Optimal Nonlinear Pricing by a Dominant Firm under Competition

Yong Chao Guofu Tan Adam Chi Leung Wong*

March 12, 2020

Abstract

We consider a nonlinear pricing problem faced by a dominant firm competing with a minor firm. The dominant firm offers a general tariff first and then the minor firm responds with a per-unit price, followed by a buyer choosing her purchases. By developing a "mechanism design approach" to solve the subgame perfect equilibrium, we characterize the dominant firm's optimal nonlinear tariff, which exhibits convexity and yet can display quantity discounts. In equilibrium the dominant firm uses a continuum of unchosen offers to constrain its rival's potential deviations and extract more surplus from the buyer. Antitrust implications are also discussed. (*JEL* L13, L42, K21)

1 Introduction

Nonlinear pricing (NLP)—total price not necessarily proportional to the quantity purchased—is ubiquitous in intermediate-goods markets. It takes the forms of various rebates and discounts conditional on volume purchased by a buyer. The competitive impact of NLP is a hotly debated antitrust topic, especially when the NLP is adopted by a dominant firm.

*Chao: College of Business, University of Louisville (yong.chao@louisville.edu); Tan: Department of Economics, University of Southern California (guofutan@usc.edu); Wong (Corresponding Author): Department of Economics, Lingnan University (adamwong@ln.edu.hk). We gratefully acknowledge helpful comments from Coeditor John Asker, two anonymous referees, Sandro Brusco, Giacomo Calzolari, Jimmy Chan, Yeon-Koo Che, Jiawei Chen, Ami Glazer, Fei Li, Giuseppe Lopomo, Preston McAfee, Alessandro Pavan, Patrick Rey, Mingjun Xiao, Huseyin Yildirim, Jidong Zhou, and seminar participants at Caltech, SHUFE, UC Davis, UCSB, HKU, USC, ZJU, UBC, U of Washington, UC Irvine, NCKU, CUHK, Nagoya U, UIUC, UNCC, UIBE, CityU, Wuhan U, 2017 Triangle Microeconomics Conference, 2017 Workshop on IOMS, 2018 NSF/NBER/CEME Decentralization Conference, 2018 CRESS-JUFE Conference, IO Workshop at U of St. Gallen, 17th IIOC, 2019 AMES, and Tsinghua BEAT 2019.

One example is three-part tariffs employed by Microsoft for its CPU licensing.[1] Another is "agreements containing individualized quantity commitments or retroactive rebate scheme" used by Tomra for the sales of reverse vending machines.[2] The complex grid of volume discounts adopted by Tetra Pak when selling its aseptic packaging materials is a third example of NLP.[3]

The vast majority of antitrust cases involving NLP is 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 schemes adopted by those asymmetric firms. Specifically, in those cases, there is a firm that is considered as "dominant" 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 relatively complex NLP whereas their small rivals usually utilize simple linear pricing (LP).[4]

Motivated by the above stylized facts of asymmetric competition, we consider the following problem faced by a dominant firm competing with 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 from both firms. We consider a three-stage game in which the dominant firm offers a NLP schedule first and then the minor firm responds with a per-unit price, followed by the buyer choosing her purchases from both firms. The dominant firm is to design a

[1]A three-part tariff entails a fixed payment, a free allowance of units up to which the marginal price is zero, and a per-unit price for additional demand beyond the allowance. See *United States* v. *Microsoft Corp.*, No. 94-1564 (D.D.C. filed July 15, 1995); Baseman, Warren-Boulton and Woroch (1995), and Chao (2013).

[2]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.

[3]On November 16, 2016, the then State Administration of Industry and Commerce of China fined Tetra Pak for abusing its dominance in China's aseptic packaging market. One of the alleged abusive practices was to exclude and limit competition through its complex grid of volume discounts. See Chao and Tan (2017), and Fu and Tan (2019) for discussions on this case.

[4]For instance, in Microsoft case, the main competitors of Microsoft's MS-DOS then were IBM's OS/2 and Digital Research Inc. (DRI)'s DR-DOS. Only Microsoft used NLP, and both IBM and DRI used LP only, as evidenced in Baseman, Warren-Boulton and Woroch (1995): "The CPU license appears to be unique to Microsoft...What options are open to an OEM who does not wish to exclusively ship its machines with MS-DOS? The OEM can negotiate a per-unit contract with Microsoft. However, Microsoft charges a price differential that is so high relative to CPU rates as to make the per-unit 'option' economically infeasible. Alternatively, the OEM can choose not to deal at all with Microsoft. In that case, it can purchase OS/2 on a per-unit basis...(and/or) DRI's per-unit license (of DR-DOS)" (see p. 276 therein). As in the Microsoft cases, the small rivals of Tomra and Tetra Pak used LP only.

general own volume-based tariff that maximizes its profit. We characterize the optimal NLP schedule for the dominant firm in the subgame-perfect equilibrium of our game, and study the properties of the optimal NLP and the implications of the equilibrium outcomes.

To find the optimal NLP schedule in our complete and perfect information game, one might conjecture that a singleton quantity-payment offer (also known as a bundle) would be sufficient and optimal for the dominant firm. This is certainly true in the case of a monopoly firm without competition. After all, there is only one buyer and no uncertainty regarding her demand. Thus, there can be only one quantity purchased by her from the dominant firm in equilibrium, regardless of how many quantity options offered by the dominant firm. Moreover, under monopoly, such a singleton is sufficient for the monopolist to extract efficient surplus fully. However, the above conjecture turns out to be not true. Indeed, as we show in Subsection 3.2 as well as in Appendix B, even offering only two bundles to a single buyer allows the dominant firm to strictly increase its profit over the best one-bundle offer, albeit in equilibrium only one bundle will be picked by the buyer. The reason hinges on the competitive pressure from the minor firm. Recall that the minor firm makes its unit-price offer after the dominant firm's move and before the buyer's choices. By offering unchosen bundles, the dominant firm can provide the buyer with extra latent choices, which in turn constrain the minor firm's potential deviations of undercutting the dominant firm. Therefore, the unchosen bundles help the dominant firm better manipulate competition against its rival and extract more surplus from the buyer. In contrast to the monopoly case, as we shall see, it is feasible for the dominant firm to fully exclude the minor firm and reach the social efficiency, but such full exclusion with social efficiency is not optimal from the dominant firm's perspective and thus will not arise in equilibrium.

Conceptually, the subgame-perfect equilibrium, in particular the dominant firm's optimal NLP schedule, is pinned down by the standard backward induction logic. However, without knowing the functional form of the NLP schedule, in practice it is difficult to solve the equilibrium by following the standard backward induction procedure. We overcome this difficulty by transforming the subgame-perfect equilibrium problem into a mechanism design problem with both hidden action and hidden information.

To understand our transformation, two features of our model should be noted: First, the dominant firm's NLP schedule is contingent on its own sales, so that the dominant firm has direct control over the quantity it sells to and the payment it receives from the buyer;[5] second, the buyer makes her purchase decision after seeing the minor firm's price and the

[5]In the context of one informed principal and N uninformed agents, Segal and Whinston (2003) considered the bilateral contracting between the principal and each agent. Similar to our setting, the bilateral contract considered there can only condition on the trade between the principal and the agent, not on the principal's trade with other agents.

minor firm makes its price decision after seeing the dominant firm's NLP schedule. It follows that the minor firm's price can be viewed as both the buyer's *hidden information* to the dominant firm, and the minor firm's *hidden action* to the dominant firm. Thus, we can imagine 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 revelation mechanism. Such a revelation mechanism requires the buyer to report the minor firm's price right after she sees it, and specifies 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 subgame-perfect equilibrium 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 revelation mechanism subject to the above constraints. Finally, after solving this constrained optimization problem, the optimal revelation mechanism can be transformed back to an optimal NLP schedule for the dominant firm by the taxation principle. This "mechanism design approach" of solving subgame-perfect equilibrium is of interest by its own. Generally speaking, for games where there is a single first mover whose action space is a function 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.[6]

In our complete and perfect information setting, we find that, 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. In other words, the optimal NLP schedule exhibits convexity. This is in stark contrast to a typical NLP tariff in the literature: in the presence of asymmetric information and under some regularity conditions, a monopolist's optimal NLP tariff involves concavity, i.e., decreasing marginal prices with increasing volumes (see Maskin and Riley (1984), and Tirole (1988), p. 156-157). Note that, a menu of two-part tariffs or incremental discounts can implement a concave tariff, but cannot implement a convex tariff (e.g., the optimal NLP in our paper). When the capacity of the minor firm is relatively small, the optimal NLP, along with the aforementioned convexity, entails a minimum quantity requirement with a positive payment. As a result, despite the convexity, the optimal NLP tariff in our setting can meanwhile display quantity discounts, i.e., decreasing average prices with increasing volumes.

[6]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.

In Chao, Tan and Wong (2018), we study a similar setup, but the dominant Firm 1 there can only offer an all-units discount tariff, parameterized by a quantity threshold, a pre-threshold per-unit price, and a post-threshold per-unit price; if the buyer's purchase from Firm 1 reaches the quantity threshold, the post-threshold price applies for all units of the purchase; otherwise the pre-threshold price applies. The main message of Chao, Tan and Wong (2018) is: If the dominance of the dominant firm is large enough, then the adoption of all-units discounts allows the dominant firm to leverage its market power from its captive (or non-contestable) portion of demand to the competitive (or contestable) portion and as a result its competitor is partially foreclosed. While this message is shared in the current paper, the study of *general* NLP offers substantially deeper insights. First, the methodological contribution of the mechanism design approach of solving subgame-perfect equilibrium is not available in our earlier work. Second, the theoretical properties of optimal general NLP (convexity, possibly together with quantity discounts) can only be proved in the current general setup. Third, while the advantage of all-units discounts uncovered in our earlier work stems from the market-power leverage effect (from captive to competitive demand), the advantage of general NLP uncovered in the current paper stems from the adoption of latent choices to constrain the competitor's potential deviations. It is worth noting that the former but not the latter advantage relies crucially on the existence of captive demand. More precisely, the adoption of all-units discounts in Chao, Tan and Wong (2018) does *not* improve the dominant firm's profit as compared with using LP, unless the dominant firm's dominance (measured by the size of captive demand) is large enough; on the other hand, as compared with offering a single bundle (which is already superior to LP), the adoption of latent choices in the current paper does improve the dominant firm's profit even when it does not enjoy a captive demand (i.e., its competitor does not have capacity constraint).

As compared to contracts that reference rivals, e.g., market-share discounts,[7] 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 might be warranted, for different reasons depending on the extent of the dominance.[8] 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

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

[8] 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.

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 with a minimum quantity requirement: The dominant firm commits not to supply the buyer *unless* she buys at least certain amount from the dominant firm. This results in a partial foreclosure to the minor firm and meanwhile hurts the 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.

