Nonlinear Pricing by a Dominant Firm under Competition

Yong Chao Guofu Tan Adam Chi Leung Wong*

January 6, 2019

Abstract

We provide a rationale for nonlinear pricing under competition in the absence of private information: The dominant firm can use unchosen offers to constrain its rival's possible deviations and extract more surplus from the buyer. When the capacity of the rival firm is constrained, as compared to linear pricing, the dominant firm can use the nonlinear pricing to partially foreclose the rival and harm the buyer. By establishing an equivalence between the subgame perfect equilibrium of our asymmetric competition game and an optimal mechanism in a "virtual" principal-agent model, we characterize the optimal nonlinear pricing.

Keywords: Nonlinear Pricing, Capacity Constraint, and Partial Foreclosure.

JEL Code: L13, L42, K21

*Chao: College of Business, University of Louisville (e-mail: yong.chao@louisville.edu); Tan: Department of Economics, University of Southern California (e-mail: guofutan@usc.edu); Wong: Department of Economics, Lingnan University (e-mail: adamwong@ln.edu.hk). We gratefully acknowledge helpful comments from Sandro Brusco, Giacomo Calzolari, Jimmy Chan, Fei Li, Giuseppe Lopomo, Preston McAfee, Alessandro Pavan, Huseyin Yildirim, Jidong Zhou, and seminar participants at Caltech, SHUFE, UC Davis, UCSB, U of Hong Kong, USC, Zhejiang U, UBC, U of Washington, UC Irvine, National Cheng Kung U, Chinese U of Hong Kong, UIUC, U of Illinois, 2017 Triangle Microeconomics Conference, 2017 Workshop on Industrial Organization and Management Strategy, Fall 2017 Midwest Economic Theory Conference, 2018 NSF/NBER/CEME Decentralization Conference, 2018 CRESS-JUFE Conference, 2018 ICIIE Conference at NJU, and IO Workshop at U of St. Gallen.

1 Introduction

Nonlinear pricing (NLP)—–total price not necessarily proportional to the quantity purchased—–is ubiquitous in intermediate-goods markets. It takes the form of various rebates and discounts conditional on volumes purchased by a buyer. The impact of NLP on competition is a hotly debated antitrust topic, especially when the NLP is adopted by a dominant firm. Examples include three-part tariffs[1] employed by *Skelly Oil*, *Pacific*, *Microsoft*,[2] and all-units discounts[3] used by *Michelin*, *British Airways*, *Canada Pipe*, *Tomra*, *Post Danmark*, and *Tetra Pak*.[4]

The vast majority of antitrust cases involving NLP is Section 2/Abuse of Dominance cases. By the very nature of abuse of dominance, a key feature shared by those cases is the *asymmetry*: the asymmetry in the size of the firms involved, and the asymmetry in terms of the complexity of pricing adopted by those asymmetric firms. Specifically, in those cases, there is a firm that is considered as "dominant"

[1] A three-part tariff is a pricing scheme consisting of a fixed fee, a free allowance of units up to which the marginal price is zero, and a positive per-unit price for additional demand beyond that allowance. For literature on three-part tariffs, see Grubb (2009, 2012, 2015), Chao (2013), and references therein.

[2] *Magnus Petroleum Co., Inc. v. Skelly Oil Co.*, 599 F.2d 196 (1979).

Barry Wright Corp. v Pacific Scientific Corp., 555 F.Supp.1264 (1983).

United States v. *Microsoft Corp.*, No. 94-1564 (D.D.C. filed July 15, 1995); Baseman, Warren-Boulton, and Woroch (1995).

[3] All-units discounts refer to a pricing practice that lowers the per-unit price on all units once the buyer's purchase crosses a volume threshold. For literature on all-units discounts, see Kolay, Shaffer, and Ordover (2004), Feess and Wohlschlegel (2010), O'Brien (2017), Chao, Tan, and Wong (2018a,b), and references therein.

[4] Case COMP/E-2/36.041/PO-Michelin, Commission Decision of 20 June 2001. Case T-203/01, *Manufacture Française des Pneumatiques Michelin v. Commission of the European Communities supported by Bandag Inc.*, Judgement of the Court of First Instance of 30 September 2003. See Motta (2009) for discussion of this case.

Case C-95/04, *British Airways plc v. Commission of the European Communities supported by Virgin Atlantic Airways Ltd.*, Judgment of the European Court of Justice, March 2007.

Canada (Commissioner of Competition) v Canada Pipe Co, 2006 FCA 233, (2007) 2 FCR.

EC, *Commission Decision of 29 March 2006 relating to proceedings under Article 82 [EC] and Article 54 of the EEA Agreement* (Case COMP/E-1/38.113 – Prokent-Tomra) [2006] OJ, C 734/07; *Tomra Systems and Others v Commission*, T-155/06 [2010] ECR II-4361; *Tomra Systems ASA and Others v Commission*, C-549/10 P, [2012] ECR I-0000.

Case C-23/14, *Post Danmark A/S v. Konkurrencerådet*, ECLI:EU:C:2015:651.

On November 16, 2016, the then State Administration of Industry and Commerce of China fined *Tetra Pak* for abusing its dominance between 2009 and 2013 in China's aseptic packaging market. One of the alleged abusive practices was to exclude and limit competition through its all-units discounts. See Chao and Tan (2017), and Fu and Tan (2018) for discussions on this case.

in market share, capacity, product lines, profits, and so on. And there is one or several smaller firms that have limited capacity, narrower product lines, or limited distribution channels. Moreover, the "dominant" firm typically offers more complex pricing schemes (e.g., rebates/discounts conditional on volumes) than its rival. In the foregoing cases, to name a few, Microsoft, Canada Pipe, Tomra and Tetra Pak are considerably larger than their rival firms, and they design complex NLP whereas their small rivals usually utilize simple linear pricing (LP).

To capture the above stylized facts of the asymmetries and propose a new rationale of NLP with minimal information requirement, we consider the following model with *complete and perfect information*. There are two firms, a dominant firm and a minor firm. Both firms can produce a homogeneous product at a constant marginal cost. However, the minor firm is capacity constrained. There is a representative downstream buyer who may purchase the product from one or both firms. We consider a three-stage game in which the dominant firm offers a general NLP schedule first and then the minor firm responds with a per-unit price, followed by the buyer choosing her purchases from both firms. We characterize subgame perfect equilibrium (SPE) outcomes of the game, and study the properties of the dominant firm's optimal NLP and the implications of the equilibrium outcomes.

It appears that, in our complete and perfect information game, a singleton quantity-payment offer (also known as a bundle) would be sufficient and optimal for the dominant firm. After all, there is only one buyer and no uncertainty regarding her demand.[5] Thus, there can be only one quantity purchased by her from the dominant firm in equilibrium, regardless of how many quantities offered by the dominant firm. However, we show in Section 4 that, by offering *multiple* bundles to a single buyer, the dominant firm can increase its profit, even though in equilibrium, only one bundle will be chosen.

The intuition for why the dominant firm can improve its profit by offering *unchosen* bundles to a single buyer is that those extra bundles provide the buyer with extra options. Such extra options, as the buyer's latent choices, constrain the minor firm's possible deviations of undercutting the dominant firm. Therefore, the unchosen bundles help the dominant firm better manipulate competition against its

[5] Klemperer and Meyer (1989) show that, as a response to demand uncertainty, firms may be forced to offer a supply function against a range of possible states.

small rival and extract more surplus from the buyer. As it turns out, the dominant firm's profit maximization requires a *continuum* of bundles, and the minimal set of such optimal bundles entails a schedule of *strictly increasing marginal prices* for increments of the buyer's purchases, *albeit the average prices could be decreasing.* We thus contribute to the literature of nonlinear pricing by providing a novel explanation for a menu of offers conditional on volumes under duopoly in the absence of private information.

As compared to contracts that reference rivals, e.g., market-share discounts,[6] a NLP schedule conditional on the supplier's *own* volume sales, such as a quantity discount scheme, is often regarded as more likely to be efficiency-enhancing. Nevertheless, our results suggest that, the antitrust scrutiny for the own volume-based NLP schedule employed by a dominant undertaking is warranted, for different reasons depending on the extent of the dominance.[7] If the dominance is very prominent (i.e., the capacity of the minor firm is small), then, as compared to LP schemes, the NLP adopted by the dominant firm reduces the price, sales, and profits of the minor firm as well as the buyer's surplus. This is because when the minor firm's capacity is limited, the dominant firm enjoys a significant non-contestable demand, and can use that demand as a stake to tie part of the contestable demand with it through a NLP schedule. This results in a partial foreclosure to the minor firm and meanwhile hurts buyer. By contrast, if the dominance is limited (i.e., the capacity of the minor firm is large), then the adoption of NLP can soften the competition and thus increase the minor firm's profits and reduce both the buyer's surplus and total surplus. In this case, the dominant firm's priority becomes to prevent the minor firm from undercutting. So now NLP is detrimental to competition not because it forecloses the minor firm, but because it acts as a competition-softening device and harms the buyer and social efficiency.

Furthermore, we have a methodological contribution of demonstrating a "mechanism design approach" to solving SPE outcomes. As we will point out in Section 3, it is considerably difficult to solve the optimal NLP schedule in our game by

[6] Majumdar and Shaffer (2009), Mills (2010), Calzolari and Denicolò (2013) and Chen and Shaffer (2014) study market-share discounts under competition.

[7] Calzolari and Denicolò (2011) show that, if two firms, competing for consumers who are privately informed about the demand, are highly asymmetric, then quantity discounts can hurt smaller firms and consumers.

applying the standard backward induction procedure, essentially because the dominant firm's action space is a functional space. Nonetheless, in Subsection 5.1 we transform the problem of determining SPE outcomes into a mechanism design problem with hidden action and hidden information that can be solved by mechanism design techniques. Generally, for games where there is a single first mover whose action space is a functional space and all the followers' action spaces are much simpler, one can apply our mechanism design approach to transform the problem of solving equilibrium outcomes into a more tractable mechanism design (constrained optimization) problem.[8]

To understand our transformation, it is important to keep in mind two features of our model: first, the dominant firm's NLP schedule is contingent on its own sales, so that the dominant firm has direct control on the quantity it sells to and the payment it receives from the buyer (but not on, say, its competitor's sales nor price); second, the buyer makes her purchase decision after seeing the minor firm's price, and the minor firm makes its price decision after seeing the dominant firm's NLP schedule. Thus, we can imagine that the dominant firm, instead of offering a NLP schedule, on one hand recommends the minor firm what price to charge, and on the other hand offers the buyer a selling mechanism. Such a mechanism requires the buyer to report the minor firm's price right after she sees it, and commits on how the quantity the buyer receives and the payment she makes depend on her report. In the spirit of the Revelation Principle, the restrictions imposed by SPE in the minor firm-buyer subgame can be captured as incentive compatibility and individual rationality constraints for the buyer, and an obedience constraint for the minor firm. The dominant firm designs the optimal recommendation and selling mechanism subject to the above constraints. Finally, after solving this constrained optimization problem, the optimal selling mechanism can be transformed back to an optimal NLP for the dominant firm. More succinctly, we transform our original problem into a "virtual" principal-agent problem that involves treating the minor firm's (an agent's) price as its hidden action meanwhile letting the buyer (another agent) to report the minor firm's price as her private information to the dominant firm (the principal). After the transformation, we can apply mechanism design

[8]For example, our mechanism design approach can be applied to transforming alternative games where (i) there are multiple minor firms and multiple buyers, and/or (ii) the minor firms and buyers have private information.

techniques to characterize the optimal NLP for the dominant firm.

Related Literature. In the literature, NLP is usually considered as a screening device in the presence of buyers' private information, e.g., Maskin and Riley (1984). The Revelation Principle indicates that a monopolist can without loss adopt a direct revelation selling mechanism (i.e., quantity-payment pairs conditional on reported types). In the absence of a buyer's private information, a direct revelation selling mechanism for the buyer reduces to a single bundle (i.e., one quantity-payment pair). By contrast, we show that the dominant firm in our setting, competing with a minor follower, has incentives to offer a NLP with a continuum of bundles. Our Equivalence Theorem (Theorem 1) shows how the idea of the Revelation Principle can be extended to our setting to greatly facilitate the equilibrium analysis.

Our setting has some similarities to sequential common agency when the two firms are regarded as principals and the buyer a common agent. The literature of common agency is large, both in theory and applications. It initially deals with simultaneous common agency, which in our context means that the two firms make offers simultaneously; the literature on sequential common agency is fast growing in recent years.[9] However, because in our setting the second-mover firm only offers LP, our analysis is very different from the analysis of sequential common agency. In fact, as it will be clear later, in our setting it is the best to view the first-mover firm as the *only* principal, whereas both the second-mover firm and the buyer should be regarded as two agents.

Unlike the existing literature on exclusionary contracts, we show that partial foreclosure resulting from NLP arises under complete information with one buyer. Thus, our exclusionary story does not need discoordinated buyers like in Rasmusen, Ramseyer, and Wiley (1991) and Segal and Whinston (2000). There is only one buyer in our model, so it is devoid of downstream competition in Simpson and Wickelgren (2007) and Asker and Bar-Isaac (2014). For the NLP contracts to have exclusionary effects in Aghion and Bolton (1987) and Choné and Linnemer (2016), it is necessary to have uncertainty about the minor firm's cost or demand. We instead provide an exclusionary theory in the absence of uncertainty.

[9] See, for example, Martimort (2006); Pavan and Calzolari (2009, 2010) for a literature review on common agency.

Several studies examine optimal pricing within specific classes of simple nonlinear pricing under asymmetric competition like in our setting. Chao (2013) considers the competitive impact of a three-part tariff employed by a leading firm. Chao, Tan, and Wong (2018a,b) study the impact of all-units discounts with one threshold. The current paper extends the analyses there to determine the general optimal NLP.

The remainder of the paper is organized as follows. In Section 2, we set up our model of asymmetric competition in intermediate-goods markets. Section 3 examines the buyer's problem and points out difficulties in applying standard backward induction procedure. Section 4 demonstrates how an extra unchosen price-quantity bundle can improve the dominant firm's profit, albeit it will not be chosen in equilibrium. Subsection 5.1 establishes an equivalence between a SPE of the game and an optimal mechanism in a "virtual" principal-agent model with hidden action and hidden information. Subsection 5.2 characterizes the equilibrium outcome of our original game. Section 6 derives the characterization by solving the "virtual" principal-agent model. Other equilibrium properties and implications (including the qualitative features of the dominant firm's optimal NLP, comparative statics, and the impact of NLP on competition) are discussed in Section 7. Section 8 concludes. Proofs are relegated to an Appendix.

2 Model

There are two firms, firm 1 and firm 2, that produce identical products, and one buyer (or downstream firm) for the product. To capture a notion of dominance, we allow for a possible capacity asymmetry between the two firms. In particular, firm 1, as a dominant firm, can produce any quantity at a unit cost $c \geq 0$. Firm 2, as a possibly smaller firm, has a capacity $k \in (0, \infty]$, up to which it can produce any quantity at the same unit cost c. If the buyer chooses to buy $Q \geq 0$ units from firm 1 and $q \in [0, k]$ units from firm 2, her payoff is the gross benefit given by $u(Q + q)$, less the payments to the two firms.

We consider a three-stage game as follows. First, firm 1 offers a nonlinear tariff $\tau(\cdot)$, which specifies the payment $\tau(Q) \in \mathbb{R} \cup \{\infty\}$ that the buyer has to make if she chooses to buy Q units from firm 1,[10] with the restriction that $\tau(0) \leq 0$. Second, after observing $\tau(\cdot)$, firm 2 offers a unit price $p \geq c$ (up to k units). Third, after observing $\tau(\cdot)$ and p, the buyer chooses the quantities she buys from the two firms. This is a sequential-move game with complete and perfect information. We use the equilibrium concept of (pure-strategy) SPE.

