G'(w(\theta))u(\theta)\cdot v(\theta) + G(w(\theta))\cdot \pi(\theta, u(\theta), q(\theta))\big]f(\theta) + \mu(\theta)a(\theta) \\ + \lambda_1(\theta)q(\theta) + \lambda_2(\theta)u(\theta)f(\theta) + \lambda_3(\theta)\pi(\theta, u(\theta), q(\theta))f(\theta),$$

we can derive the co-state equations

$$\begin{aligned} -\mu'(\theta) &= \frac{\partial \mathcal{H}}{\partial q(\theta)} \\ &= G(w(\theta))\cdot \pi_q(\theta, u(\theta), q(\theta))f(\theta) + \lambda_1(\theta) + \lambda_3(\theta)\cdot \pi_q(\theta, u(\theta), q(\theta))f(\theta) \end{aligned} \tag{10}$$

$$\begin{aligned} -\lambda_1'(\theta) &= \frac{\partial \mathcal{H}}{\partial u(\theta)} \\ &= \big[G'(w(\theta))\cdot v(\theta) + G(w(\theta))\cdot \pi_u(\theta, u(\theta), q(\theta))\big]f(\theta) \\ &\quad + \lambda_2(\theta)f(\theta) + \lambda_3(\theta)\pi_u(\theta, u(\theta), q(\theta))f(\theta) \end{aligned} \tag{11}$$

$$-\lambda_2'(\theta) = \frac{\partial \mathcal{H}}{\partial w(\theta)} = \big[G''(w(\theta))u(\theta)\cdot v(\theta) + G'(w(\theta))\cdot \pi(\theta, u(\theta), q(\theta))\big]f(\theta) \tag{12}$$

$$-\lambda_3'(\theta) = \frac{\partial \mathcal{H}}{\partial v(\theta)} = G'(w(\theta))u(\theta)f(\theta). \tag{13}$$

Solving (10)–(13) combined with the transversality conditions $\lambda_1(\overline{\theta}) = \lambda_2(\overline{\theta}) = \lambda_3(\overline{\theta}) = 0$, we have

$$\lambda_3(\theta) = G(w(\overline{\theta})) - G(w(\theta)) \tag{14}$$

$$\lambda_2(\theta) = G'(w(\overline{\theta}))v(\overline{\theta}) - G(w(\theta))v(\theta) \tag{15}$$

$$\lambda_1(\theta) = \big[G'(w(\overline{\theta}))v(\overline{\theta}) - G(w(\overline{\theta}))\big](1 - F(\theta)). \tag{16}$$

Substituting (14)–(16) into (10), we have

$$\begin{aligned} \mu'(\theta) &= -G(w(\overline{\theta}))\cdot \pi_q(\theta, u(\theta), q(\theta))f(\theta) - \big[G'(w(\overline{\theta}))v(\overline{\theta}) - G(w(\overline{\theta}))\big](1 - F(\theta)) \\ &= -G\bigg(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta)\bigg)f(\theta)\big[(\theta - (1-b)\xi(\theta)) - C'(q(\theta))\big], \end{aligned} \tag{17}$$

where b is given by (5).

Integrating (17) with the transversality conditions $\mu(\overline{\theta}) = 0$, we have

$$\mu(\theta) = G\bigg(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta)\bigg)\int_{\theta}^{\overline{\theta}} \big[(\theta - (1-b)\xi(\theta)) - C'(q(\theta))\big]\,dF(\theta).$$

Finally, the optimality condition requires that $a^*(\theta)$ maximize $\mathcal{H}$ subject to $a(\theta) \geq 0$. This implies $\mu(\theta) \leq 0$, or

$$G\bigg(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta)\bigg)\int_{\theta}^{\overline{\theta}} \big[(\theta - (1-b)\xi(\theta)) - C'(q(\theta))\big]\,dF(\theta) \leq 0.$$

Whenever $\mu(\theta) < 0$, we must have $a^*(\theta) = dq^*(\theta)/d\theta = 0$. We thus have the complementary slackness condition

$$\frac{dq^*(\theta)}{d\theta} \cdot \int_{\theta}^{\overline{\theta}} \left[(\theta - (1-b)\xi(\theta)) - C'(q(\theta))\right] dF(\theta) = 0 \tag{18}$$

for all $\theta \in [\underline{\theta}, \overline{\theta}]$.

It follows from this condition that if $q^*(\theta)$ is strictly increasing over some interval, then it must coincide with $q^s(\theta)$: if $a^*(\theta) = dq^*(\theta)/d\theta > 0$, by (18) we must have

$$\int_{\theta}^{\overline{\theta}} \left[(\theta - (1-b)\xi(\theta)) - C'(q(\theta))\right] dF(\theta) = 0$$

for all $\theta \in [\underline{\theta}, \overline{\theta}]$, which implies that

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

This is precisely the condition that defines $q^s(\theta)$.

