All-units Discounts in Retail Contracts

Sreya Kolay
University of Rochester

Greg Shaffer
University of Rochester

and

Janusz A. Ordover
*New York University**

August 2002

Abstract

All-units discounts in retail contracts refer to discounts that lower a retailer's wholesale price on *every* unit purchased when the retailer's purchases equal or exceed some quantity threshhold. These discounts pose a challenge to economic theory because, like linear pricing, the average price paid by the retailer is (almost) everywhere the same as its marginal price (the price of obtaining one more unit) but, unlike linear pricing, purchases at higher quantities may be cheaper than purchases at lower quantities. Since it is difficult to understand why a manufacturer would ever charge less for a larger order if its intentions were benign, antitrust authorities have argued that all-units discounts must be exclusionary. However, in this paper, we show that all-units discounts may profitably arise absent any exclusionary motive—in a bilateral-monopoly setting. All-units discounts can solve the problem of double marginalization when demand is known, and they can profitably induce second-degree price discrimination when retailers have private information about demand. Compared to linear pricing, all-units discounts lead to lower consumer prices and are welfare improving. Compared to two-part tariffs, all-units discounts may lead to higher or lower prices depending on demand parameters.

Key Words: *Vertical control, double marginalization, price discrimination*

*Addresses: Sreya Kolay, Department of Economics, University of Rochester, Rochester, NY 14627; Greg Shaffer, Simon School of Business, University of Rochester, Rochester, NY 14627; and Janusz A. Ordover, Department of Economics, New York University, New York, NY 10003. E-mail: srya@troi.cc.rochester.edu; shaffer@simon.rochester.edu; and janusz.ordover@nyu.edu. Corresponding author: Greg Shaffer, voice: 585-275-4497; fax: 585-273-1140.

I Introduction

The use of all-units discounts in intermediate-goods markets is common. Coca-Cola and Irish Sugar use them in their contracts with retailers. British Airways uses them in their contracts with travel agents, and Michelin uses them in their contracts with tire dealers. An all-units discount refers to a discount that lowers a retailer's wholesale price on *every* unit purchased when the retailer's purchases equal or exceed some quantity threshhold or target. The discount is usually specified in terms of some percentage off list price, and is sometimes also referred to as a 'target rebate,' because when the target is reached, the retailer receives a rebate on all units previously purchased.

All-units discounts have the property that the average price paid by a retailer is almost everywhere the same as its marginal price (the price of obtaining one more unit). Since linear pricing also has this property, one might think that the economic effects of all-units discounts and linear pricing would be similar.[1] But therein lies a puzzle: if the effects of all-units discounts are similar to those of linear pricing, where the retailer pays a constant per-unit wholesale price, why would we ever observe all-units discounts in practice? Since all-units discounts are in general more costly to administer (because the manufacturer must keep track of the retailer's purchases), we should not expect to observe them when linear pricing is optimal. And when linear pricing is not optimal, e.g., when it leads to double markups, it is not obvious why a manufacturer would ever offer a quantity-discount scheme that applies to all units when it can solve the problem of double marginalization with discounts that apply only to incremental units, or, equivalently, with a two-part tariff?[2]

Despite their prevalence, all-units discounts have for the most part been ignored in the economics and business literatures. And when they are mentioned in these literatures, it is often simply asserted that their purpose is to induce the buyer to purchase a larger quantity, or (even worse) that their use is irrational.[3] On the other hand, antitrust authorities tend to look upon all-units discounts as exclusionary.[4] They focus on the fact that, with all-units discounts, purchases at

[1] With linear pricing, the retailer's average per-unit price is independent of the amount purchased. In other words, a retailer pays the same constant per-unit price (and no fixed fee) regardless of how much it purchases.

[2] The problem of double marginalization was first pointed out by Spengler (1950). The use of two-part tariffs as a solution has been noted by Moorthy (1987), Tirole (1988; 174-176), Katz (1989; 664-665) and many others.

[3] Nahmias (2001; 216) states "The all-units schedule appears irrational in some respects. Why would .... actually charge less for a larger order? One reason would be to provide an incentive for the purchaser to buy more."

[4] See Dekeyser, "Pricing and Discounts/Rebates in Dominant Companies–The Commission's View," DG Competition, European Commission, Brussels, at 3. See also the article by Tom, Balto, and Averitt (2000).

higher quantities may be cheaper than purchases at lower quantities, and conclude from this that the manufacturer's intent is to harm its rivals by offering discounts that would be unprofitable for them to match when spread over smaller volumes. In other words, all-units discounts are viewed by the antitrust authorities with suspicion because they provide a strong incentive for a retailer to promote the sale of products on which it is eligible to earn a rebate at the expense of other products, particularly when it is close to the level of sales required to reach the rebate target.

The problem with this reasoning is that it might be used to condemn (unjustifiably in our view) as exclusionary almost any type of pricing scheme in which a manufacturer offers a quantity-based discount, because it might be argued that these schemes also induce a retailer to promote the sale of some products at the expense of other products in order to qualify for the manufacturer's best discounts. Yet antitrust authorities do not typically view other quantity-discount schemes (e.g., two-part tariffs and incremental-units discounts) with the same degree of suspicion. This raises several questions. Is the bias against all-units discounts justified? Do they unambiguously lead to lower social welfare relative to other quantity discount schemes? And, given that there is a bias against all-units discounts, why would a manufacturer ever risk prosecution by offering them?

We are the first to model the use of all-units discounts in retail contracts. In this paper we are interested in whether all-units discounts might arise for non-exclusionary reasons, and we ask how their profitability and social-welfare properties compare to other types of quantity-discounts. To rule out exclusion as a possible motive, we restrict attention to a setting in which an upstream monopolist or manufacturer sells its output to a downstream monopolist or retailer (exclusion is implicitly ruled out in a bilateral-monopoly setting because neither firm has a horizontal rival). We consider cases in which demand is known at the time of contracting, and cases in which it is not. In the former case, we compare the performance of all-units discounts against a benchmark of two-part tariffs, which are known to be optimal in simple bilateral-monopoly settings. In the latter case, we compare the performance of menus of all-units discounts and menus of two-part tariffs.

We establish three main results. First, we show that all-units discounts, like two-part tariffs, can solve the problem of double marginalization when consumers' demand is deterministic and known at the time of contracting by both parties. Instead of selling its product to the retailer at marginal cost plus a fixed fee, which is one way of solving the problem, the manufacturer can offer the retailer an all-units discount contract and eliminate double markups by choosing the target quantity to induce the joint-profit maximizing retail price and the percentage discount off list price to divide the surplus. All-units discount contracts in this case are welfare improving compared to linear pricing schedules, benefiting firms with higher profits and consumers with lower prices.

Second, we show that when the retailer has private information about consumers' demand, the manufacturer will prefer to offer the retailer a menu of all-units discounts rather than a menu of two-part tariffs. That is, when the retailer has private information about consumers' demand, the manufacturer earns higher profit under the profit-maximizing menu of all-units discount contracts than under the profit-maximizing menu of two-part tariff contracts. All-units discounts in this case are more efficient than two-part tariffs at inducing the retailer to reveal the state of demand (they are a better screening device) and hence lead to greater surplus extraction for the manufacturer.

Third, we show that when the retailer has private information, the profit-maximizing menu of all-units discounts may result in higher or lower prices for consumers than the profit-maximizing menu of two-part tariffs. Whether it is higher or lower depends on the shape of consumer demand. For example, if demand is linear, and the demand curves in the different states of nature have a common vertical intercept, or are vertical translations of each other (i.e., parallel shifts), then all-units discounts lead to lower consumer prices and are welfare improving. On the other hand, if demand is linear, and the demand curves in the different states of nature have a common horizontal intercept, then all-units discounts lead to higher consumer prices and are welfare worsening.

In summary, our results imply that all-units discounts need not be irrational, as some have asserted, nor are they necessarily used for exclusionary purposes (if at all), as antitrust authorities have asserted. We find that all-units discounts may profitably arise in retail contracts even in the absence of an exclusionary motive. The discounts can be viewed as a means of vertical control when consumers' demand is deterministic and known by both parties at the time of contracting,[5] and as a screening device to induce the retailer to reveal the state of demand and extract surplus when the retailer has private information.[6] Consumer welfare can be higher or lower as a result.

[5] Seminal articles on vertical control include Warren-Boulton (1974) and Mathewson and Winter (1984).

[6] The problem we consider is formally equivalent to implementing second-degree price discrimination among re-

The paper is organized as follows. In the next section, we introduce the model, discuss the problem of double marginalization, and present our first main result. In Section 3, we extend the model to allow for demand uncertainty and present our second main result. Our third main result and illustrative examples are presented in Section 4. We offer concluding remarks in Section 5.

II The case of demand certainty

We begin with the case of bilateral monopoly and demand certainty. There is a single upstream firm which produces a good at constant marginal cost c. The upstream firm sells the good to a single downstream firm which then resells the good to final consumers. We will refer to the upstream firm as the "manufacturer" and the downstream firm as the "retailer." For simplicity, we assume the retailer incurs no costs of distribution, and hence, the only cost the retailer bears is the amount it pays to the manufacturer for the good. Let $T(q)$ represent the payment the retailer makes to the manufacturer when it purchases $q \geq 0$ units of the good. Let p denote the retail price and $p = p(q)$ denote the consumers' inverse demand function, which is assumed to be downward sloping.

Given our assumptions, the manufacturer's profit can be expressed as $\pi^m = T(q) - cq$, the retailer's profit as $\pi^r = p(q)q - T(q)$, and the overall joint profit as $\pi^{int} = p(q)q - cq$, which is the profit an integrated firm would earn if it produced and distributed an amount q of the product.[7]

The case of bilateral monopoly with demand certainty is the simplest setting in which to examine the issue of vertical control. The vertical-control literature asks: what kinds of contracts will induce the retailer to choose the same quantity that an integrated firm would choose.[8] Comparing π^r and π^{int}, it is easy to see that $T(q) = cq$ is one such contract, and a two-part tariff contract, where $T(0) = 0$, $T(q) = cq + F$ for all $q > 0$, and $F \geq 0$ is a fixed fee, is another. Both contracts induce the retailer to maximize overall joint profit. The former contract allows the retailer to capture all the surplus. For some $F > 0$, the latter contract allows the manufacturer to capture all the surplus.

tailers with different demands. In the context of our model, the manufacturer offers a menu of options to the retailer and prices the menu so as to induce the retailer to reveal the state of demand by the option it selects.

[7]We follow the tradition of the vertical-restraints literature and say that an integrated firm is one that controls all the pricing and quantity decisions made by the vertical structure. See Tirole (1988; 170).

[8]Mathewson and Winter (1984) use the terminology of instruments and targets. One can think of the manufacturer's decision variables as instruments, and the retailer's decision variables as targets. "The control problem consists in knowing how to use the instruments to reach, or come close to, the desired values of the targets—that is, the values that maximize the vertical structure's aggregate (vertically integrated) profit (Tirole, 1988: 173)." For specific applications, see Perry and Porter (1990), O'Brien and Shaffer (1992), Winter (1993), and Reiffen (1999).