In addition, our paper is related to exclusionary contracts literature. We show that partial foreclosure resulting from NLP arises under complete information with one buyer. As such, 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 and private information.

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 demonstrates how offering an extra bundle that will not be chosen in equilibrium can improve the dominant firm's profit, and establishes an equivalence between a subgame-perfect equilibrium of the game and an optimal mechanism in a "virtual" principal-agent model with hidden action and hidden information. Section 4 characterizes the equilibrium outcome of our original game, 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 5. Section 6 discusses some of our modeling assumptions and other related literature. Proofs are relegated to Appendix A. Appendix B uses a simple two-bundle construction to illustrate why offering unchosen bundle can help improve the dominant firm's profit.

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 $k > 0$, 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, also known as the buyer's surplus BS, is the gross benefit given by $u(Q + q)$, minus 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.[9] 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) subgame-perfect equilibrium (SPE).

We say a tariff is *regular* if the subgame after Firm 1 offers such a tariff has some SPE. 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 $\tau(0) \leq 0$.[10,11] Also denote the set of feasible unit prices Firm 2 can choose from as $\mathcal{P} \equiv [c, \infty)$.[12]

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

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

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

[10] By definition, if we allow Firm 1 to choose an irregular tariff, then the whole game has no SPE. So our regularity requirement is so weak that it allows for all the tariff forms (referencing own sales) for which one may sensibly do equilibrium analysis under our setting; they include all familiar tariff forms, e.g., two-part tariffs, three-part tariffs, all-units discounts, discrete bundles, etc. Moreover, it can be shown that, under Assumption 1 below, a tariff $\tau : \mathbb{R}_+ \to \mathbb{R} \cup \{\infty\}$ is regular whenever it is lower semicontinuous and satisfies $\tau(Q) \geq cQ$ for all sufficiently large Q. Briefly speaking, after Firm 1 offers such a tariff, the continuation subgame has a SPE because (i) the buyer's optimal purchases exist no matter what unit price Firm 2 charges, and (ii) if we let the buyer use a tie-breaking rule in favor of Firm 2, Firm 2's best response exists.

[11] Since we do not formally endow the buyer with an outside option of buying nothing from and paying nothing to Firm 1, we need the requirement $\tau(0) \leq 0$ to capture that Firm 1 cannot compel the buyer to pay any positive amount when she does not buy anything from Firm 1. Equivalently, one may drop $\tau(0) \leq 0$ by formally adding such an outside option, but our current treatment makes the notation simpler.

[12] It is without loss of generality to eliminate the dominated strategies of charging $p < c$ for Firm 2. By doing so, we can avoid some uninteresting technicalities involved in our rewriting of the constraints and objective function of (OP) in Subsection 4.2.

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

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

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

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

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 Intuition, Transformation and Equivalence

To determine the SPE outcome of our sequential-move game with complete and perfect information, the standard backward induction procedure requires us to sequentially solve the buyer's problem (1), Firm 2's problem (2), and Firm 1's problem (3). But such a procedure hardly seems manageable when Firm 1's tariff offer τ is as general as in our setting. Specifically, an explicit relation between the optimal solution to the buyer's problem (1) and *arbitrary* $\tau \in \mathcal{T}$ and $p \in \mathcal{P}$ is *not* readily available. But without an explicit solution to (1) for arbitrary τ, it is considerably difficult to solve (2) and (3).

One might conjecture the optimal NLP will degenerate into a singleton quantity-payment offer, and if so, the characterization of the optimal singleton contract would then be easy.

Nevertheless, we will explain in Subsection 3.2 why such a singleton contract cannot be optimal, and as such the determination of the optimal NLP remains to be a challenging task.

To overcome this technical difficulty, we will transform our original problem of solving (1), (2), and (3) into a one-principal-two-agent mechanism design problem, and establish the equivalence between the two problems in Subsection 3.3.

3.1 Buyer's Problem

In this subsection, we study the buyer's problem (1) and introduce two useful notations. 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.,

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

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 purchased from Firm 1, the buyer chooses her purchase q from Firm 2 by solving (5); (ii) the buyer chooses her purchase Q from Firm 1, i.e.,

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

Although the sub-problem (6) still does not permit a ready solution without knowing any properties of τ, the sub-problem (5) is well behaved and tractable. Indeed, (5) has a unique maximizer given by

$$\operatorname*{argmax}_{q\in[0,k]} \left\{u(Q+q) - pq\right\} = \mathrm{Proj}_{[0,k]}(D(p) - Q), \tag{7}$$

where Proj is the projection operator.[13] By the Envelope Theorem, $V(Q,p)$ has the following properties.

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

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

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

[13] That is, for any closed interval $X \subseteq \mathbb{R}$ and any point $x \in \mathbb{R}$, $\mathrm{Proj}_X(x)$ denotes the projection of x on X, i.e., $\mathrm{Proj}_X(x) \equiv \mathrm{argmin}_{y\in X} |y - x|$. In particular, $\mathrm{Proj}_{[a,b]}(x) = \max\left\{\min\left\{x, b\right\}, a\right\}$ whenever $a \leq b$.

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

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

Next, 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.,

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

3.2 Intuition: Why an Unchosen Bundle Helps

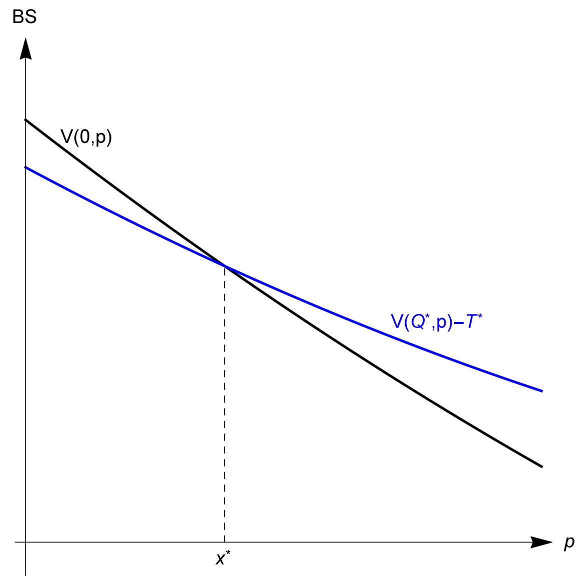
Before characterizing the optimal NLP, let us first illustrate some intuition for why the optimal NLP must *not* degenerate into a one-bundle offer as one might envision. We will show why and how Firm 1 can strictly improve its profit over its maximum profit level in the "one-bundle equilibrium" by offering an extra bundle which will not be chosen in equilibrium.[14]

Consider the "one-bundle model," i.e., suppose Firm 1 can only offer one bundle, and consider the resulting one-bundle equilibrium. Let (Q^*, T^*) be the bundle that Firm 1 would offer in this equilibrium, where Q^* denote the bundle quantity and T^* the bundle price. Clearly, $Q^* > 0$. After seeing a price offer p from Firm 2, the buyer can either reject the bundle (Q^*, T^*) or accept it; and the buyer's surplus will be $V(0,p)$ if she rejects, and $V(Q^*,p) - T^*$ if she accepts. Thanks to the properties of V from Lemma 1, the two curves of the buyer's surplus, drawn against p, are downward-sloping and cross only once at certain price x^* that equates the two surpluses, as shown in Figure 1(a).[15] The buyer will accept the bundle only when Firm 2's price p exceeds the threshold x^*. As a result, Firm 2's profit is discontinuous at x^*, as shown in Figure 1(c). It is given by $\pi(0,p)$ when $p \leq x^*$, and then drops to $\pi(Q^*,p)$ when $p > x^*$. In equilibrium, the buyer must accept Firm 1's bundle, so Firm 2 must not have incentives to offer a price below x^*. Let p^* be Firm 2's equilibrium price. Then, Firm 2's equilibrium profit $\pi(Q^*,p^*)$ must satisfy

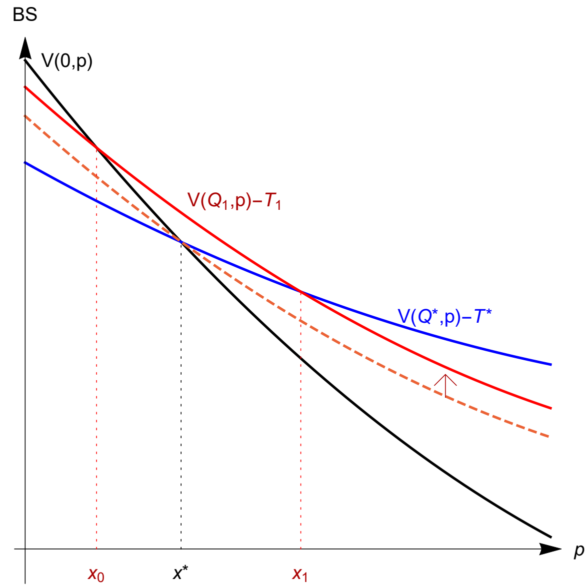
$$\pi(Q^*,p^*) = \max_{p > x^*} \pi(Q^*,p) \geq \max_{p \leq x^*} \pi(0,p), \tag{12}$$

as in Figure 1(c). Furthermore, Firm 1 must set its bundle price T^* at the highest level that does not induce Firm 2 to deviate to charge below x^*. That is, the inequality constraint in

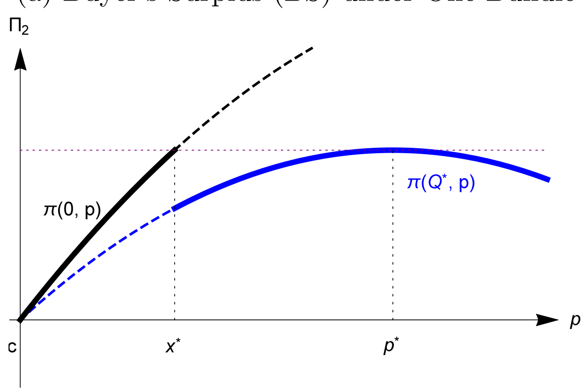
[14] The formal construction of the strictly profitable improvement can be found in Appendix B.
[15] It implies that $T^* = V(Q^*, x^*) - V(0, x^*)$.



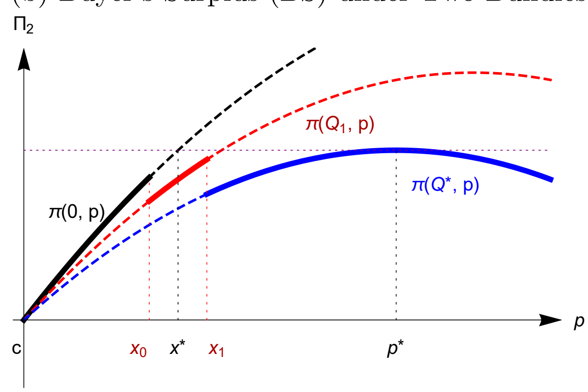
(a) Buyer's Surplus (BS) under One Bundle



(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

(12) must be binding, again as shown in Figure 1(c).