We now turn to the bunching analysis. First, we will demonstrate that bunching never occurs at the top (in the left neighborhood of $\overline{\theta}$). Suppose the negation: bunching occurs at $q^*(\theta) = \overline{q} > 0$ over some interval $[\theta_1, \overline{\theta}]$. First suppose $\overline{q} < q^s(\overline{\theta})$. Then for θ to be in the neighborhood of $\overline{\theta}$, we have $C'(\overline{q}) < C'(q^s(\overline{\theta})) = \mathrm{MR}(\overline{\theta})$. So in this neighborhood, we must have $\mu(\theta) > 0$, a contradiction. Now suppose $\overline{q} \geq q^s(\overline{\theta}) = C'^{-1}(\overline{\theta})$. Then $\theta_1 = \underline{\theta}$, as there exists no $\theta < \overline{\theta}$ for which $q^s(\theta) \geq q^s(\overline{\theta})$. But then we have $C'(\overline{q}) \geq C'(q^s(\overline{\theta})) = \overline{\theta} > \mathrm{MR}(\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. This implies $\mu(\theta) > 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, implying that the monopolist can increase his profit by reducing the quality provision, a contradiction. So the solution must exhibit perfect sorting at the top; hence $q^*(\overline{\theta}) = C'^{-1}(\overline{\theta})$.[14]

So in our model, bunching can only occur over subintervals in $(\underline{\theta}, \overline{\theta})$ (interior bunching), or in a neighborhood of $\underline{\theta}$ (bottom bunching).

It therefore only remains to determine the intervals over which $q^*(\theta)$ is constant (bunching). We first consider an interior bunching interval $[\theta_1, \theta_2] \subset (\underline{\theta}, \overline{\theta})$ as illustrated by Figure 1. To the left of θ_1 and to the right of θ_2, we have

$$\mu(\theta) = 0 \quad \text{and} \quad a^*(\theta) = \frac{dq^*(\theta)}{d\theta} = \frac{dq^s(\theta)}{d\theta} > 0.$$

By the continuity of $q^*(\theta)$, we have $q^s(\theta_1) = q^s(\theta_2)$, or (7). For any θ lying between $\theta_1$and θ_2, we have

$$\mu(\theta) < 0 \quad \text{and} \quad a^*(\theta) = 0.$$

By the continuity of $\mu(\theta)$, we have $\mu(\theta_1) = \mu(\theta_2) = 0$, which implies (8). So the bunching interval endpoints θ_1 and θ_2 are determined by solving (7) and (8). Once θ_1 and θ_2 are determined, the bunching quality is determined by $\overline{q} = q^s(\theta_1) = q^s(\theta_2)$.

For a bottom bunching interval, say, $[\underline{\theta}, \theta_2] \subseteq [\underline{\theta}, \overline{\theta})$. The right bunching endpoint θ_2 is determined by (8) with θ_1 being replaced by $\underline{\theta}$, and the bunching quality is determined by $\overline{q} = q^s(\theta_2)$.

[14] Thus efficiency at the top is a robust finding beyond the Mussa–Rosen benchmark.

Finally we show that $0 \le b \le 1$. That $b \ge 0$ is obvious as $\pi(\theta, u^*(\theta), q^*(\theta))$ must be nonnegative at the optimum. So it remains to show $b \le 1$. Since $u(\underline{\theta}) \ge 0$, $\lambda_1(\underline{\theta}) \le 0$ by the transversality condition. From (16), we have

$$\lambda_1(\underline{\theta}) = \left[G'(w(\overline{\theta}))v(\overline{\theta}) - G(w(\overline{\theta}))\right] = -G(w(\overline{\theta}))(1-b).$$

That $\lambda_1(\underline{\theta}) \le 0$ thus implies $b \le 1$. □

Proof of Lemma 3. When $b = 1$, 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 $[\theta_1, \theta_2] \subseteq [\underline{\theta}, \overline{\theta})$ in our model. Then there exists an interval $[\theta', \theta''] \subseteq [\theta_1, \theta_2]$ such that $q^{s\prime}(\theta) < 0$ for $\theta \in [\theta', \theta'']$, which in turn implies that $q^{s\prime}_{\mathrm{MR}}(\theta) < 0$ for $\theta \in [\theta', \theta'']$. So either $[\theta', \theta'']$ is not served or bunching also occurs in Mussa and Rosen over some interval, say, $[\theta_1^{\mathrm{MR}}, \theta_2^{\mathrm{MR}}] \subseteq [\underline{\theta}, \overline{\theta})$.

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

Define

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

Equations (7) and (8) can be rewritten as

$$n(\theta_1, b) - n(\theta_2, b) = 0 \tag{19}$$

$$\int_{\theta_1}^{\theta_2} F(\theta) \cdot n_\theta(\theta, b)\, d\theta = 0. \tag{20}$$

Differentiating (19) and (20) with respect to b, we have

$$n_\theta(\theta_1, b)\frac{d\theta_1}{db} + n_b(\theta_1, b) - n_\theta(\theta_2, b)\frac{d\theta_2}{db} - n_b(\theta_2, b) = 0$$