Figure 1: Two–part tariff

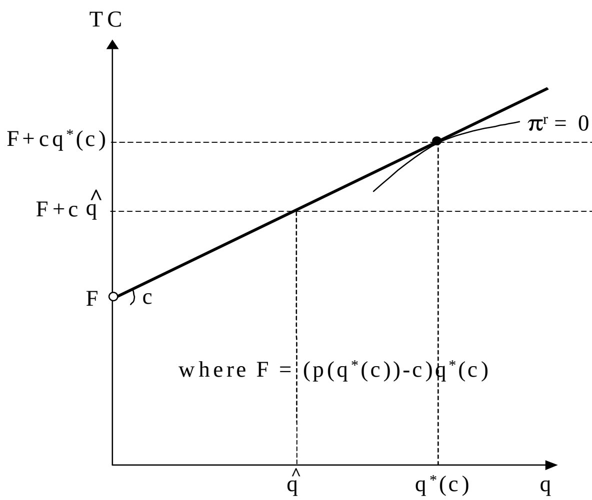
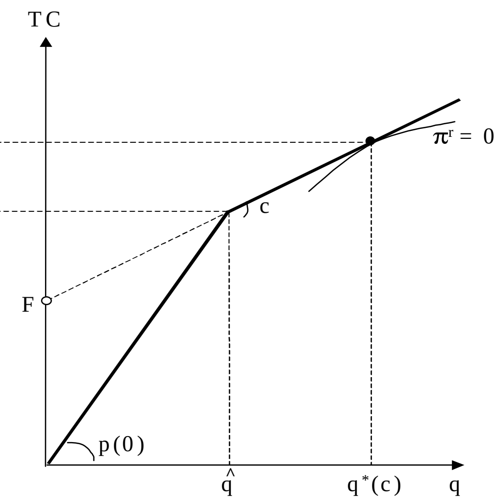


Figure 2: Incremental–units discount



To see that the manufacturer cannot do any better than this, let w be the wholesale price and F the fixed fee. Let $q^*(w)$ denote the retailer's profit-maximizing quantity assuming it pays F, i.e., $q^*(w)$ solves $\max_q p(q)q - wq$. Then the manufacturer's problem with a take-it-or-leave-it offer is

$$\max_{w,F} \; (w-c)q^*(w) + F \tag{1}$$

such that

$$(p(q^*(w)) - w)\, q^*(w) - F \geq 0, \tag{2}$$

where the inequality in (2) ensures that the retailer earns non-negative profit. Since the maximand is increasing in F, the manufacturer will choose the fixed fee to satisfy the constraint with equality. Substituting this into the maximand and solving yields $w = c$ and $F = (p(q^*(c)) - c)q^*(c)$.[9] Thus, the manufacturer sets its wholesale price equal to its marginal cost and charges a fixed fee that fully extracts the retailer's profit. Figure 1 depicts the total outlay of the retailer for any quantity it might purchase under this contract. The jump up in cost at $q = 0$ represents the fixed amount the retailer must pay if it purchases from the manufacturer. The upward-sloping line with constant slope c represents the additional payment the retailer makes as q increases. Given this contract,

[9] After substitution, the manufacturer's problem is to choose w to solve $\max_w \; (p(q^*(w)) - c)q^*(w)$. Differentiating with respect to w, and using the first-order condition from the retailer's profit maximization, we have that the profit-maximizing wholesale price solves $(w-c)\,\frac{\partial q^*(w)}{\partial w} = 0$, which implies that the manufacturer should set $w = c$.

the retailer chooses the integrated quantity, $q^*(c)$, and pays $F + cq^*(c)$ to the manufacturer. As depicted in Figure 1, this quantity occurs at the point of tangency between the retailer's isoprofit curve (the retailer earns zero profit) and the manufacturer's optimal two-part tariff.

In the case in Figure 1, and in the case of $T(q) = cq$, the manufacturer earns no profit on the per-unit sales of its product, essentially making the retailer the residual claimant to all flow profit. Alternatively, the manufacturer can induce the integrated outcome by offering a contract with a high inframarginal per-unit price, as long as the retailer's marginal price is set at cost:

$$T(q) = \begin{cases} p(0)q & \text{if } q < \hat{q}, \\ \\ c\,(q - \hat{q}) + p(0)\hat{q} & \text{if } q \geq \hat{q}, \end{cases} \tag{3}$$

where $\hat{q}$ is implicitly defined by $(p(0) - c)\,\hat{q} = p(q^*(c)) - c)q^*(c)$. In this contract, which we illustrate in Figure 2, the manufacturer offers the initial units up to $\hat{q}$ at a per-unit price of $p(0) > c$. For additional units beyond $\hat{q}$, the retailer's per-unit price falls to c. At $q = \hat{q}$, the retailer's total outlay is $p(q^*(c)) - c)q^*(c) + c\hat{q}$, which is the same outlay the retailer would make if it purchased $\hat{q}$ units and faced the contract in Figure 1. For $q > \hat{q}$, the two curves are the same. For $q < \hat{q}$, the total outlay of the retailer is $p(0)q$, which is less than the same quantity would cost in Figure 1.

Since the retailer would never purchase q less than $\hat{q}$ when faced with the contract in (3) (because at a per-unit price of $p(0)$ its cost would exceed its revenue), it follows that the retailer will be induced to purchase the same quantity in Figure 2 as in Figure 1. Since the contract in Figure 1 induces the integrated outcome, then so does the contract in Figure 2. Thus, the economic effects of the two contracts are identical: the price to consumers is the same and so is the division of profit.

The contracts in Figures 1 and 2 share the property that the retailer's average per-unit price is greater than its marginal price. This allows the manufacturer to extract surplus from the retailer without distorting its incentives at the margin. In Figure 1, the manufacturer achieves this separation of average and marginal price by specifying two parameters, a fixed fee and a constant wholesale price. In Figure 2, the manufacturer offers an initially high wholesale price and then discounts the wholesale price on the retailer's incremental purchases beyond the quantity $\hat{q}$. This latter contract is an example of an incremental-units discount, and the fact that it can achieve the same outcome as a single two-part tariff is not surprising, since one can think of an incremental-units discount as a menu of two-part tariffs. For example, the contract in Figure 2 is the lower envelope of the two-part tariff contract $T(q) = p(0)q$, where the fixed fee is zero, and the two-part tariff contract $T(0) = 0$, $T(q) = cq + F$ for all $q > 0$, where the fixed fee is the same as in Figure 1.

In contrast, contracts that offer linear pricing do not allow the retailer's average per-unit price to differ from its marginal price. This is problematic because it leads to the well-known problem of double marginalization when the manufacturer makes the offer (see Spengler, 1950). In order to extract surplus, the manufacturer must charge $w > c$, but in doing so, it necessarily distorts downward the retailer's quantity choice, $q^*(w) < q^*(c)$, for all $w > c$. The problem arises because in choosing its quantity, the retailer faces a wholesale price that is higher than the manufacturer's production marginal cost. A distortion is introduced because both firms have positive mark-ups.

Contracts that offer all-units discounts would also seem to be problematic because they too equate the retailer's average per-unit price and its marginal price. These contracts have the form

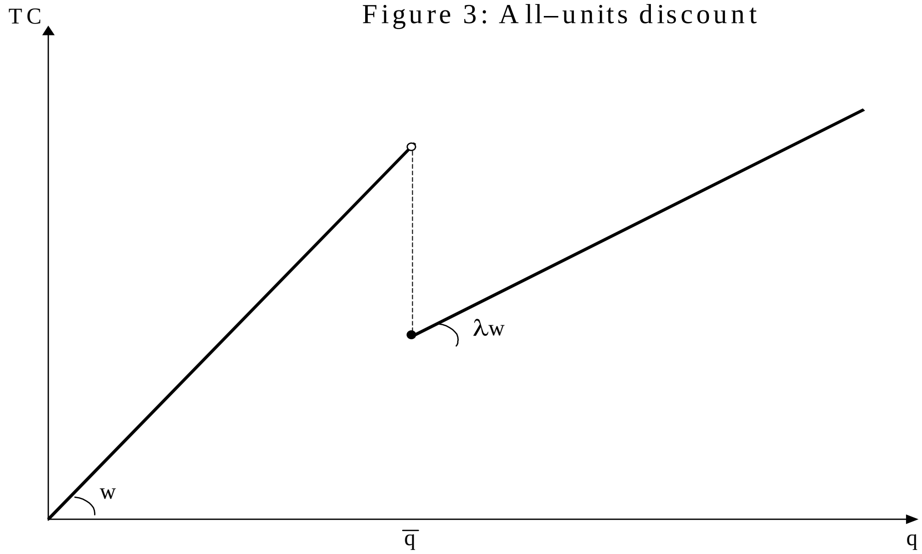
$$T(q) = \begin{cases} wq & \text{if } q < \overline{q}, \\ \lambda wq & \text{if } q \geq \overline{q}, \end{cases} \tag{4}$$

where $\lambda \in (0, 1)$. In this type of contract, if the retailer purchases $q < \overline{q}$, its average per-unit price and marginal price equals w, whereas if the retailer purchases $q \geq \overline{q}$, its average per-unit price and marginal price equals λw. Since $\lambda < 1$, the retailer is rewarded for purchasing at least $\overline{q}$ units. One can think of w as the manufacturer's list price and $1 - \lambda$ as the percentage discount off list price, which is *applied to every unit the retailer purchases* if the retailer reaches the target quantity.

Unlike linear pricing, this contract involves setting three parameters: the wholesale price w, which is valid over the range $[0, \overline{q})$, the wholesale price λw, which is valid for all units if the retailer purchases at least $\overline{q}$ units, and the target quantity $\overline{q}$, which determines the retailer's wholesale price. The most salient characteristic of contracts with all-units discounts is that the retailer's total outlay jumps down at $\overline{q}$, which implies that purchases at higher quantities can be cheaper than purchases at lower quantities. This property is easily illustrated with the help of Figure 3.

Contracts with all-units discounts have raised antitrust concerns that they may be exclusionary

Figure 3: All–units discount



when used by dominant firms because the downward discontinuity in the total outlay schedule (at $\overline{q}$ in Figure 3) allegedly unfairly induces the retailer to promote the sale of the manufacturer's product at the expense of smaller competitors who may not be able to offer such potentially large rebates.[10] Indeed, the effective marginal wholesale price in these contracts at the target quantity is negative, which under some definitions of predation might be construed as predatory pricing.

The antitrust concerns are all the more striking in this instance because, until now, no efficiency rationale has been offered to explain the use of all-units discounts. In contrast, contracts with two-part tariffs, and incremental-units discounts, which may also give rise to similar concerns, have efficiency rationales and thus tend to be treated far more leniently. In our bilateral monopoly setting, where exclusionary motives are absent, we have seen that these latter types of contracts can solve the double-marginalization problem, raising firm profits and lowering prices for consumers.

We now show that contracts with all-units discounts can also solve the double-marginalization problem. To begin, consider the set of contracts that offer an initial wholesale price of $p(0)$, a discounted wholesale price of $\lambda p(0)$, where $\lambda \in [\frac{c}{p(0)}, \frac{p(q^*(c))}{p(0)}]$, and a quantity threshhold of $q^*(c)$:

[10]The European Commission has brought several cases against upstream firms for, among other things, offering all-units discounts in their contracts. To our knowledge, the U.S. antitrust agencies have not explicitly condemned contracts with all-units discounts, although their concerns are similar (see Tom, Balto, and Averitt, 2000).