Now, given the optimal (Q^*, T^*) and the corresponding x^* under the above one-bundle model, we add an extra bundle (Q_1, T_1) with both quantity and payment slightly lower than those from the optimal one bundle, i.e., $0 < Q_1 < Q^*$ and $T_1 = V(Q_1, x^*) - V(0, x^*) - \epsilon$ for $\epsilon > 0$. The buyer now has an extra choice: accepting the newly added bundle (Q_1, T_1). By doing so, the buyer's surplus will be

$$V(Q_1, p) - T_1 = V(Q_1, p) - V(Q_1, x^*) + V(0, x^*) + \epsilon.$$

Recall that the buyer's surpluses from accepting and rejecting (Q^*, T^*) are, respectively, $V(Q^*, p) - T^* = V(Q^*, p) - V(Q^*, x^*) + V(0, x^*)$ and $V(0, p)$. Since $0 < Q_1 < Q^*$, the increasing differences property of V implies that, $V(Q_1, p) - T_1$ is flatter than $V(0, p)$, but steeper than $V(Q^*, p) - T^*$ everywhere. Moreover, it is easy to check that at $p = x^*$, $V(Q_1, p) - T_1$ is above the crossing point of $V(0, p)$ and $V(Q^*, p) - T^*$, as shown in Figure 1(b). Correspondingly, $V(Q_1, p) - T_1$ crosses $V(0, p)$ and $V(Q^*, p) - T^*$ at x_0 and x_1, respectively. The buyer will choose the new bundle if $x_0 < p \leq x_1$, the original bundle (Q^*, T^*) if $p > x_1$, and reject both bundles otherwise. Thus, Firm 2's profit is now broken into three, rather than two, discontinuous pieces as shown in Figure 1(d). It is still optimal for Firm 2 to choose the price p^* that Firm 1 desires, but the original binding constraint (12) for Firm 2 is *not* binding any more. The slack of the constraint suggests that there is a room for Firm 1 to strictly improve its profit by increasing payment T^*.

Intuitively, the extra bundle, as the buyer's latent choice, reduces the temptation for Firm 2 to undercut. Indeed, in Figure 1(c), without the extra bundle, if Firm 2 reduces 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 reducing to x^* and not. In contrast, in Figure 1(d), with the extra bundle, if Firm 2 reduces its price to x^*, the buyer still chooses the small bundle (Q_1, T_1) from Firm 1, and the resulting increase in Firm 2's sales would be so limited that Firm 2's profit would strictly decline. 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 fall. It is for this reason that Firm 1 is able to raise the price of the chosen bundle without triggering Firm 2's price cut.

The above intuition extends 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. As we shall see next, 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.

3.3 Transformation and Equivalence

Given the technical challenge in buyer's problem (1) for arbitrary τ, now we transform the original SPE problem into an equivalent mechanism design problem, which allows us to characterize the optimal NLP. 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 $\text{Proj}_{[0,k]}(D(p) - Q(p))$, 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 revelation mechanism $Q : \mathcal{P} \to \mathbb{R}_+$ and $T : \mathcal{P} \to \mathbb{R}$ to the buyer (one agent), and recommends a price $\bar{p} \in \mathcal{P}$ for Firm 2 (another agent).

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

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

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

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

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-O) is the obedience 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*

$$Q^*(p) = q_1^*(\tau^*, p) \ \ \forall p \in \mathcal{P}, \tag{13}$$

$$\text{Proj}_{[0,k]}(D(p) - Q^*(p)) = q_2^*(\tau^*, p) \ \ \forall p \in \mathcal{P}, \tag{14}$$

$$T^*(p) = \tau^*(Q^*(p)) \ \ \forall p \in \mathcal{P}, \tag{15}$$

$$\bar{p}^* = p^*(\tau^*). \tag{16}$$

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 Subsection 4.2; its solution can then be transformed back to characterize the SPE outcomes.

4 Equilibrium

4.1 Characterization of SPE

In this subsection, we state our equilibrium characterization and postpone the derivations to the next subsection.

The following theorem establishes the existence and provides a characterization of the SPE, in particular Firm 1's equilibrium tariff schedule $\tau(\cdot)$ as illustrated in Figure 2.

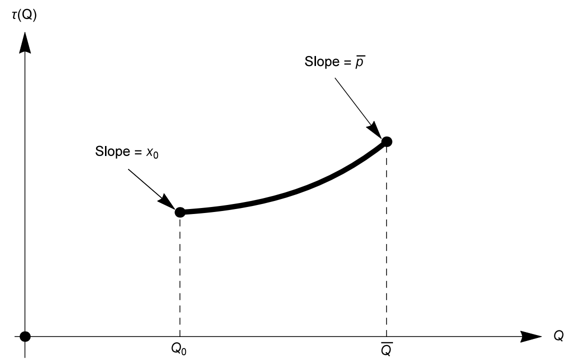


Figure 2: Equilibrium nonlinear pricing

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

(a) Firm 1's nonlinear tariff $\tau(\cdot)$ satisfies

$$\tau(Q) = u(Q_0 + k) - u(k) + \int_{Q_0}^{Q} x(\tilde{Q})d\tilde{Q} \quad \forall Q \in [Q_0, \bar{Q}], \tag{17}$$

where Q_0, $\bar{Q}$, and $x(\cdot)$ are defined below.

(b) Firm 2's price $\bar{p} \in (c, p^m)$ satisfies[16]

$$-(\bar{p} - c)D'(\bar{p}) = \frac{1}{e}\min\left\{D\left(c + \frac{\bar{p}-c}{e}\right), k\right\}. \tag{18}$$

(c) The buyer purchases $\bar{Q}$ units from Firm 1, and the residual units $D(\bar{p}) - \bar{Q}$ from Firm 2.

Given $\bar{p}$, $Q(\cdot)$ is a strictly increasing function given by

$$Q(p) = D(p) - \frac{(\bar{p}-c)[D(\bar{p}) - \pi'(\bar{p})]}{p - c} \quad \forall p \in [x_0, \bar{p}], \tag{19}$$

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

where $x_0 = c + (\bar{p} - c)/e$*,* $Q_0 = Q(x_0)$*,* $\bar{Q} = Q(\bar{p})$*, and* $x : [Q_0, \bar{Q}] \to [x_0, \bar{p}]$ *is the inverse of* $Q(\cdot)$*.*

The intuition behind Theorem 2 is as follows. In our Stackelberg-type price competition, when offering its tariff schedule, Firm 1 faces the competitive pressure from Firm 2, due to Firm 2's possible undercutting from the equilibrium. Thus, it pays for Firm 1 to offer unchosen bundles to deter potential deviations from the equilibrium strategy by Firm 2. From Firm 1's perspective, the unchosen bundles help relaxing the obedience constraint (F2-O) for Firm 2. Specifically, the unchosen bundles serve as the buyer's latent choices to constrain Firm 2's potential deviation of undercutting. Such latent bundles must be smaller than the chosen bundle, i.e., Firm 1 offers a continuum of bundles with quantities no greater than $\bar{Q}$, because Firm 2 is never tempted to raise its price above the equilibrium one $\bar{p}$.

The bundle $(\bar{Q}, \tau(\bar{Q}))$ is chosen by the buyer in equilibrium. Those $(Q, \tau(Q))$ for $Q \in [Q_0, \bar{Q})$ will not be chosen but they are the *necessary* latent options 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 absent 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. The tariff formula (17) 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 indeed indeterminate.[17]

When deviating, Firm 2 will not undercut to the marginal cost c because doing so would result in zero profit for itself. We find that, given its equilibrium price $\bar{p}$, the lowest deviation price that Firm 2 would be willing to undercut is x_0, which is strictly above c. In response to Firm 2's potential deviations in $[x_0, \bar{p}]$, Firm 1 offers a continuum of bundles with quantities in the interval $[Q_0, \bar{Q}]$ with $Q_0 = Q(x_0) = \max\{D(x_0) - k, 0\}$. If Firm 2 charges a (possibly off-equilibrium) price p in $[x_0, \bar{p}]$, the buyer's purchases from Firm 1 and Firm 2 would be $Q(p)$ and $D(p) - Q(p)$, respectively. Condition (19) says that Firm 2's profit $(p - c)[D(p) - Q(p)]$ is a constant over $[x_0, \bar{p}]$ and is equal to its equilibrium profit $(\bar{p} - c)[D(\bar{p}) - \bar{Q}]$.[18] This is because the obedience constraint (F2-O) for Firm 2 is binding over that range.

What is more interesting about the optimal NLP is that, Firm 1's marginal price evaluated at its equilibrium sales is equal to Firm 2's actual price, i.e., $\tau'(\bar{Q}) = x(\bar{Q}) = \bar{p}$, which

[17] 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 $Q \notin [Q_0, \bar{Q}] \cup \{0\}$, as in Figure 2.

[18] Theorem 2 also implies that Firm 2's equilibrium profit is positive since $(\bar{p} - c)(D(\bar{p}) - \pi'(\bar{p})) = -(\bar{p} - c)^2 D'(\bar{p}) > 0$.

can be seen from (17) and the definitions of $x(\cdot)$ and $\bar{Q}$. 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 and her purchase from Firm 2 is less than k. As we shall see in the next subsection, the shape of $Q(\cdot)$ is determined by Firm 2's iso-profit curves in the Q-p space, and hence (19) holds. Moreover, condition (18) is the first-order condition for maximizing Firm 1's profit, as we will see in Lemma 4.

Theorem 2 claims the existence but not uniqueness of the SPE outcome. Strictly speaking, the SPE outcome is never unique because of the aforementioned indeterminacy of $\tau(Q)$ for $Q \notin [Q_0, \bar{Q}]$. We say the SPE outcome is *essentially unique* if all the "relevant" components (i.e., except redundant bundles with quantities outside $[Q_0, \bar{Q}]$) of the SPE outcome are unique. For example, if the solution of (18) for $\bar{p}$ is unique, from Theorem 2 all the components of the SPE outcome except redundant bundles are uniquely determined.[19] Since the right-hand side of (18) as a function of $\bar{p}$ is non-increasing, a simple sufficient condition for the essential uniqueness is that the left-hand side of (18) as a function of $\bar{p}$ is strictly increasing.

Corollary 1 (Essential Uniqueness)**.** *The SPE outcome is essentially unique if the following condition holds:*[20]

$$-(p-c)D'(p) \text{ is strictly increasing in } p \text{ on } [c, p^m]. \tag{20}$$

4.2 Derivations of Equilibrium

This subsection outlines the proof of Theorem 2. Following Theorem 1, it suffices to solve the constrained optimization problem (OP).