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

which can be written as

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

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

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

Note that

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

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

By Cramer's rule, we have

$$\begin{aligned}\frac{d\theta_1}{db} = \frac{|B|}{|A|} &= -\frac{\int_{\theta_1}^{\theta_2} [\xi(\theta_1) - \xi(\theta)] f(\theta)\, d\theta}{[1 - (1-b)\xi'(\theta_1)] \cdot (F(\theta_2) - F(\theta_1))} \\ &= -\frac{\int_{\theta_1}^{\theta_2} [\theta_1 - \theta] f(\theta)\, d\theta}{(1-b)[1 - (1-b)\xi'(\theta_1)] \cdot (F(\theta_2) - F(\theta_1))} > 0\end{aligned} \tag{21}$$

$$\begin{aligned}\frac{d\theta_2}{db} = \frac{|C|}{|A|} &= -\frac{\int_{\theta_1}^{\theta_2} [\xi(\theta_2) - \xi(\theta)] f(\theta)\, d\theta}{[1 - (1-b)\xi'(\theta_2)] \cdot (F(\theta_2) - F(\theta_1))} \\ &= -\frac{\int_{\theta_1}^{\theta_2} [\theta_2 - \theta] f(\theta)\, d\theta}{(1-b)[1 - (1-b)\xi'(\theta_2)](F(\theta_2) - F(\theta_1))} < 0.\end{aligned} \tag{22}$$

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

Next we consider bunching at the bottom, say, over the interval $[\underline{\theta}, \theta_2]$. Substituting $\theta_1 = \underline{\theta}$ into (8) and manipulating, we have

$$\int_{\underline{\theta}}^{\theta_2} F(\theta) \cdot n_\theta(\theta, b)\, d\theta = 0. \tag{23}$$

Differentiating (23) with respect to b, we have

$$\frac{d\theta_2}{db} = -\frac{\int_{\underline{\theta}}^{\theta_2}[\xi(\theta_2)-\xi(\theta)]\,dF(\theta)}{[1-(1-b)\xi'(\theta_2)]F(\theta_2)} = -\frac{\int_{\underline{\theta}}^{\theta_2}[\theta_2-\theta]\,dF(\theta)}{(1-b)[1-(1-b)\xi'(\theta_2)]F(\theta_2)} < 0. \tag{24}$$

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

The above analysis assumes that $q^s_{\mathrm{MR}}(\theta_2^{\mathrm{MR}}) > 0$ (so bunching occurs at a positive quality level). If $q^s_{\mathrm{MR}}(\theta_2^{\mathrm{MR}}) \le 0$, there exists $\theta^*_{\mathrm{MR}} \ge \theta_2^{\mathrm{MR}} > \theta_2$, such that $q^s_{\mathrm{MR}}(\theta^*_{\mathrm{MR}}) = 0 = q^*_{\mathrm{MR}}(\theta^*_{\mathrm{MR}})$, i.e., all types below θ^*_{MR} ($> \theta_2$) are excluded from the market.

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

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 $\overline{\theta}$, q^* and q^*_{MR} must be perfect sorting. Let $[\theta_1, \overline{\theta}]$ be the longest interval in this neighborhood over which q^*_{MR} is sorting. Schedules q^* must also be sorting over $[\theta_1, \overline{\theta}]$ (otherwise contradicting Lemma 3). So by (6), we have $q^*(\theta) > q^*_{\mathrm{MR}}(\theta)$ over $[\theta_1, \overline{\theta})$. If $\theta^*_{\mathrm{MR}} = \theta_1$, we must have $\theta^* \le \theta^*_{\mathrm{MR}}$ (with equality only at $\theta^* = \theta^*_{\mathrm{MR}} = \underline{\theta}$), and we are done with the proof.

If $\theta^*_{\mathrm{MR}} < \theta_1$, 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 3, 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 $q^*(\theta) \ge q^*(\theta_2) > q^s_{\mathrm{MR}}(\theta_2) = q^*_{\mathrm{MR}}(\theta)$ for $\theta \in [\theta_2, \theta_1]$. If $\theta^*_{\mathrm{MR}} = \theta_2$, we must have $\theta^* \le \theta^*_{\mathrm{MR}}$ (with equality only at $\theta^* = \theta^*_{\mathrm{MR}} = \underline{\theta}$), and we are done with the proof.

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

Proof of Proposition 4. By Lemma 2, the monopolist who cannot charge an entry fee can implement the first-best qualify provision if and only if there exists a fixed fee $p_0 = V - P^* = G(P^*)/G'(P^*)$ that satisfies the conditions in Lemma 2: $G(P^*)/G'(P^*) = 0$ if $\underline{\theta} < C'(0)$ and $G(P^*)/G'(P^*) \in (-\infty, \underline{\theta} \cdot C'^{-1}(\underline{\theta}) - C(C'^{-1}(\underline{\theta}))]$ if $\underline{\theta} \ge C'(0)$. Since $G(P^*)/G'(P^*) > 0$, the first best cannot be achieved if $\underline{\theta} < C'(0)$. So we can focus on the case $\underline{\theta} \ge C'(0)$. In this case,

$$\begin{aligned} Ev(q^{\mathrm{fb}}) &= E\pi(q^{\mathrm{fb}}) + u(\underline{\theta}) = p_0 + u(\underline{\theta}) \\ &= p(q^{\mathrm{fb}}(\underline{\theta})) - C(q^{\mathrm{fb}}(\underline{\theta})) + u(\underline{\theta}) \\ &= \underline{\theta} \cdot q^{\mathrm{fb}}(\underline{\theta}) - C(q^{\mathrm{fb}}(\underline{\theta})) \\ &= \underline{\theta} \cdot C'^{-1}(\underline{\theta}) - C(C'^{-1}(\underline{\theta})). \end{aligned}$$

This shows that the first-best solution is achieved if and only if $G(P^*)/G'(P^*) = \eta(c^{*\mathrm{fb}}) \le Ev(q^{\mathrm{fb}})$, which is condition (9). □