We say a tariff is *regular* if the subgame after firm 1 offers such a tariff has some SPE.[11] The set of feasible tariffs firm 1 can choose from, denoted as $\mathcal{T}$, is the collection of $\tau : \mathbb{R}_+ \to \mathbb{R} \cup \{\infty\}$ that is regular and satisfies $\tau(0) \leq 0$.[12] Also denote the set of feasible unit prices firm 2 can choose from as $\mathcal{P} \equiv [c, \infty)$.

A SPE is composed of a firm 1's strategy $\tau^* \in \mathcal{T}$, a firm 2's strategy $p^* : \mathcal{T} \to \mathcal{P}$, and a buyer's strategy $q^* : \mathcal{T} \times \mathcal{P} \to \mathbb{R}_+ \times [0, k]$, such that

$$q^*(\tau, p) \in \underset{(Q,q)\in\mathbb{R}_+\times[0,k]}{\operatorname{argmax}} \{u(Q+q) - pq - \tau(Q)\} \quad \forall(\tau, p) \in \mathcal{T} \times \mathcal{P}, \tag{1}$$

$$p^*(\tau) \in \underset{p\in\mathcal{P}}{\operatorname{argmax}} \{(p-c)q_2^*(\tau, p)\} \quad \forall \tau \in \mathcal{T}, \tag{2}$$

$$\tau^* \in \underset{\tau\in\mathcal{T}}{\operatorname{argmax}} \{\tau(q_1^*(\tau, p^*(\tau))) - cq_1^*(\tau, p^*(\tau))\}. \tag{3}$$

We make the following regularity assumptions and definitions.

Assumption 1. *$u : \mathbb{R}_+ \to \mathbb{R}$ is twice continuously differentiable, satisfies $u(0) = 0$, $u''(\cdot) < 0$, $u'(0) > c$, and there exists a unique $q^e \in (0, \infty)$ such that $u'(q^e) = c$.*

Denote the quantity demanded by the buyer at any per-unit price p as $D(p) \equiv \operatorname{argmax}_{q\geq 0} \{u(q) - pq\}$, and the monopoly profit at p as $\pi(p) \equiv (p - c)D(p)$. Assumption 1 implies that $D(\cdot)$ and $\pi(\cdot)$ are continuously differentiable on $[c, u'(0))$ and $D(\cdot)$ is strictly decreasing on $[c, u'(0)]$.

We call $\max\{D(\cdot) - k, 0\}$ firm 1's *captive (or non-contestable) demand function.* From firm 1's point of view, this portion of the total demand $D(\cdot)$ is not subject to any threat of competition from firm 2, due to the latter's capacity constraint. In the Q-p space, we let

$$\Phi \equiv \{(Q, p) \in \mathbb{R}_+ \times \mathcal{P} : D(p) - k \leq Q \leq D(p)\} \tag{4}$$

[10] $\tau(Q) = \infty$ means that purchasing Q units is not allowed.

[11] It can be shown that any lower semicontinuous tariff is regular.

[12] By definition, if we allow firm 1 to choose an irregular tariff, then the whole game has no SPE.

denote the region between the original demand and the captive demand curves. Intuitively, Φ represents the *competitive (or contestable) portion* of the total demand. Note that $q^e = D(c)$ is the welfare-maximizing quantity. If $k \geq q^e$, effectively firm 2 does not have capacity constraint.

Assumption 2. *$\pi(\cdot)$ is strictly concave on $[c, u'(0)]$.*

Assumption 2 implies that there is a unique optimal monopoly price $p^m \equiv \operatorname{argmax}_p \pi(p) \in (c, u'(0))$ given by $\pi'(p^m) = 0$.

3 Buyer's Problem

Since our game is a sequential-move complete and perfect information one and we try to determine its SPE outcome, the standard backward induction method requires us to sequentially solve the buyer's problem (1), firm 2's problem (2), and firm 1's problem (3), which turns out to be an unmanageable task. Specifically, given arbitrary $\tau \in \mathcal{T}$ and $p \in \mathcal{P}$, a general characterization of the solution to the buyer's problem (1) is *not* available, since firm 1's offer τ is an endogenous function to be determined in firm 1's problem (3). But without a general explicit solution to (1), it is difficult, if not impossible, to solve (2) and (3). Thus the standard procedure of backward induction *cannot* be applied to our game.

We shall transform our original task of solving (1), (2), and (3) into a mechanism design problem, which allows us to determine the SPE outcomes. As a first step of the transformation, we introduce two important functions: $V(Q, p)$ and $\pi(Q, p)$.

Let $V(Q, p)$ denote the *buyer's conditional payoff* if she is endowed with Q units and can buy at most k more units at price p, i.e.,

$$V(Q, p) \equiv \max_{q \in [0,k]} \{u(Q + q) - pq\}. \tag{5}$$

Given the two firms' offers $\tau \in \mathcal{T}$ and $p \in \mathcal{P}$, we can decompose the buyer's maximization problem (1) into two sub-problems: (i) for any given Q, the buyer chooses q from firm 2 by solving (5); (ii) the buyer chooses Q from firm 1, i.e.,

$$\max_{Q \geq 0} \{V(Q, p) - \tau(Q)\}. \tag{6}$$

Although the sub-problem (ii) still does not permit a ready solution without knowing τ, the sub-problem (i) is well behaved and tractable: the problem in (5) has a unique maximizer given by[13]

$$\mathrm{Proj}_{[0,k]}(D(p)-Q) \equiv \max\{\min\{D(p)-Q, k\}, 0\}, \tag{7}$$

and, by the Envelope Theorem, $V(Q,p)$ has the following properties.

Lemma 1. *For every $(Q,p) \in \mathbb{R}_+ \times \mathcal{P}$,*

$$V_p(Q,p) = -\mathrm{Proj}_{[0,k]}(D(p)-Q), \tag{8}$$

$$V_Q(Q,p) = u'(\mathrm{Proj}_{[Q,Q+k]}(D(p))) = \mathrm{Proj}_{[u'(Q+k),u'(Q)]}(p), \tag{9}$$

$$V_{Qp}(Q,p) = V_{pQ}(Q,p) = \begin{cases} 1 & \text{if } D(p)-k < Q < D(p) \\ 0 & \text{if } Q < D(p)-k \text{ or } Q > D(p) \end{cases}. \tag{10}$$

Note that, from (10), V satisfies weak increasing differences. Moreover, such property of increasing differences is strict in the region Φ.

Let $\pi(Q,p)$ denote *firm 2's conditional profit* given that the buyer purchases Q units from firm 1 and firm 2 charges price p, i.e.,

$$\pi(Q,p) \equiv (p-c)\,\mathrm{Proj}_{[0,k]}(D(p)-Q). \tag{11}$$

4 Why an Unchosen Bundle Helps

Before we characterize firm 1's optimal general tariff $\tau(\cdot)$, let us first look at a specific nonlinear tariff—a finite number of bundles, each characterized by a quantity and a payment.[14]

Our starting point will be the simplest one, a one-bundle offer, which people might think as optimal for firm 1. After all, there is only one buyer and there is no demand uncertainty, so there can be only one quantity that the buyer will

[13] For any closed interval $X \subset \mathbb{R}$ and any point $x \in \mathbb{R}$, let $\mathrm{Proj}_X(x)$ denote the projection of x on X, that is, $\mathrm{argmin}_{y\in X}|y-x|$.

[14] All the results in later sections do not logically rely on this section. The purposes of this section are providing insights and helping intuitive understanding.

purchase from firm 1 in equilibrium, regardless how many quantities offered by firm 1. Nevertheless, as we shall see, firm 1 can strictly improve its profit over its optimal profit level in the "one-bundle equilibrium," by offering an extra bundle which will *not* be chosen in equilibrium.

4.1 Optimal One-Bundle Offer

In this section, we show how the optimal one-bundle offer is determined. Consider any bundle (Q, T), where $Q \geq 0$ denotes its bundle quantity and $T \in \mathbb{R}$ denotes its bundle price. Given that firm 1 offers (Q, T), the buyer may accept or reject it, after seeing firm 2's price offer p. If the buyer accepts (Q, T), the buyer's surplus is $V(Q, p) - T$; otherwise it is $V(0, p)$. Provided $Q > 0$, the increasing differences property (10) of V implies that the curve $V(Q, p) - T$, drawn against p, must cross only once the curve $V(0, p)$ from below, as shown in Figure 1(a). It follows that the buyer's acceptance/rejection decision is a cut-off policy: there exists a cut-off x such that the buyer accepts (Q, T) if $p > x$ and rejects (Q, T) if $p \leq x$, where x is determined by

$$V(Q, x) - T = V(0, x). \tag{12}$$

If the buyer accepts the bundle (Q, T) from firm 1, she would buy $\text{Proj}_{[0,k]}(D(p) - Q)$ from firm 2; otherwise she would buy $D(p)$ solely from firm 2. Then, firm 2's profit, as a function of p, consists of two pieces as

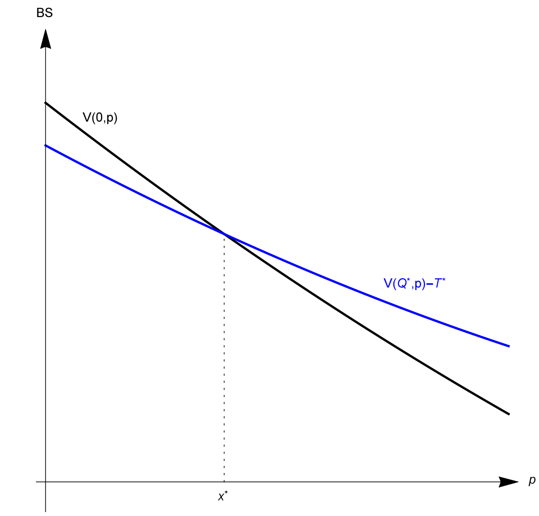
$$\begin{cases} \pi(0, p) & \text{if } p \leq x \\ \pi(Q, p) & \text{if } p > x \end{cases},$$

as shown in Figure 1(c). Since firm 1 has positive sales if and only if it can induce firm 2 to set $p > x$, firm 1 must ensure

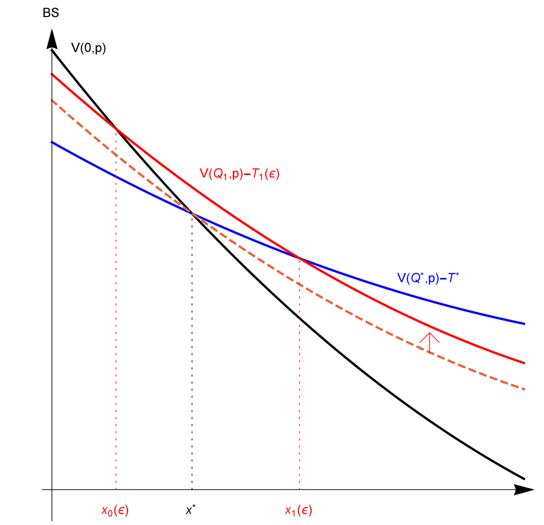
$$\max_{p > x} \pi(Q, p) \geq \max_{p \leq x} \pi(0, p). \tag{13}$$

The optimal bundle (Q^*, T^*) and cut-off x^* in a "one-bundle equilibrium" must solve the following firm 1's optimization problem

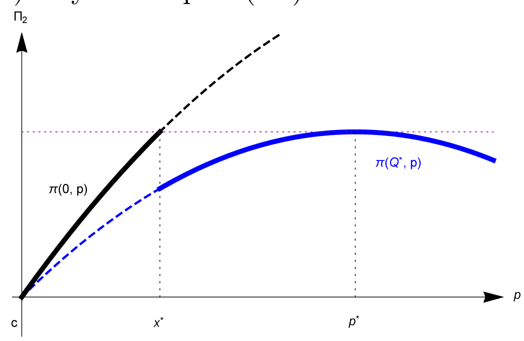
$$\underset{Q,T}{\text{Maximize}} \left\{ T - c \cdot Q \text{ s.t. } (12) \text{ and } (13) \right\}. \tag{14}$$



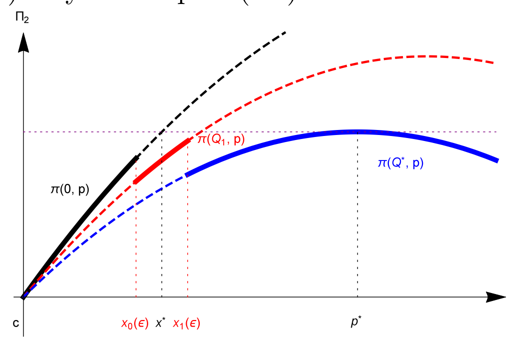
(a) Buyer's Surplus (BS) under One Bundle



(b) Buyer's Surplus (BS) under Two Bundles



(c) Firm 2's Profit (Π_2) under One Bundle



(d) Firm 2's Profit (Π_2) under Two Bundles

Figure 1: One Bundle versus Two Bundles

Using the constraint (12) to eliminate T, firm 1's profit can be written as

$$\Pi_1 = T - c \cdot Q = V(Q, x) - V(0, x) - c \cdot Q.$$

In equilibrium firm 1's sales must be positive, i.e., $Q^* > 0$. To maximize $V(Q^*, x) - V(0, x) - c \cdot Q^*$, x should be made as large as possible, because of the increasing differences property (10) of V. Consequently, at the optimal bundle (Q^*, T^*), the competitive constraint (13) from firm 2 must be binding, i.e., $\pi(Q^*, p^*) = \pi(0, x^*)$, where $p^* \in \operatorname{argmax}_{p > x^*} \pi(Q^*, p)$ is firm 2's equilibrium price. Figure 1(c) illustrates it. Also note that, at the optimum, $x^* \in (c, p^*)$.

4.2 Adding an Unchosen Bundle to Improve

We now, given the optimal one-bundle offer (Q^*, T^*) and the corresponding cut-off x^*, construct *one extra bundle to relax the originally binding no-deviation constraint (13) for firm 2*. As a result, firm 1 can strictly improve its profit over $\Pi_1^* \equiv T^* - c \cdot Q^*$ with the extra bundle.

We show the profitable improvement through the following two steps.

Step 1: Add an extra unchosen bundle.

Pick any $Q_1 \in (0, Q^*)$, and let $T_1(\epsilon) = V(Q_1, x^*) - V(0, x^*) - \epsilon$ for $\epsilon \geq 0$. From the increasing differences property (10) of V, the solid red curve $V(Q_1, p) - T_1(\epsilon)$, drawn against p, must cross once and only once the solid black curve $V(0, p)$ (the solid blue curve $V(Q^*, p) - T^*$) from below at $x_0(\epsilon)$ (above at $x_1(\epsilon)$), as illustrated in Figure 1(b). Here $x_0(\epsilon)$ and $x_1(\epsilon)$ are given by

$$V(Q_1, x_0(\epsilon)) - T_1(\epsilon) = V(0, x_0(\epsilon)), \tag{15}$$

$$V(Q_1, x_1(\epsilon)) - T_1(\epsilon) = V(Q^*, x_1(\epsilon)) - T^*. \tag{16}$$

Now, if firm 1 offers both bundles $(Q_1, T_1(\epsilon))$ and (Q^*, T^*), the buyer would pick the large bundle (Q^*, T^*) when firm 2's price is above $x_1(\epsilon)$, and would pick the small one $(Q_1, T_1(\epsilon))$ when firm 2's price is between $x_0(\epsilon)$ and $x_1(\epsilon)$, and would not pick any bundle from firm 1 when firm 2's price is below $x_0(\epsilon)$.

