

# Nonlinear Pricing

Mark Armstrong\*

February 2016

## Abstract

I survey the use of nonlinear pricing as a method of price discrimination, both with monopoly and oligopoly supply. Topics covered include an analysis of when it is profitable to offer quantity discounts and bundle discounts, connections between second- and third-degree price discrimination, the use of market demand functions to calculate nonlinear tariffs, the impact of consumers with bounded rationality, bundling arrangements between separate sellers, and the choice of prices for upgrades and add-on products.

**Keywords:** Price discrimination, nonlinear pricing, bundling, product-line pricing, screening, discrete choice.

**JEL:** D11, D21, D42, D86, L13, M31.

## 1 Introduction

In this article I discuss the form of price discrimination known as nonlinear pricing, where the price you pay for something depends on what else you buy. A walk down the aisle of a supermarket reveals how larger quantities of a product are (usually) proportionately cheaper than smaller quantities, as well as opportunities to buy “three for the price of two on selected items”. Most of the services connected to a home—electricity, telephone, water, television, broadband—are sold via tariffs where the average price for a unit depends on how much of that service you consume and/or which other services you take from the same supplier. Online retailers may waive shipping fees for orders above some threshold. A newspaper consists of a bundle of distinct articles, alongside advertisements, and may be

---

\*Department of Economics and All Souls College, University of Oxford. I am grateful to Michael Grubb, David Myatt, David Sappington, Rani Spiegler, John Vickers (who collaborated with me for much of the original research summarised in section 4), Dan Vincent, Bob Wilson and Jidong Zhou for helpful comments.

cheaper with a subscription than when casually purchased from a store. One may get a discount on car rental when you book a flight, and the flight may be cheaper if the traveller has flown frequently with the airline (or its partners) before. A “value meal” at a fast food outlet is cheaper than its components would be if purchased separately. Likewise, the “Sciences Online Collection” from *Annual Reviews* in 2016 is available at a 15% discount compared with separate purchase.

It is useful to distinguish between various scenarios along two dimensions: whether there is a single product or multiple products, and whether there is a single seller or multiple sellers. Table 1 summarises the main topics discussed in this paper.

	<b>single product</b>	<b>multiple products</b>
<b>single seller</b>	quantity discounts product-line pricing [section 2]	bundle discounts multiproduct nonlinear pricing [section 3]
<b>multiple sellers</b>	one-stop shopping product-line pricing [section 4.1]	mix-and-match shopping inter- and intra-firm bundling [section 4.2]

Table 1: The plan of the paper

With a single seller, the ability to engage in nonlinear pricing will boost profit relative to linear pricing, sometimes greatly so.<sup>1</sup> Linear pricing is just a special kind of nonlinear pricing, and the seller can do no worse when it is able to choose from a richer set of tariffs. With multiple sellers, though, as in game theory generally, the ability of agents to choose from a wider set of strategies can harm equilibrium payoffs. In section 4.2 we will see that sellers are sometimes strictly worse off when it is feasible for them to engage in nonlinear pricing.

In some cases—a series of sports events, say—it is not always clear-cut whether a consumer buys several units of one product or one unit of each of several products. In economic terms, though, the crucial distinction between the two scenarios is that with a

<sup>1</sup>For instance, a seller might obtain finite profit with linear pricing but unbounded profit with a two-part tariff. (If there is a single consumer who has gross utility  $u(q) = \log(1 + q)$  from  $q$  units, so her inverse demand function with linear pricing is  $p(q) = 1/(1 + q)$ , and the seller has no costs, then the seller cannot obtain revenue greater than 1 with linear pricing, but the consumer will agree to an “all you can eat” contract in return for *any* fixed charge.)

single product a consumer’s utility depends only on the *number* of units she buys, while with multiple products she cares about *which* units she buys. With a single product there is a natural order to the decisions the consumer makes. To buy two units of a product, a consumer can be considered to buy the “first” unit (for price  $p_1$ , say) followed by the “second” unit (for incremental price  $p_2$ ). She cannot buy the second unit on its own. This is similar to the situation where a seller offers a menu of qualities, rather than quantities, and a consumer who buys a high-quality product can be considered to purchase a “basic” product together with an “upgrade”, and there is no value to the upgrade in isolation. Another situation with this choice structure is when a consumer buys a core product, and an optional “add-on” product is available only to consumers with the core product. Although these latter situations, which concern product-line pricing, do not strictly fall under the definition of nonlinear pricing, the issues raised are in many respects similar and it makes sense to include them in this article.

This natural ordering implies that the single-product pricing problem can often be simplified by treating each incremental unit as a “separate market”. That is to say, given the distribution of incremental valuations for the  $n^{th}$  unit in the consumer population, one can calculate the most profitable price for this unit in isolation assuming that consumers buy this  $n^{th}$  unit if and only if their incremental valuation is above the incremental price. This approach does not always generate the optimal tariff—sometimes consumers might purchase initial units solely to obtain a good deal overall—but it often does, and in these cases nonlinear pricing is just an instance of third-degree price discrimination where price-cost markups reflect the elasticity of the relevant demand function. This method is known as the “demand profile” approach to nonlinear pricing.<sup>2</sup> Regardless of whether the demand profile approach is valid, we will see that a seller offers a quantity discount (say,  $p_2 < p_1$ ) when demand for the second unit is more elastic than demand for the first unit. In the related situations with product-line pricing, it does not make sense to speak of a quantity discount *per se*, but instead one can investigate how the markup for the upgrade/add-on compares with that for the basic item. Because of this natural choice order, in competitive settings it usually makes sense in these cases to suppose that consumers are “one-stop shoppers”, and do not buy the basic product from one seller and the upgrade (or similar) from another. (In the multiproduct context, substantial shopping costs will also induce

---

<sup>2</sup>Wilson (1993) demonstrates the widespread usefulness of the demand profile approach.

consumers to be one-stop shoppers.)

The multiproduct pricing problem has a different structure, without a natural order for consumer decisions. With unit demand for two underlying products, say, a consumer has three options: product 1 on its own, product 2 on its own, and the bundle of both products. The demand profile approach seems to be less useful with multiple products. Here, consumers have a choice for their “first unit” (it could be product 1 or product 2), and there is no coherent notion of the incremental price or incremental valuation for the “second unit”, since that depends on which product was purchased initially. In competitive settings, it makes sense to suppose that some consumers purchase distinct products from different sellers—that is, they “mix and match”—at least when linear pricing is used and shopping costs are not large. When competing sellers offer bundle discounts, however, artificial shopping costs are introduced which encourages more consumers to become loyal one-stop shoppers.

In the multiproduct context, we will see that offering a bundle discount is profitable when demand for the bundle is more elastic than demand for each product separately. A corollary of this is that even when products are completely unrelated it is profitable to introduce a bundle discount. A related insight is that a seller can often estimate a consumer’s value for a bundle of products more accurately than it can for an individual product, just as a broad asset portfolio allows for more predictable returns to an investor. (This effect is most marked when *many* items are in the bundle, when the law of large numbers can apply.) With a single seller, this implies that bundling often leads to higher profits and reduced consumer surplus. With multiple sellers, though, competing to supply bundles of products can sharpen competition, so sellers are harmed and consumers benefit from this form of price discrimination.

## **2 Single-Seller, Single-Product Analysis**

### **2.1 A framework with discrete choices**

It is useful initially to discuss nonlinear (and related) pricing in a simplified setting in which consumers have a binary choice from a monopoly seller, and buy either a “basic” or a “premium” product. (Analysis with continuous choices is presented in the next section.) Here, the premium product might consist of a greater quantity of the underlying product relative to the basic product, as with a café which sells coffee in two sizes. Alternatively,

the premium product might be of higher quality than the basic product (such as a car with a better engine), include an “add-on” product alongside the basic product (such as an enhanced warranty to go with a new television), or be available at an earlier date than the basic product.

Consider the following discrete choice framework. A consumer is willing to pay  $v_1$  for the basic product and  $v_1 + v_2$  for the premium product, so that  $v_2$  is her value for the “upgrade” to the premium product. The seller charges price  $p_1$  for the basic product and price  $p_1 + p_2$  for the premium product, so that  $p_2$  is the price for the upgrade.<sup>3</sup> The seller incurs constant cost  $c_1$  to supply one unit of the basic product and cost  $c_2$  to supply an upgrade, so its unit cost for the premium product is  $c_1 + c_2$ . This perspective—where a premium product is viewed as a basic product bundled with an upgrade, and where the price of this product is the sum of the basic price and the price for the upgrade—is known as the “upgrades” approach to product line pricing.<sup>4</sup>

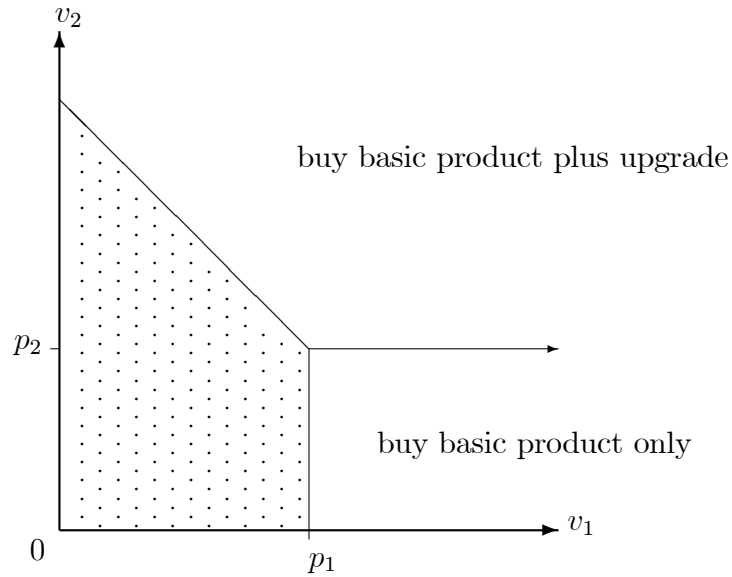


Figure 1: Pattern of demand with tariff  $(p_1, p_2)$

The type- $(v_1, v_2)$  consumer chooses the option from the three discrete choices which leaves her with the highest net surplus, so she will buy the premium product whenever

<sup>3</sup>Thus, here I assume the seller uses a deterministic strategy. See section 3.1 for a discussion of the profitability of stochastic schemes.

<sup>4</sup>The approach is presented in Johnson and Myatt (2003, 2006b). An early treatment of the issue, which implicitly uses the upgrades approach, is Itoh (1983). The approach can also be made to fit the intertemporal pricing context: if a consumer values an item now at  $V$  and the same item later at  $dV$ , where  $d \leq 1$  is her discount factor, then the described framework applies if we write  $v_1 = dV$  and  $v_2 = (1 - d)V$ .

$v_1 + v_2 - (p_1 + p_2) \geq \max\{v_1 - p_1, 0\}$ , she will buy only the basic product if  $v_1 - p_1 \geq \max\{v_1 + v_2 - (p_1 + p_2), 0\}$ , and otherwise she buys nothing. Thus, the pattern of demand is as shown in Figure 1. Note that a consumer with  $v_1 < p_1$  who buys the premium product gains negative surplus from the basic product, and only buys that product as a means with which to obtain the valued premium product. Write  $Q_1(p_1, p_2)$  for the fraction of consumers who buy something—that is, the basic product with or without the upgrade—with tariff  $(p_1, p_2)$ , and let  $Q_2(p_1, p_2)$  be the fraction who buy the upgrade. In smooth cases we have Slutsky symmetry of demand functions, so that  $\partial Q_1/\partial p_2 \equiv \partial Q_2/\partial p_1 \leq 0$ . (Loosely speaking, this negative cross-price effect is equal to the number of consumers on the diagonal line on Figure 1.) When the cross-price effect is strictly negative the basic product and the upgrade are complementary products. Profit with the tariff  $(p_1, p_2)$  is

$$\pi = (p_1 - c_1)Q_1(p_1, p_2) + (p_2 - c_2)Q_2(p_1, p_2) , \quad (1)$$

and the seller chooses its tariff to maximize this expression. To illustrate, consider the example where  $(v_1, v_2)$  is uniformly distributed on the unit square  $[0, 1]^2$  and  $c_1 = c_2 = 0$ . Then by calculating the areas of the regions on Figure 1 one finds that the most profitable tariff is  $(p_1, p_2) = (\frac{2}{3}, \frac{1}{6})$ .