Proof of Corollary 2. It can be verified that with $C(q) = q^2/2$, $Ev(q^{\mathrm{fb}}) = \underline{\theta}^2/2$. Based on this and $F(\theta) = (\theta - \underline{\theta})/(\overline{\theta} - \underline{\theta})$, (9) becomes

$$\frac{\underline{\theta}^2}{2} \ge \eta\left(\frac{\overline{\theta} + 2\underline{\theta}}{3}(\overline{\theta} - \underline{\theta})\right) \quad \text{or}$$

$$\eta^{-1}\left(\frac{\underline{\theta}^2}{2}\right) \ge \frac{\overline{\theta} + 2\underline{\theta}}{3}(\overline{\theta} - \underline{\theta}),$$

which gives rise to $0 < \Delta \le (\sqrt{(3\underline{\theta})^2 + 24\eta^{-1}(\underline{\theta}^2/2)} - 3\underline{\theta})/2$. 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 (5);[15] given b, the schedule of q^* is also uniquely determined by (6) or (7) and (8) 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 (21)–(24) established in the proof of Lemma 3, we can conclude that whenever bunching occurs over $[\theta_{11}, \theta_{12}] \subseteq [\underline{\theta}, \overline{\theta})$ in q_1^*, then either $\theta_2^* > \theta_{12}$ or bunching occurs over $[\theta_{21}, \theta_{22}] \subseteq [\underline{\theta}, \overline{\theta})$ in q_2^*, where $\theta_{21} \le \theta_{11} < \theta_{12} < \theta_{22}$. 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 $q^{\mathrm{fb}} \ge q_1^* \ge q_2^* \ge q_{\mathrm{MR}}^*$ (with equality only at $\theta = \overline{\theta}$) and $\theta_{\mathrm{MR}}^* \ge \theta_2^* \ge \theta_1^* \ge \theta^{*\mathrm{fb}}$ (with equality only at $\underline{\theta}$). More generally, we have $\partial q^*(\theta)/\partial b \ge 0$ (with equality only at $\theta = \overline{\theta}$) for $b \in [0, 1)$.

Proposition 4 has shown that when $\eta(c^{*\mathrm{fb}}) \le Ev(q^{\mathrm{fb}})$, $b = 1$ and $q^* = q^{\mathrm{fb}}$. We thus focus on the case when $\eta(c^{*\mathrm{fb}}) > Ev(q^{\mathrm{fb}})$ (and hence $b \in [0, 1)$). In this case $u^*(\underline{\theta}) = 0$ and hence $u^*(\theta)$ is uniquely determined by $q^*(\theta)$. We thus have

$$\int_{\underline{\theta}}^{\overline{\theta}} u^*(\theta)\, dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} [1 - F(\theta)] q^*(\theta)\, d\theta$$

$$\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u^*(\theta), q^*(\theta))\, dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta q^*(\theta) - C(q^*(\theta)) - \xi(\theta) q^*(\theta)\right] dF(\theta).$$

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

Note that

$$\frac{\partial}{\partial b}\int_{\underline{\theta}}^{\overline{\theta}}[1-F(\theta)]q^*(\theta)\,d\theta=\int_{\underline{\theta}}^{\overline{\theta}}[1-F(\theta)]\frac{\partial q^*(\theta)}{\partial b}\,d\theta>0 \tag{25}$$

and

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

We consider the following cases:

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

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

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

$$\begin{aligned}\int_{\theta_1}^{\theta_2}&\big[\theta-\xi(\theta)-C'(q^*(\theta))\big]\frac{\partial q^*(\theta)}{\partial b}\,dF(\theta)\\ &=\frac{\partial q^*(\theta)}{\partial b}\int_{\theta_1}^{\theta_2}\big[\theta-\xi(\theta)-C'(q^*(\theta))\big]\,dF(\theta)\\ &=\frac{\partial q^*(\theta)}{\partial b}\int_{\theta_1}^{\theta_2}\big[\theta-\xi(\theta)-(\theta-(1-b)\xi(\theta))\big]\,dF(\theta)\quad\text{(by (8))}\\ &=-b\int_{\theta_1}^{\theta_2}[1-F(\theta)]\frac{\partial q^*(\theta)}{\partial b}\,d\theta.\end{aligned}$$

Combining the two cases above, we have

$$\frac{\partial}{\partial b}\int_{\underline{\theta}}^{\overline{\theta}}H(\theta,u^*(\theta),q^*(\theta))\,dF(\theta)=-b\int_{\underline{\theta}}^{\overline{\theta}}[1-F(\theta)]\frac{\partial q^*(\theta)}{\partial b}\,d\theta<0. \tag{26}$$

We are now ready to show that if $\eta_1\le\eta_2$, then $0\le b_2\le b_1\le 1$.
Suppose in negation, we have $b_1<b_2$. By (25) and (26), we have

$$\int_{\underline{\theta}}^{\overline{\theta}}\pi(\theta,u_1^*(\theta),q_1^*(\theta))\,dF(\theta)>\int_{\underline{\theta}}^{\overline{\theta}}\pi(\theta,u_2^*(\theta),q_2^*(\theta))\,dF(\theta)$$

and

$$\int_{\underline{\theta}}^{\overline{\theta}}u_1^*(\theta)\,dF(\theta)<\int_{\underline{\theta}}^{\overline{\theta}}u_2^*(\theta)\,dF(\theta).$$