Accordingly, firm 2's profit, as a function of p, consists of three pieces as

$$\begin{cases} \pi(0,p) & \text{if } p \leq x_0(\epsilon) \\ \pi(Q_1,p) & \text{if } x_0(\epsilon) < p \leq x_1(\epsilon) \\ \pi(Q^*,p) & \text{if } p > x_1(\epsilon) \end{cases},$$

as shown in Figure 1(d). If firm 1 wants to induce the buyer to still choose the bundle (Q^*, T^*), firm 1 must ensure firm 2 to set $p > x_1(\epsilon)$, i.e.,

$$\max_{p > x_1(\epsilon)} \pi(Q^*, p) \geq \max_{p \leq x_0(\epsilon)} \pi(0, p) \tag{17}$$

and

$$\max_{p > x_1(\epsilon)} \pi(Q^*, p) \geq \max_{x(\epsilon) < p \leq x_1(\epsilon)} \pi(Q_1, p). \tag{18}$$

Thus, by offering an extra bundle, firm 1 breaks firm 2's profit function from two pieces to three pieces, and hence replaces constraints (12) and (13) with (15), (16), (17), and (18).[15]

Interestingly, while the original no-deviation constraint (13) for firm 2 is binding, after adding the extra bundle, the new ones (17) and (18) are non-binding, provided $\epsilon > 0$ is small enough. This can be seen from Figure 1(d). Since $c < x^* < p^*$, for small $\epsilon > 0$ we have $c < x_0(\epsilon) < x^* < x_1(\epsilon) < p^*$, as in Figure 1(d). Then,

$$\max_{p > x_1(\epsilon)} \pi(Q^*, p) = \pi(0, x^*) > \pi(0, x_0(\epsilon)) = \max_{p \leq x_0(\epsilon)} \pi(0, p).$$

So (17) is *not* binding. In addition, because $\pi(Q_1, x^*) < \pi(0, x^*)$, when $\epsilon > 0$ is small enough, we have $x_1(\epsilon)$ close enough to x^*, so that $\pi(Q_1, x_1(\epsilon)) < \pi(0, x^*)$. It follows that

$$\max_{p > x_1(\epsilon)} \pi(Q^*, p) = \pi(0, x^*) > \pi(Q_1, x_1(\epsilon)) = \max_{x(\epsilon) < p \leq x_1(\epsilon)} \pi(Q_1, p).$$

So (18) is *not* binding, either.

Step 2: Improve profit by increasing the price of the original chosen bundle.

Since now (17) and (18) are non-binding, firm 1 will be able to strictly increase its profit by increasing the price of the chosen Q^*-bundle from T^* to $T_2(\delta) = T^* + \delta$ with small $\delta > 0$. After that, the $x_0(\epsilon)$ defined by (15) is unaffected; the $x_1(\epsilon)$ defined by (16) is now replaced by $x_1(\epsilon, \delta)$, which is given by

$$V(Q_1, x_1(\epsilon, \delta)) - T_1(\epsilon) = V(Q^*, x_1(\epsilon, \delta)) - T^* - \delta.$$

As $\delta > 0$ is small, $x_1(\epsilon, \delta)$ is close to $x_1(\epsilon)$, so that, first, $x_1(\epsilon, \delta) < p^*$, and second, (17) and (18) still hold strictly. Accordingly, firm 2 still charges $p = p^*$ and the buyer still picks the Q^*-bundle, but now pays firm 1 a higher bundle price $T_2(\delta) > T^*$.

[15] When $\epsilon = 0$, the curve $V(Q_1, p) - T_1(0)$ (the dashed red curve in Figure 1(b)) crosses both $V(0, p)$ and $V(Q^*, p) - T^*$ at x^*, thereby $x_0(0) = x_1(0) = x^*$. Hence (15) and (16) reduce to (12); (17) and (18) reduce to (13).

4.3 Summary

From the above analysis, we see that firm 1 can improve its profit by offering the buyer an extra bundle and raising the price of the chosen bundle. (Of course, at the optimum firm 1 may also adjust the size of the chosen bundle.) The intuition is: *such extra unchosen bundle, as the buyer's latent choice, reduces the temptation for firm 2 to cut price.* Indeed, in Figure 1(c), without the extra bundle, if firm 2 cuts its price to x^*, the buyer no longer buys from firm 1, and the resulting increase in firm 2's sales makes firm 2 indifferent between cutting to x^* and not; in contrast, in Figure 1(d), with the extra bundle, if firm 2 cuts its price to x^*, the buyer still buys the small bundle from firm 1, and the resulting increase in firm 2's sales would be so limited that firm 2's profit would strictly decrease; if firm 2 wants the buyer not to buy even the small bundle from firm 1, the necessary price cut would be so deep that, once again, firm 2's profit would strictly decrease. Firm 1 can then raise the price of the chosen bundle without triggering firm 2's price cut.

Also, it is worth noting that in the above construction we require $Q_1 < Q^*$ because the extra bundle is meant to be chosen after some firm 2's price cutting, which induces the buyer to buy less from firm 1. In fact, adding an extra bundle that is larger than the chosen one can never improve firm 1's profit. Another important observation is that, at the optimum, the no-deviation constraints for firm 2 (i.e., (17) and (18)) must be binding, for otherwise there is still room for firm 1 to raise its profit.[16]

All these qualitative features extend to the case where firm 1 is allowed to offer more than two bundles. When more bundles are allowed, firm 2's profit curve is cut into more pieces, which relax firm 2's no-deviation constraints and allow firm 1 to increase its profit even further. In general, if n bundles are allowed, there are n no-deviation constraints for firm 2, which must be all binding at the optimum. As we shall see in the following section, firm 1's optimal NLP involves a continuum of quantities for the buyer to choose from, corresponding to a range of firm 2's no-deviation constraints that must be binding at the optimum.

[16] If (18) is slack, firm 1 may further raise the price of the chosen bundle (keeping the price of the smaller bundle unchanged), which increases x_1 and keeps x_0 unchanged and thus (17) intact. If (17) is slack, firm 1 may raise the prices of both the chosen and the smaller bundles by the same amount, which increases x_0 and keeps x_1 unchanged and thus (18) intact.

5 Equilibrium

As we argued in Section 3, it is difficult to apply the standard procedure of backward induction to solve the SPE. In this section, we transform the original problem into an equivalent mechanism design problem, which allows us to characterize the SPE outcomes.

5.1 Equivalence between the SPE and a Mechanism Design Problem

Observe that, for any given tariff $\tau \in \mathcal{T}$, the buyer will optimally choose some purchase $Q(p) \geq 0$ from firm 1, contingent on any possible price $p \in \mathcal{P}$ chosen by firm 2. The payment for this purchase is thus $\tau(Q(p)) \equiv T(p)$. So the buyer enjoys a net surplus $V(Q(p), p) - T(p)$. Given that the buyer's optimal purchase from firm 1 is $Q(p)$, and hence the optimal purchase from firm 2 is $\mathrm{Proj}_{[0,k]}(D(p) - Q(p))$, firm 2 would optimally choose some price $\bar{p} \in \mathcal{P}$, i.e., to maximize its profit $\pi(Q(p), p)$. Virtually, we have a one-principal-two-agent model, in which firm 1 (the principal) offers a direct revelation mechanism $Q : \mathcal{P} \to \mathbb{R}_+$ and $T : \mathcal{P} \to \mathbb{R}$ to the buyer (Agent 1), and recommends a price $\bar{p} \in \mathcal{P}$ for firm 2 (Agent 2).

In the spirit of the Revelation Principle, *imagining firm 1 asks the buyer to report firm 2's price*, solving the SPE for the whole game is equivalent to solving the following constrained optimization problem (OP):

$$\underset{Q(\cdot), T(\cdot), \bar{p}}{\text{Maximize}}\, T(\bar{p}) - c \cdot Q(\bar{p}) \tag{OP}$$

subject to

$$V(Q(p), p) - T(p) \geq V(Q(\tilde{p}), p) - T(\tilde{p}) \quad \forall p, \tilde{p} \in \mathcal{P} \tag{B-IC}$$

$$V(Q(p), p) - T(p) \geq V(0, p) \quad \forall p \in \mathcal{P} \tag{B-IR}$$

$$\pi(Q(\bar{p}), \bar{p}) \geq \pi(Q(p), p) \quad \forall p \in \mathcal{P}. \tag{F2-IC}$$

Constraint (B-IC) is the incentive compatibility constraint for the buyer, i.e., the buyer has incentive to report firm 2's price truthfully. Constraint (B-IR) is the individual rationality constraint for the buyer, i.e., the buyer is willing to participate in the mechanism rather than obtaining nothing from and paying nothing to firm 1 (and single-sourcing from firm 2). Constraint (F2-IC) is the incentive compatibility constraint for firm 2, i.e., firm 2 has an incentive to charge the recommended price $\bar{p}$, understanding that the buyer will always report its price truthfully. Finally, the objective function of (OP) is firm 1's profit provided firm 2 follows the recommendation $\bar{p}$ and the buyer reports truthfully.

The equivalence between the SPE and the optimization problem (OP) is established in the following theorem.

Theorem 1 (Equivalence)**.** *Take any $Q^* : \mathcal{P} \to \mathbb{R}_+$, $T^* : \mathcal{P} \to \mathbb{R}$, and $\bar{p}^* \in \mathcal{P}$. $(Q^*(\cdot), T^*(\cdot), \bar{p}^*)$ is a solution of (OP) if and only if there is a SPE (τ^*, p^*, q^*) such that*

$$Q^*(p) = q_1^*(\tau^*, p) \quad \forall p \in \mathcal{P}, \tag{19}$$

$$\mathrm{Proj}_{[0,k]}(D(p) - Q^*(p)) = q_2^*(\tau^*, p) \quad \forall p \in \mathcal{P}, \tag{20}$$

$$T^*(p) = \tau^*(Q^*(p)) \quad \forall p \in \mathcal{P}, \tag{21}$$

$$\bar{p}^* = p^*(\tau^*). \tag{22}$$

By virtue of Theorem 1, we reduce our task of finding the SPE to determining the solution to (OP). The optimization problem (OP) can be solved with the help of mechanism design techniques and some nice diagrams in Section 6; its solution can then be transformed back to characterize the SPE outcomes.

5.2 Characterization of the SPE

In this subsection we state our final characterization results and postpone the derivations to Section 6, so that reader not interested in the derivations may skip Section 6 and jump to the equilibrium implications in Section 7.

The following theorem establishes the existence and provides characterization of the SPE.

Theorem 2 (SPE)**.** *There exists at least one SPE. In any SPE, the following holds:*

(a) Firm 2's equilibrium price $\bar{p} \in (c, p^m)$ solves[17]

$$-(\bar{p} - c)D'(\bar{p}) = \frac{1}{e} \min \left\{ D\left(c + \frac{\bar{p} - c}{e}\right), k \right\}, \tag{23}$$

and firm 2's equilibrium profit Π_2 is given by

$$\Pi_2 = (\bar{p} - c)(D(\bar{p}) - \pi'(\bar{p})) = -(\bar{p} - c)^2 D'(\bar{p}) > 0. \tag{24}$$

(b) Given firm 1's equilibrium tariff, firm 2 is indifferent between charging its equilibrium price $\bar{p}$ and deviating to any $p \in [x_0, \bar{p})$, where

$$x_0 = c + \frac{\bar{p} - c}{e} \in (c, \bar{p}). \tag{25}$$

(c) If firm 2 charges a (possibly off-equilibrium) price p within the interval $[x_0, \bar{p}]$, the buyer would purchase $Q(p)$ from firm 1 and $D(p) - Q(p)$ from firm 2, where $Q(\cdot)$ on $[x_0, \bar{p}]$ is a strictly increasing function given by

$$Q(p) = D(p) - \frac{\Pi_2}{p - c} \quad \forall p \in [x_0, \bar{p}]. \tag{26}$$

In particular, the buyer's equilibrium purchase from firm 1 is

$$\bar{Q} = Q(\bar{p}) = \pi'(\bar{p}) \in (\max\{D(\bar{p}) - k, 0\}, D(\bar{p})) \tag{27}$$

and her equilibrium purchase from firm 2 is $D(\bar{p}) - \bar{Q} \in (0, k)$; if firm 2 deviates to x_0, the buyer's (off-equilibrium) purchase from firm 1 would be

$$Q_0 = Q(x_0) = \max\{D(x_0) - k, 0\} < \bar{Q}. \tag{28}$$

(d) Firm 1's optimal tariff $\tau(\cdot)$ on $[Q_0, \bar{Q}]$ is given by

$$\tau(Q) = u(Q_0 + k) - u(k) + \int_{Q_0}^{Q} x(\tilde{Q}) d\tilde{Q} \quad \forall Q \in [Q_0, \bar{Q}], \tag{29}$$

where $x(\cdot)$ on $[Q_0, \bar{Q}]$ is the inverse of $Q(\cdot)$ on $[x_0, \bar{p}]$.

[17] e denotes the base of natural logarithm, which is approximately 2.71828.

Noticeably, (26) implies that firm 2's profit $(p-c)(D(p)-Q(p))$ is a constant (equal to its equilibrium profit Π_2) over a range of prices in $[x_0, \bar{p}]$. It is because the incentive compatibility constraint (F2-IC) for firm 2 is binding over that range. In response to firm 2's possible deviations in $[x_0, \bar{p}]$, firm 1 offers a continuum of bundles with quantities in the interval $[Q_0, \bar{Q}]$.

The tariff formula (29) actually describes the *minimal set* of bundles that is necessary to constitute an optimal NLP for firm 1. The payments for quantities outside $[Q_0, \bar{Q}]$ are actually indeterminate. Firm 1 has some freedom to set $\tau(Q)$ for $Q \notin [Q_0, \bar{Q}]$ because, as long as they are set to be sufficiently high, firm 2 would still charge $\bar{p}$ and the buyer would still purchase $\bar{Q}$ from firm 1 and $D(\bar{p})-\bar{Q}$ from firm 2 so that the allocation and profits would be unaffected. To construct a complete optimal tariff, a simple way is to let $\tau(0) = 0$ if $0 \notin [Q_0, \bar{Q}]$, and let $\tau(Q) = \infty$ for all $Q \notin [Q_0, \bar{Q}] \cup \{0\}$.

The intuition for why only the bundles with quantities in $[Q_0, \bar{Q}]$ are needed is clear from our analysis in Section 4. Recall that the necessity of offering unchosen bundles comes from the competitive pressure from firm 2 (which is parameterized by k): they serve as the buyer's latent choices to constrain firm 2's possible deviation of undercutting, and we have seen that such latent bundles must be smaller than the chosen bundle. The bundle $(\bar{Q}, \tau(\bar{Q}))$ is chosen by the buyer in equilibrium. Those $(Q, \tau(Q))$ for $Q \in [Q_0, \bar{Q})$ are unchosen but they are the *necessary* latent choices for the purpose of preventing firm 2's deviations. Once the schedule of latent bundles extends downward up to a point at which it reaches firm 1's captive demand (i.e., $\max\{D(\cdot)-k, 0\}$), the competitive pressure is not in force and no more latent bundles are needed. In other words, those $(Q, \tau(Q))$ for $Q \notin [Q_0, \bar{Q}] \cup \{0\}$, if being offered at all, are not only unchosen but also truly redundant.

Note that, from (29), $\tau'(\bar{Q}) = x(\bar{Q}) = \bar{p}$, that is, firm 1's marginal price evaluated at its actual sales is equal to firm 2's actual price; furthermore, if firm 2 deviates to charge any price $p \in [x_0, \bar{p})$, the buyer would purchase from firm 1 the quantity Q at which firm 1's marginal price is $\tau'(Q) = x(Q) = p$. This is because the buyer's purchases are always adjusted to equate the marginal prices of the two firms provided her purchases from both firms are positive.

Theorem 2 claims the existence but not the uniqueness of the SPE outcome.

Strictly speaking, the SPE outcome is never unique because of the aforementioned freedom of choosing $\tau(Q)$ for $Q \notin [Q_0, \bar{Q}]$. But if the solution of (23) for $\bar{p}$ is unique, from Theorem 2 all the "relevant" (i.e., except redundant bundles) components of the SPE outcome are uniquely determined. Therefore, we say the SPE outcome is *essentially unique* if the solution of (23) is unique. Since the right-hand side of (23) as a function of $\bar{p}$ is non-increasing, a simple sufficient condition for the essential uniqueness is that the left-hand side of (23) as a function of $\bar{p}$ is strictly increasing.