In an important subset of cases the basic product and the upgrade are independent rather than complementary products. To understand this, for  $i = 1, 2$  write  $N_i(p_i) \equiv \Pr\{v_i \geq p_i\}$ , so that  $N_1(p_1)$  is the fraction of consumers willing to pay  $p_1$  for the basic product in isolation, while  $N_2(p_2)$  is the fraction who would pay  $p_2$  for the upgrade if they already possess the basic product. In the context of nonlinear pricing, Wilson (1993) refers to the demand functions  $\{N_1(\cdot), N_2(\cdot)\}$  as the “demand profile”, and I will also use that term in the related context of upgrades and add-ons. From Figure 1 we see in general that  $N_1(p_1) \leq Q_1(p_1, p_2)$  and  $N_2(p_2) \geq Q_2(p_1, p_2)$ . However, if the distribution of tastes is such that no consumers have taste vectors  $(v_1, v_2)$  to the “north-west” of the price vector  $(p_1, p_2)$ , then  $N_i(p_i) = Q_i(p_1, p_2)$ . Here, consumers buy as if they were myopic: they buy the basic product when  $v_1 \geq p_1$  and go on to buy the upgrade when  $v_2 \geq p_2$ . In this situation the two products can (locally) be considered as separate markets. If the tariff which maximizes the notional “demand profile” profit

$$\pi = (p_1 - c_1)N_1(p_1) + (p_2 - c_2)N_2(p_2) \quad (2)$$

is such that all consumers purchase strictly myopically—i.e.,  $v_2 \geq p_2$  implies  $v_1 > p_1$ —this

tariff satisfies the first- and second-order conditions for maximizing true profit (1).

Whenever the demand profile method can be used, second-degree price discrimination coincides with the more straightforward theory of single-product monopoly pricing, and the optimal tariff reflects standard demand elasticities. The Lerner index for each component,  $(p_i - c_i)/p_i$ , is equal to the reciprocal of the elasticity of demand for that component,  $N_i(\cdot)$ , so that each price is above its associated cost  $c_i$ .<sup>5</sup> This implies that the seller obtains more profit when it supplies a premium product than a basic product. However, the proportional markup is lower for the premium product when  $(p_1 + p_2)/(c_1 + c_2) < p_1/c_1$ , i.e., when  $p_2/c_2 < p_1/c_1$ . This is the case, broadly speaking, when demand for the upgrade,  $N_2$ , is more elastic than demand for the basic product,  $N_1$ , which is a common situation in applications.

The optimal tariff in the previous “uniform square” example cannot be derived using this demand profile method. (In that example the prices which maximize (2) are  $p_1 = p_2 = \frac{1}{2}$ , and faced with this tariff some consumers would choose to buy the premium product even though  $v_1 < p_1$ .) Other distributions for consumer tastes work well with the approach, however. For example, many models of second-degree price discrimination assume there is scalar consumer heterogeneity, in the sense that  $v_2 \equiv kv_1$  and taste vectors on Figure 1 lie on a ray from the origin of slope  $k$ .<sup>6</sup> In this case, suppose that the CDF for  $v_1$  is  $F(v_1)$ . The demand profile method chooses  $p_1$  to maximize  $(p_1 - c_1)(1 - F(p_1))$  and  $p_2$  to maximize  $(p_2 - c_2)(1 - F(p_2/k))$ . If  $c_2 > kc_1$ , so that the cost of supply rises proportionately more steeply than consumer valuations (i.e., there are “decreasing returns” to quality in the terminology of Johnson and Myatt, 2003), this procedure implies that  $p_2 > kp_1$ . This in turn implies that all consumers purchase myopically and the method yields the optimal tariff, and at this tariff some consumers buy only the basic product and some buy the upgrade.

The outcome is different when there are increasing returns to quality, so that  $c_2 \leq kc_1$ . Here, one can show it is optimal for the seller to supply *only* the premium product. For

---

<sup>5</sup>This is not necessarily the case more generally when  $Q_1$  and  $Q_2$  exhibit negative cross-price effects. As is well-understood, it is sometimes optimal for a seller which supplies complementary products to set the price for one product below cost to stimulate demand for the other.

<sup>6</sup>These preferences are sometimes referred to as “Mussa-Rosen” preferences, after Mussa and Rosen (1978) who studied a model where a consumer’s valuation for incremental units was proportional to her value for the first unit. (Such preferences apply if consumers differ not in their intrinsic preferences but only in their income, and being richer shifts a consumer’s valuation vector equiproportionately.) Mussa-Rosen preferences are used in the product-line papers of Itoh (1983), Salant (1989) and Johnson and Myatt (2003), among many others.

instance, consider the model of inter-temporal price discrimination in Stokey (1979), where all consumers have the same discount factor, applied to a two-period setting. As mentioned earlier, here  $v_1 = dV$  and  $v_2 = (1 - d)V$ , where  $V$  is the consumer’s value for the item when delivered and  $d < 1$  is the discount factor, so that  $v_2 \equiv \frac{1-d}{d}v_1$ . If the seller also has the same discount factor and incurs cost  $C$  to supply a unit when delivered, then  $c_1 = dC$ ,  $c_2 = (1 - d)C$ , and so  $c_2 = \frac{1-d}{d}c_1$ . We deduce that it is optimal to supply only the premium product, i.e., to sell the product only in the first period.

The previous paragraph showed, assuming Mussa-Rosen preferences, that if the premium product has the same production cost as the basic product (i.e., if  $c_2 = 0$ ), it is optimal to sell only the premium product. Outside this class of preferences, however, it is quite possible that the seller maximizes its profit by offering both variants, even if the better one costs no more to produce. Deneckere and McAfee (1996) provide analysis of several instances of “damaged goods”, where the firm deliberately introduces an inferior variant of its product (such as slower computer chip or slower printer), which might even cost more than its standard product (so  $c_2 < 0$ ). For instance, the “uniform square” example above had  $c_2 = 0$  and at the optimum some consumers were induced to purchase the inferior option; by continuity it would still be profitable to do so even if  $c_2$  was slightly negative. Here, it is more profitable to offer a basic product alongside a less costly premium product in order to price discriminate more finely between consumers.<sup>7</sup>

*Nonlinear pricing:* Sharper results are available in the more specific context of nonlinear pricing. Here, it is natural to suppose that the seller incurs the constant unit cost  $c$  to supply each unit of its product to a consumer, where a consumer is willing to pay  $v_1$  for a single unit the product and  $v_2$  for a second unit where there are diminishing returns to consumption so that  $v_2 \leq v_1$  for all consumers. The seller charges price  $p_1$  for the first unit and incremental price  $p_2$  for the second unit, and offers a quantity discount when  $p_2 < p_1$ .<sup>8</sup>

Given diminishing returns, Figure 1 shows that the demand profile,  $N_i(p) \equiv \Pr\{v_i \geq p\}$ , is equal to  $Q_i(p, p)$ , so that the demand profile coincides with true demand when the tariff is linear. Even when the demand profile method cannot be used to generate the optimal nonlinear tariff itself, it is profitable to offer a quantity discount whenever the demand

---

<sup>7</sup>For further discussion of this point, see Myatt and Johnson (2003, section III.B) and Anderson and Dana (2009).

<sup>8</sup>This discrete choice formulation of the nonlinear pricing problem is mentioned in Adams and Yellen (1976, pp. 488-489) and Wilson (1993, section 4.3).

profile for the second unit is more elastic than for the first unit. To see this, let  $\hat{p}$  be the most profitable linear price, i.e., which maximizes  $(p - c)(N_1(p) + N_2(p))$ . From (1), the impact on the seller's profit when it slightly reduces its price for the second unit is

$$\begin{aligned}
-\left((\hat{p} - c)\left(\frac{\partial Q_1}{\partial p_2} + \frac{\partial Q_2}{\partial p_2}\right) + Q_2\right)\Bigg|_{p_1=p_2=\hat{p}} &= -\left((\hat{p} - c)\left(\frac{\partial Q_2}{\partial p_1} + \frac{\partial Q_2}{\partial p_2}\right) + Q_2\right)\Bigg|_{p_1=p_2=\hat{p}} \\
&= -(\hat{p} - c)N_2'(\hat{p}) - N_2(\hat{p}) \\
&= \frac{N_1(\hat{p}) + N_2(\hat{p})}{N_1'(\hat{p}) + N_2'(\hat{p})}N_2'(\hat{p}) - N_2(\hat{p}) . \tag{3}
\end{aligned}$$

Here, the first equality follows from Slutsky symmetry and the final equality follows from the first-order condition for  $\hat{p}$  to be the optimal linear price. Expression (3) is positive when  $N_2(p)/N_1(p)$  strictly decreases with  $p$  at  $p = \hat{p}$ ; that is, when  $N_2$  is more elastic than  $N_1$ . (It is somewhat plausible that demand for the second unit is more elastic since diminishing returns implies that  $N_2(p) < N_1(p)$ .) Therefore, the seller obtains greater profit by offering a quantity discount than with linear pricing whenever its average demand per consumer decreases with its linear price, i.e., if participation is less elastic than usage with respect to changes in linear price.<sup>9</sup> The seller can therefore determine whether a quantity discount will boost its profit with knowledge only of its demand at linear prices. (However, it does need to know average demand per consumer, not merely aggregate demand  $N_1 + N_2$ .)

To illustrate, suppose that  $(v_1, v_2)$  is uniformly distributed with density 2 on the unit triangle  $\{(v_1, v_2) \text{ such that } 0 \leq v_2 \leq v_1 \leq 1\}$ . Then  $N_1(p) = 1 - p^2$  and  $N_2(p) = (1 - p)^2$ , so that  $N_2/N_1 = (1 - p)/(1 + p)$  decreases with  $p$  and a quantity discount is profitable. With costless production, the most profitable linear price maximizes  $p(N_1(p) + N_2(p))$ , so that  $\hat{p} = \frac{1}{2}$ . With this linear tariff, one-third of the consumers who buy something buy both units. By calculating the areas of the regions on Figure 1, one can check that the nonlinear tariff which maximizes (1) is

$$p_1 = \frac{2}{3} \text{ and } p_2 = \frac{2}{3} - \frac{1}{3}\sqrt{2} \approx 0.2 . \tag{4}$$

Now about 80% of consumers who buy something buy both units, while total output is about 20% greater than with the linear tariff. This boost in output is enough to make total welfare higher with nonlinear than linear pricing, although those consumers who place little value on a second unit are worse off since the price they pay is higher.

---

<sup>9</sup>This is an instance of the more general insight that a two-part tariff with a positive fixed charge is more profitable than linear pricing when usage is more elastic than participation. For instance, see Long (1984).

*Behavioural aspects of nonlinear pricing:* One of the most active areas of current research in nonlinear pricing concerns the exploitation of consumer biases. For instance, consider a setting in which consumers are initially over-optimistic about their eventual demand for the seller’s product. Specifically, consumers are identical *ex ante*, and a consumer will want either one or two units of the product where the value of each unit (if a consumer wants that unit) is common knowledge and denoted  $v$ .<sup>10</sup> The cost of supplying a unit is  $c < v$ . The consumer’s prior probability that she will want two units is denoted  $\alpha$ , while the seller’s prior probability that the consumer will want two units is  $\beta < \alpha$ . The seller offers a contract whereby  $p_1$  is the price for the first unit and  $p_2$  is the price for the second (if desired) given the consumer has purchased the first, and requires the consumer to accept the contract before the latter knows whether she wants one or two units.

Since the seller will choose  $p_2 \leq v$  to ensure the consumer buys the second unit if she wants it, the consumer’s anticipated surplus with contract  $(p_1, p_2)$  is

$$v - p_1 + \alpha(v - p_2) , \tag{5}$$

while the seller’s anticipated profit from this contract is

$$p_1 - c + \beta(p_2 - c) . \tag{6}$$

The seller will maximize (6) subject to (5) being non-negative, which entails

$$p_1 = (1 + \alpha)v ; p_2 = 0 . \tag{7}$$

(Here, we assumed  $p_2 < 0$  was not feasible, since otherwise a consumer would take the second unit even if she did not want it.) Thus, an optimistic consumer is offered an “all you can eat” contract where she pays a fixed fee for the right to any quantity, which she (mistakenly, from the seller’s perspective) believes to be good value. If the seller’s prior is correct, it obtains more profit with nonlinear pricing than with linear pricing, while the consumer suffers negative surplus on average. This kind of model is consistent with empirical data in DellaVigna and Malmendier (2006) showing that consumers often sign up for monthly fee contracts with exercise gyms, even though *ex post* they would be better off paying on a per-visit basis. We return to this topic in the context of oligopoly in section

---

<sup>10</sup>The following discussion is essentially taken from the illustrative example in Eliaz and Spiegler (2008, section 1). They go on to analyze a more general model in which consumers differ in their prior of whether they will have high or low demand (while the seller believes that all consumers have the same probability of having high demand).

4.1, but where consumers are over-*pessimistic* about their demand for an add-on product, in which case the add-on has a high price.