We thus have

$$\eta_1\left(\int_{\underline{\theta}}^{\overline{\theta}} u_1^*(\theta)\,dF(\theta)\right) < \eta_1\left(\int_{\underline{\theta}}^{\overline{\theta}} u_2^*(\theta)\,dF(\theta)\right) \le \eta_2\left(\int_{\underline{\theta}}^{\overline{\theta}} u_2^*(\theta)\,dF(\theta)\right).$$

But then we have

$$1 > \frac{b_1}{b_2} = \frac{\eta_2(\int_{\underline{\theta}}^{\overline{\theta}} u_2^*(\theta)\,dF(\theta))}{\eta_1(\int_{\underline{\theta}}^{\overline{\theta}} u_1^*(\theta)\,dF(\theta))}\frac{\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u_1^*(\theta), q_1^*(\theta))\,dF(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u_2^*(\theta), q_2^*(\theta))\,dF(\theta)} > 1,$$

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

Analysis when consumers can make purchases without learning their preference types

Let $\widehat{\theta} = E\theta = \int_{\underline{\theta}}^{\overline{\theta}} \theta\,dF(\theta)$ be the mean type. Consumers will earn an expected profit $u(\widehat{\theta})$ without learning their preference types in equilibrium. Given the rent provision $u(\theta)$, it is easily seen that consumers with entry cost $c < \widehat{c}$ will incur the costs to learn their preference types while those with entry cost $c > \widehat{c}$ will skip the learning and behave like a mean type, where $\widehat{c} = Eu(\theta) - u(\widehat{\theta})$.

For ease of analysis we focus on the perfect sorting equilibrium so that we can ignore the monotonicity constraint $q'(\theta) \ge 0$. The firm's problem can be formulated as

$$\begin{aligned}\max_{u(\cdot)} G\left(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta) - u(\widehat{\theta})\right) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), q(\theta))\,dF(\theta) \\ + \left(1 - G\left(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta) - u(\widehat{\theta})\right)\right) \cdot \pi(\widehat{\theta}, u(\widehat{\theta}), q(\widehat{\theta}))\end{aligned}$$

s.t. $u(\underline{\theta}) \ge 0$, $q(\theta) = u'(\theta)$.

Since

$$\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} [u(\underline{\theta}) + \xi(\theta)q(\theta)]\,dF(\theta)$$

$$u(\widehat{\theta}) = u(\underline{\theta}) + \int_{\underline{\theta}}^{\widehat{\theta}} \frac{1}{f(\theta)} q(\theta)\,dF(\theta),$$

we have

$$G\left(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\,dF(\theta) - u(\widehat{\theta})\right) = G\left(\int_{\underline{\theta}}^{\widehat{\theta}} \frac{-F(\theta)}{f(\theta)} q(\theta)\,dF(\theta) + \int_{\widehat{\theta}}^{\overline{\theta}} \xi(\theta)q(\theta)\,dF(\theta)\right). \tag{27}$$

Also note that

$$\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), q(\theta))\,dF(\theta) = \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta q(\theta) - C(q(\theta)) - \xi(\theta)q(\theta) - u(\underline{\theta})\right] dF(\theta) \tag{28}$$

$$\pi(\widehat{\theta}, u(\widehat{\theta}), q(\widehat{\theta})) = \widehat{\theta}q(\widehat{\theta}) - C(q(\widehat{\theta})) - u(\underline{\theta}) - \int_{\underline{\theta}}^{\widehat{\theta}} \frac{1}{f(\theta)} q(\theta)\, dF(\theta). \tag{29}$$

Using (27)–(29), the firm's problem can be reformulated as

$$\begin{aligned}\max_{u(\underline{\theta}), q(\cdot)} \Bigg[& G\bigg(\int_{\underline{\theta}}^{\widehat{\theta}} \frac{-F(\theta)}{f(\theta)} q(\theta)\, dF(\theta) + \int_{\widehat{\theta}}^{\overline{\theta}} \xi(\theta) q(\theta)\, dF(\theta)\bigg) \\ & \cdot \int_{\underline{\theta}}^{\overline{\theta}} \big[\theta q(\theta) - C(q(\theta)) - \xi(\theta) q(\theta) - u(\underline{\theta})\big]\, dF(\theta) \\ & + \bigg(1 - G\bigg(\int_{\underline{\theta}}^{\widehat{\theta}} \frac{-F(\theta)}{f(\theta)} q(\theta)\, dF(\theta) + \int_{\widehat{\theta}}^{\overline{\theta}} \xi(\theta) q(\theta)\, dF(\theta)\bigg)\bigg) \\ & \cdot \bigg(\widehat{\theta} q(\widehat{\theta}) - C(q(\widehat{\theta})) - \int_{\underline{\theta}}^{\widehat{\theta}} \frac{1}{f(\theta)} q(\theta)\, dF(\theta) - u(\underline{\theta})\bigg)\Bigg]\end{aligned}$$

s.t. $u(\underline{\theta}) \geq 0$, $q(\theta) \geq 0$.