Corollary 1 (Essential Uniqueness)**.** *The SPE outcome is essentially unique if the following condition is satisfied:*[18]

$$-(p-c)D'(p) \text{ is strictly increasing in } p \text{ on } [c, p^m]. \tag{30}$$

Note that $-(p-c)D'(p) = D(p) - \pi'(p)$. Thus, a graphical interpretation of condition (30) is that, for any k, the curve $Q = \pi'(p)$ and the captive demand curve $Q = D(p) - k$ cross at most once, as we draw in Figures 2 and 4 in Section 6.

6 Derivations of Equilibrium

This section outlines the derivations of Theorem 2. Following Theorem 1, we do it through solving the constrained optimization problem (OP).

6.1 Constraints for the Buyer

The following lemma characterizes the incentive compatibility constraint (B-IC) and individual rationality constraint (B-IR) for the buyer.

Lemma 2 (Constraints for Buyer)**.** *Any $Q : \mathcal{P} \to \mathbb{R}_+$ and $T : \mathcal{P} \to \mathbb{R}$ satisfy (B-IC) and (B-IR) if and only if the following conditions hold:*

$$\begin{aligned} &\forall p_1, p_2 \in \mathcal{P} \text{ with } p_1 \le p_2, \text{ either } Q(p_1) \le Q(p_2) \\ &\qquad\qquad\qquad \text{or } D(p_1) \le Q(p_2) \text{ or } Q(p_1) \le D(p_2) - k \end{aligned} \tag{Mon}$$

[18] This condition is equivalent to that $u'(q) - c$ is strictly log-concave in q on $[D(p^m), q^e]$.

$$\forall p \in \mathcal{P},\ T(p) - T(c) = V(Q(p), p) - V(Q(c), c) - \int_c^p V_p(Q(t), t)dt \tag{31}$$

$$V(Q(c), c) - T(c) \geq V(0, c) \tag{32}$$

Condition (Mon) is a weakened version of the standard monotonicity condition for mechanism design problems; it is weakened because the increasing differences property (10) of V is strict only on Φ. If (Mon) holds, $Q(\cdot)$ must be non-decreasing on $\{p \in \mathcal{P} : (Q(p), p) \in \Phi\}$, but may be decreasing when $(Q(p), p) \notin \Phi$. Condition (Mon) says that $Q(\cdot)$ may be decreasing only in a particular way: whenever $p_1 < p_2$ and $Q(p_1) > Q(p_2)$, the rectangle $[Q(p_2), Q(p_1)] \times [p_1, p_2]$ must not intersect the region Φ. Such weakened monotonicity implies the following result.

Corollary 2. *Condition (Mon) implies* $\mathrm{Proj}_{[0,k]}(D(p) - Q(p))$ *is non-increasing in* p *on* $\mathcal{P}$*.*

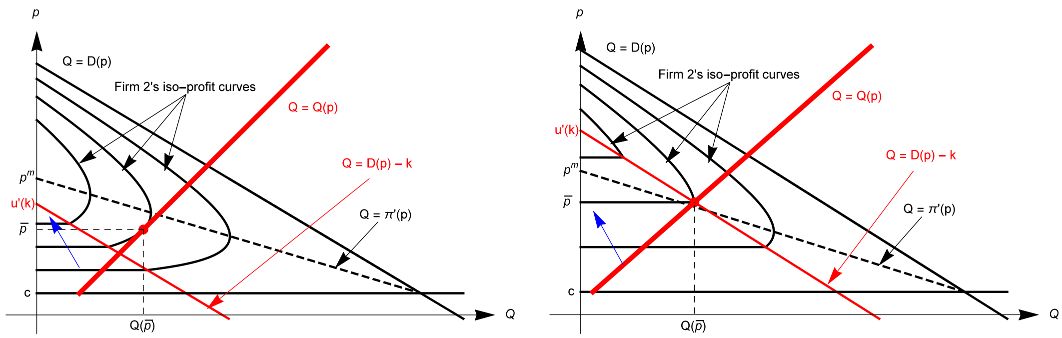
Condition (31) is the standard envelope formula for payment for mechanism design problems. Condition (Mon) and condition (31) together are necessary and sufficient conditions for (B-IC). Moreover, given (31), condition (32) is a necessary and sufficient condition for (B-IR), since (31) implies $V(Q(p), p) - T(p) - V(0, p)$ is non-decreasing in p.

Once the constraints (B-IC) and (B-IR) are replaced with (Mon), (31), and (32), we see that (32) must be binding, for otherwise firm 1 can increase its profit $T(\bar{p}) - c \cdot Q(\bar{p})$ by increasing $T(p)$ for every $p \in \mathcal{P}$ by a constant, after which all other constraints ((Mon), (31), and (F2-IC)) are intact. Therefore,

$$T(c) = V(Q(c), c) - V(0, c). \tag{33}$$

Using (31) and (33) to eliminate $T(\cdot)$, we obtain, for all $p \in \mathcal{P}$,

$$T(p) = V(Q(p), p) - V(0, c) - \int_c^p V_p(Q(t), t)dt. \tag{34}$$



(a) when $k \geq D(p^m)$ (b) when $k < D(p^m)$

Figure 2: Firm 2's Iso-profit Curves

6.2 Constraints for Firm 2

We now take a look at the incentive compatibility constraint (F2-IC) for firm 2. Given that firm 1 offers the buyer $Q(\cdot)$, firm 2 has incentives to follow the recommended price $\bar{p}$ if and only if its profit $\pi(Q(p), p)$ is maximized at $p = \bar{p}$.

Figure 2 shows firm 2's iso-profit curves (i.e., the level curves of $\pi(Q, p) = (p - c)\,\mathrm{Proj}_{[0,k]}(D(p) - Q))$ and an *arbitrary* $Q(\cdot)$ curve in the Q-p space. Firm 2 seeks to attain the highest iso-profit curve by choosing a point on the $Q(\cdot)$ curve.

Figure 2 also illustrates the general patterns of the iso-profit curves. First, each iso-profit curve associated with a positive profit level must be strictly below the demand curve $Q = D(p)$ and strictly above the cost line $p = c$. Second, Assumption 2 implies that $\pi(Q, \cdot)$ is strictly concave on $\{p : \pi(Q, p) > 0\}$ for every $Q \geq 0$. Therefore, each iso-profit curve (associated with a positive profit level) must be (horizontally) single-peaked, and thus has a unique most rightward point. Furthermore, if firm 2 does not have capacity constraint (i.e., $k \geq q^e$), its iso-profit curves are the same as the level curves of $\pi(p) - (p - c)Q$, whose slopes are $(p - c)/(\pi'(p) - Q)$. Thus, the iso-profit curves are strictly decreasing when $Q > \pi'(p)$ and strictly increasing when $Q < \pi'(p)$. When firm 2 has capacity constraint (i.e., $k < q^e$), the iso-profit curves are horizontal when (Q, p) is below the captive demand curve, i.e., $Q < D(p) - k$, and coincide the level curves of $\pi(p) - (p - c)Q$ otherwise, as shown in Figure 2. The captive demand curve $Q = D(p) - k$ may or may not cross the curve $Q = \pi'(p)$, depending on whether k is smaller or larger than $D(p^m)$ as shown in Figures 2(a) and 2(b). The direction of higher profit is indicated by the arrow pointing northwest.

Therefore, we in particular have the following results, which we will use later: (i) The largest feasible level of firm 2's profit is $\pi(\max\{p^m, u'(k)\})$; (ii) Each iso-profit curve (associated with positive profit less than $\pi(\max\{p^m, u'(k)\})$) has its unique most rightward point on the curve $Q = \max\{\pi'(p), D(p) - k\}$.

6.3 Constrained Optimization

This subsection solves (OP) and thus the SPE outcome. Recall that we have used (31) and (33) to eliminate $T(\cdot)$, as given by (34). Then, firm 1's profit can be written as

$$\begin{aligned}\Pi_1 &= T(\bar{p}) - c \cdot Q(\bar{p}) \\ &= V(Q(\bar{p}), \bar{p}) - V(0, c) - \int_c^{\bar{p}} V_p(Q(p), p)dp - c \cdot Q(\bar{p})\end{aligned} \tag{35}$$

We denote firm 2's profit as Π_2 and explicitly introduce it as a choice variable in (OP). Now, (OP) can be rewritten as

$$\underset{Q(\cdot),\bar{p},\Pi_2}{\text{Maximize}} \ (35) \tag{OP'}$$

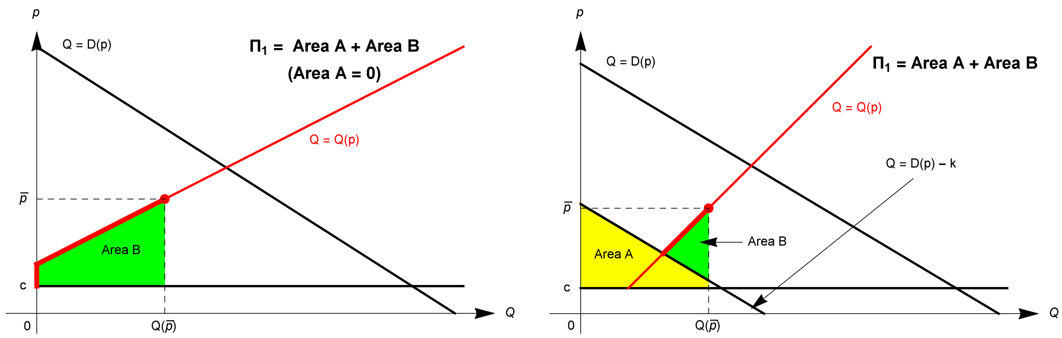
subject to

$$\text{(Mon)}$$

$$\Pi_2 \geq \pi(Q(p), p) \ \ \forall p \in \mathcal{P} \tag{F2-IC'}$$

$$\Pi_2 = \pi(Q(\bar{p}), \bar{p}). \tag{F2-Pro}$$

Our strategy of solving (OP') and hence (OP) is as follows. We decompose (OP') into two stages: in the first stage, $Q(\cdot)$ and $\bar{p}$ are optimally chosen contingent on any feasible $\Pi_2 > 0$; in the second stage, Π_2 is optimally chosen. Lemma 3 below solves the first stage contingent on Π_2, and Lemma 4 solves the second stage to pin down Π_2 and solves (OP).



(a) when $k \geq q^e$ (b) when $k < q^e$

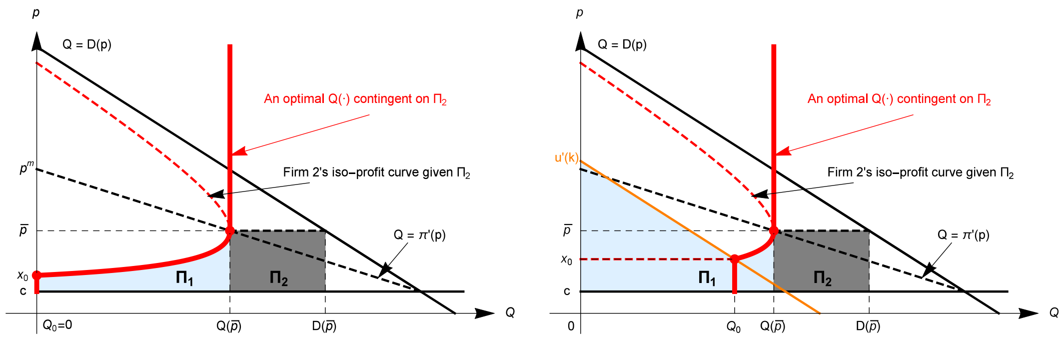
Figure 3: Firm 1's profit Π_1 contingent on $Q(\cdot)$ and $\bar{p}$

To graphically show firm 1's profit, we use (8) and (9) to rewrite (35):

$$\begin{aligned}
\Pi_1 &= \int_0^{Q(\bar{p})} [V_Q(Q,c) - c]dQ + \int_c^{\bar{p}} \left[V_p(Q(\bar{p}),p) - V_p(Q(p),p)\right] dp \\
&= \int_0^{Q(\bar{p})} [\mathrm{Proj}_{[u'(Q+k),u'(Q)]}(c) - c]dQ \\
&\quad + \int_c^{\bar{p}} [\mathrm{Proj}_{[0,k]}(D(p) - Q(p)) - \mathrm{Proj}_{[0,k]}(D(p) - Q(\bar{p}))]dp \\
&= \int_0^{Q(\bar{p})} [\mathrm{Proj}_{[u'(Q+k),u'(Q)]}(c) - c]dQ \\
&\quad + \int_c^{\bar{p}} [\mathrm{Proj}_{[D(p)-k,D(p)]}(Q(\bar{p})) - \mathrm{Proj}_{[D(p)-k,D(p)]}(Q(p))]dp
\end{aligned} \tag{36}$$

Figure 3 shows the area of Π_1 given by (36) for a given $Q(\cdot)$ and $\bar{p}$: Areas A and B correspond to the first and the second integral in (36) respectively. Area A comes from firm 1's captive demand. If firm 2 has no capacity constraint (i.e., $k \geq q^e$) as shown in Figure 3(a), the captive demand vanishes and consequently Area A is 0.

It can be seen from Figures 2 and 3 that, given a $\Pi_2 \in (0, \pi(\max\{p^m, u'(k)\}))$ and hence a firm 2's iso-profit curve, in order to maximize Π_1 subject to (Mon), (F2-IC'), and (F2-Pro), (i) the part of $Q(\cdot)$ curve below $\bar{p}$ must lie on the iso-profit curve until it reaches the boundary of Φ (i.e., either the vertical axis or the curve $Q = D(p) - k$), and (ii) the point $(Q(\bar{p}), \bar{p})$ must be chosen to be the most rightward point on the firm 2's iso-profit curve, i.e., $\bar{Q} = \max\{\pi'(\bar{p}), D(\bar{p}) - k\}$



(a) when k is large (b) when k is small

Figure 4: Optimal $Q(\cdot)$ contingent on Π_2

from Subsection 6.2. Lemma 3 below formalizes these claims and solves (OP') contingent on Π_2. Figures 4(a) and 4(b) graphically show the partial solutions contingent on Π_2 for two examples when firm 2's capacity is large and small, respectively.

Lemma 3. *Contingent on any* $\Pi_2 \in (0, \pi(\max\{p^m, u'(k)\}))$, *(OP') has a solution. For any such solution* $(Q(\cdot), \bar{p})$, $\bar{p}$ *is the unique solution of*

$$\max\{D(\bar{p}) - k, \pi'(\bar{p})\} = D(\bar{p}) - \frac{\Pi_2}{\bar{p} - c} \equiv \bar{Q}, \tag{37}$$

and $Q(p)$ *for* $p \in [x_0, \bar{p}]$ *is given by (26), where* x_0 *is the unique solution in* $(c, \bar{p}]$ *of*

$$\max\{D(x_0) - k, 0\} = D(x_0) - \frac{\Pi_2}{x_0 - c} \equiv Q_0. \tag{38}$$

Moreover, $Q(\cdot)$ *is strictly increasing on* $[x_0, \bar{p}]$.

Remark 1. (Q_0, x_0) *is the unique intersection below* $\bar{p}$ *between the iso-profit curve and the captive demand curve* $Q = \max\{D(p) - k, 0\}$. *Formula (26) gives* $Q(p)$ *only on the interval* $p \in [x_0, \bar{p}]$ *(i.e., only when* $(Q(p), p)$ *belongs to the competitive portion* Φ *of demand and is below the curve* $Q = \pi'(p)$*). How we define* $Q(p)$ *for* $p \notin [x_0, \bar{p}]$ *does not affect* Π_1, *provided* $Q(\cdot)$ *satisfies (Mon) and (F2-IC'). In particular, any monotonic extension of (26) works, e.g., we let* $Q(p) = \bar{Q}$ *for* $p > \bar{p}$ *and* $Q(p) = Q_0$ *for* $c \leq p \leq x_0$, *as shown in Figure 4.*

To solve (OP'), it remains to pin down Π_2, which should be chosen to make the Π_1 area in Figure 4 as large as possible. It turns out that the corresponding first-order condition can be simplified as (39) below. Also, when Π_1 is maximized we must have $D(\bar{p}) - k < \bar{Q}$ so that the left-hand side of (37) reduces to $\pi'(\bar{p})$. These together with Lemma 3 complete the characterization of solutions of (OP'). Once a solution $(Q(\cdot), \bar{p}, \Pi_2)$ of (OP') is obtained, we can use (34) to derive $T(\cdot)$, and then obtain the corresponding solution $(Q(\cdot), T(\cdot), \bar{p})$ of (OP). Thus, we obtain the following lemma.