Another behavioural bias is over-confidence about future demand, in the sense that a consumer’s prediction of her eventual demand is too precise and she does not adequately take into account the possibility she might want fewer or more units. The seller can exploit this bias by charging high prices when the consumer’s demand differs from her prediction. To illustrate, suppose the consumer might want 0, 1 or 2 units, and again her value for a unit, if she wants it, is  $v$  and the seller’s cost is  $c < v$ .<sup>11</sup> *Ex ante*, the consumer believes she will want one unit for sure, while the seller has a more diffuse prior and believes the consumer will want 0, 1 or 2 units with equal probability. The seller offers a contract  $(p_0, p_1, p_2)$ , where  $p_0$  is the up-front fee the consumer must pay for the right to purchase one unit for price  $p_1$  and a second unit for additional price  $p_2$ .

The consumer’s anticipated surplus if she accepts the contract  $(p_0, p_1, p_2)$  is

$$v - (p_0 + p_1) , \tag{8}$$

while the seller’s anticipated profit is

$$p_0 + \frac{2}{3}(p_1 - c) + \frac{1}{3}(p_2 - c) . \tag{9}$$

Maximizing (9) subject to (8) being non-negative, as well as the constraint  $p_2 \leq v$  which ensures the consumer buys the second unit if she wants it, implies that the most profitable contract is

$$p_0 = v ; p_1 = 0 ; p_2 = v . \tag{10}$$

Thus, the consumer is asked to pay for one unit in advance, is given no rebate if she ends up not needing this unit, and pays her reservation value if she ends up wanting a second unit.<sup>12</sup> Again, if its prior is accurate, the seller obtains greater profit with this nonlinear tariff than with the best linear contract, while the consumer suffers negative surplus on average. Tariff (10) is an example of a so-called “three-part” tariff, as often used in telephony or car rental, where the consumer pays a fixed charge for a specified quantity (or less) and an “overage”

---

<sup>11</sup>This example is taken from Grubb (2009, section II). Grubb goes on to solve a more general formulation of this problem and also studies a model with competition. See Grubb (2015) for a relatively non-technical account of how sellers choose nonlinear tariffs when consumers are either over-optimistic or over-confident.

<sup>12</sup>The extreme assumption that the consumer foresees no uncertainty in her demand is not important here, and the same outcome is optimal whenever the consumer’s prior is such that the probabilities she needs no unit and needs two units are both below one-third.

charge when she consumes beyond the threshold. With a three-part tariff, the highest-demand consumer pays the highest marginal price, which is the opposite pattern to the standard nonlinear pricing model where quantity discounts are the norm.

Additional issues arise when consumers suffer from weak will. For instance, a (sophisticated) person keen to become physically fit may foresee that when the time comes she will be unwilling to go to the gym, and so choose a fixed-fee gym contract with zero marginal price so that financial disincentives do not add to her self-control problem. Opposite effects apply when a consumer is tempted to purchase too much of a bad product: a sophisticated shopper who is aware of his temptation to over-spend, say, may opt for a bank account which levies steep overdraft charges.<sup>13</sup>

Finally, it may be that offering a second unit at discount makes the deal look like a “bargain”, which stimulates demand by more than the model with rational buyers suggests. Jahedi (2013) conducts lab experiments in which a seller offers two units for little more than the price for a single unit, and shows how some consumers are less likely to buy two units when faced with the choice set {buy nothing, buy two units for \$1} than they are with the expanded choice set {buy nothing, buy one unit for \$0.96, buy two units for \$1}. Likewise, inattentive consumers might follow the heuristic that “large size = good value”, and buy a large box of detergent, say, even if the smaller size has a lower price per-unit.<sup>14</sup>

## 2.2 A framework with continuous choices

In contrast to section 2.1, most textbook treatments of nonlinear pricing treat quantity as a continuous variable and suppose that consumer tastes vary just by a scalar parameter,  $\theta$ . To recapitulate this classical analysis, which will also be useful later, suppose that the type- $\theta$  consumer gains gross utility  $u(q, \theta)$  from  $q$  units of the product. (This discussion can be adapted so that  $q$  represents product quality instead of quantity.) There is diminishing marginal utility, so  $u$  is concave in  $q$ , and  $u(0, \theta) \equiv 0$ . We suppose that  $u$  increases with  $\theta$ , and also that there are increasing differences in  $u(q, \theta)$ , so that  $u_q(q, \theta)$  increases with  $\theta$ , where  $u_q$  denotes the partial derivative of  $u$ . Since  $u_q(q, \theta)$  is the linear price which induces

---

<sup>13</sup>A key early paper in this literature is DellaVigna and Malmendier (2004). See Spiegler (2011, chapters 2–4) for an overview of contract design with dynamically inconsistent consumers, and Koszegi (2014) for a survey of contract design in the presence of biased agents.

<sup>14</sup>Clerides and Courty (2015) observe empirically that the same brand of detergent is simultaneously sold in two sizes, where the large size contains twice as much as the smaller. Sometimes the large size is more than twice as expensive as the smaller, and yet significant numbers of consumers buy it.

the type- $\theta$  consumer to purchase  $q$  units of the product—i.e., it is this consumer’s inverse demand curve—this assumption implies that demand functions from different consumers do not cross.<sup>15</sup>

If the seller offers the nonlinear tariff  $T(q)$ , the type- $\theta$  consumer will choose quantity  $q(\theta)$  which maximizes her surplus  $u(q, \theta) - T(q)$  and so obtain net surplus

$$s(\theta) \equiv \max_{q \geq 0} : u(q, \theta) - T(q) . \quad (11)$$

Regardless of the shape of  $T(\cdot)$ , increasing differences in  $u(q, \theta)$  implies that consumers with higher  $\theta$  will choose weakly greater quantity, and conversely, any weakly increasing function  $q(\theta)$  can be implemented with a suitable nonlinear tariff.<sup>16</sup> The taste parameter  $\theta$  is distributed in the consumer population according to the smooth CDF  $F(\theta)$ , which has support  $[\theta_1, \theta_2]$  and associated density  $f(\theta) = F'(\theta)$ . To maximize its profit, the seller will leave the lowest type consumer with zero surplus, i.e.,  $s(\theta_1) = 0$ , since otherwise it could increase its tariff  $T$  by an additive constant without driving any consumers away. Since  $s(\theta)$  weakly increases with  $\theta$  in (11), if the lowest- $\theta$  consumer is willing to participate so are all consumers. If the seller’s unit cost is  $c$ , its profit with tariff  $T$  is

$$\int_{\theta_1}^{\theta_2} [T(q(\theta)) - cq(\theta)] f(\theta) d\theta = \int_{\theta_1}^{\theta_2} [u(q(\theta), \theta) - s(\theta) - cq(\theta)] f(\theta) d\theta \quad (12)$$

$$= \int_{\theta_1}^{\theta_2} \left[ u(q(\theta), \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(q(\theta), \theta) - cq(\theta) \right] f(\theta) d\theta , \quad (13)$$

where the first equality follows from (11) and the second follows by integrating  $\int s f d\theta$  by parts and using the envelope condition  $s'(\theta) \equiv u_\theta(q(\theta), \theta)$  together with the participation constraint  $s(\theta_1) = 0$ .

A candidate solution to the seller’s problem is to maximize the integral (13) pointwise with respect to  $q(\theta)$ , so that

$$q(\theta) \text{ maximizes}_{q \geq 0} : u(q, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(q, \theta) - cq \quad (14)$$

for each  $\theta$ . Provided the procedure (14) results in a weakly increasing function  $q(\theta)$ , which is so if the right-hand side of (14) has increasing differences in  $(q, \theta)$ , it yields the seller’s

---

<sup>15</sup>Prominent early contributions to this classical analysis include Spence (1977), Mussa and Rosen (1978), Goldman, Leland, and Sibley (1984) and Maskin and Riley (1984). The important trick of using the envelope condition and integration by parts—see expression (13)—is due to Mirrlees (1971). A useful overview of this analysis is presented in Rochet and Stole (2003, section 2).

<sup>16</sup>For instance, see Maskin and Riley (1984).

most profitable strategy. If the seller could observe a consumer's type directly, it would offer the type- $\theta$  consumer the quantity which maximizes total surplus  $u(q, \theta) - cq$ . When  $\theta$  is private information, though, (14) implies that the type- $\theta$  consumer is supplied with a lower quantity than the first-best quantity. The exception is for the consumer with the strongest tastes (if such a consumer exists), and if  $\theta_2$  is finite the type- $\theta_2$  consumer is served with the efficient quantity and there is "no distortion at the top".

To illustrate, consider utility of the form  $u(q, \theta) = \theta q + U(q)$  for concave  $U(\cdot)$ , so that the parameter  $\theta$  represents an additive shift to inverse demand. In this case,  $q(\theta)$  in (14) increases with  $\theta$  if

$$\theta - \frac{1 - F(\theta)}{f(\theta)} \text{ increases with } \theta . \quad (15)$$

Since the type- $\theta$  consumer chooses quantity to maximize  $u(q, \theta) - T(q)$ , the incremental price of the  $q$ th unit, denoted  $p(q) \equiv T'(q)$ , satisfies  $p(q(\theta)) = u_q(q(\theta), \theta)$ . In this specification, (14) implies that

$$p(q(\theta)) = c + \frac{1 - F(\theta)}{f(\theta)} .$$

In particular, if the hazard rate  $f(\theta)/(1 - F(\theta))$  increases with  $\theta$ , which implies (15), then the marginal price decreases with quantity.<sup>17</sup> Alternatively, with the "Mussa-Rosen" specification,  $u(q, \theta) = \theta U(q)$ , the parameter  $\theta$  represents a multiplicative shift to inverse demand. Again,  $q(\theta)$  in (14) increases with  $\theta$  if (15) holds.

An alternative and very elegant approach, which can apply to a broader class of situations than the scalar parametric framework just outlined, uses the demand profile method discussed in section 2.1.<sup>18</sup> Suppose the seller faces a heterogeneous population of consumers, some of whom are excluded at the optimum, and offers a smooth tariff  $T(q)$  such that  $T(0) = 0$ , where the marginal price schedule is  $p(q) \equiv T'(q)$ . When faced with this tariff, suppose the proportion of consumers who buy no more than  $q$  units is  $H(q)$ , assumed to be a smooth function with density  $h(q) \equiv H'(q)$ . Then the seller's profit is

$$\int_0^\infty (T(q) - cq)h(q)dq = \int_0^\infty (p(q) - c)(1 - H(q))dq ,$$

where the second expression follows after integrating by parts. (Since the seller obtains no profit when consumers choose  $q = 0$ , the consumers who buy nothing do not play a role in these expressions.) Suppose that the tariff  $T$  is such that every consumer's problem

<sup>17</sup>See Maskin and Riley (1984, Proposition 6) for more general analysis along these lines.

<sup>18</sup>For further discussion see Goldman, Leland, and Sibley (1984), Wilson (1993, section 4.1 and elsewhere) and Rochet and Choné (1998, section 2.3.1).

of maximizing her surplus  $u(q) - T(q)$  is single-peaked in  $q$ , so that the marginal price schedule  $p(\cdot)$  crosses each consumer's inverse demand function  $u'(\cdot)$  at most once and from below. Figure 2 below depicts this situation; the bold curve is the marginal price schedule  $p$ , and the inverse demand curves for a variety of consumers are shown as faint curves. Then the fraction of consumers who buy at least  $q$  units with this tariff,  $1 - H(q)$ , is equal to the fraction of consumers whose inverse demand curve  $u'(\cdot)$  is above  $p(\cdot)$  at the point  $q$  (see Figure 2), which in turn is the fraction who purchase at least  $q$  units when faced with the *linear* tariff with constant price  $p = p(q)$ . If  $N_q(p)$  is the demand profile, i.e., the fraction of consumers who purchase at least  $q$  units with a linear tariff with price  $p$ , the seller's profit is therefore

$$\int_0^\infty (p(q) - c)N_q(p(q))dq . \quad (16)$$

(This is the continuous version of the previous expression (2).) A candidate solution therefore has marginal price schedule  $p(\cdot)$  which maximizes this notional “demand profile” profit pointwise, so that

$$p(q) \text{ maximizes } (p - c)N_q(p) . \quad (17)$$

Provided that the resulting price schedule indeed cuts each consumer's inverse demand schedule from below, something which needs to be checked *ex post*, this method generated the optimal nonlinear tariff.

When the demand profile method applies, the shape of the nonlinear tariff is therefore governed by how the elasticity of the demand profile varies with  $q$ . Since  $N_q(p)$  necessarily decreases with  $q$ , this elasticity will “often” increase with  $q$ , in which case a concave tariff with decreasing marginal price is optimal.<sup>19</sup> If there is an upper bound on the quantity supplied to any consumer at the optimum, say  $\bar{q}$ , then this unit is supplied efficiently so that  $p(\bar{q}) = c$ . (If it exists, this maximum quantity  $\bar{q}$  is the smallest  $q$  which just allows the seller to break even in (17), so that  $N_{\bar{q}}(c) = 0$ , and for that quantity the price which just breaks even is  $p(\bar{q}) = c$ .)