4.2.1 Constraints for the Buyer

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

[19] It is worth noting that multiplicity of solutions to (18), if any, does *not* necessarily imply essential multiplicity of the SPE outcomes. This is because only the solution candidate that yields the highest profit for Firm 1 will be chosen by Firm 1. Of course, there is still a possibility that multiple candidates give the same highest profit. Condition (20) precludes this possibility and makes (18) not only necessary but also sufficient for the equilibrium $\bar{p}$.

[20] This condition is slightly stronger than the strict concavity of $\pi(\cdot)$ (Assumption 2), because $\pi'(p) = D(p) + (p-c)D'(p)$ being decreasing does not necessarily require $-(p-c)D'(p)$ to be strictly increasing in p. Besides, the condition is weaker than concave demand, and is equivalent to that $u'(q) - c$ is strictly log-concave in q on $[D(p^m), q^e]$.

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 hold:*

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

$$\forall p \in \mathcal{P},\ T(p) - T(c) = V(Q(p), p) - V(Q(c), c) - \int_c^p V_p(Q(t), t)dt \tag{21}$$

$$V(Q(c), c) - T(c) \ge V(0, c) \tag{22}$$

Condition (Mon) is a weakened version of the standard monotonicity condition in mechanism design problems, since the increasing differences property (10) of V is strict only on Φ defined in (4). 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 a 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 (21) is the standard envelope formula for payment in mechanism design problems. Condition (Mon) and condition (21) together are necessary and sufficient conditions for (B-IC). Moreover, given (21), condition (22) is a necessary and sufficient condition for (B-IR), since (21) 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), (21), and (22), we see that (22) 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), (21), and (F2-O)) are intact. Therefore,

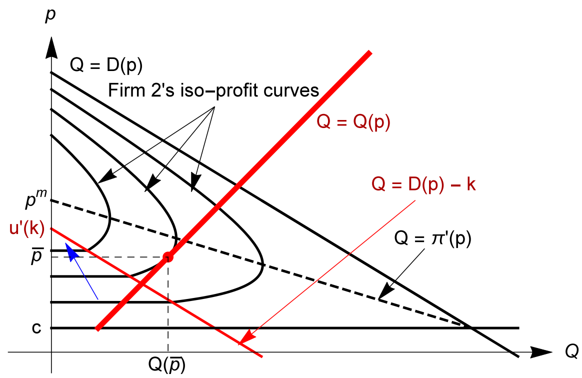
$$T(c) = V(Q(c), c) - V(0, c). \tag{23}$$

Combining (21) and (23), we obtain, for all $p \in \mathcal{P}$,

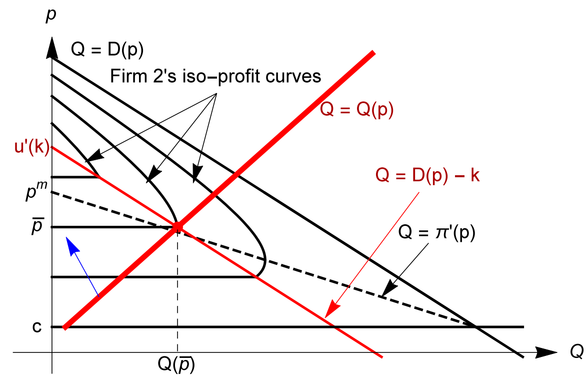
$$T(p) = V(Q(p), p) - V(0, c) - \int_c^p V_p(Q(t), t)dt. \tag{24}$$

4.2.2 Constraints for Firm 2

We now take a closer look at the obedience constraint (F2-O) 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



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



(b) when $k < D(p^m)$

Figure 3: Firm 2's Iso-profit Curves

if its profit $\pi(Q(p), p)$ is maximized at $p = \bar{p}$.

Figure 3 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 3 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 3. 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 3(a) and 3(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 $Q = \max\{\pi'(p), D(p) - k\}$.

4.2.3 Constrained Optimization

This subsection solves (OP) and thus the SPE outcome. Recall that we have used (21) and (23) to eliminate $T(\cdot)$, as given by (24). Then, Firm 1's profit can be written as

$$\begin{aligned}\Pi_1 &= T(\bar{p}) - c \cdot Q(\bar{p}) \\ &= V(Q(\bar{p}), \bar{p}) - V(0, c) - \int_c^{\bar{p}} V_p(Q(p), p)dp - c \cdot Q(\bar{p})\end{aligned} \tag{25}$$

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

$$\underset{Q(\cdot),\bar{p},\Pi_2}{\text{Maximize}} \ (25) \tag{OP'}$$

subject to

$$(\text{Mon})$$

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

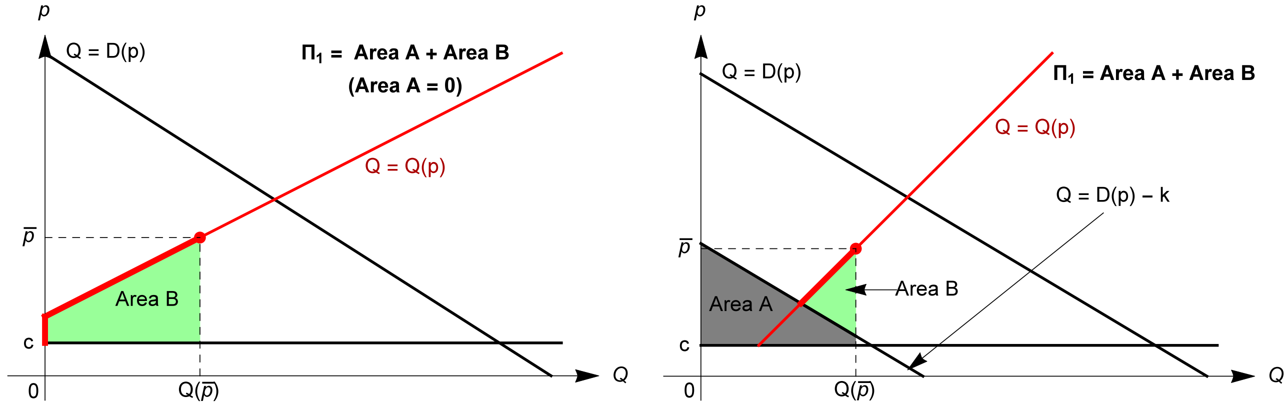
$$\Pi_2 = \pi(Q(\bar{p}), \bar{p}). \tag{F2-Pro}$$

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).

To graphically show Firm 1's profit, we use (8) and (9) to rewrite (25):

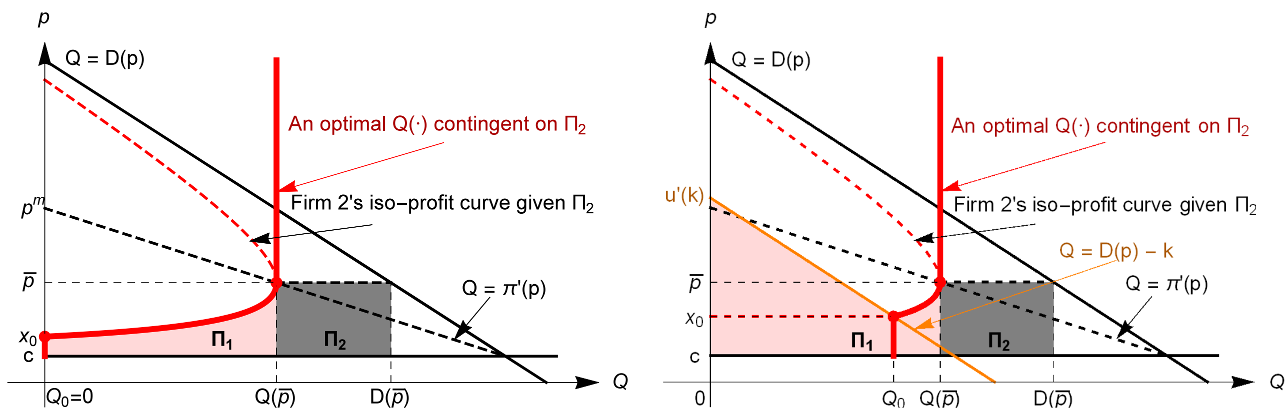
$$\begin{aligned}\Pi_1 &= \int_0^{Q(\bar{p})} [V_Q(Q, c) - c]dQ + \int_c^{\bar{p}} \left[V_p(Q(\bar{p}), p) - V_p(Q(p), p)\right] dp \\ &= \int_0^{Q(\bar{p})} [\text{Proj}_{[u'(Q+k), u'(Q)]}(c) - c]dQ \\ &\quad + \int_c^{\bar{p}} [\text{Proj}_{[0,k]}(D(p) - Q(p)) - \text{Proj}_{[0,k]}(D(p) - Q(\bar{p}))]dp \\ &= \int_0^{Q(\bar{p})} [\text{Proj}_{[u'(Q+k), u'(Q)]}(c) - c]dQ \\ &\quad + \int_c^{\bar{p}} [\text{Proj}_{[D(p)-k, D(p)]}(Q(\bar{p})) - \text{Proj}_{[D(p)-k, D(p)]}(Q(p))]dp\end{aligned} \tag{26}$$

Figure 4 shows the area of Π_1 given by (26) for a given $Q(\cdot)$ and $\bar{p}$: Areas A and B correspond to the first and the second integral in (26) 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 4(a), the captive demand vanishes and consequently Area A is 0.



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

Figure 4: Firm 1's profit Π_1 contingent on $Q(\cdot)$ and $\bar{p}$



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

Figure 5: Optimal $Q(\cdot)$ contingent on Π_2

It can be seen from Figures 3 and 4 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-O'), 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., $\bar{Q} = \max\{\pi'(\bar{p}), D(\bar{p}) - k\}$ from Subsection 4.2.2. Lemma 3 below formalizes these claims and solves (OP') contingent on Π_2. Figures 5(a) and 5(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*

$$\max\{D(\bar{p}) - k, \pi'(\bar{p})\} = D(\bar{p}) - \frac{\Pi_2}{\bar{p} - c} \equiv \bar{Q}, \tag{27}$$

and $Q(\cdot)$ *satisfies*

$$Q(p) = D(p) - \frac{\Pi_2}{p - c} \quad \forall p \in [x_0, \bar{p}], \tag{28}$$

where x_0 *is the unique solution in* $(c, \bar{p}]$ *of*

$$\max\{D(x_0) - k, 0\} = D(x_0) - \frac{\Pi_2}{x_0 - c} \equiv Q_0. \tag{29}$$

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 (28) 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-O'). In particular, any monotonic extension of (28) works, e.g., we may 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 5.*

To solve (OP'), it remains to pin down Π_2, which should be chosen to make the Π_1 area in Figure 5 as large as possible. It turns out that the corresponding first-order condition can be simplified as (30) below. Also, when Π_1 is maximized we must have $D(\bar{p}) - k < \bar{Q}$ so that the left-hand side of (27) 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 (24) 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. For any such solution* $(Q(\cdot), T(\cdot), \bar{p})$*,* $(\bar{p}, x_0)$ *satisfies*[21]

$$\bar{p} - c = e \cdot (x_0 - c) > 0, \tag{30}$$

$$(\bar{p} - c)[D(\bar{p}) - \pi'(\bar{p})] = (x_0 - c)\min\{D(x_0), k\}, \tag{31}$$

$Q(p)$ *satisfies (19), and*

$$T(p) = u(Q_0 + k) - u(k) + \int_{x_0}^{p} t dQ(t) \quad \forall p \in [x_0, \bar{p}], \tag{32}$$

where $Q_0 = \max\{D(x_0) - k, 0\}$*.*

[21]Remark 1 also applies here.