Lemma 4. *Problem (OP) has at least one solution. Any such solution* $(Q(\cdot), T(\cdot), \bar{p})$, *together with the corresponding* $x_0, \Pi_2, \bar{Q}, Q_0$, *satisfies*[19]

$$\bar{p} - c = e \cdot (x_0 - c) > 0, \tag{39}$$

$$\Pi_2 = (\bar{p} - c)(D(\bar{p}) - \bar{Q}) = (x_0 - c)\min\{D(x_0), k\}, \tag{40}$$

(26)–(28), and

$$T(p) = u(Q_0 + k) - u(k) + \int_{x_0}^{p} t dQ(t) \quad \forall p \in [x_0, \bar{p}]. \tag{41}$$

Finally, we can apply Theorem 1 to characterize the equilibrium outcome of the original game. Theorem 2 follows.

7 Equilibrium Implications

In this section, we first examine the properties of the optimal NLP, including the monotonicity of its marginal and average prices and the impact of the minor firm's capacity constraint on equilibrium outcomes. Then through comparing to a situation where NLP is banned, we demonstrate the (anti)competitive effects of the NLP when employed by the dominant firm.

7.1 Marginal and Average Prices of the Optimal NLP

Proposition 1 (Convex Tariff and Quantity Discounts)**.** *In any equilibrium, firm 1's tariff* τ *is continuously differentiable, strictly increasing, and strictly convex on* $[Q_0, \bar{Q}]$. *Moreover, it exhibits quantity discounts in the sense that* $\tau(Q)/Q$ *is*

[19] Remark 1 also applies here.

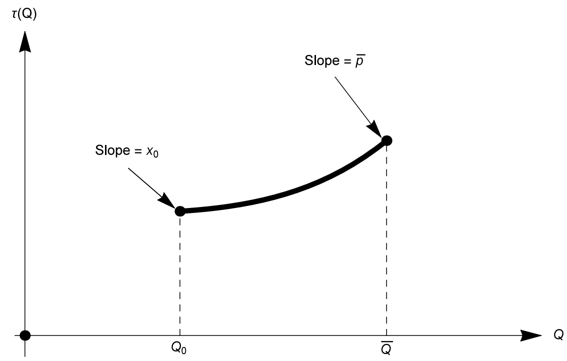


Figure 5: Equilibrium nonlinear pricing

strictly decreasing for $Q \in [Q_0, \bar{Q}]$ *if and only if* $\tau(\bar{Q})/\bar{Q} \geq \bar{p}$*, which is true for all small* $k > 0$*.*

In the absence of asymmetric information, we find that the dominant firm's optimal nonlinear tariff exhibits convexity (or increasing marginal price) in the relevant quantity range $[Q_0, \bar{Q}]$. A typical equilibrium tariff is shown in Figure 5. This is in stark contrast to a typical nonlinear tariff in Maskin and Riley (1984): in Maskin and Riley (1984), under asymmetric information, a monopolist's optimal nonlinear tariff often involves concavity (or decreasing marginal price).

The convexity of the optimal tariff can be understood as follows. Under competition, even though there is no asymmetric information, firm 2 can undercut firm 1's pricing in the relevant price range $[x_0, \bar{p}]$. Since firm 1 wants to induce firm 2 to set $\bar{p}$ and sell $\bar{Q}$ to the buyer in equilibrium, to prevent firm 2 from undercutting below $\bar{p}$ and hence the buyer from buying less than $\bar{Q}$, firm 1 must offer lower and lower marginal prices in case the buyer buys less and less from firm 1. This is why firm 1's optimal tariff's marginal price is increasing on $[Q_0, \bar{Q}]$.

In spite of the convexity, the optimal NLP tariff can meanwhile exhibit quantity discounts (or decreasing average price). Such quantity discounts property holds in the *whole* relevant range $[Q_0, \bar{Q}]$ when the dominant firm's actual average price $\tau(\bar{Q})/\bar{Q}$ is at least as high as the minor firm's average price $\bar{p}$, which is true when k is small.[20] When $k > \hat{k}$ (so that $Q_0 = 0$ and $\tau(Q_0) = 0$), it is obvious to see the optimal NLP must manifest a quantity premium because it is a convex

[20] The proof of Proposition 1 actually reveals that $\tau(Q)/Q$ has a single trough on $[Q_0, \bar{Q}]$. That $\tau(\bar{Q})/\bar{Q} \geq \bar{p}$ is the condition under which the trough is at $\bar{Q}$.

curve passing through the origin. When k is in the intermediate range, there will be a quantity cutoff below which the convex NLP tariff will still display quantity discounts, and above which it will display quantity premiums.

In the real world, NLP has to be simple enough for practical reasons, and thus may not coincide with the optimal NLP we derive there. However, some of them do exhibit the convexity as shown in Proposition 1, such as three-part tariffs and all-units discounts in the neighborhood of the quantity threshold point. Meanwhile, all-units discounts display quantity discounts.

7.2 Impact of Capacity Asymmetry

Now we study how the equilibrium objects change as the minor firm's capacity varies. When k is large enough such that $D(x_0) \leq k$ (or equivalently $Q_0 = 0$), the equilibrium outcome does not vary with k.[21] It is because whenever $D(x_0) < k$ firm 2 does not supply to its full capacity even if it deviates to the lowest relevant deviating price x_0. So the equilibrium objects will vary with k only when k is small. The comparative statics results are as follows.

Proposition 2 (Comparative Statics on k)**.** *There is a unique $\hat{k} \in (D(p^m), q^e)$ such that $Q_0 = 0$ in equilibrium if and only if $k \geq \hat{k}$. The set of equilibrium outcomes is independent of k on $[\hat{k}, \infty]$.*

For $k \in (0, \hat{k}]$, as k increases, the followings hold,[22]

(a) Firm 1's equilibrium profit Π_1 and output $\bar{Q}$ decrease;

(b) Firm 2's equilibrium profit Π_2, price $\bar{p}$ increase; if (30) holds then its output $D(\bar{p}) - \bar{Q}$ increase;

(c) The buyer's equilibrium surplus BS increases when k is small, and decreases when k is close to but below $\hat{k}$;

(d) The equilibrium total surplus TS decreases.

For $k \leq \hat{k}$, an increase in firm 2's capacity always benefits firm 2, and harms firm 1. This is not surprising because firm 2's capacity represents its competitive

[21] This can be seen from (23): The right-hand side of (23) becomes $\frac{1}{e}D\left(c + \frac{\bar{p}-c}{e}\right)$ when $D(x_0) = D\left(c + \frac{\bar{p}-c}{e}\right) \leq k$. It follows that the equilibrium $\bar{p}$ becomes independent of k, so are all other equilibrium objects.

[22] In the proof, we also show that, the equilibrium Q_0 (x_0) is decreasing (increasing) in k; the equilibrium $\Pi_2 + BS$ is decreasing in k when k is close to but below $\hat{k}$.

threat on firm 1. Total surplus of the industry decreases in k. In the limiting case that k approaches zero, firm 1 becomes a perfectly discriminatory monopoly and supplies q^e, and thus the equilibrium total surplus approaches the first-best level. Interestingly, the buyer benefits from the increase in k when k is small, whereas gets harmed from it when k is close to but below $\hat{k}$. This implies, even though firm 2 always has an incentive to increase its capacity as long as $k \leq \hat{k}$, firm 1 will not want this to happen, and the buyer will be on the same stance with firm 2 when k is small, whereas with firm 1 when k is close to but below $\hat{k}$.

7.3 Comparing with LP

In this subsection, let us look at a benchmark case where NLP is banned, and then see why the NLP adopted by the dominant firm can be a harmful practice by comparing the equilibrium outcomes with and without NLP.

Consider a game that is similar to the one we presented in Section 2, except that firm 1 is forced to choose a uniform price. Call it the *LP vs LP game*, and the game presented in Section 2 the *NLP vs LP game*. The LP vs LP equilibrium outcomes are as follows.

Proposition 3 (LP vs LP Equilibrium)**.** *Consider the LP vs LP game. If $k < q^e$, then there is a unique SPE outcome, in which both firms offer $\bar{p}^{LP}$, where $\pi'(\bar{p}^{LP}) = k$, and the buyer purchases $q_1^{LP} = D(\bar{p}^{LP}) - k$ and $q_2^{LP} = k$ units from firm 1 and firm 2 respectively. If $k \geq q^e$, then there are multiple SPE outcomes, in which the prevailing price can be any $\bar{p}^{LP} \in [c, p^m]$ (either $p_1 = p_2 = \bar{p}^{LP} \in [c, p^m]$ or $p_1 \geq p^m = p_2$) and firm 1 makes no sales.*

In the LP vs LP game, the uniform per-unit price from firm 1 is available for the buyer's entire demand, which invites firm 2 to undercut/match if it wants to have some sales. Accordingly, firm 1 will serve the residual demand after the buyer buys all k from firm 2 under LP vs LP.

From Proposition 3, it is easy to see that $\bar{p}^{LP}$ decreases with k for $k < q^e$.[23] Recall from Proposition 2 that $\bar{p}$ increases with k for $k \in (0, \hat{k}]$ and then stays constant for $k \in (\hat{k}, \infty)$. Because $\lim_{k \to 0} \bar{p} = c < p^m = \lim_{k \to 0} \bar{p}^{LP}$ and, when

[23]Other comparative statics results straightforwardly follow. For a full description of the comparative statics for the LP vs LP game, see Corollary 1 in Chao, Tan, and Wong (2018a).

$k = \hat{k}$, $\bar{p} > x_0 = u'(\hat{k}) > \bar{p}^{LP}$, there must exist a unique $\check{k} \in (0, \hat{k})$ such that $\bar{p} \lesseqqgtr \bar{p}^{LP}$ if and only if $k \lesseqqgtr \check{k}$.

Proposition 4 (Comparison)**.** *Let $k \in (0, q^e)$ and compare any SPE outcome of the NLP vs LP game with the unique SPE outcome of the LP vs LP game.*

(a) Quantities: $\bar{Q} > q_1^{LP}$ when $k \in (0, \check{k}]$ or k is close to q^e; $D(\bar{p}) - \bar{Q} < q_2^{LP} = k$;

(b) Profits: $\Pi_1 > \Pi_1^{LP}$; $\Pi_2 < \Pi_2^{LP}$ when $k \in (0, \check{k}]$, and $\Pi_2 > \Pi_2^{LP}$ when $k \in [\hat{k}, q^e)$;

(c) Buyer's Surpluses: $BS < BS^{LP}$ when k is small or $k \in [\hat{k}, q^e)$;

(d) Joint Surpluses: $\Pi_2 + BS < \Pi_2^{LP} + BS^{LP}$; $TS \gtreqqless TS^{LP}$ if and only if $k \lesseqqgtr \check{k}$.

Proposition 4 demonstrates the competitive effects of NLP. As compared with LP, NLP adopted by firm 1 always benefits firm 1, whereas harms firm 2 and the buyer jointly. NLP allows the marginal price for each unit to vary, in contrast to the uniformity under LP. Such flexibility in pricing has two effects: one is the surplus-extraction effect, and the other is the competition-manipulating effect. Thanks to the better instrument in surplus extraction from NLP than from LP, firm 1 has an incentive to expand quantity supplied, which tends to increase the total surplus. Meanwhile, because firm 1 under NLP can customize marginal price for every single unit accordingly to the competitive pressure from firm 2, it will better manipulate competition, which tends to reduce total surplus. Which effect dominates depends on the extent of the dominance.

When k is relatively small, the competitive threat from firm 2 does not concern firm 1 that much, and firm 1's NLP will intensify the competition and extract surplus from firm 2 and the buyer, notwithstanding total surplus is increased. Through general linear demands in the next subsection, we demonstrate that for NLP to have the above exclusionary effect, the k does not have to be really small. But as k becomes sufficiently large, LP vs LP competition would result in an outcome close to zero profits for both firms. Thus, firm 1 will employ the NLP to soften the competition, which benefits firm 2, but hurts the buyer and total surplus.

7.4 An Example of Linear Demands

To demonstrate our analysis above, now we consider an example of linear demands. Suppose that $u(q) = q - q^2/2$ and $c = 0$. Then $D(p) = 1 - p$, $\pi(p) = p \cdot (1 - p)$, and $\pi'(p) = 1 - 2p$ for all $p \in [0, 1]$. Assumptions 1, 2, as well as condition (30) are satisfied, so that the equilibrium is essentially unique and thus the equilibrium conditions in Theorem 2 are not only necessary but also sufficient. Applying Theorem 2 and Proposition 3, we can perform our comparative statics analyses for the full range of $k \in (0, 1]$. It is easy to compute the cutoff above which $Q_0 = 0$ is $\hat{k} = \frac{e^2}{1+e^2}$. All the calculated results are listed in Table A1 in the Appendix.

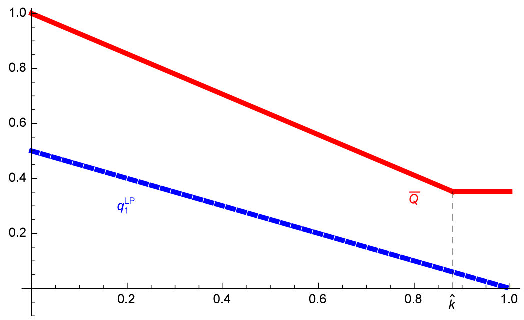
As shown in Figures 6(a) and 6(b), NLP adopted by the dominant firm always increases the dominant firm's sales volume, and decreases the minor firm's. Moreover, when the minor firm is relatively small, e.g., $k \leq e^2/(2 + e^2) \approx 0.79$ in our linear demand example, the minor firm gets partially foreclosed by the dominant firm's NLP, in terms of lower profits, volume sales, and market shares than under LP vs LP equilibrium. This can be seen from Figure 6(c).

As we claim generally in Proposition 2 and as shown in Table A1, all the equilibrium objects, except the buyer's surplus BS in NLP vs LP equilibrium, are monotone in k. Figure 6(d) demonstrates these non-monotone patterns. Due to its non-monotonicity, the buyer gets harmed by the NLP when the minor firm is either relatively small, or sufficiently large in our linear demand example, e.g., either $k \leq e \cdot \left(3e - 2\sqrt{e^2 - 4}\right) / (16 + 5e^2) \approx 0.23$, or $k \geq e \cdot \left(3e + 2\sqrt{e^2 - 4}\right) / (16 + 5e^2) \approx 0.61$.

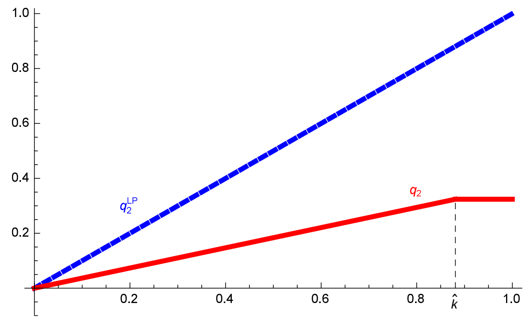
From Figure 6, when the minor firm is capacity constrained, both the minor firm and the buyer are harmed by the dominant firm's NLP. Our results provide some supports to the antitrust concerns about conditional discounts adopted by some dominant firms when competing against small rival firms.