Differentiating the objective function, L, with respect to $u(\underline{\theta})$, we have

$$\frac{\partial L}{\partial u(\underline{\theta})} = -G(\cdot) - [1 - G(\cdot)] = -1 < 0,$$

where

$$G(\cdot) = G\bigg(\int_{\underline{\theta}}^{\widehat{\theta}} \frac{-F(\theta)}{f(\theta)} q(\theta)\, dF(\theta) + \int_{\widehat{\theta}}^{\overline{\theta}} \xi(\theta) q(\theta)\, dF(\theta)\bigg).$$

Thus we have $u(\underline{\theta}) = 0$. Next we consider the following cases in order:

(i) We have $\theta \in [\underline{\theta}, \widehat{\theta})$:

$$\begin{aligned}\frac{\partial L}{\partial q(\theta)} \frac{1}{f(\theta)} = {} & G'(\cdot) \frac{-F(\theta)}{f(\theta)} \cdot \int_{\underline{\theta}}^{\overline{\theta}} \big[\theta q(\theta) - C(q(\theta)) - \xi(\theta) q(\theta) - u(\underline{\theta})\big]\, dF(\theta) \\ & + G(\cdot)\big[\theta - C'(q(\theta)) - \xi(\theta)\big] \\ & - G'(\cdot) \frac{-F(\theta)}{f(\theta)} \cdot \bigg(\widehat{\theta} q(\widehat{\theta}) - C(q(\widehat{\theta})) - \int_{\underline{\theta}}^{\widehat{\theta}} \frac{1}{f(\theta)} q(\theta)\, dF(\theta) - u(\underline{\theta})\bigg) \\ & + (1 - G(\cdot)) \frac{-1}{f(\theta)} \\ = {} & 0.\end{aligned}$$

Simplifying, we have

$$C'(q(\theta)) = \theta - \xi(\theta) - \widetilde{b} \frac{F(\theta)}{f(\theta)} - \frac{1 - G(\cdot)}{G(\cdot)} \frac{1}{f(\theta)},$$

where

$$
\begin{aligned}
\widetilde{b} = \frac{G'(\cdot)}{G(\cdot)} \bigg[\int_{\underline{\theta}}^{\overline{\theta}} \big[\theta q(\theta) - C(q(\theta)) - \xi(\theta) q(\theta)\big]\, dF(\theta) \\
- \bigg(\widehat{\theta} q(\widehat{\theta}) - C(q(\widehat{\theta})) - \int_{\underline{\theta}}^{\widehat{\theta}} \frac{1}{f(\theta)} q(\theta)\, dF(\theta) \bigg) \bigg] \qquad (30) \\
= \frac{G'(\cdot)}{G(\cdot)} \bigg(\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), q(\theta))\, dF(\theta) - H(\widehat{\theta}, u(\widehat{\theta}), q(\widehat{\theta})) \bigg).
\end{aligned}
$$

(ii) We have $\theta = \widehat{\theta}$:

$$
\begin{aligned}
\frac{\partial L}{\partial q(\theta)} \frac{1}{f(\widehat{\theta})} = {} & G'(\cdot)\xi(\widehat{\theta}) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \big[\theta q(\theta) - C(q(\theta)) - \xi(\theta) q(\theta) - u(\underline{\theta})\big]\, dF(\theta) \\
& + G(\cdot)\big[\widehat{\theta} - C'(q(\widehat{\theta})) - \xi(\widehat{\theta})\big] \\
& - G'(\cdot)\xi(\widehat{\theta}) \cdot \bigg(\widehat{\theta} q(\widehat{\theta}) - C(q(\widehat{\theta})) - \int_{\underline{\theta}}^{\widehat{\theta}} \frac{1}{f(\theta)} q(\theta)\, dF(\theta) - u(\underline{\theta}) \bigg) \\
& + (1 - G(\cdot))\big(\widehat{\theta} - C'(q(\widehat{\theta}))\big) \frac{1}{f(\widehat{\theta})} \\
= {} & 0.
\end{aligned}
$$

Simplifying, we have

$$
C'(q(\widehat{\theta})) = \widehat{\theta} - \frac{G(\cdot) f(\widehat{\theta})}{1 - G(\cdot)} (1 - \widetilde{b}) \xi(\widehat{\theta}),
$$

where $\widetilde{b}$ is given by (30).

(iii) We have $\theta \in (\widehat{\theta}, \overline{\theta}]$:

$$
\begin{aligned}
\frac{\partial L}{\partial q(\theta)} \frac{1}{f(\theta)} = {} & G'(\cdot)\xi(\theta) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \big[\theta q(\theta) - C(q(\theta)) - \xi(\theta) q(\theta) - u(\underline{\theta})\big]\, dF(\theta) \\
& + G(\cdot)\big[\theta - C'(q(\theta)) - \xi(\theta)\big] \\
& - G'(\cdot)\xi(\theta) \cdot \bigg(\widehat{\theta} q(\widehat{\theta}) - C(q(\widehat{\theta})) - \int_{\underline{\theta}}^{\widehat{\theta}} \frac{1}{f(\theta)} q(\theta)\, dF(\theta) - u(\underline{\theta}) \bigg) \\
= {} & 0.
\end{aligned}
$$

Simplifying, we have

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

where $\widetilde{b}$ is given by (30).