With the Mussa-Rosen specification  $u(q, \theta) = \theta U(q)$ , for instance, a consumer buys at least  $q$  units with linear price  $p$  if  $\theta U'(q) \geq p$ , and so  $N_q(p) = 1 - F(p/U'(q))$  where  $F$  is the CDF for  $\theta$ . By examining the first-order conditions, one can check that the marginal

---

<sup>19</sup>If, as with the Mussa-Rosen specification, demand for the product is determined by the consumer's income, a concave tariff implies that on a per-unit base “the poor pay more”. (This will often be true not just relative to richer consumers, but relative to the price the poor would pay in a linear pricing regime.) See Attanasio and Pastorino (2015), and the references there, for further discussion.

price schedule defined in (17) does cut each consumer's inverse demand function  $\theta U'(\cdot)$  from below, provided that the monotonicity condition (15) is satisfied, and so the demand profile approach is valid in this case.<sup>20</sup> The Mussa-Rosen class has an especially neat solution when (15) is linear in  $\theta$ , so that  $\theta - (1 - F(\theta))/f(\theta) = b\theta - a$  for constants  $a$  and  $b > 0$ .<sup>21</sup> (A variety of familiar distributions fall into this category, including the uniform and exponential.) In this case, (17) implies that  $bp(q) = c + aU'(q)$ . Provided that it is optimal for some consumers to be excluded, this can be integrated to obtain the closed-form expression for the optimal nonlinear tariff  $T(q) = \frac{1}{b}(cq + aU(q))$ . A curiosity here is that this payment  $T(q)$  is the charge the seller would levy if it *only* sold its product as a bundle of  $q$  units. If her sole option was to buy  $q$  units at price  $T$ , the type- $\theta$  consumer would accept if  $\theta U(q) \geq T$ , and so  $T$  is chosen to maximize  $(1 - F(T/U(q)))(T - cq)$ , which is solved by choosing the above  $T(q)$ . Thus, the seller chooses its tariff as if each bundle was a separate market, even though there are actually cross-price effects in consumer demand for the various bundles. In the context of product line pricing, this implies that a seller can calculate its price for one product variant (business class air travel, say) by supposing that that was the only product it supplied.

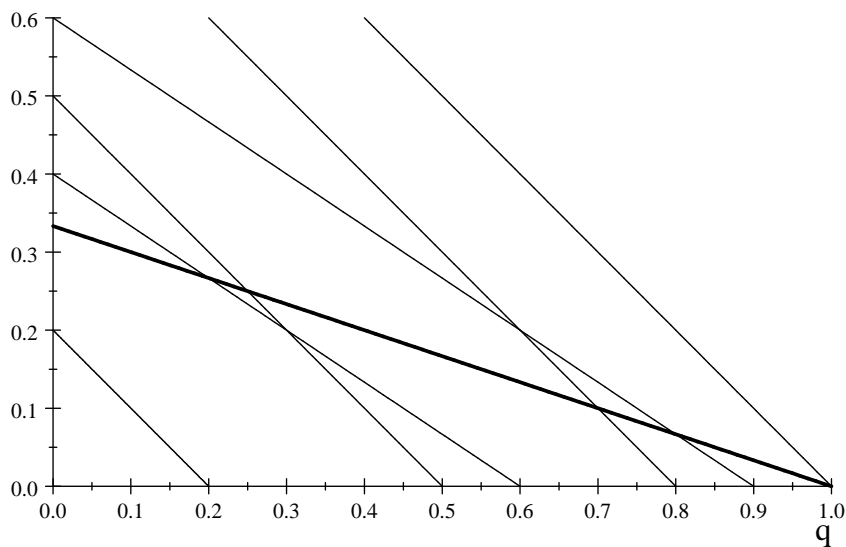


Figure 2: Illustrating the demand profile method

In some special cases the elasticity of  $N_q(p)$  does not depend on  $q$ , in which case the

<sup>20</sup>In the Mussa-Rosen context when the support of  $\theta$  includes zero, when  $q(\theta)$  is increasing one can change variables from  $\theta$  to  $q(\theta)$  in (13) to obtain expression (16).

<sup>21</sup>See Itoh (1983), Verboven (2002) and Johnson and Myatt (2015).

optimal nonlinear tariff is just a linear tariff.<sup>22</sup> (Note that when  $p(q)$  is constant, it surely cuts each inverse demand curve from below.) An important situation where this occurs is where  $u(q, \theta) = \theta q$  and  $q$  is constrained to lie in the interval  $[0, 1]$ . This corresponds to the situation where a risk-neutral consumer wants a single indivisible unit of the product, which she values at  $\theta$ , and where the seller considers offering the product stochastically, so that a consumer can obtain the product with probability  $q$  in return for payment  $T(q)$ . Provided that  $q \leq 1$  the demand profile is  $N_q(p) = 1 - F(p)$ , which does not depend on  $q$ , and so (17) implies that the optimal strategy is to offer the product for sure in return for the price  $p^*$  which maximizes  $(p - c)(1 - F(p))$ . Thus, when the buyer is risk-neutral and has unit demand, the seller has no incentive to offer its product stochastically, even though such a strategy might enable it to screen between consumer types more finely.<sup>23</sup> As discussed later in section 3.1, this result need not hold in a multiproduct context.

As well as its superior economic intuition—compare the two expressions (14) and (17), for instance—a major advantage of the demand profile method is that it applies more widely than the scalar parametric specification where consumer demand curves were not allowed to cross. For example, consumer demand curves might be linear and consumers differ both in the slope and the intercept of demand, as depicted as faint lines in Figure 2. Such problems can often be easily solved using the demand profile approach, but are very difficult to solve using a “mechanism design” approach which focusses on preference parameters.<sup>24</sup> This suggests that it is not so much the lack of a natural order on “types” *per se* which makes nonlinear pricing problems hard. Rather, as discussed in the introduction, with multiple products there is a lack of a natural order on decisions, and this makes it

---

<sup>22</sup>For example, suppose that the type- $\theta$  consumer has demand  $\theta - p$  when faced with linear price  $p$ , and the intercept  $\theta$  is exponentially distributed in the consumer population with mean  $\mu$ . Then  $N_q(p) = e^{-(p+q)/\mu}$  and (17) implies that  $p(q) \equiv c + \mu$ .

<sup>23</sup>This argument does not require that the hazard rate condition (15) be satisfied. See Myerson (1981, section 6), Riley and Zeckhauser (1983) and Manelli and Vincent (2007) for further discussion.

<sup>24</sup>Wilson (1993, section 8.4) and Rochet and Stole (2003, section 5) present examples where consumers have linear demand with heterogeneous slopes and intercepts. Earlier, Laffont, Maskin, and Rochet (1987) solved such an example using a mechanism design approach, which was far more laborious. To illustrate the demand profile approach, suppose that the type- $(\theta_1, \theta_2)$  consumer has linear demand function  $q = \theta_1 - \theta_2 p$ , where  $(\theta_1, \theta_2)$  is uniformly distributed on  $[0, 1] \times [1, 2]$ . (We bound  $\theta_2$  away from zero so that arbitrarily high profit cannot be extracted from consumers with perfectly inelastic demand.) One can calculate  $N_q(p)$ , and with costless production one can check that the price which maximizes profit from the  $q$ th unit,  $p \times N_q(p)$ , is  $p(q) = \frac{1}{3}(1 - q)$ . This price schedule is depicted as the bold line on Figure 2. This price schedule cuts each consumer’s inverse demand curve from below. Deneckere and Severinov (2014) suggest that the example in Wilson (1993, page 196) violates the requirement that the marginal price schedule cuts each inverse demand curve from below, and present a detailed analysis of this kind of screening problem when the demand profile method cannot be used.

hard to derive multiproduct tariffs except in specific cases (as in section 3.2).

*Nonlinear pricing with an uncertain outside option:* This analysis has so far assumed consumers have an outside option of zero if they do not purchase from the seller. Here we discuss how the analysis is affected if consumers are heterogeneous in their outside option. This analysis provides a bridge between the situations with a single seller and with oligopoly supply in section 4.1: in a market with several sellers where consumers are one-stop shoppers, a consumer’s outside option if she does not accept one seller’s offer could be an offer from a rival seller.

Outside options can make the seller’s problem more complicated, due to the two-dimensional nature of consumer heterogeneity combined with the fact that the demand profile approach is unlikely to work well. (With an attractive outside option, a consumer will only obtain a benefit from buying from the seller after some threshold quantity is purchased, and so they do not purchase myopically.) Consider the following framework studied by Rochet and Stole (2002).<sup>25</sup> As above, with the nonlinear tariff  $T(q)$  the consumer obtains maximum surplus  $s(\theta) = \max_q u(q, \theta) - T(q)$  if she buys from the seller. However, the consumer has an uncertain outside option, say  $x \geq 0$ , if she does not buy from the seller, where the conditional CDF for  $x$  given the consumer’s type  $\theta$  is  $G(x, \theta)$ . Thus, the type- $(\theta, x)$  consumer will buy from the seller if  $x \leq s(\theta)$ , and the seller’s profit is modified from expression (12) to

$$\int_{\theta_1}^{\theta_2} G(s(\theta), \theta) [u(q(\theta), \theta) - s(\theta) - cq(\theta)] f(\theta) d\theta . \quad (18)$$

Here, the seller operates on an extensive margin for all consumer types, and if it offers higher surplus  $s(\theta)$  it attracts more type- $\theta$  consumers to buy from it. The extra term  $G(s(\theta), \theta)$  in (18) means we can no longer use the trick of integrating by parts to eliminate  $s(\theta)$  as we did in expression (13), and instead one must often resort to using optimal control methods or similar which rarely generate closed-form solutions.

On the other hand, outside options can also simplify the seller’s problem when, as is plausible, a consumer’s outside option is positively correlated with her tastes for the seller’s product. For instance, a consumer with a strong taste for the seller’s product may have a strong taste for a rival’s product too. In these situations, “countervailing incentives” may mean that the seller is unable to exploit a consumer even if it knew she had strong

---

<sup>25</sup>See also Yang and Ye (2008).

preferences for its product.<sup>26</sup> Countervailing incentives sometimes entail particularly simple schemes being optimal, and in the context of nonlinear pricing it may be that the seller sets price equal to marginal cost regardless of the quantity purchased.

To illustrate, write  $\hat{s}(\theta) = \max_q u(q, \theta) - cq$  to be the surplus generated for a type- $\theta$  consumer with marginal-cost pricing, and suppose the outside option for a type- $\theta$  consumer takes the form  $x = \hat{s}(\theta) + \epsilon$ , where  $\epsilon$  is an additive shock to  $\hat{s}(\theta)$  which is independent of  $\theta$ . (While this seems a very particular specification with monopoly supply, it emerges more naturally in the duopoly context of section 4.1.) Suppose, hypothetically, that the seller can observe a consumer's parameter  $\theta$  (but not  $\epsilon$ ). The most profitable way to deliver a given surplus to this consumer is to charge marginal price equal to marginal cost, and then extract the desired profit via a fixed charge, say  $P$ . Doing so means that the consumer will buy from the seller if  $\hat{s}(\theta) - P \geq x = \hat{s}(\theta) + \epsilon$ , and the seller therefore chooses the fixed charge  $P$  to maximize  $P \times \Pr\{\epsilon \leq -P\}$ . Since  $\epsilon$  is independent of  $\theta$ , the optimal fixed charge does not depend on  $\theta$ , and this is the optimal scheme even if the seller cannot observe  $\theta$ . Therefore, the most profitable nonlinear tariff is a two-part tariff with price equal to marginal cost, and countervailing incentives imply that the seller cannot profitably screen consumers on the basis of their tastes for its product. Broadly speaking, the independence of  $\theta$  and  $\epsilon$  implies that the seller's demand from infra-marginal buyers is the same as its demand from marginal buyers, and in these cases we expect efficient outcomes to be chosen.<sup>27</sup>

### 3 Single-Seller, Multi-Product Analysis

#### 3.1 A framework with discrete choices

Consider a monopolist which supplies two products, labeled 1 and 2, where a consumer buys either zero or one unit of each product (and maybe a unit of each). A consumer's preferences are described by the vector  $(v_1, v_2, v_b)$ , and this consumer is willing to pay  $v_i$  for product  $i = 1, 2$  on its own, and to pay  $v_b$  for the bundle of both products. A consumer's valuations are additive if  $v_b = v_1 + v_2$ , while she views the two products as partial substitutes when  $v_b < v_1 + v_2$  and as partial complements if  $v_b > v_1 + v_2$ . The

---

<sup>26</sup>See Lewis and Sappington (1989) for an early exploration of this issue in the context of monopoly regulation, where a regulated firm with a high marginal cost is known to have a low fixed cost.