8 Conclusion

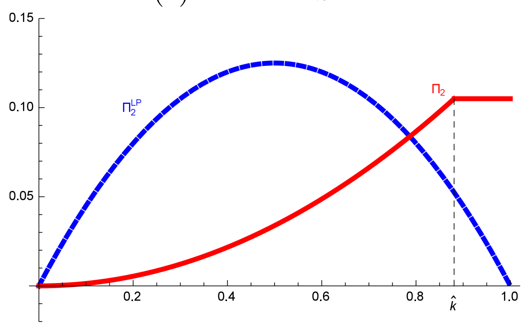
The NLP schedule is naturally viewed as a menu of inducing buyers with private information to self-select. Nevertheless, the main antitrust concern about NLP, when used by dominant firms, is its potential exclusionary effects, which has been



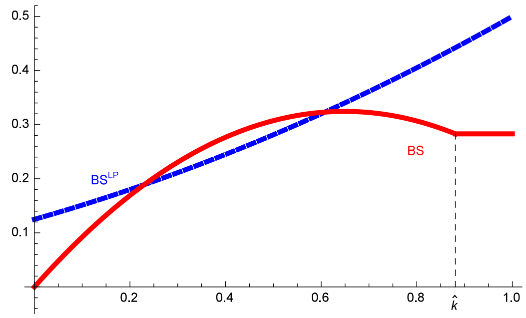
(a) Firm 1's Sales



(b) Firm 2's Sales



(c) Firm 2's Profits



(d) Buyer's Surplus

Figure 6: NLP vs LP and LP vs LP Equilibria for Linear Demand

one of the most controversial issues in antitrust policy and enforcement.

Absent asymmetric information about the buyer, downstream competition, or demand uncertainty, we offer a new motive for NLP when a dominant firm competes with a capacity-constrained rival for one known type buyer: By offering unchosen bundles, the dominant firm can constrain its rival's possible deviations and extract more surplus from the buyer. We characterize the optimal NLP schedule employed by the dominant firm, and find that, such NLP schedule can be both profitable and anticompetitive. The anticompetitive effects of the NLP schedule employed by the dominant firm depend on the extent of the dominant firm's dominance.

Appendix

Proof of Lemma 1. Fix any $Q \in \mathbb{R}_+$. Note that the unique maximizer $\text{Proj}_{[0,k]}(D(p)-Q)$ of the value function $V(Q,p)$ is piecewise continuously differentiable. For any $p \in \mathcal{P}$ at which $\text{Proj}_{[0,k]}(D(p)-Q)$ is differentiable (i.e., $D(p)-Q \neq 0$ and $D(p)-Q \neq k$), clearly $V(Q,p)$ is also differentiable at p and the derivative $V_p(Q,p)$ computed from the Envelope Theorem is given by (8). Moreover, even for $p \in \mathcal{P}$ at which $\text{Proj}_{[0,k]}(D(p)-Q)$ is not differentiable (i.e., $D(p)-Q=0$ or $D(p)-Q=k$), $\text{Proj}_{[0,k]}(D(p)-Q)$ is still continuous; it is clear that the left-derivative and right-derivative of $V(Q,\cdot)$ exist and both are equal to the right-hand side of (8). Thus, $V(Q,\cdot)$ is differentiable and (8) holds. The same logic proves that $V(\cdot,p)$ is differentiable and (9) holds. From (8) and (9), we know $V_p(Q,\cdot)$, $V_p(\cdot,p)$, $V_Q(Q,\cdot)$, and $V_Q(\cdot,p)$ are all piecewise continuously differentiable. In particular, whenever differentiable (i.e., $D(p)-Q \neq 0$ and $D(p)-Q \neq k$), the cross derivatives V_{Qp} and V_{pQ} are given by (10). ∎

The proof of Theorem 1 requires the following two lemmas.

Lemma A1. *For any $Q : \mathcal{P} \to \mathbb{R}_+$, $T : \mathcal{P} \to \mathbb{R}$, and $\bar{p} \in \mathcal{P}$ that satisfy (B-IC), (B-IR), and (F2-IC), there is a $\tau \in \mathcal{T}$ and a SPE of the subgame after firm 1 offers τ such that*

(i) in this SPE of the subgame, firm 2 chooses $p = \bar{p}$, and the buyer, contingent on any firm 2's unit price $p \in \mathcal{P}$, chooses to buy $Q(p)$ and $\text{Proj}_{[0,k]}(D(p)-Q(p))$ units from firm 1 and firm 2 respectively, and

(ii) $\tau(Q(p)) = T(p)$ for all $p \in \mathcal{P}$.

Proof. Suppose that $Q : \mathcal{P} \to \mathbb{R}_+$, $T : \mathcal{P} \to \mathbb{R}$, and $\bar{p} \in \mathcal{P}$ satisfy (B-IC), (B-IR), and (F2-IC). Define

$$\tau(Q) = \begin{cases} T(p) & \text{if } \exists p \in \mathcal{P} \text{ s.t. } Q(p) = Q \\ 0 & \text{if } Q = 0 \text{ and } \nexists p \in \mathcal{P} \text{ s.t. } Q(p) = 0 \\ \infty & \text{otherwise} \end{cases}. \tag{A1}$$

Note that the above τ is well defined because (B-IC) implies $T(p) = T(\tilde{p})$ whenever $Q(p) = Q(\tilde{p})$. Clearly, (ii) holds. To see that $\tau(0) \leq 0$, note that if $\nexists p \in \mathcal{P}$ s.t.

$Q(p) = 0$, then $\tau(0) = 0$; if $Q(\hat{p}) = 0$ for some $\hat{p} \in \mathcal{P}$, then $\tau(0) = T(\hat{p}) \leq 0$, where the inequality follows from (B-IR). Thus, $\tau(0) \leq 0$.

Given this τ and any $p \in \mathcal{P}$, (B-IC) and (B-IR) imply that a buyer's optimal action is to buy $Q(p)$ and $\text{Proj}_{[0,k]}(D(p) - Q(p))$ units from firm 1 and firm 2 respectively. Given τ and that the buyer uses the above strategy, (F2-IC) implies that a firm 2's optimal action is to choose $p = \bar{p}$. Therefore, the strategies in (i) constitute a SPE of the subgame after firm 1 offers τ. It follows that τ is regular and hence $\tau \in \mathcal{T}$. ■

Lemma A2. *For any $\tau \in \mathcal{T}$ and any SPE of the subgame after firm 1 offers τ, if $Q : \mathcal{P} \to \mathbb{R}_+$, $T : \mathcal{P} \to \mathbb{R}$, and $\bar{p} \in \mathcal{P}$ satisfy (i) and (ii) in Lemma A1, then $Q(\cdot), T(\cdot), \bar{p}$ also satisfy (B-IC), (B-IR), and (F2-IC).*

Proof. Take any $\tau \in \mathcal{T}$ and any SPE of the subgame after firm 1 offers τ. Suppose that $Q : \mathcal{P} \to \mathbb{R}_+$, $T : \mathcal{P} \to \mathbb{R}$, and $\bar{p} \in \mathcal{P}$ satisfy (i) and (ii) in Lemma A1. Since the strategies described in (i) constitute a SPE of the subgame after firm 1 offers τ, we have (F2-IC) and

$$V(Q(p), p) - \tau(Q(p)) \geq V(Q, p) - \tau(Q) \quad \forall (Q, p) \in \mathbb{R}_+ \times \mathcal{P}. \tag{A2}$$

To see (B-IC), take $Q = Q(\tilde{p})$ for arbitrary $\tilde{p} \in \mathcal{P}$ in (A2) and use (ii). To see (B-IR), take $Q = 0$ in (A2) and use $\tau(0) \leq 0$ and (ii). ■

Proof of Theorem 1. "Only if" part. Suppose that $(Q^*(\cdot), T^*(\cdot), \bar{p}^*)$ is a solution of (OP). Then $Q^*(\cdot), T^*(\cdot), \bar{p}^*$ satisfy (B-IC), (B-IR), and (F2-IC). From Lemma A1, there is a $\tau^* \in \mathcal{T}$ (defined by (A1) with $\tau(\cdot), Q(\cdot), T(\cdot)$ replaced by $\tau^*(\cdot), Q^*(\cdot), T^*(\cdot)$) such that (21) holds and a SPE $(p^*(\tau^*), q^*(\tau^*, \cdot))$ of the subgame after firm 1 offers τ^* is described by (19), (20), and (22).

In the subgame after firm 1 offers this τ^*, we let firm 2 and the buyer play the SPE $(p^*(\tau^*), q^*(\tau^*, \cdot))$, so that firm 1's profit is $T^*(\bar{p}^*) - c \cdot Q^*(\bar{p}^*)$. In the subgame after firm 1 offers any other $\tau \in \mathcal{T} \backslash \{\tau^*\}$, we let firm 2 and the buyer play any SPE $(p^*(\tau), q^*(\tau, \cdot))$, which exists because every $\tau \in \mathcal{T}$ is regular. By such constructions, p^*, q^* satisfy (1) and (2).

From Lemma A2, the SPE outcome of the subgame after firm 1 offers an arbitrary $\tau \in \mathcal{T}$ must be characterized by some $Q(\cdot), T(\cdot), \bar{p}$ that satisfy (B-IC), (B-IR), and (F2-IC), and the associated firm 1's profit is $T(\bar{p}) - c \cdot Q(\bar{p})$. Since $(Q^*(\cdot), T^*(\cdot), \bar{p}^*)$ is a solution of (OP), firm 1 cannot make strictly higher profit than $T^*(\bar{p}^*) - c \cdot Q^*(\bar{p}^*)$ by offering any $\tau \in \mathcal{T}$. That is, (τ^*, p^*, q^*) satisfies (3) and hence is a SPE of the whole game.

"If" part. Let $(Q^{**}(\cdot), T^{**}(\cdot), \bar{p}^{**})$ denote the solution of (OP) given by Lemma 4, and Π_1^{**} the maximum value of (OP).[24] Suppose that (τ^*, p^*, q^*) is a SPE and $Q^*(\cdot), T^*(\cdot), \bar{p}^*$ satisfy (19), (20), (21), and (22). From Lemma A2, $Q^*(\cdot), T^*(\cdot), \bar{p}^*$ satisfy (B-IC), (B-IR), and (F2-IC). In the SPE (τ^*, p^*, q^*), firm 1's profit is $\Pi_1^* = T^*(\bar{p}^*) - cQ^*(\bar{p}^*)$. Also suppose, by way of contradiction, that $(Q^*(\cdot), T^*(\cdot), \bar{p}^*)$ is not a solution of (OP). It follows that $\Pi_1^* < \Pi_1^{**}$. We shall show that firm 1 then can offer a tariff in $\mathcal{T}$ that guarantees itself a profit arbitrarily close to Π_1^{**} in every SPE of the firm 2-buyer subgame that follows. Once this is proved, offering such a tariff is a firm 1's profitable deviation in the SPE (τ^*, p^*, q^*), which is a contradiction.[25]

To do that, we perturb the solution $(Q^{**}(\cdot), T^{**}(\cdot), \bar{p}^{**})$ so that firm 2 would have to lower its price a bit more if it wishes to increase its sales by any given amount. We can keep $\bar{p}^{**}$ unchanged and, for any $\varepsilon > 0$, let

$$Q_\varepsilon(p) = \begin{cases} Q^{**}(p) & \text{if } p \geq \bar{p}^{**} \\ Q^{**}(\bar{p}^{**}) & \text{if } \bar{p}^{**} - \varepsilon < p < \bar{p}^{**} \\ Q^{**}(p + \varepsilon) & \text{if } p \leq \bar{p}^{**} - \varepsilon \end{cases},$$

$$T_\varepsilon(p) = V(Q_\varepsilon(p), p) - V(0, c) - \int_c^p V_p(Q_\varepsilon(t), t) dt.$$

From the analyses in Section 6, $(Q_\varepsilon(\cdot), T_\varepsilon(\cdot), \bar{p}^{**})$ satisfies all the constraints of (OP); the (F2-IC) constraint holds strictly at every $p \neq \bar{p}^{**}$; the value of (OP) evaluated at $(Q_\varepsilon(\cdot), T_\varepsilon(\cdot), \bar{p}^{**})$ is arbitrarily close to the maximum value Π_1^{**} when

[24] Admittedly, we use the results in Section 6, which appear later than Theorem 1, to prove the "if" part of Theorem 1. However, there is no circularity of reasoning because the analyses in Section 6 do not rely on Theorem 1. (One can always formally analyze (OP) as in Section 6 even if Theorem 1 is not true.)

[25] Here we cannot simply use the tariff implied by the solution $(Q^{**}(\cdot), T^{**}(\cdot), \bar{p}^{**})$ of (OP) (see Lemma 4) because (i) if firm 1 offers this tariff, then firm 2 is indifferent between offering $\bar{p}^{**}$ and offering some lower price (note that (F2-IC) is binding over a range), and (ii) if firm 2 does offer some lower price, then firm 1's profit is strictly lower than Π_1^{**}.

ε is made arbitrarily small.

Define $\tau_\varepsilon(\cdot)$ by the right-hand side of (A1) with $Q(\cdot)$ and $T(\cdot)$ replaced by $Q_\varepsilon(\cdot)$ and $T_\varepsilon(\cdot)$. Now, if firm 1 offers τ_ε, the best responses of the buyer and firm 2 are unique. In particular, firm 2 would surely offer $\bar{p}^{**}$; the buyer would surely purchase $Q_\varepsilon(\bar{p}^{**})$ from firm 1; firm 1's profit would surely be the value of (OP) evaluated at $(Q_\varepsilon(\cdot), T_\varepsilon(\cdot), \bar{p}^{**})$. Therefore, offering τ_ε with small enough $\varepsilon > 0$ is a firm 1's profitable deviation as desired. ■

Proof of Lemma 2. We shall first show that (B-IC) is equivalent to (Mon) and (31), then establish that, given (Mon), (B-IR) is equivalent to (32). Let $U(p) \equiv V(Q(p), p) - T(p)$. Then (B-IC) can be written as

$$U(p) - U(\tilde{p}) \geq V(Q(\tilde{p}), p) - V(Q(\tilde{p}), \tilde{p}) \;\; \forall p, \tilde{p} \in \mathcal{P}, \tag{A3}$$

and (31) can be written as

$$U(p) - U(c) = \int_c^p V_p(Q(t), t) dt \;\; \forall p \in \mathcal{P}. \tag{A4}$$

Step 1. (B-IC) implies (Mon) and (31).

Suppose (B-IC) is satisfied. Then (A3) implies that, for any $p_1, p_2 \in \mathcal{P}$,

$$V(Q(p_1), p_2) - V(Q(p_1), p_1) \leq U(p_2) - U(p_1) \leq V(Q(p_2), p_2) - V(Q(p_2), p_1). \tag{A5}$$

If (Mon) does not hold, then there exist $p_1, p_2 \in \mathcal{P}$ such that $p_1 < p_2$ and $Q(p_1) > Q(p_2)$ and $D(p_1) > Q(p_2)$ and $Q(p_1) > D(p_2) - k$. But then (A5) implies

$$\begin{aligned} 0 &\geq [V(Q(p_1), p_2) - V(Q(p_1), p_1)] - [V(Q(p_2), p_2) - V(Q(p_2), p_1)] \\ &= \int_{p_1}^{p_2} \int_{Q(p_2)}^{Q(p_1)} V_{pQ}(Q, p) dQ dp > 0, \end{aligned}$$

which is a contradiction. The above equality holds because, from Lemma 1, $V(Q, \cdot)$ is continuously differentiable and $V_p(\cdot, p)$ is piecewise continuously differentiable (and hence they are absolutely continuous on any compact interval). The last inequality holds because, first, $V_{pQ} \geq 0$ almost everywhere and $V_{pQ} = 1$ on the interior of Φ; second, in the Q-p space, the point $(Q(p_2), p_1)$ is strictly below the curve $Q = D(p)$ (from $D(p_1) > Q(p_2)$) and the point $(Q(p_1), p_2)$ is strictly above the curve $Q = D(p) - k$ (from $Q(p_1) > D(p_2) - k$), so the rectangle $[Q(p_2), Q(p_1)] \times [p_1, p_2]$ must intersect the interior of Φ, on which $V_{pQ} > 0$. Therefore, (Mon) must hold.