To sum up, we have

$$C'(q(\theta)) = \begin{cases} \theta - (1-\widetilde{b})\xi(\theta) - (\frac{1-G(\cdot)}{G(\cdot)} + \widetilde{b})\frac{1}{f(\theta)} & \text{for } \theta \in [\underline{\theta}, \widehat{\theta}) \\ \widehat{\theta} - \frac{G(\cdot)f(\widehat{\theta})}{1-G(\cdot)}(1-\widetilde{b})\xi(\widehat{\theta}) & \text{for } \theta = \widehat{\theta} \\ \theta - (1-\widetilde{b})\xi(\theta) & \text{for } \theta \in (\widehat{\theta}, \overline{\theta}], \end{cases}$$

where $\widetilde{b}$ is given by (30).

For $\theta \in (\widehat{\theta}, \overline{\theta}]$, $C'(q(\theta)) = \theta - (1-\widetilde{b})\xi(\theta) \geq (\leq)\, \theta - \xi(\theta) = C'_{\mathrm{MR}}(q(\theta))$ (equality holds only at $\overline{\theta}$) if $\widetilde{b} > (<) 0$, and for $\theta \in [\underline{\theta}, \widehat{\theta})$, $C'(q(\theta)) = \theta - \xi(\theta) - (1-G(\cdot))/(G(\cdot)f(\theta)) - \widetilde{b}F(\theta)/f(\theta) < \theta - \xi(\theta) = C'_{\mathrm{MR}}(q(\theta))$ if $\widetilde{b} > 0$ (and the comparison is ambiguous if $\widetilde{b} < 0$). So if $\widetilde{b} > 0$ (i.e., learning is beneficial to the firm, as an individual consumer contributes more expected profit to the firm when she learns than when she does not learn), then the firm will reduce quality distortion for high types ($\theta > \widehat{\theta}$) but *increase* quality distortion for low types ($\theta < \widehat{\theta}$). If $\widetilde{b} < 0$ (i.e., learning is not beneficial to the firm, as an individual consumer contributes less expected profit to the firm when she learns than when she does not learn), the firm makes quality distortion *larger*, rather than *smaller*, for the high types ($\theta > \widehat{\theta}$); for types below $\widehat{\theta}$, however, the comparison is ambiguous.

Appendix B: Sufficient conditions for optimality

We will show that the (strict) log concavity of G (Assumption 1) is sufficient to guarantee that the necessary conditions for optimality derived in the main text using the optimal control technique are also sufficient for optimization.

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

$$\begin{aligned} &\max_{u(\cdot)} \log\biggl(G\biggl(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\, dF(\theta)\biggr) \cdot \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), u'(\theta))\, dF(\theta)\biggr) \\ &\qquad = \log G\biggl(\int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\, dF(\theta)\biggr) + \log \int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta, u(\theta), u'(\theta))\, dF(\theta) \\ &\quad \text{s.t. } u(\theta) \geq 0, u'(\theta) \geq 0, u''(\theta) \geq 0 \end{aligned}$$

or

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

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

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

$$\begin{aligned}
&\log\biggl(G\biggl(\int_{\underline{\theta}}^{\overline{\theta}} \widehat{u}(\theta)\,dF(\theta)\biggr)\biggr)\\
&\qquad = \log\biggl(G\biggl(\alpha\int_{\underline{\theta}}^{\overline{\theta}} u_1(\theta)\,dF(\theta) + (1-\alpha)\int_{\underline{\theta}}^{\overline{\theta}} u_2(\theta)\,dF(\theta)\biggr)\biggr)\\
&\qquad > \alpha\log\biggl(G\biggl(\int_{\underline{\theta}}^{\overline{\theta}} u_1(\theta)\,dF(\theta)\biggr)\biggr) + (1-\alpha)\log\biggl(G\biggl(\int_{\underline{\theta}}^{\overline{\theta}} u_2(\theta)\,dF(\theta)\biggr)\biggr).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\log\biggl(\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta,\widehat{u}(\theta),\widehat{u}'(\theta))\,dF(\theta)\biggr)\\
&= \log\biggl(\int_{\underline{\theta}}^{\overline{\theta}} \bigl[\theta(\widehat{u}'(\theta)) - (\widehat{u}(\theta)) - C(\widehat{u}'(\theta))\bigr]\,dF(\theta)\biggr)\\
&> \log\biggl(\int_{\underline{\theta}}^{\overline{\theta}} \bigl[\alpha\bigl(\theta u_1'(\theta) - u_1(\theta) - C(u_1'(\theta))\bigr)\\
&\qquad\qquad + (1-\alpha)\bigl(\theta u_2'(\theta) - u_2(\theta) - C(u_2'(\theta))\bigr)\bigr]\,dF(\theta)\biggr)\\
&= \log\biggl(\alpha\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta,u_1(\theta),u_1'(\theta))\,dF(\theta) + (1-\alpha)\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta,u_2(\theta),u_2'(\theta))\,dF(\theta)\biggr)\\
&> \alpha\log\biggl(\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta,u_1(\theta),u_1'(\theta))\,dF(\theta)\biggr) + (1-\alpha)\log\biggl(\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta,u_2(\theta),u_2'(\theta))\,dF(\theta)\biggr).
\end{aligned}$$

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 $C^2[\underline{\theta},\overline{\theta}]$ and P is a convex cone in $C^2[\underline{\theta},\overline{\theta}]$. By Theorem 1 in Luenberger (1969, pp. 256–257) and Lemma 1 in Luenberger (1969, p. 227), the necessary conditions derived in the main text are also sufficient for optimization.