Figure 4: All–units discount
(λw = p(q*(c)))

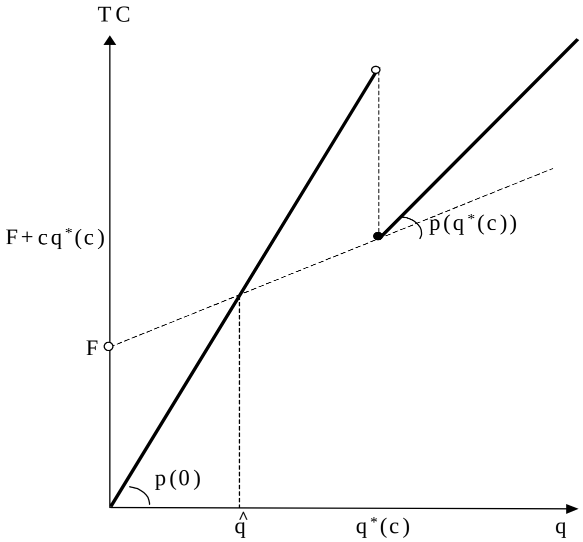
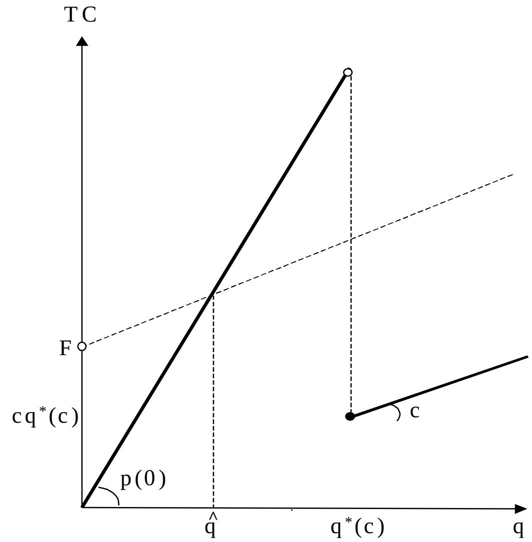


Figure 5: All–units discount
(λw = c)



$$T(q) = \begin{cases} p(0)q & \text{if } q < q^*(c), \\ \\ \lambda p(0)q & \text{if } q \geq q^*(c). \end{cases} \tag{5}$$

With these contracts, the retailer would never purchase less than $q^*(c)$ because at a per-unit price of $p(0)$ its cost would exceed its revenue, and we know by the definition of $q^*(c)$ that the retailer would never purchase more than $q^*(c)$ because $\lambda p(0) \geq c$ implies that $q^*(\lambda p(0)) \leq q^*(c)$. Thus, the contracts in (5) induce the retailer to choose $q = q^*(c)$, realizing the integrated outcome. When λ is at its upper bound of $\frac{p(q^*(c))}{p(0)}$, as we illustrate in Figure 4, the manufacturer earns all the surplus. When λ is at its lower bound of $\frac{c}{p(0)}$, as we illustrate in Figure 5, the retailer earns all the surplus. Intermediate levels of λ reflect a more balanced sharing of the overall surplus. In all cases, quantity is higher and consumer prices are lower than under linear pricing with double marginalization.

Proposition 1 *Contracts with all-units discounts can solve the problem of double marginalization.*

Proposition 1 implies that the integrated outcome can be achieved even if the retailer's average per-unit price is the same as its marginal price. This feature of the contract makes all-units discounts appear more similar to linear pricing, which cannot induce the integrated outcome, than other quantity discount schemes, which distinguish between inframarginal and marginal payments.

The contracts in (5) induce the integrated outcome because the quantity threshhold $q^*(c)$ effectively caps the retailer's price to consumers at $p(q^*(c))$.[11] The price is effectively capped because if the retailer were to charge more than $p(q^*(c))$, it would not be able to sell $q^*(c)$ units and thus would not qualify for the target rebate. Of course, the retailer also does not want to charge less than $p(q^*(c))$ because, under the contracts in (5), it never wants to sell more than $q^*(c)$ units (this follows because the retailer's wholesale price is everywhere weakly larger than c). If the manufacturer makes a take-it-or-leave-it offer, the double-marginalization problem is solved by eliminating the retailer's markup. If the retailer makes a take-it-or-leave-it offer, the double marginalization problem is solved by eliminating the manufacturer's markup. Otherwise, the contracts in (5) ensure that the sum of the firms' markups is equal to the markup that an integrated firm would have. To the extent that both firms have some bargaining power, there will be agreement on the quantity threshhold, $q^*(c)$, but disagreement on the percentage discount off list price. The manufacturer will prefer a lower percentage discount, while the retailer will prefer a higher percentage discount.

III The case of uncertain demand

The analysis thus far has assumed that consumer demand is known by both firms at the time of contracting. In this section, we extend the model to allow for demand uncertainty. This is an important case to consider for at least two reasons. First, in practice, consumer demand often varies depending on the state of nature, and thus it is realistic to assume that demand may sometimes be unpredictable. For example, sales of ice-cream and cold drinks will be higher when the weather is warm, and lower when the weather is cold, but *apriori* neither a manufacturer nor its retailers may know at the time of contracting which state will occur. Second, the model as yet does not have predictive power to explain why one contract form may be chosen over another, whereas the extended model with demand uncertainty does. As we shall see, the equivalence among two-part tariffs, incremental-units discounts, and all-units discounts no longer holds in this latter case.

We assume, for simplicity, that consumer demand can take one of two forms. It may be high or low depending on the state of nature. The low-demand state of nature occurs with probability

[11] A similar idea motivates the use of maximum resale price maintenance (a vertical restraint in which a resale price-ceiling is imposed on the retailer) as a solution to the problem of double marginalization.

$\alpha \in [0,1]$ and the high-demand state of nature occurs with probability $(1-\alpha)$. We denote the low-demand state by "L" and the high-demand state by "H". The retailer faces the inverse demand $p = p_L(q)$ under state L and the inverse demand $p = p_H(q)$ under state H, with $p_H(\cdot) \geq p_L(\cdot)$.

Let $R_i(q) \equiv p_i(q)q$ denote the retailer's revenue from the sale of $q \geq 0$ units of the good in the ith state of nature, where $i \in \{L, H\}$. Our assumptions on inverse demands imply that the retailer's revenue in the high-demand state is weakly larger than the retailer's revenue in the low-demand state for a given quantity sold: $R_H(q) \geq R_L(q)$. We also make the following assumptions on $R_i(q)$:

$$\frac{\partial^2 R_i(q)}{\partial q^2} < 0, \quad \frac{\partial R_H(q)}{\partial q} > \frac{\partial R_L(q)}{\partial q}. \tag{6}$$

The first assumption in (6) implies that the retailer's revenue in each state is concave (this ensures that firm i's marginal revenue is downward sloping). The second assumption in (6) implies that the retailer's marginal revenue in the high-demand state is always greater than its marginal revenue in the low-demand state (this property is sometimes referred to as the single-crossing condition).

The timing of the game is as follows. In the first stage, the manufacturer specifies the terms at which it will sell its good to the retailer. It does so without knowing whether consumer demand will be high or low. In the second stage, the uncertainty is resolved and the retailer realizes the nature of the demand it faces. The retailer then chooses whether to purchase and how much to purchase from the manufacturer, and pays the manufacturer according to the terms of its contract. The retailer resells this quantity to final consumers, earning revenue $R_i(q)$, where $i \in H, L$ is the state of nature. We solve the game backwards, using subgame perfection as our solution concept.

We assume the manufacturer cannot contract on the state of nature (if it could, we are back to the model in the previous section), and so cannot prevent a retailer facing one demand state from pretending that another demand state has occurred. Thus, the manufacturer must choose its contract in stage one to induce the retailer to reveal the true state of demand in stage two.[12]

This situation has been studied in the literature on self-selection and price discrimination (in our model, the manufacturer offers a menu of options to the retailer and prices the menu so as

[12] In what follows, we assume that the manufacturer wants to serve both retailer types. That is, we assume that selling to the high-demand retailer only is less profitable than selling to both types. See Salant (1989) for a characterization of the necessary and sufficient conditions for discrimination to be optimal for the two-type case.

to induce the retailer to reveal the state of demand by the option it selects).[13] Conventional wisdom is that (1) the retailer will earn zero surplus in the low-demand state and positive surplus in the high-demand state; and (2) the manufacturer will distort downward the quantity chosen by a low-demand retailer, but not the quantity chosen by a high-demand retailer.[14] We will show these results using the manufacturer's profit-maximizing menu of two-part tariffs, or, equivalently, incremental-units discount, as a benchmark. We will then compare consumer prices and social welfare under this benchmark with those of the manufacturer's optimal menu of all-units discounts.

Menu of two-part tariffs, incremental-units discount

Suppose the manufacturer offers a menu of two-part tariffs $((w_L, F_L), (w_H, F_H))$, where w_i denotes the per-unit price and F_i is the fixed fee, with (w_L, F_L) meant for the retailer facing low demand and (w_H, F_H) meant for the retailer facing high demand. Let $q_L^*(w)$ denote the low-demand retailer's quantity choice if it faces a per-unit price of w and purchases a positive quantity, and let $q_H^*(w)$ be defined similarly, i.e., $q_i^*(w)$ solves $\frac{\partial R_i(q)}{\partial q} = w$. Then the manufacturer's problem is

$$\max_{w_L, w_H, F_L, F_H} \alpha(w_L - c)q_L^*(w_L) + (1-\alpha)(w_H - c)q_H^*(w_H) + \alpha F_L + (1-\alpha)F_H, \tag{7}$$

subject to the low-demand retailer choosing a positive quantity under (w_L, F_L),

$$R_L(q_L^*(w_L)) - w_L q_L^*(w_L) - F_L \geq 0, \tag{8}$$

and the high-demand retailer choosing to purchase under (w_H, F_H) rather than (w_L, F_L),[15]

$$R_H(q_H^*(w_H)) - w_H q_H^*(w_H) - F_H \geq R_H(q_H^*(w_L)) - w_L q_H^*(w_L) - F_L. \tag{9}$$

Our assumptions on the revenue functions, and the fact that the maximand in (7) is increasing in F_L and F_H imply that these two constraints will be binding. Specifically, it follows that F_L will be chosen to satisfy (8) with equality, and that F_H will be chosen to satisfy (9) with equality,

[13] See, for example, Willig (1978), Spence (1977, 1980), Roberts (1979), Goldman, Leland, and Sibley (1984), Maskin and Riley (1984), and the excellent survey on second-degree price discrimination in Tirole (1988), pp. 142-158.

[14] This result depends on the assumption that the retailer's demand in the high-demand state is independent of its demand in the low-demand state. Ordover and Panzar (1982) show that if the demands across states of nature, or among different types of consumers, are interrelated, then a distortion may arise even in the high-demand state.

[15] Two constraints are suppressed: a high-demand retailer must purchase a positive quantity, and a low-demand retailer must purchase under (w_L, F_L) rather than (w_H, F_H) (these constraints are trivially satisfied).

i.e., the low-demand retailer will pushed to indifference between purchasing a positive quantity and not purchasing, and the high-demand retailer will be pushed to indifference between choosing a quantity under contract (w_H, F_H) and choosing a quantity under contract (w_L, F_L). Thus, a low-demand retailer will earn zero profit while a high-demand retailer will earn positive profit.[16]

Substituting the fixed fees that satisfy (8) and (9) with equality into the maximand, and assuming the new maximand is concave in w_L and w_H, we have that the optimal w_H and w_L solve

$$\left(\frac{\partial R_H(q_H^*(w_H))}{\partial q} - c\right)\frac{\partial q_H^*(w_H)}{\partial w_H} = 0, \tag{10}$$

$$\alpha\left(\frac{\partial R_L(q_L^*(w_L))}{\partial q} - c\right)\frac{\partial q_L^*(w_L)}{\partial w_L} + (1-\alpha)(q_H^*(w_L) - q_L^*(w_L)) = 0. \tag{11}$$

We see from the expression in (10) that the manufacturer should charge $w_H = c$ to the high-demand retailer, thereby inducing it to purchase and resell $q_H^*(c)$, the integrated quantity when demand is high. However, because of the term $(1-\alpha)(q_H^*(w_L) - q_L^*(w_L))$, the expression in (11) implies that the manufacturer will want to distort the low-demand retailer's quantity. Since the high-demand retailer's marginal revenue is everywhere above the low-demand retailer's marginal revenue, it follows that $q_H^*(w_L) > q_L^*(w_L)$, for all $w_L \geq 0$, which implies that the left side of (11) is positive when evaluated at $w_L = c$. Thus, the manufacturer should charge $w_L > c$ and induce the retailer to purchase and resell less than $q_L^*(c)$, the integrated quantity when demand is low.

Let $((w_L^*, F_L^*), (w_H^*, F_H^*))$ solve the manufacturer's problem in (7) - (9), and let $T_i^* = w_i^* q_i^*(w_i^*) + F_i^*$ denote the amount a retailer facing demand state i pays to the manufacturer, $i = H, L$. Then, the manufacturer's maximized profit given the profit-maximizing menu of two-part tariffs is:

$$\pi^{2PT} = \alpha(T_L^* - cq_L^*(w_L^*)) + (1-\alpha)(T_H^* - cq_H^*(w_H^*)).$$

Figure 6 depicts the total outlay of the retailer for any quantity it might purchase under the manufacturer's profit-maximizing menu of two-part tariffs. The two-part tariff meant for the low-demand retailer is given by the upward-sloping line beginning at F_L^* on the vertical axis, with slope w_L^*. The two-part tariff meant for the high-demand retailer is given by the upward-sloping

[16] To see this, note that the high-demand retailer's profit is $R_H(q_H^*(w_L)) - w_L q_H^*(w_L) - F_L$, which is strictly positive because $R_H(q_H^*(w_L)) - w_L q_H^*(w_L) - F_L \geq R_H(q_L^*(w_L)) - w_L q_L^*(w_L) - F_L > R_L(q_L^*(w_L)) - w_L q_L^*(w_L) - F_L = 0$.