Moreover, (A5) implies (A4). Therefore, (31) holds.

Step 2. (Mon) and (31) imply (B-IC).

First, (Mon) implies that, for all $p_1, p_2 \in \mathcal{P}$ with $p_1 \leq p_2$, we have

$$\mathrm{Proj}_{[0,k]}(D(p_2) - Q(p_1)) \geq \mathrm{Proj}_{[0,k]}(D(p_2) - Q(p_2)), \tag{A6}$$

$$\mathrm{Proj}_{[0,k]}(D(p_1) - Q(p_1)) \geq \mathrm{Proj}_{[0,k]}(D(p_1) - Q(p_2)). \tag{A7}$$

Indeed, $p_1 \leq p_2$ and (Mon) imply either (i) $Q(p_1) \leq Q(p_2)$, or (ii) $D(p_1) \leq Q(p_2)$, or (iii) $Q(p_1) \leq D(p_2) - k$. In case (i), clearly (A6) and (A7) hold. In case (ii), we have $D(p_2) \leq D(p_1) \leq Q(p_2)$ so that the right-hand sides of (A6) and (A7) are 0. In case (iii), we have $Q(p_1) + k \leq D(p_2) \leq D(p_1)$ so that the left-hand sides of (A6) and (A7) are $k > 0$. Therefore, (A6) and (A7) hold in each case.

Recall that (31) is equivalent to (A4). Therefore, for any $p_1, p_2 \in \mathcal{P}$ (no matter whether $p_1 \leq p_2$ or not), we have

$$\begin{aligned}
U(p_2) - U(p_1) &= \int_{p_1}^{p_2} V_p(Q(p), p)dp \\
&= -\int_{p_1}^{p_2} \mathrm{Proj}_{[0,k]}(D(p) - Q(p))dp \\
&\geq -\int_{p_1}^{p_2} \mathrm{Proj}_{[0,k]}(D(p) - Q(p_1))dp \\
&= \int_{p_1}^{p_2} V_p(Q(p_1), p)dp \\
&= V(Q(p_1), p_2) - V(Q(p_1), p_1),
\end{aligned}$$

where the inequality is from (A6) when $p_1 \leq p_2$ and from (A7) when $p_1 \geq p_2$. It proves (A3) and hence (B-IC).

Step 3. Given (B-IC) (in fact, (31) only), (B-IR) is equivalent to (32).

It suffices to show that $V(Q(p), p) - T(p) - V(0, p) = U(p) - V(0, p)$ is non-decreasing in p on $\mathcal{P}$. Indeed, from (31), which is equivalent to (A4), and Lemma

1, we know both $U(\cdot)$ and $V(0,\cdot)$ are differentiable, and $U'(p) = V_p(Q(p),p) \geq V_p(0,p)$. Therefore, (B-IR) is equivalent to (32). ∎

Proof of Corollary 2. It is implied by (A6) in the proof of Lemma 2. ∎

Proof of Lemma 3. Fix any $\Pi_2 \in (0, \pi(\max\{p^m, u'(k)\}))$ and hence a firm 2's iso-profit curve in the Q-p space (see Figure 2). Constraint (F2-Pro) requires that $(Q(\bar{p}), \bar{p})$ must be on the iso-profit curve. Constraint (F2-IC') requires that the graph of $Q(\cdot)$ must not cut into the left side of the iso-profit curve.

From Subsection 6.2, we know: the iso-profit curve, which contains $(Q(\bar{p}), \bar{p})$, is strictly below the demand curve and strictly above the cost line, so that $Q(\bar{p}) < D(\bar{p})$ and $\bar{p} > c$. Also, the iso-profit curve is (horizontally) single-peaked, with its unique most rightward point satisfying $Q = \max\{D(p) - k, \pi'(p)\}$.

Here we prove by contradiction that $(Q(\bar{p}), \bar{p})$ must be the most rightward point (horizontal peak) of the iso-profit curve, i.e.,

$$Q(\bar{p}) = \max\{D(\bar{p}) - k, \pi'(\bar{p})\}). \tag{A8}$$

Suppose not. Consider the case that $Q(\bar{p}) > \max\{D(\bar{p}) - k, \pi'(\bar{p})\}$ (i.e., $(Q(\bar{p}), \bar{p})$ lies on the strictly decreasing portion of the iso-profit curve). Pick a small $\varepsilon > 0$ such that $Q(\bar{p}) > \max\{D(\bar{p} - \varepsilon) - k, \pi'(\bar{p} - \varepsilon)\}$ and $\bar{p} - \varepsilon > c$. To satisfy constraint (F2-IC'), $Q(\bar{p} - \varepsilon)$ must satisfy $Q(\bar{p} - \varepsilon) > Q(\bar{p})$. But then constraint (Mon) is violated because now $Q(\bar{p} - \varepsilon) > Q(\bar{p}) > D(\bar{p}) - k$ and $D(\bar{p} - \varepsilon) \geq D(\bar{p}) > Q(\bar{p})$. Now consider the case that $Q(\bar{p}) < \max\{D(\bar{p}) - k, \pi'(\bar{p})\}$ (i.e., $(Q(\bar{p}), \bar{p})$ lies on the non-decreasing portion of the iso-profit curve). Then, Π_1 can be raised by increasing both $Q(\bar{p})$ and $\bar{p}$ along the iso-profit curve toward the horizontal peak (see Figures 2 and 3). It proves (A8).

Since $(Q(\bar{p}), \bar{p})$ is on the iso-profit curve (i.e., $\Pi_2 = (\bar{p} - c)\,\mathrm{Proj}_{[0,k]}(D(\bar{p}) - Q(\bar{p}))$) and $D(\bar{p}) - k \leq Q(\bar{p}) < D(\bar{p})$, we have

$$\Pi_2 = (\bar{p} - c)(D(\bar{p}) - Q(\bar{p})), \tag{A9}$$

which is equivalent to (26) at $p = \bar{p}$. It, together with (A8), proves (37) and $\bar{Q} = Q(\bar{p})$.

Let x_0 and Q_0 be defined as in the lemma (i.e., defined by $x_0 \leq \bar{p}$ and (38)).

That is, (Q_0, x_0) is the intersection below $\bar{p}$ between the iso-profit curve and the curve $Q = \max\{D(p) - k, 0\}$.[26] In particular, $c < x_0 \leq \bar{p}$ and $Q_0 \leq \bar{Q}$. Now, recall that Π_1 can be written as (36) and visualized in Figure 3. In order to maximize Π_1, it remains to maximize the second integral in (36) (or Area B in Figure 3) subject to (Mon) and (F2-IC'). Neglect (Mon) for a moment. Then, $Q(\cdot)$ on $[x_0, \bar{p}]$ must coincide the iso-profit curve, i.e., (26) holds, for otherwise Π_1 can be improved by shifting the part of $Q(\cdot)$ on $[x_0, \bar{p}]$ that does not match with the iso-profit curve toward the latter. How we define $Q(p)$ for $p \notin [x_0, \bar{p}]$ does not affect Π_1, but those values have to be defined such that $Q(\cdot)$ satisfies (Mon) and (F2-IC') on $\mathcal{P}$. One way is: let $Q(p) = \bar{Q}$ for $p > \bar{p}$ and $Q(p) = Q_0$ for $c \leq p \leq x_0$, as shown in Figure 4. Then, $Q(\cdot)$ is non-decreasing on $\mathcal{P}$ so that (Mon) is satisfied. It is also clear from Figure 4 that (F2-IC') is satisfied. It proves (26). Finally, (26) and Assumption 2 imply that $Q(\cdot)$ is strictly increasing on $[x_0, \bar{p}]$. ■

Proof of Lemma 4. Lemma 3 has characterized the optimal $(Q(\cdot), \bar{p})$ contingent on any $\Pi_2 \in (0, \pi(\max\{p^m, u'(k)\}))$. Clearly, the maximum Π_1 contingent on $\Pi_2 = 0$ is equal to the limiting contingent maximum Π_1 as $\Pi_2 \downarrow 0$ (which is equal to $u(\max\{q^e - k, 0\}) - c \cdot \max\{q^e - k, 0\}$), and the maximum Π_1 contingent on $\Pi_2 = \pi(\max\{p^m, u'(k)\})$ is equal to the limiting contingent maximum Π_1 as $\Pi_2 \uparrow \pi(\max\{p^m, u'(k)\})$ (which is equal to 0). After reducing the first stage (where $(Q(\cdot), \bar{p})$ is chosen contingent on Π_2), (OP') has only one choice variable, Π_2, and the reduced objective function is continuous in Π_2 on $[0, \pi(\max\{p^m, u'(k)\})]$. Thus, (OP') and hence (OP) has at least one solution.

If $\Pi_2 = 0$, then the contingent maximum can be raised by increasing Π_2 (contemplating an upward-and-leftward shift of $Q(\cdot)$ to a higher firm 2's iso-profit curve in Figure 4). Thus, at any optimum, $\Pi_2 > 0$, which implies $\bar{p} > c$ and $D(\bar{p}) > \bar{Q}$. On the other hand, if Π_2 is $\pi(\max\{p^m, u'(k)\})$ or is so large that the contingent solution exhibits $D(\bar{p}) - \bar{Q} = k$, then the contingent maximum can be raised by decreasing Π_2 (contemplating a downward-and-rightward shift of $Q(\cdot)$ to a lower firm 2's iso-profit curve in Figure 4). Thus, at any optimum, it holds that $\Pi_2 < \pi(\max\{p^m, u'(k)\})$ and $\bar{Q} > \max\{D(\bar{p}) - k, 0\}$, which in turn imply $Q_0 < \bar{Q}$, and $c < x_0 < \bar{p} < p^m$ (see Figure 4 again). So (40) and (26)–(28) follow from

[26] There is another intersection above $\bar{p}$ (which is on the vertical axis), but the intersection below $\bar{p}$ is uniquely given by (38).

Lemma 3.

Next, we show (39). From Figure 3, (36) can be rewritten as

$$\Pi_1 = \int_0^{Q_0} u'(Q+k)dQ + x_0 \cdot \left(\bar{Q} - Q_0\right) + \int_{x_0}^{\bar{p}} \left(\bar{Q} - Q(p)\right) dp - c\bar{Q} \tag{A10}$$
$$= \int_0^{Q_0} \left[u'(Q+k) - x_0\right] dQ + (\bar{p} - c)\bar{Q} - \int_{x_0}^{\bar{p}} Q(p)dp$$
$$= \int_{x_0}^{\infty} \max\{D(p) - k, 0\}dp + (\bar{p} - c)\bar{Q} - \int_{x_0}^{\bar{p}} \left[D(p) - \frac{\Pi_2}{p - c}\right] dp,$$

where the last equality is due to (28) and (26). Let

$$TS(p) \equiv u(D(p)) - c \cdot D(p) = \int_p^{\infty} D(t)dt + (p - c)D(p) \tag{A11}$$

denote the total surplus under linear pricing p. From (27) and the first equality of (40),

$$TS(\bar{p}) = \int_{\bar{p}}^{\infty} D(p)dp + (\bar{p} - c)D(\bar{p}) = \int_{\bar{p}}^{\infty} D(p)dp + (\bar{p} - c)\bar{Q} + \Pi_2,$$

so that we can further rewrite Π_1 as

$$\Pi_1 = \int_{x_0}^{\infty} \max\{D(p) - k, 0\}dp + TS(\bar{p}) - \Pi_2 - \int_{x_0}^{\infty} D(p)dp + \int_{x_0}^{\bar{p}} \frac{\Pi_2}{p - c}dp$$
$$= TS(\bar{p}) - \int_{x_0}^{\infty} \min\{D(p), k\}dp + \left(\ln \frac{\bar{p} - c}{x_0 - c} - 1\right) \Pi_2. \tag{A12}$$

The partial derivatives of (A12) are

$$\frac{\partial \Pi_1}{\partial \bar{p}} = (\bar{p} - c)D'(\bar{p}) + \frac{\Pi_2}{\bar{p} - c},$$

$$\frac{\partial \Pi_1}{\partial x_0} = \min\{D(x_0), k\} - \frac{\Pi_2}{x_0 - c},$$

$$\frac{\partial \Pi_1}{\partial \Pi_2} = \ln \frac{\bar{p} - c}{x_0 - c} - 1.$$

Note that (40) and (27) imply that $\partial \Pi_1 / \partial \bar{p} = \partial \Pi_1 / \partial x_0 = 0$. Hence, the total derivative of (A12) with respect to Π_2 is

$$\frac{d\Pi_1}{d\Pi_2} = \ln \frac{\bar{p} - c}{x_0 - c} - 1. \tag{A13}$$

Therefore, the first-order condition $d\Pi_1/d\Pi_2 = 0$ implies (39). Substituting this first-order condition into (A12), the maximum Π_1 can be written as

$$\Pi_1 = TS(\bar{p}) - \int_{x_0}^{\infty} \min\{D(p), k\} dp. \tag{A14}$$

Last, we derive (41). Note that $T(\bar{p}) = \Pi_1 + c\bar{Q}$. Use the expression (A10) for Π_1, we have

$$\begin{aligned} T(\bar{p}) &= \int_0^{Q_0} u'(Q+k)dQ + x_0 \cdot \left(\bar{Q} - Q_0\right) + \int_{x_0}^{\bar{p}} \left(\bar{Q} - Q(p)\right) dp \\ &= \int_0^{Q_0} u'(Q+k)dQ + \bar{p}\bar{Q} - x_0 Q_0 - \int_{x_0}^{\bar{p}} Q(p)dp \\ &= u(Q_0 + k) - u(k) + \int_{x_0}^{\bar{p}} p dQ(p). \end{aligned} \tag{A15}$$

From (26), we know $Q(\cdot)$ is differentiable on $(x_0, \bar{p})$. Then, from (34), $T(\cdot)$ is differentiable on $(x_0, \bar{p})$ as well, and for all $p \in (x_0, \bar{p})$,

$$\begin{aligned} T'(p) &= V_Q(Q(p), p) \cdot Q'(p) \\ &= u'(\mathrm{Proj}_{[Q(p), Q(p)+k]}(D(p))) \cdot Q'(p) \\ &= u'(D(p)) \cdot Q'(p) \\ &= p \cdot Q'(p), \end{aligned} \tag{A16}$$

where the second equality is from (9); the third equality is from

$$\begin{aligned} D(p) - k &\leq Q(x_0) \ (\because p > x_0 \text{ and } (28)) \\ &\leq Q(p) \ (\because Q(\cdot) \text{ is increasing on } [x_0, \bar{p}]) \end{aligned}$$

and

$$\begin{aligned} D(p) &\geq Q(\bar{p}) \ (\because p < \bar{p} \text{ and } D(\bar{p}) > Q(\bar{p})) \\ &\geq Q(p) \ (\because Q(\cdot) \text{ is increasing on } [x_0, \bar{p}]). \end{aligned}$$

From (34), we know $T(\cdot)$ is absolutely continuous on $[x_0, \bar{p}]$, so that $T(p) = T(\bar{p}) - \int_p^{\bar{p}} T'(t)dt$ for all $p \in [x_0, \bar{p}]$. It, together with (A15) and (A16), proves (41). ■

Proof of Theorem 2. The results are from Lemma 4 and Theorem 1. (26)–(28) are already stated in Lemma 4. (25) is from (39). (24) follows from (27) and the first equality of (40). Substituting (25) and (27) into the second equality of (40) and cancelling the positive factor $\bar{p} - c$, we obtain (23). Finally, the formula (29) for $\tau(Q)$ with $Q \in [Q_0, \bar{Q}]$ is derived from (41) through changing of variable: $x(\cdot)$ for $Q(\cdot)$. ■

Proof of Corollary 1. Under condition (30), the left-hand side of (23) is strictly increasing in $\bar{p}$, and the right-hand side is non-increasing in $\bar{p}$. Therefore, (23) has at most one solution under (30). ■

Proof of Proposition 1. From (29), $\tau'(\cdot) = x(\cdot)$ on $[Q_0, \bar{Q}]$. Since $x(\cdot)$, define on $[Q_0, \bar{Q}]$, is the inverse of $Q(\cdot)$ on $[x_0, \bar{p}]$ and the latter is continuous and strictly increasing and on $[Q_0, \bar{Q}]$, $x(\cdot)$ is also continuous and strictly increasing. The range of $x(\cdot)$ is $[x_0, \bar{p}]$ and $x_0 > c \geq 0$. The first part of the proposition follows.