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

5 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.

5.1 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 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 2. This is in stark contrast to a typical nonlinear tariff in the literature: in Maskin and Riley (1984) and Tirole (1988) (p. 156-157), under asymmetric information, a monopolist's optimal nonlinear tariff often involves concavity, i.e., 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, i.e., 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.[22] 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

[22] 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 $\bar{Q}$.

because it is a convex 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.

5.2 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.[23] 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 following hold:[24]

(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 (20) holds then its output $D(\bar{p}) - \bar{Q}$ increases;

(c) The equilibrium total surplus, denoted as $\mathrm{TS} \equiv \Pi_1 + \Pi_2 + \mathrm{BS}$, 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 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.

[23] This can be seen from (18): The right-hand side of (18) 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.

[24] In the proof, we also show that, the equilibrium Q_0 (x_0) is decreasing (increasing) in k.

5.3 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*. We use superscript "LP" to denote various variables for the LP 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 $k < q^e$.[25] 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 $k = \hat{k}$, $\bar{p} > x_0 = u'(\hat{k}) > \bar{p}^{LP}$, 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) Joint Surpluses: $\Pi_2 + \mathrm{BS} < \Pi_2^{LP} + \mathrm{BS}^{LP}$; $\mathrm{TS} \gtreqqless \mathrm{TS}^{LP}$ 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,

[25] 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 (2018).

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.

5.4 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 (20) 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 1.

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 1, 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

Table 1: Linear Demand Example

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	$\mathbf{\Pi_1^{LP}}$		$\mathbf{\Pi_2^{LP}}$	
	$\frac{(1-k)^2}{4}$		$\frac{(1-k)\cdot k}{2}$	
	$\mathbf{BS^{LP}}$		$\mathbf{TS^{LP}}$	
	$\frac{(1+k)^2}{8}$		$\frac{(1+k)(3-k)}{8}$	

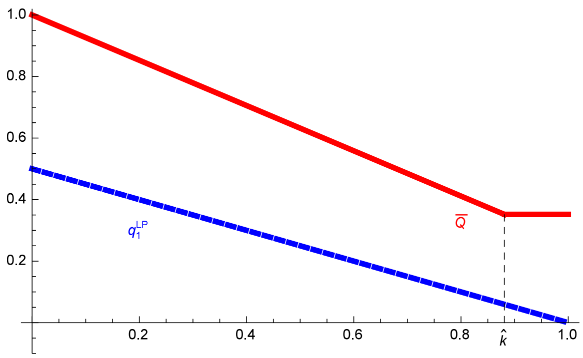
Note: $\hat{k}=e^2/(1+e^2)$.

the antitrust concerns about conditional discounts adopted by some dominant firms when competing against small rival firms.

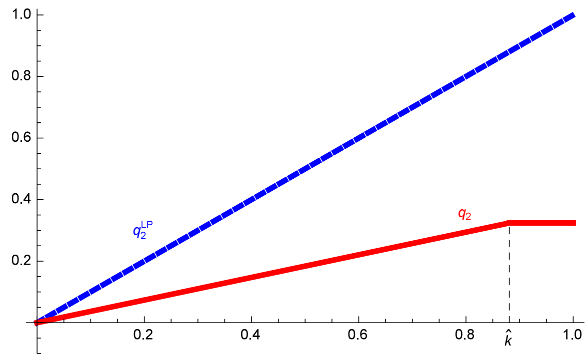
6 Discussions and Related Literature

In the literature, NLP is usually considered as a screening device in the presence of buyers' private information, e.g., Maskin and Riley (1984), Wilson (1993). *Absent* private information of the buyer, downstream competition, or demand uncertainty,[26] 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 potential undercutting and extract more surplus from the buyer. By developing a mechanism design approach to solve the subgame perfect equilibrium of our sequential offering game, 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

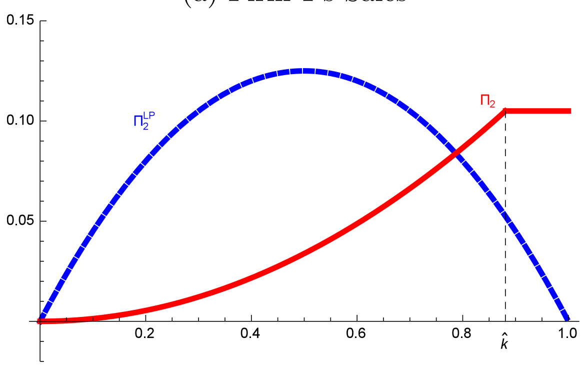
[26] Klemperer and Meyer (1989) show that, as a response to demand uncertainty, firms may be forced to offer a supply function against a range of possible states.



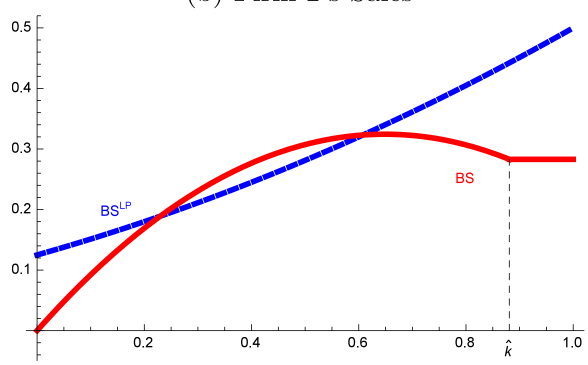
(a) Firm 1's Sales



(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

of the NLP schedule employed by the dominant firm depend on the extent of the dominant firm's dominance.

The setting of our sequential-move game features asymmetries between the dominant Firm 1 and the minor Firm 2 in terms of contract spaces and order of offering. Specifically, Firm 2 can only use LP while Firm 1 can use NLP; and Firm 1 leads and Firm 2 follows. We assume asymmetric contract spaces in order to capture the fact that small firms in practice usually do not match the contract as complex as offered by a dominant firm, e.g., LP is used mostly by small rivals of the defendants in our aforementioned antitrust cases. It could be due to the fact that the dominant firm is more experienced in dealing with downstream buyers than new entrants to the market. Regarding the order of offering, in practice, when there is a dominant firm in the industry, it is the dominant firm that often moves first.[27]

However, one might still wonder, given Firm 1's equilibrium NLP, how large is Firm 2's incentive to deviate to use NLP if it could?[28] Such an incentive does exist because, in our equilibrium outcome (where Firm 1 uses NLP and Firm 2 uses LP), Firm 2 is facing

[27] *"Price leadership probably works best and arises most frequently in industries in which a single firm is outstanding by virtue of large size or recognized high quality of management"* (see Oxenfeldt (1951), p. 296).

[28] We are grateful to an anonymous referee for raising this interesting question.

a downward-sloping residual demand from which NLP could help capture more surplus.[29] However, on the other hand, it is also commonly believed that using NLP involves higher transaction costs (designing cost, contracting cost, implementation cost, enforcement cost, etc.) than using LP. So it is interesting to ask, if we modify our three-stage game such that both firms can choose to use NLP or LP but using NLP (as opposed to LP) requires the firm to pay some additional fixed "NLP cost," then, is our equilibrium outcome still valid under a plausible NLP cost (common to both firms)?

To answer this question, we need to compute Firm 2's gain from NLP (i.e., the increase of Firm 2's profit if it deviates to an optimal NLP, given Firm 1's equilibrium NLP), and Firm 1's gain from NLP (i.e., the decrease of Firm 1's profit if it restricts to LP[30]). If Firm 2's gain from NLP is smaller than Firm 1's, then our equilibrium outcome is still valid in the aforementioned modified game provided that the NLP cost lies in between the two levels of gains.

It is easy to prove that, when Firm 2's capacity k is small enough, Firm 2's gain from NLP is smaller than Firm 1's, since Firm 2's gain vanishes as $k \to 0$ while Firm 1's does not. In the linear demand example in Subsection 5.4, Firm 2's gain from NLP is smaller than Firm 1's whenever $k < 0.66$ (recall that the efficient quantity q^e here is 1). When $k = 0.2$ for instance, Firm 1's gain from NLP is 30 times Firm 2's; if the NLP cost is 2% of Firm 1's equilibrium profit, our NLP vs LP equilibrium outcome holds.[31] The intuition is that adopting NLP, which is costly, is worth it only for large firms. Therefore, we believe our assumption that Firm 2 only uses LP while Firm 1 uses NLP can be reasonably justified by a plausible NLP cost, especially when one is concerned with an industry where minor firms are significantly constrained by their technologies.

An alternative way to justify our asymmetric treatment for the two firms is to use the strategic argument provided by Chao, Tan and Wong (2019). In that paper, we focus on a particular form of NLP, namely, all-units discounts (AUDs), and allow both the full-capacity dominant firm (Firm 1) and the capacity-constrained minor firm (Firm 2) to be able to commit upfront (i.e., before their price competition) to use LP, and let the timing of the price competition be determined endogenously. More precisely, in that paper we consider the following four-stage duopoly game. In stage 0, each of the two firms simultaneously decides

[29]This residual demand is $p \mapsto \text{Proj}_{[0,k]}(D(p) - Q(p))$.

[30]It is easy to show that, if Firm 1 restricts to use LP, then its highest profit (regardless of whether Firm 2 uses NLP or LP) would then be its "captive demand monopoly profit" $\max_p (p - c)D^{cap}(p)$, where $D^{cap}(\cdot)$ denotes Firm 1's captive demand $\max\{D(\cdot) - k, 0\}$.