Figure 6: Menu of two-part tariffs

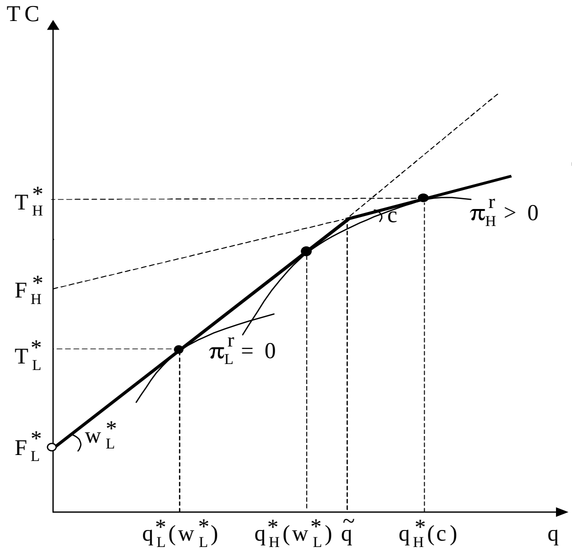
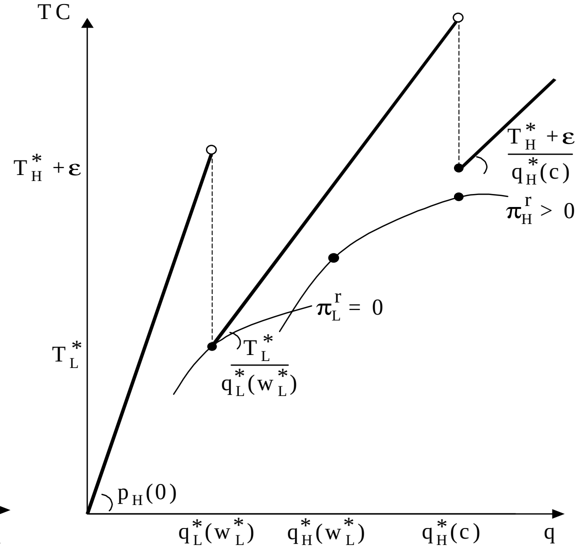


Figure 7: Menu of all-units discounts



line beginning at F^*_H on the vertical axis, with slope $w^*_H = c$. The cost-minimizing outlay for any quantity the retailer might purchase is given by the lower envelope of these two lines, with a kink-point at $\tilde{q}$, which is implicitly defined by $w^*_L\tilde{q} + F^*_L = c\tilde{q} + F^*_H$. If demand turns out to be low, the retailer will choose $q^*_L(w^*_L)$ from this envelope and pay the manufacturer T^*_L. Since (8) is satisfied with equality, it follows that the retailer in this case is indifferent between purchasing $q^*_L(w^*_L)$ and zero, earning profit $\pi^r_L = 0$. If demand turns out to be high, the retailer will choose $q^*_H(c)$ from this envelope and pay the manufacturer T^*_H. Since (9) is satisfied with equality, it follows that the retailer in this case is indifferent between purchasing $q^*_H(c)$ and $q^*_H(w^*_L)$, earning profit $\pi^r_H > 0$.

Since the retailer will *only* choose points from along the lower envelope of the menu of two-part tariffs, as depicted in Figure 6 (the cost-minimizing outlay), the manufacturer could instead have induced the same profit-maximizing outcome with a single contract that traces out this curve:

$$T^{IU}(q) \equiv \begin{cases} w^*_L q + F^*_L & \text{if } q < \tilde{q}, \\ \\ c(q - \tilde{q}) + w^*_L\tilde{q} + F^*_L & \text{if } q \geq \tilde{q}. \end{cases} \tag{12}$$

With this contract, the retailer faces a per-unit price of w^*_L for all units purchased up to $\tilde{q}$. For incremental units purchased beyond $\tilde{q}$, the retailer pays a per-unit price of c. If demand is low, the retailer purchases $q_L^*(w_L^*)$ and pays the manufacturer T_L^*. If demand is high, the retailer purchases $q_H^*(c)$ and pays the manufacturer T_H^*. Since the quantities and payments are the same as in the case of the profit-maximizing menu of two-part tariffs, it follows that each firm's profit will also be the same. This establishes that contracts with incremental-units discounts can obtain the same outcome as contracts with menus of two-part tariffs. Since any contract with incremental-units discounts can be replicated with a menu of two-part tariffs, the converse can also be shown. Thus, the equivalence between these two types of contracts extends to the case of demand uncertainty.

All-units discounts

We now show that there exists a contract with all-units discounts that induces the retailer to choose the same quantities as in Figure 6, but yields strictly higher profit than π^{2PT}. Thus, the equivalence among different types of contracts does *not* extend to contracts with all-units discounts.

Proposition 2 *When the retailer has private information about demand, the manufacturer can earn higher profit with a menu of all-units discounts than with a menu of two-part tariffs.*

Proof: Let $\epsilon > 0$ be arbitrarily small, and consider the menu of contracts $T_1(q)$ and $T_2(q)$, where

$$T_1(q) \equiv \begin{cases} p_H(0)q & \text{if } q < q_L^*(w_L^*) \\ \frac{T_L^*}{q_L^*(w_L^*)}q & \text{if } q \geq q_L^*(w_L^*) \end{cases}, \quad T_2(q) \equiv \begin{cases} p_H(0)q & \text{if } q < q_H^*(c) \\ \frac{T_H^*+\epsilon}{q_H^*(c)}q & \text{if } q \geq q_H^*(c) \end{cases}, \tag{13}$$

or, equivalently, the contract that corresponds to the cost-minimizing outlay given this menu:

$$T^{AU}(q) \equiv \begin{cases} p_H(0)q & \text{if } q < q_L^*(w_L^*), \\ \frac{T_L^*}{q_L^*(w_L^*)}q & \text{if } q_L^*(w_L^*) \leq q < q_H^*(c), \\ \frac{T_H^*+\epsilon}{q_H^*(c)}q & \text{if } q \geq q_H^*(c). \end{cases} \tag{14}$$

Figure 7 depicts the retailer's total cost for any quantity it might purchase under $T^{AU}(q)$. For all units purchased up to $q_L^*(w_L^*)$, the retailer's per-unit price is $p_H(0)$. For purchase quantities of $q_L^*(w_L^*)$ or between $q_L^*(w_L^*)$ and $q_H^*(c)$, the retailer's per-unit price on *all units* is $\frac{T_L^*}{q_L^*(w_L^*)}$, and for purchase quantities that equal or exceed $q_H^*(c)$, the retailer's per-unit price on *all units* is $\frac{T_H^*+\epsilon}{q_H^*(c)}$.

To expedite the proof, it is useful to begin with a couple of observations. First, we note that, when faced with contract $T^{AU}(q)$, the retailer will never purchase $q < q_L^*(w_L^*)$ because at a per-unit price of $p_H(0)$ its marginal revenue is everywhere below its marginal cost. Second, we note that $T^{AU}(q) = T^{IU}(q)$ at $q = q_L^*(w_L^*)$. Third, we note that for all $q > q_L^*(w_L^*)$, $T^{AU}(q) > T^{IU}(q)$.[17]

These observations imply that the retailer will choose $q_L^*(w_L^*)$ and pay T_L^* when the low-demand state occurs. Recall that $q_L^*(w_L^*) = \arg\max_q R_L(q) - T^{IU}(q)$. Hence, it must also be the case that $q_L^*(w_L^*) = \arg\max_q R_L(q) - T^{AU}(q)$. At this quantity, the low-demand retailer earns zero profit.

These observations also imply that the retailer will choose $q_H^*(c)$ and pay $T_H^* + \epsilon$ when the high-demand state occurs. To see this, recall that $q_H^*(c) = \arg\max_q R_H(q) - T^{IU}(q)$ and $T^{IU}(q_H^*(c)) = T_H^*$. Then, it follows that $R_H(q_H^*(c)) - T^{AU}(q_H^*(c)) > R_H(q) - T^{AU}(q)$, for all $q \neq q_H^*(c)$, because

$$\begin{aligned} R_H(q_H^*(c)) - T^{AU}(q_H^*(c)) &= R_H(q_H^*(c)) - T^{IU}(q_H^*(c)) - \epsilon \\ &> R_H(q) - T^{IU}(q) \ \text{ if } q \neq q_H^*(c) \\ &\geq R_H(q) - T^{AU}(q) \ \text{ for all } q \geq q_L^*(w_L^*), \end{aligned}$$

where the first inequality follows because ϵ is arbitrarily small, and the second inequality follows because $T^{AU}(q) \geq T^{IU}(q)$ in the relevant range. At $q_H^*(c)$, the high-demand retailer earns positive profit, but not as much as it would have earned if the manufacturer had offered the contract $T^{IU}(q)$.

Since the manufacturer earns $T_L^* - cq_L^*(w_L^*)$ when the low-demand state occurs, and $T_H^* + \epsilon - cq_H^*(w_H^*)$ when the high-demand state occurs, its expected payoff under contract $T^{AU}(q)$ is

$$\pi^{AU} = \alpha(T_L^* - cq_L^*(w_L^*)) + (1-\alpha)(T_H^* + \epsilon - cq_H^*(w_H^*)).$$

Comparing π^{AU} and π^{2PT}, we see that $\pi^{AU} = \pi^{2PT} + (1-\alpha)\epsilon$, which implies that the manufacturer earns higher profit under the all-units discount contract in (14). Q.E.D.

The manufacturer can extract a higher profit with an all-units discount contract because such contracts are more efficient at inducing the retailer to reveal the state of demand. Put simply, the flexibility afforded by the discontinuous outlay schedule makes it possible for the manufacturer to

[17] To see this, note that for all units between $q_L^*(w_L^*)$ and $q_H^*(c)$, the per-unit price in $T^{IU}(q)$ is either w_L^* or c, which is less than the per-unit price of $\frac{w_L^* q_L^*(w_L^*) + F_L^*}{q_L^*(w_L^*)}$ in $T^{AU}(q)$. And, for all units greater than or equal to $q_H^*(c)$, the per-unit price in $T^{IU}(q)$, which is c, is less than the per-unit price in $T^{AU}(q)$, which is $\frac{cq_H^*(c) + F_H^* + \epsilon}{q_H^*(c)}$.

charge higher prices for all quantities greater than $q_L^*(w_L^*)$ while still inducing a retailer in the high-demand state to purchase $q_H^*(c)$. As can be seen from Figure 7, the retailer in the high-demand state cannot reach the same isoprofit curve that it did in Figure 6 because the profitability of purchasing any quantity between $q_L^*(w_L^*)$ and $q_H^*(c)$, and in particular of purchasing $q_H^*(w_L^*)$, is lower when the high-demand retailer faces the contract $T^{AU}(q)$ than when it faces the manufacturer's profit-maximizing menu of two-part tariffs. Thus, choosing to purchase under the contract option meant for the low-demand retailer—$T_1(q)$—is less profitable than it was before, implying that the manufacturer does not have to leave the retailer as large an informational rent in this case.