<sup>27</sup>Spence (1975) makes this point in the context of quality choice.

great majority of articles written on bundling as a form of price discrimination assume valuations are additive, as much for tractability as for realism.

Consumers face three prices:  $p_1$  is the price for product 1 on its own,  $p_2$  is the price for product 2 on its own, and  $p_1 + p_2 - \delta$  is the price for the bundle of both products. Thus,  $\delta$  is the bundle discount, which is zero with linear pricing. A consumer chooses the option from the four discrete choices which leaves her with the highest net surplus, so she will buy both items whenever  $v_b - (p_1 + p_2 - \delta) \geq \max\{v_1 - p_1, v_2 - p_2, 0\}$ , she will buy product  $i = 1, 2$  on its own whenever  $v_i - p_i \geq \max\{v_b - (p_1 + p_2 - \delta), v_j - p_j, 0\}$ , and otherwise she buys neither product. Figure 3 shows the pattern of demand in the case of additive valuations ( $v_b \equiv v_1 + v_2$ ) with a positive bundle discount.

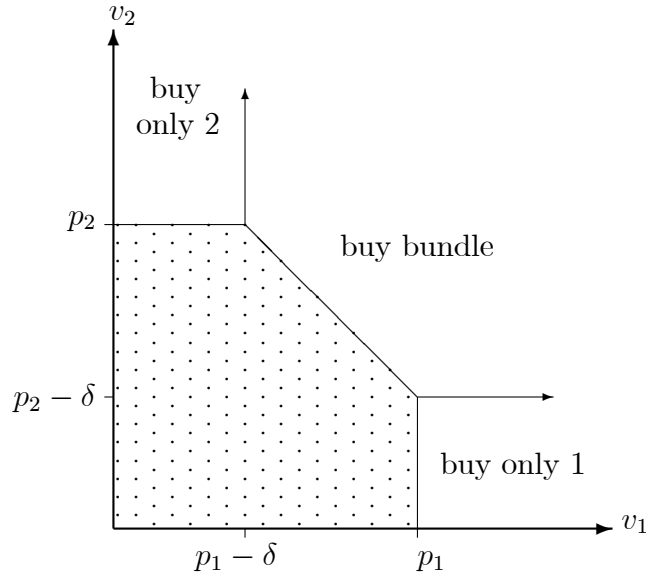


Figure 3: Pattern of demand with additive valuations and tariff  $(p_1, p_2, \delta)$

As a function of the tariff  $(p_1, p_2, \delta)$ , denote by  $Q_i$  the proportion of consumers who buy product  $i = 1, 2$  (either on its own or as part of the bundle) while  $Q_b$  is the proportion who buy both products. As can be checked from Figure 3 in the special case with additive valuations, in general we have symmetry of cross-price effects, so that  $\partial Q_i / \partial \delta \equiv -\partial Q_b / \partial p_i$  for  $i = 1, 2$ . Similarly to the demand profile functions used in section 2, it is useful to make use of the demand functions that correspond to linear pricing (i.e., with  $\delta = 0$ ), and let  $N_i(p_1, p_2) \equiv Q_i(p_1, p_2, 0)$  and  $N_b(p_1, p_2) \equiv Q_b(p_1, p_2, 0)$ . As one would expect, the sign of cross-price effects with linear pricing depends on product substitutability or complementarity: if  $v_b \leq v_1 + v_2$  (respectively,  $v_b \geq v_1 + v_2$ ) for all consumers, then  $N_i$

weakly increases (respectively, decreases) with the other product’s price  $p_j$ . Importantly, though, when products are substitutes and a bundle discount is offered, the cross-price effect can be reversed. That is to say, a bundle discount can convert products which intrinsically are substitutes into complements. This insight implies that separate sellers sometimes can relax competition by agreeing to offer a coordinated bundle discount, as discussed in section 4.2.

The seller’s incentive to introduce a bundle discount can be analyzed in a similar manner to the approach in section 2.1.<sup>28</sup> If the constant marginal cost of supplying product  $i = 1, 2$  is  $c_i$ , the seller’s profit with tariff  $(p_1, p_2, \delta)$  is

$$\pi = (p_1 - c_1)Q_1 + (p_2 - c_2)Q_2 - \delta Q_b . \quad (19)$$

One can calculate the most profitable linear prices, denoted  $(\hat{p}_1, \hat{p}_2)$ , and then derive the impact on profit (19) of introducing a small bundle discount  $\delta > 0$ . Doing so reveals that offering a discount is profitable when demand for a product is less elastic than demand for the bundle, in the sense that

$$\left. \frac{d N_b(\hat{p}_1 + t[\hat{p}_1 - c_1], \hat{p}_2 + t[\hat{p}_2 - c_2])}{dt N_i(\hat{p}_1 + t[\hat{p}_1 - c_1], \hat{p}_2 + t[\hat{p}_2 - c_2])} \right|_{t=0} < 0 , \quad (20)$$

so that a small “amplification” in price/cost markups causes demand for the bundle to fall proportionally more than demand for an individual product. (If this inequality holds for one product  $i$  it holds for the other.) As with nonlinear pricing, then, to determine whether offering a bundle discount is profitable the seller needs knowledge only of its demand system with linear prices.

Consider the case where consumer valuations are additive, i.e.,  $v_b \equiv v_1 + v_2$ , in which case  $N_i$  is a function only of  $p_i$ . Moreover, if  $v_1$  and  $v_2$  are stochastically independent, bundle demand with linear prices is just  $N_b \equiv N_1 \times N_2$ , and the left-hand side of (20) has the same sign as  $N'_j(\hat{p}_j)$ , which is negative. Thus we obtain the striking result that when products are “doubly independent”—that is, valuations are additive and stochastically independent—it is profitable for the seller to offer a bundle discount.<sup>29</sup>

To illustrate, suppose that  $c_1 = c_2 = 0$ , valuations are additive, and  $(v_1, v_2)$  is uniformly distributed on the unit square  $[0, 1]^2$ . Due to the symmetry of the two products, the seller

<sup>28</sup>For details see Long (1984) and Armstrong (2013).

<sup>29</sup>Long (1984) and McAfee, McMillan, and Whinston (1989) independently derived this result. Menicucci, Hurkens, and Jeon (2015) show that with stochastically independent valuations, the seller wishes to offer the two products as a pure bundle, rather than to use a mixed bundling tariff, when virtual valuations always exceed cost for each product.

offers the same price for either product on its own. This example is essentially the same as the “unit triangle” example presented in section 2.1 where the most profitable nonlinear tariff was (4), and the seller’s most profitable bundling tariff is to charge a price  $p_1 = p_2 = \frac{2}{3}$  for a single product and to offer the bundle discount  $\delta = \sqrt{2}/3 \approx 0.47$ .<sup>30</sup> For comparison, the most profitable linear prices are  $p_1 = p_2 = \frac{1}{2}$ , and so the use of nonlinear pricing causes the price for a single product to rise, but the charge for the bundle to fall.

When valuations are additive, the revenue from selling two products using a nonlinear tariff is greater than the sum of selling the two products separately with linear prices. Thus, even with products with no cross-effects in demand, there is an incentive for two sellers, each supplying one product, to merge, if that means that bundling can be employed. This demand-side economy of scope is the force behind the finding in Bakos and Brynjolfsson (2000) that a content provider with an existing wide portfolio of content is willing to bid more for new content than a smaller content provider.

Beyond the “doubly independent” case, it is somewhat intuitive that, all else equal, negative correlation in (additive) valuations makes it more likely that bundle discounts are profitable. Negative correlation puts less weight in the “north-east” part of the distribution on Figure 3, which tends to make bundle demand  $N_b$  more elastic.<sup>31</sup> Likewise, making products more substitutable also reduces the scale of bundle demand, which will often increase its elasticity.<sup>32</sup> Subadditive preferences give a direct reason, in addition to screening motives, to engage in nonlinear pricing. For example if a seller with costless production knows a consumer’s valuation vector is  $(v_1, v_2, v_b) = (2, 2, 3)$ , it cannot extract the maximum surplus of 3 with linear pricing but can do so with a bundle discount.

The early literature on bundling, such as Stigler (1963) and Adams and Yellen (1976), often emphasized the case where valuations were additive and very negatively correlated. If taste vectors lie on the diagonal line segment  $v_1 + v_2 \equiv 1$ , say, and production is costless, the seller can extract all consumer surplus by making only the bundle available (for total price 1), while if it used linear pricing consumers would be left with positive surplus but

---

<sup>30</sup>Manelli and Vincent (2006, Theorem 5) show that the optimal tariff in the corresponding “unit cube” example with three products has price 0.75 for any one product, price 1.14 for any two products, and price 1.22 for all three products.

<sup>31</sup>Chen and Riordan (2013) explore with copula techniques how changing the correlation in valuations, while keeping marginal distributions unchanged, affects the profitability of bundling. They find that introducing a bundle discount is usually profitable with negative correlation, and also with moderate positive correlation.

<sup>32</sup>See the discussion in Armstrong (2013, pages 457-8).

profit and total welfare would be lower. In such cases, the seller accurately knows each consumer's value for the bundle, but not for either product individually. (One then has the counter-intuitive result that consumers can be harmed when the second product is introduced, since that enables the seller to fully extract their surplus.)

A similar effect can be seen when consumer valuations are independent across products, arguably a more natural situation, provided the number of products is relatively *large*. (Examples where buyers have demand for many products include online retailing of music, video content or e-books, publishers which supply a large collection of academic journals, or software companies which sell site licenses to institutions with many users.) To illustrate, suppose there are  $n$  symmetric products, each with cost  $c$ . Valuations are additive, and each consumer's valuation for each product  $i$ ,  $v_i$ , is an independent draw from a common distribution with CDF  $G(\cdot)$ , say. Faced with a particular consumer whose list of valuations is  $(v_1, \dots, v_n)$ , the seller's ideal strategy is to supply this consumer with product  $i$  if  $v_i \geq c$  and to extract the consumer's total valuation from all such products. This can be done by offering the consumer a two-part tariff whereby she can purchase any product for price  $p = c$  by paying a fixed charge  $P$  equal to her resulting surplus, which is

$$P = \sum_{i=1}^n \max\{v_i - c, 0\} . \quad (21)$$

The obvious problem with this strategy is that this fixed charge  $P$  depends on the consumer's valuations, which are not observed by the seller. However,  $P$  in (21) is the independent sum of the random variables  $\max\{v_i - c, 0\}$ . If we write  $\mu$  for the expected value of  $\max\{v_i - c, 0\}$ , the weak law of large numbers tells us that the fraction of consumers for whom  $\frac{1}{n}P \geq \mu - \varepsilon$ , where  $\varepsilon > 0$ , converges to one as the number of products becomes large. That is, even though the seller does not know a consumer's valuation for any individual product, it can accurately estimate her total surplus from the option to purchase every product at cost.

We deduce that the seller can approximately obtain its first-best profit by offering a two-part tariff with price equal to cost and fixed charge a little lower than the average consumer surplus generated with this price. Such a tariff is approximately efficient—it allows almost all consumers to participate and when a consumer participates she buys a product whenever her valuation is above cost—but almost fully extracts each consumer's surplus. (By contrast, with linear pricing the seller offers a price strictly above cost, which

inefficiently excludes consumers from some products, but allows consumers to retain some surplus.) When  $c = 0$ , as with information goods such as music downloads or electronic journals, the seller offers the grand bundle of all products in return for a fixed charge.<sup>33</sup>

This analysis assumes that consumers exhibit no systematic differences, only idiosyncratic product preferences, which is clearly too strong to be plausible in most situations.<sup>34</sup> Suppose instead that consumers differ systematically in their overall taste for the set of products, captured by the scalar parameter  $\theta$  which is distributed in the population according to CDF  $F(\theta)$ . The type- $\theta$  consumer's valuation for each product,  $v_i$ , is now an independent draw from the CDF  $G(v | \theta)$ , which has associated density  $g(v | \theta)$ . Suppose that  $G(v | \theta)$  decreases with  $\theta$ , so that higher- $\theta$  consumers tend to have higher valuations for each product.