[31]Firm 2's gain from NLP actually depend on how Firm 1 offers the "redundant bundles" (in the sense as in Subsection 4.1). The values reported here are based on the assumption that Firm 1 offers the redundant bundles to minimize Firm 2's gain from deviating to NLP. Equivalently, we let $Q(p) \geq D(\bar{p})$ for $p > \bar{p}$, so that Firm 2's residual demand $\text{Proj}_{[0,k]}(D(p) - Q(p)) = 0$ for $p > \bar{p}$.

whether to commit itself to use LP or to use AUDs. In stage 1, each firm can either offer a pricing scheme from the feasible set determined by its choice in stage 0, or wait until stage 2. In stage 2, any firm who chose to wait in stage 1 offers a pricing scheme (again from the feasible set determined by its choice in stage 0). In stage 3, the buyer chooses the quantities she purchases from the two firms. The main result there is that, in equilibrium, only Firm 2 commits upfront to use LP and the timing is that the dominant firm leads and the minor firm follows. Intuitively, Firm 2 restricts itself to LP upfront in order to soften the competition between the two firms, partially through encouraging Firm 1 to make offer first. Although some adaptations are needed for our current purpose, because here we consider general NLP rather than AUDs, the main message in that paper maintains here, in the sense that *the equilibrium outcome in the current NLP paper, including the asymmetries in contract spaces and order of offering, can be regarded as an equilibrium outcome of an extended game in which the contract spaces and order of offering are endogenous.*

Another crucial feature of our setting is that the minor firm is strategic and has significant market power. To see the importance of this feature, consider an alternative setting in which Firm 2 is replaced by a competitive fringe. Then, the strategic interactions between Firm 1 and its rivals are essentially removed, because from Firm 1's perspective the buyer and the price-taking fringe firms can be viewed as a single player who can produce the product by itself at marginal cost c up to k. Accordingly, Firm 1 will behave as a monopolist facing the residual demand, and one single bundle, as opposed to a NLP schedule, suffices to extract full surplus from the residual demand. The equilibrium pattern and implications would thus be starkly different from those of our model.[32]

Our setting also features a single, multi-sourcing, downstream buyer, which can be regarded as the common agent of the two firms. The common agency literature, initiated by Bernheim and Whinston (1986*a*,*b*), has a strand on complete information setting with simultaneous moves since then. This strand mainly concerns whether the equilibrium outcome under simultaneous-move common agency is efficient or not, e.g., Bernheim and Whinston (1986*a*, 1998). Supplementarily, we have explored a different timing of common agency with complete information, in which two upstream firms offer their contracts sequentially and the downstream buyer makes purchase decision only after observing both offers. Our game allows us to pin down the optimal NLP and examine its properties and antitrust implications.

Another strand of simultaneous-move common agency literature studies competing contracts with incomplete information. The complication of the common agency comes from the failure of the standard revelation principle, because the agent's preferences over a contract from one principal may depend on contracts offered by other principals. Epstein and Peters

[32]We are gratefully indebted to an anonymous referee for bringing up this point.

(1999) conceptually proposed a universal message space for which the revelation principle still works with type space appropriately enlarged. However, as acknowledged in Peters (2001) and Martimort and Stole (2002), such universal mechanism approach is "complex and hard to apply because the agent's type must describe the mechanisms being used by the other principals, whether these other mechanisms depend on other principals' mechanisms, and so on" (Peters 2001, p. 1349). Peters (2001) and Martimort and Stole (2002) showed that, under certain conditions, it is without loss of generality to restrict competing principals to offer payoff-relevant menus, rather than consider any indirect communication mechanisms (the so-called menu theorem in the literature). Our Equivalence Theorem (Theorem 1) can be viewed as enlarging the buyer's type space in the spirit of Epstein and Peters (1999)'s idea.

Recently, Calzolari and Pavan (2008) and Pavan and Calzolari (2009) considered sequential contracting games in which the common agent contracts with different principals, respectively, in different periods. One of the major contributions of their two papers is to examine under what conditions "the menu theorems of simultaneous common agency extend to" their sequential setting (Pavan and Calzolari 2009, p. 524). Again, our main objective is to characterize the optimal NLP and examine its antitrust implications in an extensive form of game different from theirs. Whether a simultaneous-move game, a sequential-offering game, or a sequential-contracting game is more appropriate may depend on the context of applications. Our paper complements to the (sequential) common agency literature.

The idea that unchosen options can affect the equilibrium outcome was first mentioned by Bernheim and Whinston (1986*b*). In Bernheim and Whinston (1986*b*), the unchosen options cause multiple equilibria and inefficiency under complete information. As a result, they restored the efficiency by introducing the "truthful strategy" refinement. Our equilibrium outcomes, with or without unchosen options, are always inefficient. Moreover, Bernheim and Whinston (1986*a*,*b*) mainly considered "intrinsic common agency," whereas our paper studies delegated common agency, which is more complex, because the formation of the multi-sourcing common agency is endogenous. Martimort and Stole (2003) provided a similar insight that unchosen offers can discipline rivals in both intrinsic and delegated common agency under complete information. Due to the simultaneous-move nature of the principals in the game they considered, a wide range of equilibrium outcomes can be sustained, including the one without unchosen offers. Conversely, the equilibrium in our sequential-move setting always involves unchosen offers. In the offer game studied by Segal and Whinston (2003) in which one principal makes simultaneous offers to the agents, the problem of multiplicity of equilibria arises with unchosen offers, because the agents' beliefs about the principal's contracts with other agents can be arbitrary. Such arbitrary beliefs are absent from our sequential-move game with complete information. More importantly, in Segal and Whinston (2003), it is the informed principal that offers a menu to screen herself, in order to mitigate the negative inferences/beliefs from the uninformed agents. In our transformed one-principal-two-agent model, it is the uninformed principal that offers a menu to screen informed agents.

Appendix A Proofs

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 A.1. *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-O), 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-O). Define

$$\tau(Q) = \begin{cases} T(p) & \text{if } \exists p \in \mathcal{P} \text{ s.t. } Q(p) = Q \\ 0 & \text{if } Q = 0 \text{ and } \nexists p \in \mathcal{P} \text{ s.t. } Q(p) = 0 \\ \infty & \text{otherwise} \end{cases} . \tag{A.1}$$

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 $\tau(0) = T(\hat{p}) \le 0$, where the inequality follows from (B-IR). Thus, $\tau(0) \le 0$.

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-O) 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 A.2. *For any $\tau \in \mathcal{T}$ and any SPE of the subgame after Firm 1 offers τ, $Q : \mathcal{P} \to \mathbb{R}_+$, $T : \mathcal{P} \to \mathbb{R}$, and $\bar{p} \in \mathcal{P}$ given in (i) and (ii) in Lemma A.1 satisfy (B-IC), (B-IR), and (F2-O).*

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

$$V(Q(p), p) - \tau(Q(p)) \ge V(Q, p) - \tau(Q) \quad \forall (Q, p) \in \mathbb{R}_+ \times \mathcal{P}. \tag{A.2}$$

To see (B-IC), take $Q = Q(\tilde{p})$ for arbitrary $\tilde{p} \in \mathcal{P}$ in (A.2) and use (ii). To see (B-IR), take $Q = 0$ in (A.2) and use $\tau(0) \le 0$ 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-O). From Lemma A.1, there is a $\tau^* \in \mathcal{T}$ (defined by (A.1) with $\tau(\cdot), Q(\cdot), T(\cdot)$ replaced by $\tau^*(\cdot), Q^*(\cdot), T^*(\cdot)$) such that (15) holds and a SPE $(p^*(\tau^*), q^*(\tau^*, \cdot))$ of the subgame after Firm 1 offers τ^* is described by (13), (14), and (16).

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 A.2, 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-O), 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).[33] Suppose that (τ^*, p^*, q^*) is a SPE and $Q^*(\cdot), T^*(\cdot), \bar{p}^*$ satisfy (13), (14), (15), and (16). From Lemma A.2, $Q^*(\cdot), T^*(\cdot), \bar{p}^*$ satisfy (B-IC), (B-IR), and (F2-O). 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.[34]

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 Subsection 4.2, $(Q_\varepsilon(\cdot), T_\varepsilon(\cdot), \bar{p}^{**})$ satisfies all the constraints of (OP); the (F2-O) 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 ε is made arbitrarily small.

Define $\tau_\varepsilon(\cdot)$ by the right-hand side of (A.1) 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 (21), then establish that, given (21), (B-IR) is equivalent to (22).

[33] Admittedly, we use the results in Subsection 4.2, 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 Subsection 4.2 do not rely on Theorem 1. (One can always formally analyze (OP) as in Subsection 4.2 even if Theorem 1 is not true.)

[34] 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-O) 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^{**}.

Let $U(p) \equiv V(Q(p), p) - T(p)$. Then (B-IC) can be written as

$$U(p) - U(\tilde{p}) \geq V(Q(\tilde{p}), p) - V(Q(\tilde{p}), \tilde{p}) \ \ \forall p, \tilde{p} \in \mathcal{P}, \tag{A.3}$$

and (21) can be written as

$$U(p) - U(c) = \int_c^p V_p(Q(t), t)dt \ \ \forall p \in \mathcal{P}. \tag{A.4}$$

Step 1. (B-IC) implies (Mon) and (21).

Suppose (B-IC) is satisfied. Then (A.3) implies that, for any $p_1, p_2 \in \mathcal{P}$,

$$V(Q(p_1), p_2) - V(Q(p_1), p_1) \leq U(p_2) - U(p_1) \leq V(Q(p_2), p_2) - V(Q(p_2), p_1). \tag{A.5}$$

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 (A.5) implies

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

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, (A.5) implies (A.4). Therefore, (21) holds.

Step 2. (Mon) and (21) imply (B-IC).

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

$$\text{Proj}_{[0,k]}(D(p_2) - Q(p_1)) \geq \text{Proj}_{[0,k]}(D(p_2) - Q(p_2)), \tag{A.6}$$

$$\text{Proj}_{[0,k]}(D(p_1) - Q(p_1)) \geq \text{Proj}_{[0,k]}(D(p_1) - Q(p_2)). \tag{A.7}$$

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 (A.6) and (A.7) hold. In case (ii), we have

$D(p_2) \leq D(p_1) \leq Q(p_2)$ so that the right-hand sides of (A.6) and (A.7) 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 (A.6) and (A.7) are $k > 0$. Therefore, (A.6) and (A.7) hold in each case.

Recall that (21) is equivalent to (A.4). Therefore, for any $p_1, p_2 \in \mathcal{P}$ (no matter whether $p_1 \leq p_2$ or not), we have

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

where the inequality is from (A.6) when $p_1 \leq p_2$ and from (A.7) when $p_1 \geq p_2$. It proves (A.3) and hence (B-IC).

Step 3. Given (B-IC) (in fact, (21) only), (B-IR) is equivalent to (22).

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 (21), which is equivalent to (A.4), 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 (22). ■

Proof of Corollary 2. It is implied by (A.6) 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 3). Constraint (F2-Pro) requires that $(Q(\bar{p}), \bar{p})$ must be on the iso-profit curve. Constraint (F2-O') requires that the graph of $Q(\cdot)$ must not cut into the left side of the iso-profit curve.