IV Profit-maximizing quantities with all-units discounts

Menus of two-part tariffs are viewed more favorably in antitrust law than menus of all-units discounts because the former are thought to have efficiency justifications while the latter are thought only to be exclusionary. Yet, as we showed in section II, all-units discounts can solve the double-marginalization problem, and as we showed in section III, all-units discount contracts may even be the preferred choice of the manufacturer relative to other types of quantity-discount contracts in certain settings, e.g., in a bilateral-monopoly setting when the retailer has private information about consumers' demand. Moreover, all-units discounts need not lower welfare in these settings. In our example in Figure 7, the all-units discount contract hurts the retailer (relative to the two-part tariff contract in Figure 6) but does not hurt consumers, since the quantity sold by the retailer in each state is unchanged. This implies that, in our example, there is no effect on social welfare.

In this section, we extend the analysis to consider the welfare effects of the manufacturer's *profit-maximizing* all-units discount contract. Our main result is that that the direction of change is ambiguous. Depending on the functional form of demand, it is possible for consumers to be better or worse off with all-units discounts than with two-part tariffs or incremental-units discounts.

Let $(T_L(q), T_H(q))$ denote the menu of all-units discount contracts offered by the manufacturer, where T_L is meant for the low-demand retailer and T_H is meant for the high-demand retailer, and

$$T_L(q) = \begin{cases} w_L q & \text{if } q < \overline{q}_L \\ \\ \lambda_L w_L q & \text{if } q \geq \overline{q}_L \end{cases}, \quad T_H(q) = \begin{cases} w_H q & \text{if } q < \overline{q}_H \\ \\ \lambda_H w_H q & \text{if } q \geq \overline{q}_H \end{cases}.$$

Let $q_L^{**}(w_i, \lambda_i, \overline{q}_i) \in \arg\max_q R_L(q) - T_i(q)$ denote the low-demand retailer's quantity choice if it purchases under contract $T_i(q)$, and define $q_H^{**}(w_i, \lambda_i, \overline{q}_i)$ similarly. Then the manufacturer solves

$$\max_{w_L,\lambda_L,\overline{q}_L,w_H,\lambda_H,\overline{q}_H} \alpha(T_L(q_L^{**}(w_L,\lambda_L,\overline{q}_L)) - cq_L^{**}(w_L,\lambda_L,\overline{q}_L)) \\ + (1-\alpha)\left(T_H(q_H^{**}(w_H,\lambda_H,\overline{q}_H)) - cq_H^{**}(w_H,\lambda_H,\overline{q}_H)\right), \tag{15}$$

subject to the low-demand retailer choosing a positive quantity under $T_L(q)$,

$$R_L(q_L^{**}(w_L,\lambda_L,\overline{q}_L)) - T_L(q_L^{**}(w_L,\lambda_L,\overline{q}_L)) \geq 0, \tag{16}$$

and the high-demand retailer choosing to purchase under $T_H(q)$ rather than $T_L(q)$,[18]

$$R_H(q_H^{**}(w_H,\lambda_H,\overline{q}_H)) - T_H(q_H^{**}(w_H,\lambda_H,\overline{q}_H)) \geq R_H(q_H^{**}(w_L,\lambda_L,\overline{q}_L)) - T_L(q_H^{**}(w_L,\lambda_L,\overline{q}_L)). \tag{17}$$

The solution to the manufacturer's problem is not as straightforward here as it is in the case of two-part tariffs because, with all-units discounts, there are no fixed fees to equate the two sides of (16) and (17), and hence seemingly no way to extract the maximum possible surplus from the retailer in each demand state. Nevertheless, as we now show, even without fixed fees, it is possible for the manufacturer to choose the terms of its contracts to satisfy (16) and (17) with equality.

Lemma 1 *Let $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ be a solution to the manufacturer's problem. Then it must be that (16) and (17) are satisfied with equality at $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$. The low-demand retailer earns zero profit and the high-demand retailer is indifferent between purchasing $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ under contract $T_H(q)$ and purchasing $q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ under contract $T_L(q)$:*

$$R_L(q_L^{**}(w_L^{**},\lambda_L^{**},\overline{q}_L^{**})) - T_L(q_L^{**}(w_L^{**},\lambda_L^{**},\overline{q}_L^{**})) = 0,$$

$$R_H(q_H^{**}(w_H^{**},\lambda_H^{**},\overline{q}_H^{**})) - T_H(q_H^{**}(w_H^{**},\lambda_H^{**},\overline{q}_H^{**})) = R_H(q_H^{**}(w_L^{**},\lambda_L^{**},\overline{q}_L^{**})) - T_L(q_H^{**}(w_L^{**},\lambda_L^{**},\overline{q}_L^{**})).$$

Proof: See the appendix.

We can understand Lemma 1 as follows. The manufacturer can induce a retailer in the low-demand state to purchase zero or at least $\overline{q}_L$ by choosing w_L sufficiently high. And, for a finite

[18]We suppress the other constraints. As in section III, the high-demand retailer's participation is assured given (16), and the constraint that the low-demand retailer purchase under $T_L(q)$ rather than $T_H(q)$ is trivially satisfied.

w_L, the manufacturer can induce the retailer to purchase at most $\overline{q}_L$ by choosing λ_L sufficiently high. It follows that the manufacturer can induce the low-demand retailer to purchase exactly $\overline{q}_L$ and extract all surplus by choosing $\lambda_L w_L$ appropriately. For example, the manufacturer can extract all surplus and induce the low-demand retailer to purchase $q_L^{**}(w_L, \lambda_L, \overline{q}_L) = \overline{q}_L$ by choosing $w_L \geq p_H(0)$ and $\lambda_L w_L = \frac{R_L(\overline{q}_L)}{\overline{q}_L} = p_L(\overline{q}_L)$. The extraction of surplus from a retailer in the high-demand state can be achieved similarly, except that instead of choosing $\lambda_H w_H$ to extract all of the high-demand retailer's surplus, the manufacturer chooses $\lambda_H w_H$ to make the retailer indifferent between purchasing $q_H^{**}(w_H, \lambda_H, \overline{q}_H) = \overline{q}_H$ under $T_H(q)$ and purchasing $q_H^{**}(w_L, \lambda_L, \overline{q}_L)$ under $T_L(q)$.

The following lemma shows that the manufacturer must choose its contract terms in this way.

Lemma 2 *Let* $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ *be a solution to the manufacturer's problem. Then it must be that* $w_L^{**} \geq p_L(0)$, $\lambda_L^{**} w_L^{**} = p_L(\overline{q}_L^{**})$, $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \overline{q}_L^{**}$ *and* $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \overline{q}_H^{**}$.

Proof: See the appendix.

Lemma 2 implies that at the optimum the manufacturer must choose its contract terms to induce the retailer to purchase at the quantity threshhold that corresponds to each demand state. Since we know that for any $\overline{q}_L$ the manufacturer can choose $w_L \geq p_L(0)$ and $\lambda_L w_L = p_L(\overline{q}_L)$ to induce $q_L^{**}(w_L, \lambda_L, \overline{q}_L) = \overline{q}_L$ and satisfy (16) with equality, and since we know that for any $\overline{q}_H$ the manufacturer can choose w_H and $\lambda_H w_H$ to induce $q_H^{**}(w_H, \lambda_H, \overline{q}_H) = \overline{q}_H$ and satisfy (17) with equality,[19] we can use Lemmas 1 and 2 to rewrite the manufacturer's problem in (15)–(17) as

$$\max_{w_L, \lambda_L, \overline{q}_L, \overline{q}_H} \quad \alpha(R_L(\overline{q}_L) - c\overline{q}_L) - (1-\alpha)\left(R_H(q_H^{**}(w_L, \lambda_L, \overline{q}_L)) - T_L(q_H^{**}(w_L, \lambda_L, \overline{q}_L))\right)$$

$$+ (1-\alpha)\left(R_H(\overline{q}_H) - c\overline{q}_H\right), \tag{18}$$

such that

$$w_L \geq p_L(0), \ \lambda_L w_L = p_L(\overline{q}_L). \tag{19}$$

It follows that the profit-maximizing $\overline{q}_H$ solves $\max_q (1-\alpha)\left(R_H(q) - cq\right)$, implying that, for $\alpha \neq 1$, there is no distortion in the quantity purchased by a retailer in the high-demand state.

[19] For example, the manufacturer can choose w_H such that $R_H(q_H^*(w_H)) - w_H q_H^*(w_H) < R_H(q_H^{**}(w_L, \lambda_L, \overline{q}_L)) - T_L(q_H^{**}(w_L, \lambda_L, \overline{q}_L))$, and $\lambda_H w_H$ such that $\lambda_H w_H \overline{q}_H = R_H(\overline{q}_H) - (R_H(q_H^{**}(w_L, \lambda_L, \overline{q}_L)) - T_L(q_H^{**}(w_L, \lambda_L, \overline{q}_L)))$.

Proposition 3 *Let* $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ *be a solution to the manufacturer's problem. Then it must be that a retailer in the high-demand state will choose* $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = q_H^*(c)$.

Proof: See the appendix.

Proposition 3 establishes that a retailer in the high-demand state will choose the integrated quantity, which is the same quantity that a high-demand retailer would choose under the manufacturer's profit-maximizing menu of two-part tariffs. Thus, the welfare comparison between the profit-maximizing menu of two-part tariffs and the profit-maximizing menu of all-units discounts will depend solely on the relation between the quantities purchased by the low-demand retailer in the two cases. From the manufacturer's problem in (18)–(19), we see that, in the case of the profit-maximizing menu of all-units discounts, the manufacturer will want to distort downward the quantity purchased by a retailer in the low-demand state. There are two subcases to consider.

Proposition 4 *Let* $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ *be a solution to the manufacturer's problem. Then it must be that if* $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$, *then* $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ *solves*

$$\alpha(\frac{\partial R_L(q)}{\partial q} - c) + (1-\alpha)\frac{\partial p_L(q)}{\partial q} q_H^*(p_L(q)) = 0, \tag{20}$$

and if $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, *then* $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ *solves*

$$\alpha(\frac{\partial R_L(q)}{\partial q} - c) + (1-\alpha)(\frac{\partial R_L(q)}{\partial q} - \frac{\partial R_H(q)}{\partial q}) = 0. \tag{21}$$

Proof: See the appendix.

Proposition 4 characterizes the manufacturer's choice of $\overline{q}_L^{**}$, and hence the retailer's induced choice $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, for the two subcases that arise depending on whether or not a retailer in the high-demand state would be constrained by $\overline{q}_L^{**}$ if it purchased under contract $T_L(q)$. As the linear-demand examples below make clear, this will depend on exogenous parameters that determine the gap between consumers' inverse demands in the low and high demand states. If the high-demand retailer would be unconstrained by $\overline{q}_L^{**}$, then $\overline{q}_L^{**}$ must solve (20). If the high-demand retailer would be constrained by $\overline{q}_L^{**}$, then $\overline{q}_L^{**}$ must solve (21). In both cases, we see that $\overline{q}_L^{**}$, and hence $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, will be less than $q^*(c)$, the integrated quantity in the low-demand state.

The welfare effects of all-units discounts

To determine whether welfare is higher with all-units discounts or with two-part tariffs, we compare the low-demand retailer's quantity choice as characterized in Proposition 4 with its quantity choice under the profit-maximizing menu of two-part tariffs, $q_L^*(w_L^*)$, where w_L^* is characterized in (11).

Subcase 1: $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$

If we evaluate the left-hand sides of (20) and (11) at $q_L^*(w_L^*)$ and w_L^* respectively, subtract (11) from (20), and use the fact that retailer optimality implies $w_L^* = \frac{\partial R_L(q_L^*(w_L^*))}{\partial q}$, so that

$$\frac{\partial q_L^*(w_L^*)}{\partial w_L^*} = \frac{1}{\frac{\partial^2 R_L(q_L^*(w_L^*))}{\partial q^2}}, \tag{22}$$

then the left-hand side of (20) can be written as

$$-(1-\alpha)\left(q_H^*(w_L^*) - q_L^*(w_L^*)\right)\frac{\partial^2 R_L(q_L^*(w_L^*))}{\partial q^2} + (1-\alpha)\frac{\partial p_L(q_L^*(w_L^*))}{\partial q} q_H^*(p_L(q_L^*(w_L^*))). \tag{23}$$

It follows that if $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$, so that the condition in (20) applies, and if the left-hand side of (20) is decreasing in q, then $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) > q_L^*(w_L^*)$ if and only if (23) is positive.