With many products, the sample CDF of a type- $\theta$  consumer's list of valuations—that is, the fraction of her products with valuation below  $v$ —is usually approximated by the underlying CDF,  $G(v | \theta)$ .<sup>35</sup> Since products are symmetric, suppose the seller chooses a tariff which depends only on the number of products purchased. Since there are many products, normalize their number to 1 and let  $q$  denote the fraction of available products purchased and  $T(q)$  denote the seller's charge for purchasing any such bundle. If we take the type- $\theta$  consumer's sample CDF of valuations to be exactly  $G(v | \theta)$ , her gross utility from consuming the highest-value fraction  $q$  of products is

$$u(q, \theta) = \int_{V(q, \theta)}^{\infty} v g(v | \theta) dv , \quad (22)$$

where  $V(q, \theta)$  is the type- $\theta$  consumer's value such that  $1 - G(V(q, \theta) | \theta) \equiv q$ . Then  $u_q(q, \theta) = V(q, \theta)$ , which decreases with  $q$  and increases with  $\theta$  since  $1 - G(v | \theta)$  decreases with  $q$  and increases with  $\theta$ . Therefore, the utility function (22) satisfies the requirements for the textbook analysis of single-product nonlinear pricing presented in section 2.2, and provide it induces an increasing function  $q(\theta)$  the seller's optimal strategy is given in

---

<sup>33</sup>See Armstrong (1999) and Bakos and Brynjolfsson (1999) for details of this analysis. Geng, Stinchcombe, and Whinston (2005) focus on the situation where products have very asymmetric distributions, in the sense that the distribution for valuation  $v_i$  becomes concentrated at zero as  $i$  becomes large. They show that pure bundling is optimal provided that the distributions do not converge to zero too fast.

<sup>34</sup>Some consumers like music more than others, for instance. Shiller and Waldfoegel (2011) obtain survey data about valuations for various songs from a population of students, and in section III(ii) they discuss how most of the variation in valuations is associated with individuals rather than songs. Because of this strong positive correlation in valuations, they estimate that the best pure bundling tariff cannot extract a high proportion of the available surplus.

<sup>35</sup>This is the Glivenko-Cantelli Theorem.

expression (14).<sup>36</sup> For instance, if the type- $\theta$  consumer has  $v$  exponentially distributed with mean  $\theta$ , then utility in (22) takes the Mussa-Rosen form  $u(q, \theta) = \theta q(1 - \log q)$ , and with costless production the seller should then sell the grand bundle of all products for a fixed fee  $P$  (where  $P$  maximizes  $P(1 - F(P))$ ). In sum, when consumers have unit demands for many products, an approximately optimal bundling tariff can often be derived using the standard approach to single-product nonlinear pricing.

This discussion of nonlinear pricing with many products is an instance of an approach to price discrimination which tries to approximate the optimal tariff by means of an in some sense “simple” tariff, or to gauge the loss incurred when simple tariffs are used. For instance, with two products and consumer valuations which are identically and independently distributed, Hart and Nisan (2014) show how a symmetric linear tariff can achieve at least 73% of the possible revenue from any mechanism, including the stochastic schemes discussed shortly. (By contrast, though, if valuations can be correlated then linear pricing cannot ensure any positive fraction of optimal profit.) Chu, Leslie, and Sorensen (2011) examine the profitability of “bundle size pricing”, whereby the price for a bundle depends only on the number of items in the bundle. (With  $n$  products, bundle size pricing involves  $n$  prices, while mixed bundling might involve  $2^n - 1$  prices.) They show numerically that this simple tariff often obtains a large fraction of the maximum profits obtained from (deterministic) mixed bundling. Using market data from a theatre company, they estimate the joint distribution of valuations for the various shows. (They pin down correlations in valuations by observing which shows appear together when consumers choose the “any five shows” bundle offered by the company.) They estimate in this case that bundle-size pricing would obtain 98.5% of the profits which could be generated with mixed bundling.<sup>37</sup>

*Stochastic selling schemes:* So far in this section I have considered deterministic selling schemes. A multiproduct monopolist (or a single-product monopolist offering different quantities or qualities) can often increase its profit by using a stochastic scheme, whereby a consumer is uncertain about which bundle she will receive. To illustrate, consider the following example with three risk-neutral consumers each with additive valuations. Consumer 1 has valuation pair  $(v_1, v_2) = (10, 1)$ , consumer 2 has  $(v_1, v_2) = (1, 10)$ , while consumer

---

<sup>36</sup>A version of this discussion was contained in Armstrong (1999, section 3).

<sup>37</sup>Shi (2016) conducts a related analysis using data from an online book retailer. He estimates that a simple scheme whereby the retailer waives the shipping fee with a large enough order obtains a large proportion of the profit obtained with a more ornate mixed bundling scheme.

3 has  $(v_1, v_2) = (9, 9)$ . Thus consumer 3 has strong tastes for both products, while consumer  $i = 1, 2$  has slightly stronger tastes for product  $i$  but very weak tastes for the other product. With costless production, the best deterministic selling strategy is to offer any single product at price 10 and the bundle of both products at price 18, which makes each consumer's participation constraint bind while the incentive compatibility constraints are slack. However, some surplus is wasted since consumers 1 and 2 do not obtain both products, and the seller can extract more revenue by offering these consumers the deal whereby for price  $10 + \varepsilon$  she obtains her preferred product for sure but also obtains her less preferred product with probability  $\varepsilon$ . This keeps these consumers to their participation constraint, and if  $\varepsilon$  is small enough (if  $\varepsilon < \frac{1}{8}$ ) then consumer 3 continues to suffer negative surplus if she chooses one of these stochastic contracts. This scheme therefore yields higher profit for the seller, and does so by permitting local surplus-increasing changes which do not interact with the slack incentive constraints.

Pavlov (2011, section 4) studies situations with additive valuations, and finds that a stochastic scheme can be optimal even in the most regular cases where valuations are uniformly distributed on a square with lower boundary above zero. Manelli and Vincent (2006) find restrictive conditions under which a deterministic mechanism is optimal. With two products and stochastically independent and additive valuations with lower boundary of support equal to zero, a sufficient condition to rule out stochastic schemes is that  $v_i f'_i(v_i)/f_i(v_i)$  increases with  $v_i$  where  $f_i$  is the density for the valuation for product  $i$ . Manelli and Vincent (2007) discuss the underlying reason why stochastic schemes are so often optimal with multiple products but not with one product. When consumers are risk-neutral and have unit demands, the seller's problem is a linear function of the probability schedule, and hence is maximized at the "extreme points" of the feasible set of such schedules. With a single product, the extreme points simply correspond to selling the item for sure in return for a fixed price. With several products, the extreme points can take many forms, including those which involve random assignment.

### 3.2 An example with continuous choices

In most cases where a seller engages in nonlinear pricing, a consumer either purchases multiple units of one product (as in section 2) or a single unit of several products (section 3.1). Nevertheless, there are important situations in which consumers purchase multiple

units of several products. For instance, an energy company might supply electricity and gas, and offer a tariff which depends on the chosen quantities of both fuels. Alternatively, the consumer might choose both the quality and the quantity of a product.

Unfortunately, the economics needed to understand the optimal form of these multiproduct nonlinear tariffs is not yet well developed. In this section, I merely describe how the most profitable such tariff can sometimes be derived, and in a relatively simple manner. Suppose the seller supplies two symmetric products, each with marginal cost  $c$ , and suppose that the type- $(w_1, w_2)$  consumer's demand for product  $i = 1, 2$  with linear prices  $(p_1, p_2)$  is  $w_i x(p_i)$ . Thus, there are no cross-price effect in a consumer's demands,  $x(\cdot)$  is an underlying demand function, and the taste parameter  $w_i$  shifts her demand for product  $i$  multiplicatively. This demand system corresponds to a gross utility function for the type- $(w_1, w_2)$  consumer given by

$$w_1 U\left(\frac{q_1}{w_1}\right) + w_2 U\left(\frac{q_2}{w_2}\right), \quad (23)$$

where  $U$  is the concave utility function corresponding to the demand function  $x(\cdot)$ , i.e.,  $U'(x(p)) \equiv p$ .

It is useful to change variables so that

$$\theta = w_1 + w_2; \quad r = \frac{w_1}{w_1 + w_2}.$$

Here,  $\theta$  represents the average scale of the consumer's demands and  $r$  represents her relative demand for product 1 compared to product 2. A property of the utility specification (23) is that the maximum utility which a consumer can obtain from total quantity  $q = q_1 + q_2$  depends only on  $\theta$ , not on  $r$ . More precisely, we have

$$\max \left\{ w_1 U\left(\frac{q_1}{w_1}\right) + w_2 U\left(\frac{q_2}{w_2}\right) \text{ subject to } q_1 + q_2 = q \right\} = \theta U\left(\frac{q}{\theta}\right), \quad (24)$$

where the individual quantities which solve (24) are  $q_1 = r q$  and  $q_2 = (1 - r)q$ .

Suppose hypothetically that the seller can directly observe a consumer's relative taste parameter  $r$ , but not her scale parameter  $\theta$ , and offers this particular group of consumers a nonlinear tariff. Conditional on  $r$ , suppose the CDF for  $\theta$  is  $F(\theta | r)$  with associated density  $f(\theta | r)$ . Since the seller's cost depends only on the total quantity supplied to a consumer, it will make the tariff to this group depend only on total quantity  $q$ , not on the individual quantities,  $q_i$ , and let the nonlinear tariff offered to this group be  $T(q | r)$ . The

function  $u(q, \theta) \equiv \theta U(q/\theta)$  in expression (24) has increasing differences in  $(q, \theta)$ , and we can apply the textbook method described in section 2.2. Provided it results in an increasing total quantity schedule  $q(\theta)$ , expression (14) reveals that the seller's optimal policy for this group of consumers is such that the type- $\theta$  consumer in this group is allocated total quantity  $q(\theta)$  which maximizes

$$u(q, \theta) - \frac{1 - F(\theta | r)}{f(\theta | r)} u_\theta(q, \theta) - cq . \quad (25)$$

Here, the type- $\theta$  consumer in this group obtains quantity  $q_1 = rq(\theta)$  of product 1 and  $q_2 = (1 - r)q(\theta)$  of product 2. If  $\theta$  and  $r$  are independent random variables, then the conditional distribution  $F(\theta | r)$  in (25) does not depend on  $r$ . In this case, the optimal tariff  $T(\cdot)$  contingent on being able to observe  $r$  does not depend on  $r$ , and (25) represents the optimal strategy even when the seller cannot observe  $r$ . In such cases, we have derived the optimal multiproduct nonlinear tariff.

Intuitively, when  $\theta$  and  $r$  are independent the consumer is happy for the seller to observe her relative tastes  $r$  but not her average tastes  $\theta$ . That is, for a given total quantity, the consumer and seller have aligned preferences with respect to relative quantities. The seller is willing to delegate to the consumer the choice of relative quantities, to enable her to make use of her private information about relative tastes, and it does this by offering a tariff which depends only on the total quantity supplied. However, if  $\theta$  and  $r$  were correlated, observing the consumer's choice of relative quantities is informative about her scale parameter  $\theta$ , and the seller will wish to make its tariff depend on the individual quantities, not just total quantity. I do not know how to solve such cases analytically.<sup>38</sup>

To illustrate the method, suppose that  $w_1$  and  $w_2$  are independent exponential variables with support  $[0, \infty)$  and parameter  $\lambda$ . A property of i.i.d. exponential variables is that their sum and ratio are also independent random variables, where  $\theta \equiv w_1 + w_2$  is a Gamma variable with CDF  $F(\theta) = 1 - (1 + \lambda\theta)e^{-\lambda\theta}$  while  $r \equiv w_1/(w_1 + w_2)$  is uniformly distributed on  $[0, 1]$ . Thus, the above method can be applied. To obtain an explicit solution, suppose that  $c = 0$ ,  $\lambda = 1$ , and  $x(p) = 1 - p$ , so that  $U(x) = x - \frac{1}{2}x^2$  and  $u(q, \theta) = q - q^2/(2\theta)$ . Then formula (25) implies that  $q(\theta) = \theta^3/(1 + \theta + \theta^2)$ , which increases with  $\theta$  as required. In terms of the original preference parameters  $(w_1, w_2)$ , the optimal quantity of product

---

<sup>38</sup>One can instead study a discrete-types model. For instance, Armstrong and Rochet (1999) examine a model with two products and where consumers can have strong or weak tastes for each product (so there are four types of consumer in all). By considering the various ways in which the participation and incentive compatibility constraints might bind, one can completely solve such a model.

$i = 1, 2$  for the type- $(w_1, w_2)$  consumer is

$$q_i(w) = w_i \frac{(w_1 + w_2)^2}{1 + (w_1 + w_2) + (w_1 + w_2)^2}. \quad (26)$$

The quantities the seller would allocate if it knew each consumer’s tastes perfectly are  $q_i = w_i$ , and expression (26) shows that the seller allocates lower quantities than this when tastes are private information (although the allocation is approximately efficient when  $w$  is large). Here,  $q_i$  in (26) increases with  $w_j$ , and even though utility is additive across products and taste parameters  $w_1$  and  $w_2$  are stochastically independent—i.e., preferences are “doubly independent”—it is optimal to induce bundling in this manner.