To see the second part, first recall that $\tau'(\bar{Q}) = x(\bar{Q}) = \bar{p}$ and observe that

$$\frac{d}{dQ}\left(\frac{\tau(Q)}{Q}\right) = \frac{f(Q)}{Q},$$

where $f(Q) \equiv \tau'(Q) - \tau(Q)/Q$. Since

$$f'(Q) = \tau''(Q) - \frac{d}{dQ}\left(\frac{\tau(Q)}{Q}\right) = \tau''(Q) - \frac{f(Q)}{Q},$$

we have $f'(Q) = \tau''(Q) > 0$ whenever $f(Q) = 0$. In other words, $f(Q)$ crosses zero from below once and only once, if it does. So a necessary and sufficient condition for $f(Q) < 0$ on $[Q_0, \bar{Q})$ is $f(\bar{Q}) \leq 0$, so is the case for $d(\tau(Q)/Q)/dQ < 0$ on $[Q_0, \bar{Q})$.

Finally, as $k \to 0$, both $\bar{p}$ and x_0 approach c (from (23) and (25)), both $\bar{Q}$ and Q_0 approach q^e (from (27) and (28)), and $\tau(\bar{Q})$ approaches $u(q^e)$ (from (29)). It follows that

$$\lim_{k\to 0}\left(\frac{\tau(\bar{Q})}{\bar{Q}} - \bar{p}\right) = \frac{u(q^e)}{q^e} - c > 0.$$

It shows that $\tau(\bar{Q})/\bar{Q} > \bar{p}$ for all small $k > 0$.[27] ■

The proof of Proposition 2 and Proposition 4 require the following lemma, which states that what firm 2 and the buyer jointly earn in equilibrium is equal to their joint outside option under the counterfactual situation that firm 2's unit cost was raised to x_0.

Lemma A3. *In any equilibrium,*

$$\begin{aligned}\Pi_2 + BS &= \int_{x_0}^{\infty} \min\{D(p), k\}dp \qquad \text{(A17)}\\ &= u(\min\{D(x_0), k\}) - x_0 \cdot \min\{D(x_0), k\}\\ &= u(D(x_0) - Q_0) - x_0 \cdot (D(x_0) - Q_0).\end{aligned}$$

Proof. The first equality follows from (A14) and the equilibrium buyer's surplus as $BS \equiv u(D(\bar{p})) - \bar{p}(D(\bar{p}) - \bar{Q}) - \tau(\bar{Q}) = TS - \Pi_1 - \Pi_2$. The second equality is clear. The third equality is from (28). ■

Proof of Proposition 2. Let $\hat{x}_0$ be the minimum equilibrium x_0 when $k = \infty$, given by (23) and (25) with $\min\{D(x_0), k\} = D(x_0)$. Define $\hat{k} \equiv D(\hat{x}_0)$. From Theorem 2, $\hat{k}$ satisfies the first two claims (see Figure 4).

The rest of the proof considers comparative statics for $k \in (0, \hat{k}]$. Following the proof of Lemma 4, we regard $\Pi_1, \bar{p}, \bar{Q}, x_0, Q_0, x(\cdot), BS, TS$ as functions of Π_2. Here we also regard them as functions of k. In particular, we write $\Pi_1(\Pi_2; k)$.

Fix Π_2 and let k increase on $(0, \hat{k}]$. Note that $Q_0 = \max\{D(x_0) - k, 0\} > 0$ before the increase, so that we have $D(x_0) > k$ before the increase. The $\bar{p}$ and $\bar{Q}$ determined by (27) and the first equality of (40) do not change. The x_0, Q_0,

[27] As a matter of fact, when $k \geq \hat{k}$ or k is smaller than but close to $\hat{k}$, the tariff τ exhibits quantity premiums, i.e., $\tau'(Q)/Q$ is strictly increasing on $[Q_0, \bar{Q}]$. It follows from the fact that τ is strictly convex on $[Q_0, \bar{Q}]$, $\tau(0) = 0$, and that $Q_0 = 0$ when $k \geq \hat{k}$.

and Π_1 determined by the second equality of (40), (28), and (A12) decrease as k increases (see Figure 4).

In equilibrium, $\Pi_1 = \max_{\Pi_2}\{\Pi_1(\Pi_2; k)\}$ decreases in k, because $\Pi_1(\cdot; k)$ shifts down as k increases. From (A13), we see that $\partial\Pi_1(\Pi_2; k)/\partial\Pi_2$ increases, because $\bar{p}$ is unchanged whereas x_0 decreases when we fix Π_2 and let k increase. In other words, $\Pi_1(\Pi_2; k)$ satisfies strict increasing differences. Therefore, the Π_2 that maximizes Π_1 must increase when k increases. Then, from (24) and Assumption 2, $\bar{p}$ must increase, and hence $\bar{Q}$ decreases follows from (27). Then from (A11), TS decreases. From (39), x_0 increases. From (28), Q_0 decreases. The result for $D(\bar{p})-\bar{Q}$ can be immediately seen from $D(\bar{p})-\bar{Q} = -(\bar{p}-c)D'(\bar{p})$. This completes the proof of parts (a), (b), and (d).

Last, we prove part (c). To see the first half of part (c), note that BS is positive and tends to zero as $k \to 0$. To see the second half of part (c), first note that, as shown above, we have $\min\{D(x_0), k\} = k$ when $k \leq \hat{k}$. From Lemma A3, $\Pi_2 + BS = u(k) - x_0 k$ whenever $k \leq \hat{k}$. Hence,

$$\left.\frac{d(\Pi_2 + BS)}{dk}\right|_{k\nearrow\hat{k}} = u'(\hat{k}) - x_0 - \hat{k}\cdot\left.\frac{dx_0}{dk}\right|_{k\nearrow\hat{k}} < 0.$$

The last inequality follows from $u'(\hat{k}) - x_0 \leq u'(\hat{k}) - \hat{x}_0 = 0$ and $\left.\frac{dx_0}{dk}\right|_{k\nearrow\hat{k}} > 0$. Therefore, $\Pi_2 + BS$ is decreasing in k when k is close to but below $\hat{k}$. This is true for BS as well, because Π_2 is increasing in k. ■

Proof of Proposition 3. Straightforward and omitted. ■

Proof of Proposition 4. Recall $D(\bar{p})-\bar{Q} < k$ from Theorem 2. This, together with the definition of $\check{k}$, implies that $\bar{Q} > D(\bar{p}) - k \geq D(\bar{p}^{LP}) - k$ when $k \in (0, \check{k}]$. As $k \nearrow q^e$, $D(\bar{p}^{LP}) - k$ tends to zero and $\bar{Q} > 0$. Also recall $q_1^{LP} = D(\bar{p}^{LP}) - k$ and $q_2^{LP} = k$ from Proposition 3. Part (a) follows.

Clearly, $\Pi_1 > \Pi_1^{LP}$ hold. From Theorem 2 and Proposition 3, we know $\Pi_2 = (\bar{p}-c)(D(\bar{p}) - \bar{Q})$ and $\Pi_2^{LP} = (\bar{p}^{LP} - c)k$. Then part (a) and the definition of $\check{k}$ imply $\Pi_2 < \Pi_2^{LP}$ when $k \in (0, \check{k}]$.

When $k \in [\hat{k}, q^e)$, (26) evaluated at $p = x_0$ implies $\Pi_2 = (\hat{x}_0 - c)k$, where $\hat{x}_0$ is (as in the proof of Proposition 2) the minimum equilibrium x_0 when $k = \infty$.

Therefore, $\Pi_2^{LP} = (\bar{p}^{LP} - c)k < \Pi_2$ since $\bar{p}^{LP} < u'(k) \leq u'(\hat{k}) = \hat{x}_0$. It proves the result for Π_2, Π_2^{LP} when $k \in [\hat{k}, q^e)$. It completes the proof of part (b).

The result for TS, TS^{LP} follows from the definition of $\check{k}$ and the fact that the total output is equal to the demand $D(\cdot)$ evaluated at firm 2's price. Moreover,

$$\begin{aligned}
\Pi_2^{LP} + BS^{LP} &= \left(\bar{p}^{LP} - c\right) k + \int_{\bar{p}^{LP}}^{\infty} D(p)dp \\
&\geq \int_{c}^{\infty} \min\left\{D(p), k\right\} dp \\
&> \int_{x_0}^{\infty} \min\left\{D(p), k\right\} dp \ (\because x_0 > c) \\
&= \Pi_2 + BS \ (\because (A17)).
\end{aligned}$$

This completes the proof of part (d).

Compare $BS = TS - \Pi_1 - \Pi_2$ and $BS^{LP} = TS^{LP} - \Pi_1^{LP} - \Pi_2^{LP}$. When $k \in [\hat{k}, q^e)$, our previous results that $TS < TS^{LP}$, $\Pi_1 > \Pi_1^{LP}$, and $\Pi_2 > \Pi_2^{LP}$ together imply $BS < BS^{LP}$. As $k \searrow 0$, from (A17), BS tends to zero but BS^{LP} is positive. Therefore, we also have $BS < BS^{LP}$ when k is small. It completes the proof of part (c). ■

Table A1: Linear Demand Example (Note: $\hat{k} = \frac{e^2}{1+e^2}$.)

	NLP vs LP Equilibrium			
Pricing	$\mathbf{x_0}$	$\mathbf{Q_0}$	$\bar{\mathbf{p}}$	$\bar{\mathbf{Q}}$
	$\frac{1}{e^2}\min\{k,\hat{k}\}$	$\frac{1+e^2}{e^2}\max\{\hat{k}-k,0\}$	$\frac{1}{e}\min\{k,\hat{k}\}$	$1-\frac{2}{e}\min\{k,\hat{k}\}$
Surplus	$\mathbf{\Pi_1}$		$\mathbf{\Pi_2}$	
	$\frac{1}{2(1+e^2)}+\frac{1+e^2}{2e^2}(\max\{\hat{k}-k,0\})^2$		$\frac{1}{e^2}(\min\{k,\hat{k}\})^2$	
	BS		**TS**	
	$\min\{k,\hat{k}\}-\frac{4+e^2}{2e^2}(\min\{k,\hat{k}\})^2$		$\frac{1}{2}-\frac{1}{2e^2}(\min\{k,\hat{k}\})^2$	
	LP vs LP Equilibrium			
Pricing	$\mathbf{p_1^{LP}}$	$\mathbf{p_2^{LP}}$	$\mathbf{q_1^{LP}}$	$\mathbf{q_2^{LP}}$
	$\frac{1-k}{2}$	$\frac{1-k}{2}$	$\frac{1-k}{2}$	k
Surplus	$\mathbf{\Pi_1^{LP}}$		$\mathbf{\Pi_2^{LP}}$	
	$\frac{(1-k)^2}{4}$		$\frac{(1-k)\cdot k}{2}$	
	$\mathbf{BS^{LP}}$		$\mathbf{TS^{LP}}$	
	$\frac{(1+k)^2}{8}$		$\frac{(1+k)(3-k)}{8}$	

References

P. Aghion and P. Bolton. Contracts as a barrier to entry. *The American Economic Review*, 77(3):388–401, 1987.

J. Asker and H. Bar-Isaac. Raising retailers' profits: On vertical practices and the exclusion of rivals. *The American Economic Review*, 104(2):672–86, February 2014.

K. C. Baseman, F. R. Warren-Boulton, and G. A. Woroch. Microsoft plays hardball: The use of exclusionary pricing and technical incompatibility to maintain monopoly power in markets for operating system software. *Antitrust Bulletin*, 40:265–315, 1995.

G. Calzolari and V. Denicolò. On the anti-competitive effects of quantity discounts. *International Journal of Industrial Organization*, 29(3):337 – 341, 2011.

G. Calzolari and V. Denicolò. Competition with exclusive contracts and market-share discounts. *The American Economic Review*, 103(6):2384–2411, 2013.

Y. Chao. Strategic effects of three-part tariffs under oligopoly. *International Economic Review*, 54(3):977–1015, 2013.

Y. Chao and G. Tan. All-units discounts: Leverage and partial foreclosure in single-product markets. *Canadian Competition Law Review*, 30(1):93–111, 2017.

Y. Chao, G. Tan, and A. C. L. Wong. All-units discounts as a partial foreclosure device. *The RAND Journal of Economics*, 49(1):155–180, 2018a.

Y. Chao, G. Tan, and A. C. L. Wong. Asymmetry in capacity and the adoption of all-units discounts. 2018b.

Z. Chen and G. Shaffer. Naked exclusion with minimum-share requirements. *The RAND Journal of Economics*, 45(1):64–91, 2014.

P. Choné and L. Linnemer. Nonlinear pricing and exclusion: II. must-stock products. *The RAND Journal of Economics*, 47(3):631–660, 2016.

E. Feess and A. Wohlschlegel. All-unit discounts and the problem of surplus division. *Review of Industrial Organization*, 37:161–178, 2010.

X. Fu and G. Tan. Abuse of market dominance under China's Anti-monopoly Law: The case of Tetra Pak. *Review of Industrial Organization*, 2018.

M. D. Grubb. Selling to overconfident consumers. *The American Economic Review*, 99(5):1770–1807, 2009.

M. D. Grubb. Dynamic nonlinear pricing: Biased expectations, inattention, and bill shock. *International Journal of Industrial Organization*, 30(3):287–290, 2012.

M. D. Grubb. Overconfident consumers in the marketplace. *Journal of Economic Perspectives*, 29(4):9–36, November 2015.

P. D. Klemperer and M. A. Meyer. Supply function equilibria in oligopoly under uncertainty. *Econometrica*, 57(6):1243–1277, 1989.

S. Kolay, G. Shaffer, and J. A. Ordover. All-units discounts in retail contracts. *Journal of Economics & Management Strategy*, 13(3):429–459, 2004.

A. Majumdar and G. Shaffer. Market-share contracts with asymmetric information. *Journal of Economics & Management Strategy*, 18(2):393–421, 2009.

D. Martimort. Multi-contracting mechanism design. In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume I, pages 57–101, 2006.

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

D. E. Mills. Inducing downstream selling effort with market share discounts. *International Journal of the Economics of Business*, 17(2):129–146, 2010.

M. Motta. *Michelin II: The treatment of rebates*, pages 29–49. Cambridge University Press, 2009.

D. P. O'Brien. All-units discounts and double moral hazard. *Journal of Economic Theory*, 170:1–28, 2017.

A. Pavan and G. Calzolari. Sequential contracting with multiple principals. *Journal of Economic Theory*, 144(2):503–531, 2009.

A. Pavan and G. Calzolari. Truthful revelation mechanisms for simultaneous common agency games. *American Economic Journal: Microeconomics*, 2(2):132–90, 2010.

E. B. Rasmusen, J. M. Ramseyer, and J. S. Wiley, Jr. Naked exclusion. *The American Economic Review*, 81(5):1137–1145, 1991.

I. R. Segal and M. D. Whinston. Naked exclusion: Comment. *The American Economic Review*, 90(1):296–309, 2000.

J. Simpson and A. L. Wickelgren. Naked exclusion, efficient breach, and downstream competition. *The American Economic Review*, 97(4):1305–1320, 2007.