Subcase 2: $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$

If we evaluate the left-hand sides of (21) and (11) at $q_L^*(w_L^*)$ and w_L^* respectively, subtract (11) from (21), and use (22), and the fact that retailer optimality implies

$$\frac{\partial R_L(q_L^*(w_L^*))}{\partial q} = \frac{\partial R_H(q_H^*(w_L^*))}{\partial q},$$

then the left-hand side of (21) can be written as

$$(1-\alpha)\int_{q_L^*(w_L^*)}^{q_H^*(w_L^*)} \left(\frac{\partial^2 R_H(q)}{\partial q^2} - \frac{\partial^2 R_L(q_L^*(w_L^*))}{\partial q^2}\right) dq. \tag{24}$$

It follows that if $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, so that the condition in (21) applies, and if the left-hand side of (21) is decreasing in q, then $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) > q_L^*(w_L^*)$ if and only if (24) is positive.

Linear Demands

With linear demands, the left-hand sides of (20) and (21) are decreasing in q, implying that the signs of (23) and (24) suffice to determine whether the low-demand retailer's quantity choice under the profit-maximizing menu of all-units discounts is greater than, less than, or equal to the low-demand retailer's quantity choice under the profit-maximizing menu of two-part tariffs. Since the high-demand retailer's quantity choice is the same for both, we have the following proposition.

Proposition 5 *Suppose consumer demands are linear and $\alpha \neq 1$. Then, if $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$, consumer welfare is higher (lower) with all-units discounts than with two-part tariffs if and only if*

$$2(q_H^*(w_L^*) - q_L^*(w_L^*)) \;\; > \; (<) \;\; q_H^*(p_L(q_L^*(w_L^*))). \tag{25}$$

Otherwise, if $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, consumer welfare is higher (lower) if and only if

$$\frac{\partial^2 R_H(q)}{\partial q^2} \;\; > \; (<) \;\; \frac{\partial^2 R_L(q)}{\partial q^2}. \tag{26}$$

Proof: With linear demands, (25) follows by substituting $\frac{\partial^2 R_L(q)}{\partial q^2} = 2\frac{\partial p_L(q)}{\partial q}$ into (23). And, with linear demands, (26) follows by noting that $\frac{\partial^2 R_i(q)}{\partial q^2}$, $i \in \{H, L\}$, are constants in (24). Q.E.D.

As we now show, consumer welfare can be higher, the same, or lower with all-units discounts depending on the functional form of demand. This can be most easily seen when consumers' demand in the high and low demand states are sufficiently close that $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$. In this case, (26) implies that consumer welfare is higher (lower) if and only if the slope of the marginal-revenue curve in the high-demand state is larger (smaller) than the slope of the marginal-revenue curve in the low-demand state. The three possible outcomes are illustrated below in Figures 8, 9, and 10.

Example in which welfare is higher with all-units discounts

Suppose the inverse demands are given by $p_i(q) = a - \frac{q}{\theta_i}$, where $i \in \{H, L\}$ and $\theta_H > \theta_L$, so that, when plotted (see Figure 8), the inverse demand is everywhere flatter in the high-demand state. Since this implies that the corresponding marginal-revenue curve is everywhere flatter in the high-demand state, it follows immediately from (26) that, when $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, consumer prices are

Figure 8 : Welfare is higher with all-units discounts

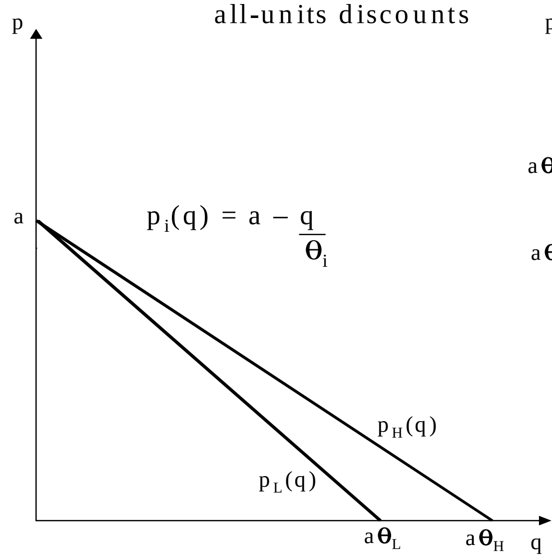
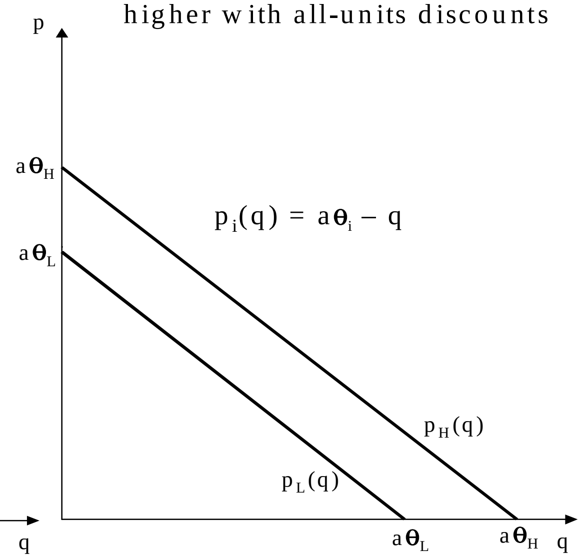


Figure 9 : Welfare is the same or higher with all-units discounts



lower and social welfare is higher with all-units discounts than with two-part tariffs. As we show in the appendix, a similar conclusion also holds when (25) applies, i.e., when $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$.

This case might arise, for example, if the differences in the high and low-demand states correspond to differences in the population size of a market with no change in the elasticity of demand. For example, suppose the manufacturer sells a product in a locale that depends on tourism, which in turn depends on the weather (tourism is higher when the weather is good and lower when the weather is bad). Then one can think of the high-demand state as corresponding to an increase in the population size of a market relative to the low-demand state with no change in the underlying distribution of consumer preferences. At the choke price, the quantity demanded in either state is zero, but for any price with positive demand, the high-demand quantity is a multiple of the low-demand quantity reflecting its larger population size. It is plausible that at the time of contracting, the manufacturer may not know whether the weather will be good or bad, and so it must offer a menu of options to the retailer to induce self-selection. Our findings imply that antitrust authorities would be wrong to forbid all-units discounts in this case because, in addition to the manufacturer's profit being higher with all-units discounts, consumer surplus and social welfare will also be higher.

Example in which welfare is the same or higher with all-units discounts

Suppose the inverse demands are given by $p_i(q) = a\theta_i - q$, where $i \in \{H, L\}$ and $\theta_H > \theta_L$. When plotted (see Figure 9), we see that the inverse demands in the two states of nature are parallel. Since this implies that the corresponding marginal-revenue curves will also be parallel, it follows immediately from (26) that, when $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, consumer prices and social welfare will be unchanged with all-units discounts. However, as we show in the appendix, if the differences between demand states are sufficiently large that $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$, then consumer prices will always be lower and social welfare will always be higher with all-units discounts than with two-part tariffs.

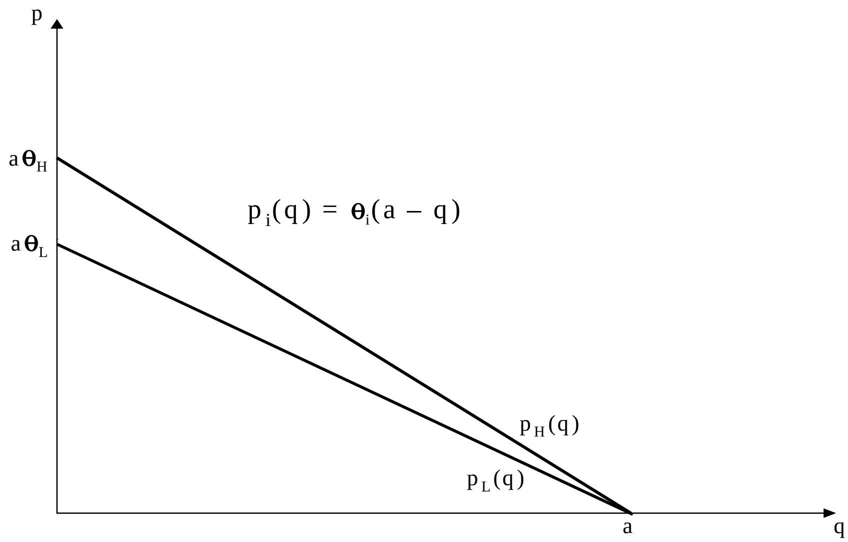
This case might arise, for example, if both the market size and consumers' willingness-to-pay are affected by the state of nature. For example, with cold drinks and ice cream, consumers may be willing to consume more and pay a higher price when the weather is warm than when it is cold. Our findings imply that antitrust authorities would be wrong to forbid all-units discounts in this case because then consumer surplus and social welfare would either be unaffected or would decrease.

Example in which welfare is lower with all-units discounts

Suppose the inverse demands are given by $p_i(q) = \theta_i(a - q_i)$, where $i \in \{H, L\}$ and $\theta_H > \theta_L$, so that, when plotted (see Figure 10), the inverse demand is everywhere steeper in the high-demand state. Since this implies that the corresponding marginal-revenue curve is everywhere steeper in the high-demand state, it follows immediately from (26) that, when $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, consumer prices are higher and social welfare is lower with all-units discounts than with two-part tariffs. As we show in the appendix, a similar conclusion holds when (25) applies, i.e., when $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$.

This case might arise, for example, if the differences in the high and low-demand states correspond to differences in consumers' incomes. For example, suppose the manufacturer sells a product for which demand depends primarily on consumers' discretionary income. In good times, consumers will have more to spend, while in recessionary times, consumers will have less to spend. Thus, when the product's price is zero, demand is the same in the two states of nature, but as the product's price increases, demand is everywhere higher in the state of the world where consumers in aggregate have higher incomes. In this case, it is plausible that at the time of contracting, the manu-

Figure 10 : Welfare is lower with all-units discounts



facturer may not know which state of the world will occur. Our findings imply that, in this instance, all-units discounts will lead to lower consumer surplus and social welfare relative to two-part tariffs.

Thus, with these three examples, we have shown that social welfare can be higher, the same, or lower with all-units discounts compared to a benchmark of two-part tariffs. If demand is linear, and the demand curves in the different states of nature have a common vertical intercept, or are vertical translations of each other (i.e. parallel shifts), then we have shown that all-units discounts lead to the same or lower consumer prices and are welfare improving. In contrast, if demand is linear, and the demand curves in the different states of nature have a common horizontal intercept, then we have shown that all-units discounts lead to higher consumer prices and decrease welfare.

V Conclusion

A discussion of the competitive effects of all-units discounts has until now been left to policymakers and legal scholars, who invariably conclude that the discounts are exclusionary. On the one hand, it is not surprising that all-units discounts are viewed with suspicion because it seems odd that a manufacturer would ever want to charge less for a larger order if its intention is benign. On the other hand, policymakers and legal scholars have overlooked plausible efficiency rationales.

In this paper, we have offered an efficiency rationale for the use of all-units discounts in retail contracts. In particular, we showed in a bilateral monopoly setting that all-units discounts can arise in the absence of any exclusionary motive. We compared and contrasted all-units discounts to other quantity-discount schemes, such as two-part tariffs and incremental-units discounts, and showed that, when demand is deterministic, all-units discounts are equally adept at solving the double marginalization problem that arises with linear pricing when retailers have market power. And, when retailers have private information about demand, we showed that all-units discounts can yield strictly higher profit for the manufacturer (lower profit for the retailer) than the profit-maximizing menu of two-part tariffs or incremental-units discount. Compared to linear pricing, all-units discounts are welfare improving. Compared to a benchmark of two-part tariffs, there exist environments in which all-units discounts can lead to higher or lower consumer prices in equilibrium.