This analysis is loosely based on Armstrong (1996, section 4.4). However, the way that taste parameters enter a consumer’s gross utility function in (23) differs from the specification in Armstrong (1996), and avoids the inevitability of “exclusion”—whereby a positive measure of consumers do not buy anything from the firm—which was emphasized in Armstrong (1996, section 3).<sup>39</sup> Indeed, in (26) we see that all consumers (except when  $w_1 = w_2 = 0$ ) buy something in this optimal scheme. Rochet and Choné (1998, section 7.2) analyze a related example with taste parameters being exponentially distributed, but with the support shifted away from the origin. However, taste parameters enter differently into utility in their model: in (23) the type- $(w_1, w_2)$  consumer has gross utility  $u = q_1 + q_2 - q_1^2/(2w_1) - q_2^2/(2w_2)$ , while their consumers have gross utility  $u = w_1q_1 + w_1q_2 - q_1^2/2 - q_2^2/2$ . Rochet and Choné’s example is solved numerically, and they find that consumers with small  $(w_1 + w_2)$  buy nothing, consumers with intermediate  $(w_1 + w_2)$  are “bunched” in the sense that a consumer in this region buys the same quantity of both products, while consumers with large  $(w_1 + w_2)$  each buy a distinct bundle.

## 4 Multi-Seller Analysis

### 4.1 One-stop shopping

In this section we discuss competition between oligopolists when there is one-stop shopping, so that a consumer buys all her units/products from one seller or another. (In the next section I discuss the alternative scenario where consumers can “mix and match” from

---

<sup>39</sup>The argument in Armstrong (1996) assumed that gross utility was convex and homogeneous degree 1 in taste parameters, while in (23) neither of these requirements are satisfied. See Barelli, Basov, Bugarin, and King (2014) for further discussion.

several suppliers.) For instance, competing bookstores might stock broadly the same range of books, and consumers all else equal buy from their nearest store (but might buy more there with a “three for the price of two” deal). Shopping costs might mean that consumers have a strong preference for making all their purchases from a single seller. (When dining out people rarely consume a main course from one restaurant and dessert from another.) And in settings with quality choice or add-on products, it will usually make sense to suppose that the upgrade (or similar) is purchased from the supplier of the basic product.

As in section 2.1, suppose a seller offers consumers two options, a basic product and a premium product, and we can think of the premium product as consisting of the basic product bundled with an upgrade. The one-stop shopping assumption implies that a consumer cannot buy the basic product from one seller and the upgrade from another. Each seller incurs the cost  $c_1$  to supply the basic product and an extra cost  $c_2$  to supply the upgrade.

Perhaps the most straightforward situation is when sellers offer homogenous product lines and compete in Cournot fashion.<sup>40</sup> Here, the type- $(v_1, v_2)$  consumer is willing to pay  $v_1$  for the basic product (from either seller) and an extra amount  $v_2$  for the upgrade to the premium product (from either seller). One can in principle calculate demand functions for the two product variants from Figure 1 and invert these to obtain the inverse demand functions, and go on to determine the multiproduct Cournot equilibrium.<sup>41</sup> As usual, though, the analysis is much simplified when the demand profile approach is valid, in which case the basic product and the upgrade can be considered as separate markets without cross-price effects.<sup>42</sup> In these cases, the equilibrium price for the upgrade is determined as in a standard single-product Cournot market; in particular, the equilibrium price for the upgrade is above its cost.

Consider next situations in which sellers compete in tariffs rather than in quantities. Here, it makes sense to suppose that consumers care about which seller they buy from—that is, they have a brand preference—for otherwise Bertrand competition entails prices equal to costs for both product variants.<sup>43</sup> This scenario is potentially more complicated

<sup>40</sup>For further discussion, see Wilson (1993, section 12.3) and Johnson and Myatt (2006a).

<sup>41</sup>For instance, for the “uniform square” example from section 2.1, calculations reveal that the Cournot duopoly equilibrium tariff is  $(p_1, p_2) = (\frac{3}{5}, \frac{1}{30})$ .

<sup>42</sup>The demand profile approach is valid in this oligopoly context when there is scalar consumer heterogeneity combined with a suitable hazard rate condition, as assumed in Johnson and Myatt (2006a).

<sup>43</sup>Champsaur and Rochet (1989) study an alternative two-stage interaction, where sellers, which are initially symmetric, first choose their product lines and then compete in tariffs. Focussing on the case

than monopoly supply, due to the additional margin on which a seller operates: if it increases the price for one option an existing customer might (i) choose the other option from that seller, (ii) exit the market altogether, or (iii) switch to the best option available from a rival. To simplify the analysis somewhat, suppose there is full coverage over the relevant range of prices, so that all consumers buy at least the basic product and margin (ii) does not apply.<sup>44</sup> From an individual seller’s perspective, perhaps the main difference when (iii) rather than (ii) constitutes the consumer’s outside option is that there will likely be countervailing incentives. As discussed in section 2.2, countervailing incentives affect, and often simplify, a seller’s optimal tariff.

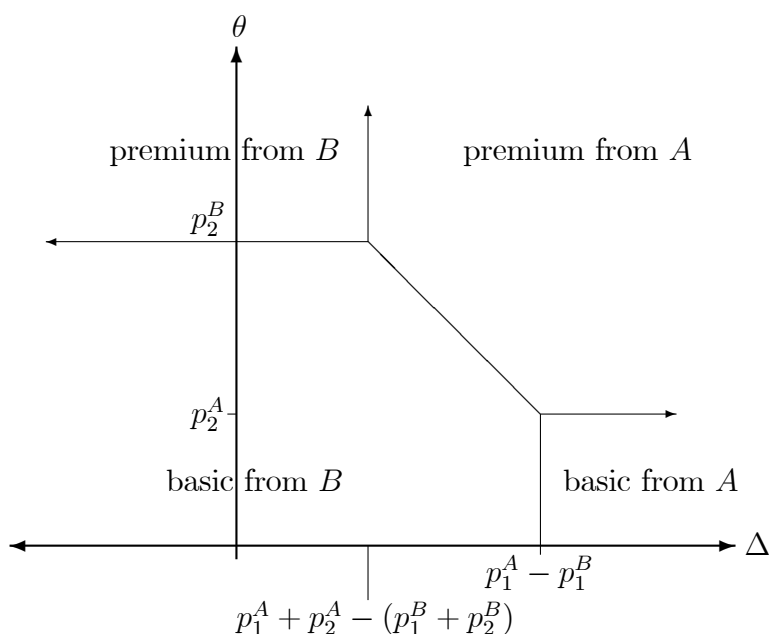


Figure 4: Pattern of demand with one-stop shopping ( $p_2^A < p_2^B$ )

In more detail, suppose there are two sellers, denoted  $A$  and  $B$ , and that consumers are heterogeneous in two dimensions: their “horizontal” brand preference for seller  $A$  over seller  $B$ , represented by parameter  $\Delta$ , and their “vertical” valuation for the upgrade to the premium product (from either seller), represented by  $\theta$ . That is, a consumer’s valuation for the basic product from seller  $i = A, B$  is  $V^i$ , say, and her total valuation for the premium product from this seller is  $V^i + \theta$ . Since there is full consumer coverage, only the difference

---

where sellers choose products lines in an interval, they show that sellers choose product lines which do not overlap, and hence in the second stage there is vertical differentiation between sellers.

<sup>44</sup>See Rochet and Stole (2002) and Yang and Ye (2008) for analyses of the situation with partial consumer participation.

$\Delta \equiv V^A - V^B$  matters for consumer choices.<sup>45</sup> Suppose that seller  $i$  charges  $p_1^i$  for its basic product and an incremental price  $p_2^i$  for the upgrade to its premium variant. With this pair of tariffs, the pattern of consumer demand is shown on Figure 4. Notice that intra-firm options are complements—a decrease in a seller’s basic price boosts demand for that seller’s upgrade—while inter-firm options are substitutes in that any price rise from one seller increases the number of consumers who buy from the rival.

One can calculate the proportion of consumers who choose each of the four options in Figure 4 and so derive the equilibrium tariffs. The equilibrium is particularly simple when  $\Delta$  and  $\theta$  are stochastically independent, so that knowledge of a consumer’s preference for seller  $A$  over seller  $B$  carries no information about her willingness to pay for the upgrade. In this case, the discussion of countervailing incentives in section 2.2 demonstrates that if one seller sets its upgrade price  $p_2$  equal to cost  $c_2$ , the rival’s best response is to do the same. Thus, subject to mild regularity conditions on the distribution of  $\Delta$ , an equilibrium exists in which both sellers price the upgrade at cost and obtain all their profit from selling the basic product.<sup>46</sup> The reason, as with monopoly, is that the preferences of a seller’s marginal customers do not differ systematically from the preferences of its infra-marginal customers. This is another instance where the demand profile perspective is valid: sellers can be considered to compete separately to supply the basic product and the upgrade, and since the upgrade is an undifferentiated product its price is forced down to cost.<sup>47</sup>

In the context of nonlinear pricing, where  $c_1 = c_2 = c$  say, the issue arises whether sellers make more or less profit if they engage in nonlinear pricing relative to linear pricing. (With monopoly supply, the seller must benefit from price discrimination since it has more instruments to work with, but in oligopoly the issue is less clear-cut.) When  $\Delta$  and  $\theta$  are independent, and when sellers are symmetric in the sense that  $\Delta$  has density  $f(\Delta)$  which is symmetric about zero, the answer can be seen from Figure 4. With nonlinear pricing, we know that in equilibrium  $p_2 = c$ , and one can check that the basic price is  $p_1 = c + 1/(2f(0))$  so that industry profit is  $1/(2f(0))$ . If instead sellers compete with linear prices, so that  $p_1^A = p_2^A$  and  $p_1^B = p_2^B$ , Figure 4 shows that a price cut by a seller

<sup>45</sup>This set-up is similar to that studied by Verboven (1999). He assumes that  $V^A$  and  $V^B$  are extreme value random variables, so that the difference  $\Delta$  has a logistic distribution.

<sup>46</sup>With distributional assumptions for  $\Delta$  and  $\theta$ , this result is shown in Verboven (1999, Proposition 1).

<sup>47</sup>McManus (2007, Table 4) studies how the incremental price to obtain a larger cup of coffee compares with the associated cost of ingredients, and finds the two to be similar in the case of “sweet expresso” drinks, while other styles of coffee have larger incremental markups.

yields an advantageous change in the composition of that seller’s demand, and the lower price attracts a disproportionate number of two-unit buyers since the lower price is enjoyed over more units. Since two-unit buyers are more profitable than single-unit buyers when linear prices are used, a seller has a strong incentive to undercut its rival. Confirming this intuition, one can show that industry profit with linear pricing is strictly lower than with nonlinear pricing, so long as there is a mixture of one- and two-unit demand with linear pricing.<sup>48</sup>

Similar analysis applies when consumers make continuous choices as in section 2.2. Suppose that the type- $(\Delta, \theta)$  consumer has gross utility  $u(q, \theta) + \Delta$  if she buys  $q$  units from seller  $A$  and gross utility  $u(q, \theta)$  if she buys these units from  $B$ . Then if the parameters  $\Delta$  and  $\theta$  are stochastically independent and over the relevant range of tariffs there is full coverage, it is an equilibrium for each seller to offer a two-part tariff with price equal to marginal cost.<sup>49</sup> Note that both  $q$  and  $\theta$  could be multidimensional here, with a distinct marginal cost for each product, in which case equilibrium multiproduct nonlinear tariffs take the very simple form of a two-part tariff with each price equal to the associated cost. The requirement that  $\Delta$  and  $\theta$  be stochastically independent plays essentially the same role as the requirement in section 3.2 that “average” and “relative” demands be independent. In either setting, the strategy to calculate the equilibrium tariff is to suppose hypothetically that a seller can observe all-but-one aspect of the consumer’s private information, solve for the optimal tariff conditional on that observation (an easy one-dimensional problem which screens on the basis of the remaining dimension of private information) and then note that with stochastic independence this tariff does not depend on the observed parameter.

Returning to the discrete choice setting, Ellison (2005) suggests that one reason why price is above cost for upgrades is correlation between horizontal and vertical preferences: consumers with a high marginal utility of income are plausibly willing to pay less for an upgrade *and* willing to pay less to buy from their preferred brand. In this situation we expect that consumers with small  $\theta$  will be more likely to have  $\Delta$  close to zero. When this is the case, it is plausible that an equilibrium exists in which sellers charge more than cost for the upgrade, since they enjoy greater market power over those consumers who have a strong taste for the premium version.<sup>50</sup>

---

<sup>48</sup>For a proof of this claim in a richer model, see Armstrong and Vickers (2010, Proposition 5).