References

Armstrong, Mark and John Vickers (2001), "Competitive price discrimination." *Rand Journal of Economics*, 32, 579–605. [144, 145]

Armstrong, Mark and John Vickers (2010), "Competitive nonlinear pricing and bundling." *Review of Economic Studies*, 77, 30–60. [144]

Baron, David P. and Roger B. Myerson (1982), "Regulating a monopolist with unknown costs." *Econometrica*, 50, 911–930. [145]

Bergemann, Dirk and Juuso Välimäki (2006), "Information in mechanism design." In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Vol. 1 (Richard Blundell, Whitney K. Newey, and Torsten Persson, eds.), 186–221, Cambridge University Press, Cambridge. [145]

Bergemann, Dirk and Achim Wambach (2015), "Sequential information disclosure in auctions." *Journal of Economic Theory: Symposium Issue on Dynamic Contracts and Mechanism Design*, 159, 1074–1095. [146]

Champsaur, Paul and Jean-Charles Rochet (1989), "Multiproduct duopolists." *Econometrica*, 57, 533–557. [144]

Courty, Pascal and Li Hao (2000), "Sequential screening." *Review of Economic Studies*, 67, 697–717. [146]

Crémer, Jacques and Fahad Khalil (1992), "Gathering information before signing a contract." *American Economic Review*, 82, 566–578. [145]

Crémer, Jacques, Fahad Khalil, and Jean-Charles Rochet (1998a), "Contracts and productive information gathering." *Games and Economic Behavior*, 25, 174–193. [145]

Crémer, Jacques, Fahad Khalil, and Jean-Charles Rochet (1998b), "Strategic information gathering before a contract is offered." *Journal of Economic Theory*, 81, 163–200. [145]

Ellison, Glenn (2005), "A model of add-on pricing." *Quarterly Journal of Economics*, 120, 585–637. [144]

Engelbrecht-Wiggans, Richard (1993), "Optimal auctions revisited." *Games and Economic Behavior*, 5, 227–239. [145]

Esö, Péter and Balázs Szentes (2007), "Optimal information disclosure in auctions and the handicap auction." *Review of Economic Studies*, 74, 705–731. [146]

Gilbert, Richard J. and Carmen Matutes (1993), "Product line rivalry with brand differentiation." *Journal of Industrial Economics*, 41, 223–240. [144]

Kamien, Morton L. and Nancy L. Schwartz (2012), *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, second edition. Dover Publications, Inc., Mineola, New York. [150]

Levin, Dan and James L. Smith (1994), "Equilibrium in auctions with entry." *American Economic Review*, 84, 585–599. [145]

Lu, Jingfeng (2010), "Entry coordination and auction design with private costs of information acquisition." *Economic Inquiry*, 48, 274–289. [145]

Luenberger, David G. (1969), *Optimization by Vector Space Methods*, Wiley Professional Paperback Series. John Wiley & Sons, New York. [171]

Maskin, Eric S. and John Riley (1984), "Monopoly with incomplete information." *Rand Journal of Economics*, 15, 171–196. [141, 150, 152]

McAfee, R. Preston and John McMillan (1987), "Auctions with entry." *Economic Letters*, 23, 343–347. [145]

Moreno, Diego and John Wooders (2011), "Auctions with heterogeneous entry costs." *Rand Journal of Economics*, 42, 313–336. [145]

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317. [141, 142, 143]

Myerson, Roger B. (1981), "Optimal auction design." *Mathematics of Operations Research*, 6, 58–73. [143, 150, 152, 155]

Pavan, Alessandro, Ilya Segal, and Juuso Toikka (2014), "Dynamic mechanism design: A myersonian approach." *Econometrica*, 82, 601–653. [146]

Pontryagin, Lev Semyonovich, Vladimir Grigorevich Boltyanskii, Revaz Valerianovich Gamkrelidze, and Evgenii Frolovich Mishchenko (1962), *The Mathematical Theory of Optimal Processes*. John Wiley & Sons, New York. [150]

Rochet, Jean-Charles and Lars A. Stole (2002), "Nonlinear pricing with random participation." *Review of Economic Studies*, 69, 277–311. [144]

Spulber, Daniel F. (1989), "Product variety and competitive discounts." *Journal of Economic Theory*, 48, 510–525. [144]

Stole, Lars A. (1995), "Nonlinear pricing and oligopoly." *Journal of Economics and Management Strategy*, 4, 529–562. [144]

Tan, Guofu (1992), "Entry and R&D in procurement contracting." *Journal of Economic Theory*, 58, 41–60. [145]

Verboven, Frank (1999), "Product line rivalry and market segmentation—with an application to automobile optional engine pricing." *Journal of Industrial Economics*, 47, 399–425. [144]

Villas-Boas, J. Miguel and Udo Schmidt-Mohr (1999), "Oligopoly with asymmetric information: Differentiation in credit markets." *Rand Journal of Economics*, 30, 375–396. [144]

Wilson, Robert B. (1993), *Nonlinear Pricing*. Oxford University Press, Oxford. [144]

Yang, Huanxing and Lixin Ye (2008), "Nonlinear pricing, market coverage, and competition." *Theoretical Economics*, 3, 123–153. [144, 145]

Co-editor Johannes Hörner handled this manuscript.

Manuscript received 21 August, 2014; final version accepted 27 December, 2015; available online 24 January, 2016.