Our results suggest that it might be possible to determine when all-units discounts are likely to be welfare improving and when not. In the context of our bilateral monopoly setting, we find that with linear demands the former case is more likely when the uncertainty about demand is over population size (tourism), or when one demand state is a vertical translation of another, while the latter case is more likely when the uncertainty applies to aggregate consumers' incomes.

These results suggest that a more cautious approach should be taken in antitrust enforcement against all-units discounts. The bias that currently exists against all-units discounts appears to be unjustified. In the absence of traditional percursors of potential market foreclosure, such as the manufacturer engaging in predatory pricing by pricing all-units below marginal cost (Marx and Shaffer, 1998), or large fixed costs that give rise to significant economies of scale (Rasmusen et. al. 1991; and Segal and Whinston, 2000), or the existence of long-term contracts that explicitly specify exclusive dealing (Aghion and Bolton, 1987) or otherwise condition quantity sold on the retailer's purchases of competitors' products, we suggest that these discounts can be welfare improving in a wide variety of circumstances. As a solution to the double marginalization problem they unambiguously lead to lower prices and higher welfare. As a means of inducing second-degree price discrimination, they may lead to higher or lower prices than other quantity-discount schemes.

Appendix

Proof of Lemma 1

To prove Lemma 1, we must show that (16) and (17) are satisfied with equality at $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$. We proceed by contradiction. Suppose first that $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ are such that (17) is not satisfied with equality. Then it must be that the profit the manufacturer earns in the high-demand state, $T_H(q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})) - cq_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$, is less than Ω_H, where $\Omega_H \equiv R_H(q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})) - cq_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) - R_H(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) + T_L(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}))$.

To show the manufacturer has a profitable deviation, consider an alternative menu of contracts $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(p_H(0), \hat{\lambda}_H, q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}))$, where

$$\hat{\lambda}_H = \frac{R_H(q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})) - R_H(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) + T_L(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}))}{p_H(0) q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})}. \tag{A.1}$$

With these contracts, there is no change in the manufacturer's profit in the low-demand state and the constraint in (16) is satisfied (the option meant for the low-demand retailer is unchanged). To see that the constraint in (17) is satisfied, note that because $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ is the same the right-hand side is unchanged, and because the retailer can always choose $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$, the left-hand side must be at least as large as $R_H(q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})) - p_H(0)\hat{\lambda}_H q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$. It follows that the constraint in (17) is satisfied and thus that the manufacturer's profit in the high-demand state is $(p_H(0)\hat{\lambda}_H - c)q = \Omega_H \frac{q}{q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})} \geq \Omega_H$, for all $q \geq q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$. Since the high-demand retailer would never purchase less than $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ under the new menu of contracts (because its per-unit price of $p_H(0)$ would be prohibitively high), it follows that the manufacturer will earn profit of at least Ω_H in the high-demand state. Thus, the deviation is profitable, contradicting the supposition that $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ are such that (17) is not satisfied with equality.

Now suppose $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ are such that (16) is not satisfied with equality. Then it must be that the profit the manufacturer earns in the low-demand state, $T_L(q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) - cq_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, is less than Ω_L, where $\Omega_L \equiv R_L(q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) - cq_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$.

To show the manufacturer has a profitable deviation, consider an alternative menu of contracts $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$, $(p_H(0), \hat{\lambda}_L, q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}))$, where

$$\hat{\lambda}_L = \frac{R_L(q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}))}{p_H(0) q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})}. \tag{A.2}$$

With these contracts, there is no change in the manufacturer's profit in the high-demand state and the constraint in (17) is satisfied (the menu meant for the high-demand retailer is unchanged, and it is now more costly to mimic the low-demand retailer). To see that the constraint in (16) is satisfied, note that because the retailer can always choose $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, the left-hand side must be at least as large as $R_L(q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) - p_H(0)\hat{\lambda}_L q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$. It follows that the constraint in (16) is satisfied and thus that the manufacturer's profit in the low-demand state is $(p_H(0)\hat{\lambda}_L - c)q = \Omega_L \frac{q}{q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})} \geq \Omega_L$, for all $q \geq q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$. Since the low-demand retailer would never purchase less than $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ under the new menu of contracts (because its per-unit price of $p_H(0)$ would be prohibitively high), it follows that the manufacturer will earn profit of at least Ω_L in the low-demand state. Thus, the deviation is profitable, contradicting the supposition that $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ are such that (16) is not satisfied with equality. Q.E.D.

Proof of Lemma 2

We know from Lemma 1 that (16) must be satisfied with equality at $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$. This implies that $w_L^{**} \geq p_L(0)$ because otherwise a retailer in the low-demand state could earn positive profit by purchasing $q < \overline{q}_L^{**}$ for some q. Given $w_L^{**} \geq p_L(0)$, Lemma 1 also implies that $\lambda_L^{**} w_L^{**} = p_L(\overline{q}_L^{**})$ because if $\lambda_L^{**} w_L^{**} < p_L(\overline{q}_L^{**})$, a retailer in the low-demand state could earn positive profit by purchasing $q = \overline{q}_L^{**}$, and if $\lambda_L^{**} w_L^{**} > p_L(\overline{q}_L^{**})$, a retailer in the low-demand state would earn negative profit for all $q > 0$. Given $w_L^{**} \geq p_L(0)$ and $\lambda_L^{**} w_L^{**} = p_L(\overline{q}_L^{**})$, it follows that the retailer chooses $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \overline{q}_L^{**}$ because for all other $q > 0$ it would earn negative profit.

To finish the proof, it remains only to establish that $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \overline{q}_H^{**}$. We know from Lemma 1 that (17) must be satisfied with equality at $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$. This implies that w_H^{**} must be such that $\max_{q < \overline{q}_H^{**}} R_H(q) - w_H^{**} q \leq K$, where K is the right-hand side of (17) evaluated at $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, because otherwise a retailer in the high-demand state would strictly prefer to purchase $q < \overline{q}_H^{**}$ for some q under contract $T_H(q)$ than purchase under contract $T_L(q)$. Given w_H^{**}, Lemma 1 also implies that $\lambda_H^{**} w_H^{**}$ must be such that $\max_{q \geq \overline{q}_H^{**}} R_H(q) - \lambda_H^{**} w_H^{**} q \leq K$ because otherwise a retailer in the high-demand state would strictly prefer to purchase $q = \max\{\overline{q}_H^{**}, q_H^*(\lambda_H^{**} w_H^{**})\}$ under contract $T_H(q)$ than purchase under contract $T_L(q)$. Finally, Lemma

1 implies that at least one of these two inequalities must be satisfied with equality because otherwise a high-demand retailer would strictly prefer to purchase under contract $T_L(q)$. It follows that $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) \in \{q_H^*(w_H^{**}), \max\{\overline{q}_H^{**}, q_H^*(\lambda_H^{**} w_H^{**})\}\}$ if both inequalities are satisfied with equality, $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = q_H^*(w_H^{**})$ if only the first inequality is satisfied with equality, and $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \max\{\overline{q}_H^{**}, q_H^*(\lambda_H^{**} w_H^{**})\}$ if only the second inequality is satisfied with equality.

To narrow the possibilities, suppose $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = q_H^*(w_H^{**})$. Then the manufacturer earns profit $w_H^{**} q_H^*(w_H^{**}) - c q_H^*(w_H^{**}) = R_H(q_H^*(w_H^{**})) - c q_H^*(w_H^{**}) - K$ in the high-demand state. Since $q_H^*(w_H^{**}) < q_H^*(c)$ because $w_H^{**} > c$, this profit is less than $R_H(q_H^*(c)) - c q_H^*(c) - K$, the profit it could earn in the high-demand state by choosing w_H such that $\max_{q<\overline{q}_H} R_H(q) - w_H q < K$, $\lambda_H w_H$ such that $R_H(q_H^*(c)) - \lambda_H w_H q_H^*(c) = K$, and $\overline{q}_H = q_H^*(c)$. Thus, there is a profitable deviation, contradicting the supposition that $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = q_H^*(w_H^{**})$. Suppose $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = q_H^*(\lambda_H^{**} w_H^{**})$. Then, since $\lambda_H^{**} w_H^{**} > c$, implying that $q_H^*(\lambda_H^{**} w_H^{**}) < q_H^*(c)$, the same argument also establishes the existence of a profitable deviation. Thus, it must be that $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \overline{q}_H^{**}$. Q.E.D.

Proof of Proposition 3

To prove Proposition 3, we must show that $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = q_H^*(c)$. The proof is by contradiction. Using Lemma 1, we can write the manufacturer's maximized profit as

$$\alpha(R_L(q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) - c q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) + (1-\alpha)\left(R_H(q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})) - c q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})\right)$$
$$- (1-\alpha)\left(R_H(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) - T_L(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}))\right). \tag{A.3}$$

Suppose $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) \neq q_H^*(c)$. Then, since $q_H^*(c) = \arg\max_q R_H(q) - cq$, it must be that the manufacturer's profit in (A.3) is less than $\tilde{\Omega}_H$, where

$$\tilde{\Omega}_H \equiv \alpha(R_L(q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) - c q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) + (1-\alpha)\left(R_H(q_H^*(c)) - c q_H^*(c)\right)$$
$$- (1-\alpha)\left(R_H(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) - T_L(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}))\right). \tag{A.4}$$

To show the manufacturer has a profitable deviation, consider an alternative menu of contracts $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(p_H(0), \tilde{\lambda}_H, q_H^*(c))$, where

$$\tilde{\lambda}_H = \frac{R_H(q_H^*(c)) - R_H(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})) + T_L(q_H^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}))}{p_H(0) q_H^*(c)}. \tag{A.5}$$

With these contracts, there is no change in the manufacturer's profit in the low-demand state and the constraint in (16) is satisfied (the option meant for the low-demand retailer is unchanged). Also, because $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ is the same, the right-hand side of (17) is unchanged, and because the retailer can always choose $q_H^*(c)$, the left-hand side must be at least as large as $R_H(q_H^*(c)) - p_H(0)\hat{\lambda}_H q_H^*(c)$. It follows that the constraint in (17) is satisfied and thus that the manufacturer's profit in the high-demand state is $(p_H(0)\tilde{\lambda}_H - c)q$, for all $q \geq q_H^*(c)$. Since the high-demand retailer will never purchase less than $q_H^*(c)$ under the new menu of contracts (because its per-unit price of $p_H(0)$ would be prohibitively high), it follows that the sum of the manufacturer's profit in the low-demand state and high demand state will be at least $\tilde{\Omega}_H$. Thus, the deviation is profitable, contradicting the supposition that $(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$, $(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**})$ are such that $q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) \neq q_H^*(c)$. Q.E.D.

Proof of Proposition 4

To prove Proposition 4, we begin by considering the high-demand retailer's quantity choice if it were to purchase under the contract meant for the low-demand retailer. Since $w_L \geq p_H(0)$ is optimal, the high-demand retailer will never purchase $q < \overline{q}_L$.[20] This means that the high-demand retailer will either choose $q = \overline{q}_L$ if the threshhold is binding, or $q > \overline{q}_L$ if the threshhold is not binding. If the threshhold is not binding, then the high-demand retailer will choose $q_H^{**}(w_L, \lambda_L, \overline{q}_L) = q_H^*(p_L(\overline{q}_L))$. Substituting the retailer's quantity choice into (18), we can write the manufacturer's problem as

$$\max_{\overline{q}_L} \ \alpha(R_L(\overline{q}_L) - c\overline{q}_L) - (1-\alpha)\left(R_H(q_H^*(p_L(\overline{q}_L))) - p_L(\overline{q}_L)q_H^*(p_L(\overline{q}_L))\right). \tag{A.6}$$

Differentiating (A.6) with respect to $\overline{q}_L$, and using the envelope theorem, we have that $\overline{q}_L^{**}$ solves the first-order condition in (20). If the threshhold binds, then the high-demand retailer will choose $q_H^{**}(w_L, \lambda_L, \overline{q}_L) = \overline{q}_L$. Substituting this into (18), we can write the manufacturer's problem as

$$\max_{\overline{q}_L} \ \alpha(R_L(\overline{q}_L) - c\overline{q}_L) - (1-\alpha)\left(R_H(\overline{q}_L) - p_L(\overline{q}_L)\overline{q}_L\right), \tag{A.7}$$

Differentiating (A.7) with respect to $\overline{q}_L$, we have that $\overline{q}_L^{**}$ solves the first-order condition in (21). We conclude the proof by noting that Lemma 2 implies that $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \overline{q}_L^{**}$. Q.E.D.