From Subsection 4.2.2, 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.,

$$Q(\bar{p}) = \max\{D(\bar{p}) - k, \pi'(\bar{p})\}). \tag{A.8}$$

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-O'), $Q(\bar{p} - \varepsilon)$ 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 3 and 4). It proves (A.8).

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

$$\Pi_2 = (\bar{p} - c)(D(\bar{p}) - Q(\bar{p})), \tag{A.9}$$

which is equivalent to (28) at $p = \bar{p}$. It, together with (A.8), proves (27) 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 (29)). That is, (Q_0, x_0) is the intersection below $\bar{p}$ between the iso-profit curve and the curve $Q = \max\{D(p) - k, 0\}$.[35] In particular, $c < x_0 \leq \bar{p}$ and $Q_0 \leq \bar{Q}$. Now, recall that Π_1 can be written as (26) and visualized in Figure 4. In order to maximize Π_1, it remains to maximize the second integral in (26) (or Area B in Figure 4) subject to (Mon) and (F2-O'). Neglect (Mon) for a moment. Then, $Q(\cdot)$ on $[x_0, \bar{p}]$ must coincide the iso-profit curve, i.e., (28) 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-O') 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 5. Then, $Q(\cdot)$ is non-decreasing on $\mathcal{P}$ so that (Mon) is satisfied. It is also clear from Figure 5 that (F2-O') is satisfied. It proves (28). Finally, (28) 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 5). Thus,

[35] There is another intersection above $\bar{p}$ (which is on the vertical axis), but the intersection below $\bar{p}$ is uniquely given by (29).

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 5). 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 $Q_0 < \bar{Q}$, $c < x_0 < \bar{p} < p^m$, and $\pi'(\bar{p}) > D(\bar{p}) - k$ (see Figure 5 again). So Lemma 3 implies $\bar{Q} = \pi'(\bar{p})$, $Q_0 = \max\{D(x_0) - k, 0\}$, (19), and that both sides of (31) equal Π_2.

Next, we show (30). From Figure 4, (26) can be rewritten as

$$\begin{aligned}
\Pi_1 &= \int_0^{Q_0} u'(Q+k)dQ + x_0 \cdot \left(\bar{Q} - Q_0\right) + \int_{x_0}^{\bar{p}} \left(\bar{Q} - Q(p)\right) dp - c\bar{Q} \qquad \text{(A.10)} \\
&= \int_0^{Q_0} \left[u'(Q+k) - x_0\right] dQ + (\bar{p} - c)\bar{Q} - \int_{x_0}^{\bar{p}} Q(p)dp \\
&= \int_{x_0}^{\infty} \max\{D(p) - k, 0\}dp + (\bar{p} - c)\bar{Q} - \int_{x_0}^{\bar{p}} \left[D(p) - \frac{\Pi_2}{p - c}\right] dp,
\end{aligned}$$

where the last equality is due to (28) and $Q_0 = \max\{D(x_0) - k, 0\}$. Let

$$\text{TS}(p) \equiv u(D(p)) - c \cdot D(p) = \int_p^{\infty} D(t)dt + (p - c)D(p) \qquad \text{(A.11)}$$

denote the total surplus under linear pricing p. From (27),

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

so that we can further rewrite Π_1 as

$$\begin{aligned}
\Pi_1 &= \int_{x_0}^{\infty} \max\{D(p) - k, 0\}dp + \text{TS}(\bar{p}) - \Pi_2 - \int_{x_0}^{\infty} D(p)dp + \int_{x_0}^{\bar{p}} \frac{\Pi_2}{p - c}dp \\
&= \text{TS}(\bar{p}) - \int_{x_0}^{\infty} \min\{D(p), k\}dp + \left(\ln \frac{\bar{p} - c}{x_0 - c} - 1\right) \Pi_2. \qquad \text{(A.12)}
\end{aligned}$$

The partial derivatives of (A.12) 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 (27) (together with $\bar{Q} = \pi'(\bar{p})$) and (29) imply that $\partial \Pi_1/\partial \bar{p} = \partial \Pi_1/\partial x_0 = 0$. Hence, the total derivative of (A.12) with respect to Π_2 is

$$\frac{d\Pi_1}{d\Pi_2} = \ln \frac{\bar{p} - c}{x_0 - c} - 1. \tag{A.13}$$

Therefore, the first-order condition $d\Pi_1/d\Pi_2 = 0$ implies (30). Substituting this first-order condition into (A.12), the maximum Π_1 can be written as

$$\Pi_1 = \mathrm{TS}(\bar{p}) - \int_{x_0}^{\infty} \min\{D(p), k\} dp. \tag{A.14}$$

Last, we derive (32). Note that $T(\bar{p}) = \Pi_1 + c\bar{Q}$. Use the expression (A.10) for Π_1, we have

$$\begin{aligned} T(\bar{p}) &= \int_0^{Q_0} u'(Q+k)dQ + x_0 \cdot \left(\bar{Q} - Q_0\right) + \int_{x_0}^{\bar{p}} \left(\bar{Q} - Q(p)\right) dp \\ &= \int_0^{Q_0} u'(Q+k)dQ + \bar{p}\bar{Q} - x_0 Q_0 - \int_{x_0}^{\bar{p}} Q(p)dp \\ &= u(Q_0 + k) - u(k) + \int_{x_0}^{\bar{p}} p dQ(p). \end{aligned} \tag{A.15}$$

From (19), we know $Q(\cdot)$ is differentiable on $(x_0, \bar{p})$. Then, from (24), $T(\cdot)$ is differentiable on $(x_0, \bar{p})$ as well, and for all $p \in (x_0, \bar{p})$,

$$\begin{aligned} T'(p) &= V_Q(Q(p), p) \cdot Q'(p) = u'(\mathrm{Proj}_{[Q(p), Q(p)+k]}(D(p))) \cdot Q'(p) \\ &= u'(D(p)) \cdot Q'(p) = p \cdot Q'(p), \end{aligned} \tag{A.16}$$

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

$$\begin{aligned} D(p) - k &\leq Q(x_0) \ (\because p > x_0 \text{ and } Q_0 = \max\{D(x_0) - k, 0\}) \\ &\leq Q(p) \ (\because Q(\cdot) \text{ is increasing on } [x_0, \bar{p}]) \end{aligned}$$

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 (24), 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 (A.15) and (A.16), proves (32). ∎

Proof of Theorem 2. The results are from Lemma 4 and Theorem 1. First, (19) is already stated in Lemma 4. That $x_0 = c + (\bar{p} - c)/e$ follows from (30). (31) and (19) imply $Q(x_0) = \max\{D(x_0) - k, 0\}$ and $\bar{Q} = \pi'(\bar{p})$. Substituting $x_0 = c + (\bar{p} - c)/e$ into (31), canceling the positive factor $\bar{p} - c$, and using the fact that $\pi'(\bar{p}) = D(\bar{p}) + (\bar{p} - c)D'(\bar{p})$, we obtain (18). Finally, the formula (17) for $\tau(Q)$ with $Q \in [Q_0, \bar{Q}]$ is derived from (32) through changing of variable: $x(\cdot)$ for $Q(\cdot)$. ■

Proof of Corollary 1. Under condition (20), the left-hand side of (18) is strictly increasing in $\bar{p}$, and the right-hand side is non-increasing in $\bar{p}$. Therefore, (18) has at most one solution under (20). Theorem 2 then implies the essential uniqueness. ■

Proof of Proposition 1. From (17), $\tau'(\cdot) = x(\cdot)$ 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 (18) and $x_0 = c + (\bar{p} - c)/e$), both $\bar{Q}$ and Q_0 approach q^e because $\bar{Q} = Q(\bar{p})$ and $Q_0 = Q(x_0)$, and $\tau(\bar{Q})$ approaches $u(q^e)$ (from (17)). 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 $k > 0$.[36] ■

Proof of Proposition 2. Let $\hat{x}_0$ be the minimum equilibrium x_0 when $k \geq q^e$, given by (18) and $x_0 = c + (\bar{p} - c)/e$ with $\min\{D(x_0), k\} = D(x_0)$. Define $\hat{k} \equiv D(\hat{x}_0)$. From Theorem 2,

[36] 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 $k \geq \hat{k}$.

$\hat{k}$ satisfies the first two claims (see Figure 5).

The rest of the proof considers comparative statics for $k \in (0, \hat{k}]$. Following the proof of Lemma 4, we regard $\Pi_1, \bar{p}, \bar{Q}, x_0, Q_0, x(\cdot), \mathrm{TS}$ 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 $\Pi_2 = (\bar{p} - c)[D(\bar{p}) - \bar{Q}]$ and $\bar{Q} = \pi'(\bar{p})$ do not change. The x_0, Q_0, and Π_1 determined by $\Pi_2 = (x_0 - c)\min\{D(x_0), k\} = (x_0 - c)k$, $Q_0 = \max\{D(x_0) - k, 0\} = D(x_0) - k$, and (A.12) decrease as k increases (see Figure 5).

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 (A.13), 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 $\Pi_2 = (\bar{p}-c)[D(\bar{p})-\pi'(\bar{p})] = \pi(\bar{p})-(\bar{p}-c)\pi'(\bar{p})$ and the strict concavity of $\pi(\cdot)$ (Assumption 2), $\bar{p}$ must increase, and hence $\bar{Q}$ decreases following from $\bar{Q} = \pi'(\bar{p})$. Then $\mathrm{TS} = u(\bar{Q}) - c\bar{Q}$ decreases. Then $x_0 = c + (\bar{p} - c)/e$ increases and $Q_0 = D(x_0) - k$ decreases. The result for $D(\bar{p}) - \bar{Q}$ can be immediately seen from $D(\bar{p}) - \bar{Q} = D(\bar{p}) - \pi'(\bar{p}) = -(\bar{p} - c)D'(\bar{p})$.

■

Proof of Proposition 3. Straightforward and omitted. ■

The proof of Proposition 4 requires 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 A.3. *In any equilibrium,*

$$\begin{aligned}\Pi_2 + \mathrm{BS} &= \int_{x_0}^{\infty} \min\{D(p), k\}dp \qquad \text{(A.17)}\\ &= u(\min\{D(x_0), k\}) - x_0 \cdot \min\{D(x_0), k\}\\ &= u(D(x_0) - Q_0) - x_0 \cdot (D(x_0) - Q_0).\end{aligned}$$

Proof. The first equality follows from (A.14) and the equilibrium buyer's surplus as $\mathrm{BS} \equiv u(D(\bar{p})) - \bar{p}(D(\bar{p}) - \bar{Q}) - \tau(\bar{Q}) = \mathrm{TS} - \Pi_1 - \Pi_2$. The second equality is clear. The third equality is from $Q_0 = \max\{D(x_0) - k, 0\}$. ■

Proof of Proposition 4. Recall $\bar{Q} > \max\{D(\bar{p}) - k, 0\}$ from the proof of Lemma 4. 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, $\Pi_1 > \Pi_1^{LP}$ holds. 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 $k \in (0, \check{k}]$.