[20] Note that w_L affects the maximand in (18) only through the term $R_H(q_H^{**}(w_L, \lambda_L, \overline{q}_L)) - T_L(q_H^{**}(w_L, \lambda_L, \overline{q}_L))$. Since the maximand is decreasing in this term, and since this term is weakly decreasing in w_L, it follows that the manufacturer will choose w_L such that a high-demand retailer would never purchase $q < \overline{q}_L$, e.g., $w_L \geq p_H(0)$.

Example in which welfare is higher with all-units discounts

Suppose the inverse demands are given by $p_i(q) = a - \frac{q}{\theta_i}$, where $i \in \{H, L\}$ and $\theta_H > \theta_L$, so that, when plotted (see Figure 8), the inverse demand is everywhere flatter in the high-demand state. Solving for $q_H^*(\cdot)$, $q_L^*(\cdot)$, $\frac{\partial R_i(\cdot)}{\partial q}$, and $\frac{\partial q_i^*(\cdot)}{\partial q}$, and substituting these expressions into (10) and (11), we obtain the retailer's quantity choices under the profit-maximizing menu of two-part tariffs:

$$q_H^*(c) = \frac{\theta_H(a-c)}{2} \;\; ; \quad q_L^*(w_L^*) = \frac{\theta_L^2 \alpha(a-c)}{2[(1-\alpha)\theta_H - (1-2\alpha)\theta_L]}.$$

In the case of all-units discounts, we will have different solutions for the two subcases that may arise. If $q_H^*(\lambda_L^{**} w_L^{**}) > \overline{q}_L^{**}$, then we can use Proposition 4 and condition (20) to solve for the retailer's quantity choices under the profit-maximizing menu of all-units discounts:

$$q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \frac{\theta_H(a-c)}{2} \; ; \; q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \frac{\theta_L^2 \alpha(a-c)}{2[\alpha\theta_L + \frac{(1-\alpha)}{4}\theta_H]}.$$

In this subcase, the high-demand retailer's quantity choice is the same in both cases (as expected), while the low-demand retailer's quantity choice is higher if and only if $\theta_H > \frac{4}{3}\theta_L$. Since this subcase arises if and only if $\theta_H > 2\theta_L > \frac{4}{3}\theta_L$, it follows that welfare is higher with all-units discounts.

If $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, then we can use Proposition 4 and condition (21) to solve for the retailer's quantity choices under the profit-maximizing menu of all-units discounts:

$$q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \frac{\theta_H(a-c)}{2} \;\; ; \;\; q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \frac{\theta_H \theta_L \alpha(a-c)}{2[\theta_H - (1-\alpha)\theta_L]}.$$

Comparing $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ and $q_L^*(w_L^*)$, it can be shown that $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ is higher than $q_L^*(w_L^*)$ for all $\theta_H \leq 2\theta_L$, implying that, once again, welfare is higher with all-units discounts.

Example in which welfare is the same or higher with all-units discounts

Suppose the inverse demands are given by $p_i(q) = a\theta_i - q$, where $i \in \{H, L\}$, and $\theta_H > \theta_L$, so that, when plotted (see Figure 9), the inverse demands in the states of nature are parallel. Solving for $q_H^*(\cdot)$, $q_L^*(\cdot)$, $\frac{\partial R_i(\cdot)}{\partial q}$, and $\frac{\partial q_i^*(\cdot)}{\partial q}$, and substituting these expressions into (10) and (11), we obtain the retailer's quantity choices under the profit-maximizing menu of two-part tariffs:

$$q_H^*(c) = \frac{a\theta_H - c}{2} \;\; ; \quad q_L^*(w_L^*) = \frac{\alpha(a\theta_L - c) - (1-\alpha)a(\theta_H - \theta_L)}{2\alpha}.$$

In the case of all-units discounts, we will have different solutions for the two subcases that may arise. If $q_H^*(\lambda_L^{**}w_L^{**}) > \overline{q}_L^{**}$, then we can use Proposition 4 and condition (20) to solve for the retailer's quantity choices under the profit-maximizing menu of all-units discounts:

$$q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \frac{a\theta_H - c}{2} \; ; \; q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \frac{(1+\alpha)a\theta_L - (1-\alpha)a\theta_H - 2\alpha c}{1+3\alpha}.$$

In this subcase, the high-demand retailer's quantity choice is the same in both cases (as expected), while the low-demand retailer's quantity choice is higher if and only if $(1+\alpha)a(\theta_H - \theta_L) > \alpha(a\theta_L - c)$. Since it can be shown that this subcase arises if and only if this condition is satisfied, it follows that welfare is higher with all-units discounts when $q_H^*(\lambda_L^{**}w_L^{**}) > \overline{q}_L^{**}$.

If $q_H^*(\lambda_L^{**}w_L^{**}) \leq \overline{q}_L^{**}$, then we can use Proposition 4 and condition (21) to solve for the retailer's quantity choices under the profit-maximizing menu of all-units discounts:

$$q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \frac{a\theta_H - c}{2} \; ; \; q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \frac{\alpha(a\theta_L - c) - (1-\alpha)a(\theta_H - \theta_L)}{2\alpha}.$$

We see that $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = q_L^*(w_L^*)$ whenever this subcase applies, i.e., whenever $(1+\alpha)a(\theta_H - \theta_L) \leq \alpha(a\theta_L - c)$, implying that, in this case, welfare is unchanged with all-units discounts.

Example in which welfare is lower with all-units discounts

Suppose the inverse demands are given by $p_i(q) = \theta_i(a - q_i)$, where $i \in \{H, L\}$, and $\theta_H > \theta_L$, so that, when plotted (see Figure 10), the inverse demand is everywhere steeper in the high-demand state. Solving for $q_H^*(\cdot)$, $q_L^*(\cdot)$, $\frac{\partial R_i(\cdot)}{\partial q}$, and $\frac{\partial q_i^*(\cdot)}{\partial q}$, and substituting these expressions into (10) and (11), we obtain the retailer's quantity choices under the profit-maximizing menu of two-part tariffs:

$$q_H^*(c) = \frac{1}{2}\left(a - \frac{c}{\theta_H}\right) \quad ; \quad q_L^*(w_L^*) = \frac{1}{2}\left(a - \frac{\alpha\theta_H c}{\theta_L[(1-\alpha)\theta_L - (1-2\alpha)\theta_H]}\right).$$

In the case of all-units discounts, we will have different solutions for the two subcases that may arise. If $q_H^*(\lambda_L^{**}w_L^{**}) > \overline{q}_L^{**}$, then we can use Proposition 4 and condition (20) to solve for the retailer's quantity choices under the profit-maximizing menu of all-units discounts:

$$q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \frac{1}{2}\left(a - \frac{c}{\theta_H}\right) \quad ; \quad q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \frac{a\theta_L[(3\alpha - 1)\theta_H + (1-\alpha)\theta_L] - 2\alpha\theta_H c}{\theta_L[4\alpha\theta_H + (1-\alpha)\theta_L]}.$$

In this subcase, the high-demand retailer's quantity choice is the same in both cases (as expected), while the low-demand retailer's quantity choice is higher if and only if

$$\frac{a}{c} < \frac{\alpha(2\theta_H - \theta_L)}{\theta_L(\theta_L - (1-\alpha)\theta_H)}.$$

If this condition holds, then welfare is lower with lower with all-units discounts.

If $q_H^*(\lambda_L^{**} w_L^{**}) \leq \overline{q}_L^{**}$, then we can use Proposition 4 and condition (21) to solve for the retailer's quantity choices under the profit-maximizing menu of all-units discounts:

$$q_H^{**}(w_H^{**}, \lambda_H^{**}, \overline{q}_H^{**}) = \frac{1}{2}\left(a - \frac{c}{\theta_H}\right) \quad ; \quad q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**}) = \frac{1}{2}\left(a - \frac{\alpha c}{\theta_L - (1-\alpha)\theta_H}\right).$$

Comparing $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ and $q_L^*(w_L^*)$, it can be shown that $q_L^{**}(w_L^{**}, \lambda_L^{**}, \overline{q}_L^{**})$ is always lower than $q_L^*(w_L^*)$, implying that, once again, welfare is lower with all-units discounts.

REFERENCES

Aghion, P. and P. Bolton (1987), "Contracts as a Barrier to Entry," *American Economic Review*, 77: 388-401.

Goldman, M., H. Leland, and D. Sibley (1984), "Optimal Nonuniform Pricing," *Review of Economic Studies*, 51: 305-320.

Katz, M. (1989), "Vertical Contractual Relations," in *Handbook of Industrial Organization*, R. Schmalensee and R. Willig, eds., Amsterdam, The Netherlands: Elsevier Science Publishers.

Marx, L. and G. Shaffer (1998), "Predatory Accomodation: Below-Cost Pricing Without Exclusion in Intermediate Goods Markets," *Rand Journal of Economics*, 30: 22-43.

Maskin, E. and J. Riley (1984), "Monopoly with Incomplete Information," *Rand Journal of Economics*, 15: 171-196.

Mathewson, G.F. and R. Winter (1984), "An Economic Theory of Vertical Restraints," *Rand Journal of Economics*, 15: 27-38.

Moorthy, K.S. (1987), *Managing Channel Profits: Comment*, Marketing Science, 6: 375-379.

Nahmias, S. (2001), *Production and Operations Analysis*, 4th ed., New York, NY: McGraw Hill.

O'Brien, D.P. and G. Shaffer (1992), "Vertical Control with Bilateral Contracts," *Rand Journal of Economics*, 23: 299-308.

Ordover, J. and J. Panzar (1982), "On the Nonlinear Pricing of Inputs," *International Economic Review*, 23: 659-675.

Perry, M.K. and R. Porter (1990), "Can Resale Price Maintenance and Franchise Fees Correct Sub-Optimal Levels of Retail Service," *International Journal of Industrial Organization*, 8: 115-141.

Rasmusen, E., Ramseyer, J., and J. Wiley (1991), "Naked Exclusion," *American Economic Review*, 81: 1137-1145.

Reiffen, D. (1999), "On the Equivalence of Resale Price Maintenance and Quantity Restrictions," *International Journal of Industrial Organization*, 17: 277-288.

Roberts, K. (1979), "Welfare Considerations of Nonlinear Pricing," *Economic Journal*, 89: 66-83.

Salant, S. (1989), "When is Inducing Self Selection Suboptimal for a Monopolist," *Quarterly Journal of Economics*, 104: 391-398.

Segal, I. and M. Whinston (2000), "Naked Exclusion: Comment," *American Economic Review*, 90: 296-311.

Spence, A.M. (1977), "Nonlinear Prices and Welfare," *Journal of Public Economics*, 8: 1-18.

Spence, A.M. (1980), "Multiproduct Quantity Dependent Prices and Profitability Constraints," *Review of Economic Studies*, 47: 821-841.

Spengler, J. (1950), "Vertical Integration and Antitrust Policy," *Journal of Political Economy*, 58: 347-352.

Tirole, J. (1988), *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.

Tom, W., Balto, D., and N. Averitt (2000), "Anticompetitive Aspects of Market-Share Discounts and Other Incentives to Exclusive Dealing," *Antitrust Law Journal*, 67: 615-639.

Warren-Boulton, R. (1974), "Vertical Control with Variable Proportions," *Journal of Political Economy*, 82: 783-802.

Willig, R. (1978), "Pareto-Superior Nonlinear Outlay Schedules," *Bell Journal of Economics*, 9: 56-69.

Winter, R. (1993), "Vertical Control and Price Versus Non-Price Competition," *Quarterly Journal of Economics*, 108: 61-76.