Now consider $k \in [\hat{k}, q^e)$. It follows that $D(x_0) \leq k < D(\bar{p}^{LP}) < q^e$, where the first inequality is from $k \geq \hat{k}$ and the definition of $\hat{k}$, and the second and third inequalities are from $k < q^e$ and Proposition 3. This can be rewritten as $x_0 \geq u'(k) > \bar{p}^{LP} > c$. In the NLP vs LP equilibrium, if Firm 2 deviates to charge $u'(k)$, which is weakly below x_0, then Firm 2's output would be k and its profit would be $(u'(k) - c)k$. In equilibrium this resulting profit must not exceed Firm 2's equilibrium profit Π_2. Therefore, $\Pi_2 \geq (u'(k) - c)k > (\bar{p}^{LP} - c)k = \Pi_2^{LP}$. 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,

$$
\begin{aligned}
\Pi_2^{LP} + \text{BS}^{LP} &= (\bar{p}^{LP} - c)k + \int_{\bar{p}^{LP}}^{\infty} D(p)dp \geq \int_c^{\infty} \min\{D(p), k\}dp \\
&> \int_{x_0}^{\infty} \min\{D(p), k\}dp \ (\because x_0 > c) \\
&= \Pi_2 + \text{BS} \ (\because (A.17)).
\end{aligned}
$$

This completes the proof of part (c). ■

Appendix B Why an Unchosen Bundle Helps

We provide a formal proof for the intuition on why adding an unchosen bundle can strictly improve Firm 1's profit over the "one-bundle equilibrium," as discussed in Subsection 3.2.

First, we determine the one-bundle equilibrium. Given that Firm 1 offers (Q, T), after seeing Firm 2's price offer p, the buyer's surplus is $V(Q, p) - T$ if she accepts (Q, T); and $V(0, p)$ otherwise. Provided $Q > 0$, the increasing differences property (10) of V implies that curve $V(Q, p) - T$, drawn against p, must cross only once curve $V(0, p)$ from below at x,[37] where x is determined by

$$V(Q, x) - T = V(0, x). \tag{B.1}$$

[37] The weak increasing differences property (10) of V does not guarantee $V(Q, p) - T$ and $V(0, p)$ always cross. However, *in equilibrium*, they must cross with each other, for otherwise either Firm 1 would have no sales when curve $V(0, p)$ is everywhere above curve $V(Q, p) - T$, or Firm 1 can always increase its profits by increasing T when curve $V(0, p)$ is everywhere below curve $V(Q, p) - T$.

It follows that the buyer accepts (Q,T) and buys $\mathrm{Proj}_{[0,k]}(D(p)-Q)$ from Firm 2 if $p>x$, and rejects (Q,T) and buys $\mathrm{Proj}_{[0,k]}(D(p))$ from Firm 2 if $p \le x$. Then, Firm 2's profit consists of two pieces as

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

Since Firm 1 has positive sales only if it can induce Firm 2 to set $p > x$, Firm 1 must ensure

$$\max_{p>x} \pi(Q,p) \ge \max_{p \le x} \pi(0,p). \tag{B.2}$$

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

$$\underset{Q,T}{\text{Maximize}} \left\{T - c\cdot Q \text{ s.t. (B.1) and (B.2)}\right\}. \tag{B.3}$$

Using (B.1) 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^*), (B.2) must be binding, i.e., $\pi(Q^*,p^*) = \pi(0,x^*)$, where $p^* \in \mathrm{argmax}_{p>x^*} \pi(Q^*,p)$ is Firm 2's equilibrium price, as illustrated in Figure 1(c). Also note that, at the optimum, $x^* \in (c,p^*)$.

Next, given the optimal one-bundle offer (Q^*,T^*) and the corresponding cut-off x^*, we construct *one extra bundle to relax the originally binding no-deviation constraint (B.2) for Firm 2*, so that Firm 1 can strictly improve its profit over $\Pi_1^* \equiv T^* - c\cdot Q^*$.

We add an extra bundle by picking any $Q_1 \in (0,Q^*)$, and letting $T_1(\epsilon) = V(Q_1,x^*) - V(0,x^*) - \epsilon$ for $\epsilon \ge 0$. 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

$$V(Q_1, x_0(\epsilon)) - T_1(\epsilon) = V(0, x_0(\epsilon)), \tag{B.4}$$

$$V(Q_1, x_1(\epsilon)) - T_1(\epsilon) = V(Q^*, x_1(\epsilon)) - T^*. \tag{B.5}$$

Accordingly, Firm 2's profit, as shown in Figure 1(d), consists of three pieces:

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

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.,

$$\max_{p > x_1(\epsilon)} \pi(Q^*, p) \ge \max_{p \le x_0(\epsilon)} \pi(0, p) \tag{B.6}$$

and

$$\max_{p > x_1(\epsilon)} \pi(Q^*, p) \ge \max_{x(\epsilon) < p \le x_1(\epsilon)} \pi(Q_1, p). \tag{B.7}$$

While the original no-deviation constraint (B.2) for Firm 2 is binding, after adding the extra bundle, the new ones (B.6) and (B.7) are *non-binding*, provided $\epsilon > 0$ is small enough. This can be seen from Figure 1(d), since for small $\epsilon > 0$ we have $c < x_0(\epsilon) < x^* < x_1(\epsilon) < p^*$. Then,

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

So (B.6) 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

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

So (B.7) is *not* binding, either.

Now we increase the price of the chosen Q^*-bundle from T^* to $T_2(\delta) = T^* + \delta$ with small $\delta > 0$. After that, $x_0(\epsilon)$ defined by (B.4) is unaffected; $x_1(\epsilon)$ defined by (B.5) 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, (B.6) and (B.7) still hold strictly. Accordingly, Firm 2 still charges p^* and the buyer still picks the Q^*-bundle, but now pays Firm 1 a higher bundle price $T_2(\delta) > T^*$. The resulting (sub-optimal) two-bundle profit $T_2(\delta) - c \cdot Q^*$ is thus strictly higher than Π_1^*.

References

Aghion, Philippe, and Patrick Bolton. 1987. "Contracts as a Barrier to Entry." *The American Economic Review*, 77(3): 388–401.

Asker, John, and Heski Bar-Isaac. 2014. "Raising Retailers' Profits: On Vertical Practices and the Exclusion of Rivals." *The American Economic Review*, 104(2): 672–86.

Baseman, Kenneth C., Frederick R. Warren-Boulton, and Glenn A. Woroch. 1995. "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.

Bernheim, B. Douglas, and Michael D. Whinston. 1986*a*. "Common Agency." *Econometrica*, 54(4): 923–942.

Bernheim, B Douglas, and Michael D Whinston. 1986*b*. "Menu auctions, resource allocation, and economic influence." *The Quarterly Journal of Economics*, 101(1): 1–31.

Bernheim, B. Douglas, and Michael D. Whinston. 1998. "Exclusive Dealing." *Journal of Political Economy*, 106(1): 64–103.

Calzolari, Giacomo, and Alessandro Pavan. 2008. "On the use of menus in sequential common agency." *Games and Economic Behavior*, 64(1): 329–334.

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

Calzolari, Giacomo, and Vincenzo Denicolò. 2013. "Competition with Exclusive Contracts and Market-Share Discounts." *The American Economic Review*, 103(6): 2384–2411.

Chao, Yong. 2013. "Strategic Effects of Three-Part Tariffs Under Oligopoly." *International Economic Review*, 54(3): 977–1015.

Chao, Yong, and Guofu Tan. 2017. "All-Units Discounts: Leverage and Partial Foreclosure in Single-Product Markets." *Canadian Competition Law Review*, 30(1): 93–111.

Chao, Yong, Guofu Tan, and Adam Chi Leung Wong. 2018. "All-units discounts as a partial foreclosure device." *The RAND Journal of Economics*, 49(1): 155–180.

Chao, Yong, Guofu Tan, and Adam Chi Leung Wong. 2019. "Asymmetry in capacity and the adoption of all-units discounts." *International Journal of Industrial Organization*, 65: 152–172.

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

Choné, Philippe, and Laurent Linnemer. 2016. "Nonlinear pricing and exclusion: II. Must-stock products." *The RAND Journal of Economics*, 47(3): 631–660.

Epstein, Larry G, and Michael Peters. 1999. "A revelation principle for competing mechanisms." *Journal of Economic Theory*, 88(1): 119–160.

Fu, Xiao, and Guofu Tan. 2019. "Abuse of Market Dominance Under China's Anti-Monopoly Law: The Case of Tetra Pak." *Review of Industrial Organization*, 54(2): 409–434.

Klemperer, Paul D., and Margaret A. Meyer. 1989. "Supply Function Equilibria in Oligopoly under Uncertainty." *Econometrica*, 57(6): 1243–1277.

Majumdar, Adrian, and Greg Shaffer. 2009. "Market-Share Contracts with Asymmetric Information." *Journal of Economics & Management Strategy*, 18(2): 393–421.

Martimort, David, and Lars Stole. 2002. "The Revelation and Delegation Principles in Common Agency Games." *Econometrica*, 70(4): 1659–1673.

Martimort, David, and Lars Stole. 2003. "Contractual externalities and common agency equilibria." *The B.E. Journal of Theoretical Economics*, 3(1).

Maskin, Eric, and John Riley. 1984. "Monopoly with Incomplete Information." *The RAND Journal of Economics*, 15(2): 171–196.

Mills, David E. 2010. "Inducing Downstream Selling Effort with Market Share Discounts." *International Journal of the Economics of Business*, 17(2): 129–146.

Oxenfeldt, Alfred R. 1951. *Industrial Pricing and Market Practices.* New York: Prentice-Hall.

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

Peters, Michael. 2001. "Common Agency and the Revelation Principle." *Econometrica*, 69(5): 1349–1372.

Rasmusen, Eric B., J. Mark Ramseyer, and John S. Wiley, Jr. 1991. "Naked Exclusion." *The American Economic Review*, 81(5): 1137–1145.

Segal, Ilya, and Michael D. Whinston. 2003. "Robust Predictions for Bilateral Contracting with Externalities." *Econometrica*, 71(3): 757–791.

Segal, Ilya R., and Michael D. Whinston. 2000. "Naked Exclusion: Comment." *The American Economic Review*, 90(1): 296–309.

Simpson, John, and Abraham L. Wickelgren. 2007. "Naked Exclusion, Efficient Breach, and Downstream Competition." *The American Economic Review*, 97(4): 1305–1320.

Tirole, Jean. 1988. *The Theory of Industrial Organization.* MIT Press.

Wilson, Robert B. 1993. *Nonlinear Pricing.* Oxford University Press, New York.