<sup>49</sup>See Armstrong and Vickers (2001, Proposition 5) and Rochet and Stole (2002, Propositions 2 and 6).

<sup>50</sup>In fact, in Ellison’s model the previous equilibrium in which the upgrade is priced at cost continues to exist as well. Bonatti (2011) studies an alternative source of market power of upgrades, which is that

A second reason why upgrades are often expensive is that consumers do not pay sufficient attention to the add-on price until they are locked in to their chosen seller. This might be because sellers “shroud” their add-on price until the consumer has committed to purchase. (For instance, it is not easy to check mini-bar prices in advance when choosing a hotel.) Alternatively, consumers might be over-pessimistic about their eventual demand for the add-on product at the time they choose supplier, and even if add-on prices are clearly displayed they do not give them adequate consideration. (Learner drivers might underestimate the likelihood they will fail their test on the first attempt, and so do not value a tariff from a driving school of the form “free second course of lessons if you fail”.)

To discuss this issue in more detail, follow Verboven (1999)’s model and suppose sellers do not reveal their upgrade price  $p_2$  until a consumer has committed to purchase the basic product from a seller, while consumers are rational and anticipate their seller’s incentives to choose this price.<sup>51</sup> If  $\Delta$  and  $\theta$  are stochastically independent, each seller will choose its upgrade price to maximize upgrade profit  $(p_2 - c_2) \times \Pr\{\theta \geq p_2\}$ . Thus, upgrade prices are set at the monopoly level, despite competition between sellers to supply the basic product. The monopoly profit generated in the upgrade market stimulates competition for consumers, and depending on the distribution for brand preferences  $\Delta$ , it may be that the basic product in equilibrium is subsidised ( $p_1 < c_1$ ). This “bargain-then-ripoff” pattern of pricing is a common feature of markets with lock-in when subsequent price is not revealed until the buyer is committed to the seller. Comparing this outcome to the previous situation where both sellers advertise add-on prices, in which case the add-on is priced at cost, industry profit is unchanged although the balance between the basic price and add-on price is very different. Welfare and consumer surplus is higher in the transparent market since the add-on is then priced efficiently.

By contrast, Ellison (2005) assumes positive correlation between  $\theta$  and  $\Delta$  which implies that profit is higher in the opaque “add-on pricing game” in which sellers coordinate to conceal upgrade prices. Ellison (2005, section V.C) and Gabaix and Laibson (2006) study a related model in which naive consumers are unaware of the add-on product if sellers

---

consumers have brand preferences over the supplier of the upgrade.

<sup>51</sup>More generally, the following argument carries over if consumers merely incur a cost for discovering the add-on price from each seller. Consumers anticipate that both sellers will set monopoly add-on prices, and so it is not worthwhile for them to incur the cost to discover this price in advance. When there is full coverage, this argument also carries over if consumers are naive in the sense that they do not anticipate they will need the add-on service and so choose a seller only by comparing  $p_1^A$  and  $p_1^B$ .

shroud their add-on prices, but if a single seller unshrouds its add-on price, all consumers become aware of their demand for this service (and of a seller’s incentive to set its add-on price if it keeps this price shrouded). It can then be an equilibrium for all sellers to shroud their add-on prices, in contrast to Ellison’s main model with rational consumers where each seller unilaterally wishes to advertise all its prices.

## 4.2 Mix-and-match shopping

We next consider situations in which to obtain her desired bundle of products a consumer might “mix and match” from different sellers. With linear pricing, some consumers choose to source products from several sellers, such as when a traveller chooses the airline for each trip which offers the most suitable departure time. Here, a seller may have an incentive to offer a bundle discount if a consumer buys several products from it, as with a frequent flier program, and this encourage consumers to become loyal one-stop shoppers.

To discuss this issue, suppose that two sellers, denoted  $A$  and  $B$ , each supply the same pair of products, denoted 1 and 2, and each consumer wants at most one unit of each product. (In the airline context, these might be two routes, say.) Even in this stylized setting, a consumer has a large number of shopping possibilities: she can buy each product  $i$  from seller  $A$  or seller  $B$  or not at all, so there are nine possibilities in total, each of which has an associated value to the consumer. To simplify the analysis somewhat, suppose that valuations are additive, in that if  $v_i^j$  is a consumer’s valuation for buying product  $i$  from seller  $j$ , her value for the bundle where product 1 is purchased from seller  $j$  and product 2 is purchased from seller  $k$  is  $v_1^j + v_2^k$ . In particular, this implies there is no intrinsic benefit, such as saving on shopping costs, in buying both items from the same seller. In addition, suppose that consumer valuations are such that, over the relevant range of tariffs, every consumer wishes to buy both products. Together, these assumptions imply that consumer decisions are determined by two horizontal taste parameters,  $\Delta_1$  and  $\Delta_2$ , where  $\Delta_i$  is a consumer’s brand preference for seller  $A$ ’s product  $i$  over  $B$ ’s version.

Suppose seller  $j$  chooses price  $p_1^j$  for product 1 on its own, price  $p_2^j$  for product 2 on its own, and offers a discount  $\delta^j \geq 0$  if a consumer buys both items. Thus, a consumer prefers to buy both items from  $A$  to buying product 1 from  $A$  and product 2 from  $B$ , say, if

$$v_1^A + v_2^A - (p_1^A + p_2^A - \delta^A) \geq v_1^A + v_2^B - (p_1^A + p_2^B) ,$$

i.e., if  $\Delta_2 \geq p_2^A - p_2^B - \delta^A$ . The resulting pattern of consumer demand is depicted on Figure

5. Consumers with very asymmetric brand preferences—who prefer one seller for product 1 but the other for product 2—will mix and match, while others are one-stop shoppers. The number of one-stop shoppers increases with the size of bundle discounts, which convert otherwise independent products into complements.

The analysis is most transparent when the two sellers are symmetric, in the sense that their costs are the same and consumer tastes are symmetric so that the joint density for taste parameters,  $f$ , satisfies  $f(\Delta_1, \Delta_2) = f(-\Delta_1, -\Delta_2)$ . In this scenario, it is natural to suppose each seller offers the same tariff in equilibrium, say  $(p_1, p_2, \delta)$ . If sellers do not engage in bundling ( $\delta = 0$ ) then every consumer buys each product from her preferred seller, i.e., a consumer buys product  $i$  from  $A$  if and only if  $\Delta_i > 0$ . This implies total welfare is maximized. In this setting with full coverage, then, bundling is sure to lower welfare, since it induces excessive one-stop shopping. (Frequent flier programmes encourage travellers sometimes to fly at inconvenient times, say, in order to make use of air miles.) The fraction of consumers who mix-and-match with the symmetric tariff  $(p_1, p_2, \delta)$  depends only on  $\delta$ , and denote the number of two-stop shoppers by the decreasing function  $\Phi(\delta)$ . Here,  $\Phi$  can be considered to be the demand for two-stop shopping as a function of the implicit price of two-stop shopping,  $\delta$ .

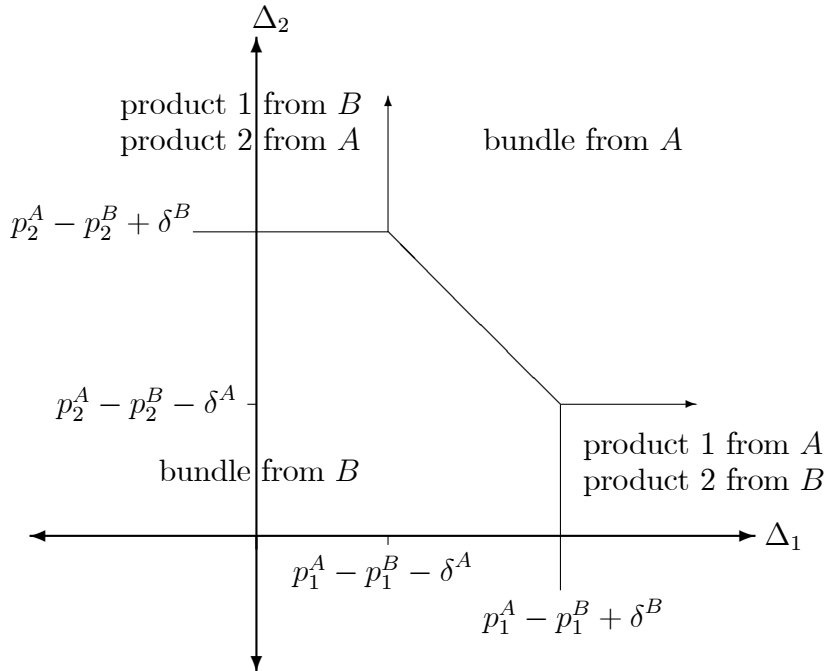


Figure 5: Pattern of demand when duopolists engage in bundling

There is an intuitive formula for the size of the equilibrium bundle discount. Relative to

the symmetric situation where both sellers choose tariff  $(p_1, p_2, \delta)$ , suppose a seller increases both of its stand-alone prices by some small amount  $\varepsilon$  and increases its bundle discount by  $2\varepsilon$ . This deviation keeps the seller’s bundle price unchanged but increases a consumer’s cost of two-stop shopping by  $\varepsilon$ . Using Figure 5, one can check that, regardless of the stand-alone prices  $(p_1, p_2)$ , the impact of this change on the seller’s profit to first order is  $\varepsilon \times [\Phi(\delta) + \frac{1}{2}\delta\Phi'(\delta)]$ . Thus, the equilibrium discount makes the term  $[\cdot]$  vanish, so that

$$-\frac{\delta\Phi'(\delta)}{\Phi(\delta)} = 2 \tag{27}$$

and the elasticity of two-stop shopping is equal to 2.<sup>52</sup> In particular, so long as  $\Phi(0) > 0$ , so there are some two-stop shoppers with linear pricing, the equilibrium involves a positive bundle discount. This contrasts with the monopoly analysis, when a monopolist only had an incentive to offer a bundle discount if expression (20) was satisfied. Expression (27) can be interpreted as an instance of the situation where two single-product sellers costlessly supply perfect complements, in which case the combined product has elasticity 2 in equilibrium. In the current context, a mix-and-match shopping bundle involves each seller supplying one complementary component, and each seller independently chooses the combined price of two-stop shopping,  $\delta$ , via its choice of stand-alone prices. Just as sellers who supply complementary products set an inefficiently high total price, here sellers in equilibrium choose an inefficiently deep bundle discount.

To illustrate, suppose that  $(\Delta_1, \Delta_2)$  is uniformly distributed with density  $\frac{1}{4}$  on the square  $[-1, 1]^2$ , in which case  $\Phi(\delta) = \frac{1}{2}(1 - \delta)^2$  and (27) implies that the equilibrium bundle discount is  $\delta = \frac{1}{2}$  and only one in eight consumers choose to mix and match. With costless production, one can then show using Figure 5 that equilibrium stand-alone prices are  $p_1 = p_2 = \frac{11}{12}$ . By contrast, when the sellers compete with linear tariffs (so  $\delta = 0$ ), half the consumers mix and match and the equilibrium linear prices are  $p_1 = p_2 = 1$ . In this example, then, when sellers engage in nonlinear pricing *all* prices fall relative to the situation with linear pricing. It follows that all consumers are better off with nonlinear pricing (even though they sometimes buy a less preferred product), while sellers obtain lower profit. Sellers are forced to play a prisoner’s dilemma: each has a unilateral incentive to offer a bundle discount—for the same reason as the monopolist in section 3.1 does—but when both do so their profit falls.<sup>53</sup>

<sup>52</sup>See Armstrong and Vickers (2010, Proposition 1), who show the formula is also valid when consumers have an intrinsic shopping cost when buying from two sellers.

<sup>53</sup>Papers which discuss this “competition intensifying” property of competitive mixed bundling include

Thus we see a contrast between situations with one-stop shopping, where nonlinear pricing usually boosts profit relative to linear pricing, and with mix-and-match shopping, where bundling sometimes acts to intensify competition. Intuition for why bundling can intensify competition is most transparent in the case of *pure* bundling versus linear pricing.<sup>54</sup> In more detail, suppose that the parameters  $\Delta_1$  and  $\Delta_2$  are independent draws from some common distribution with density  $f(\Delta)$  which is symmetric about  $\Delta = 0$ . When sellers compete in linear prices, product by product, and the cost of either product is  $c$ , the equilibrium price for each product is  $p_1 = p_2 = c + 1/(2f(0))$ . Here, the density  $f(0)$  captures the size of the competitive margin between these symmetric sellers, and determines a seller's own-price elasticity of demand. Suppose instead that sellers compete only in bundles, so that a consumer must buy both products from one or other seller. If  $P^i$  is seller  $i$ 's price for the bundle, a consumer will buy from  $A$  if  $\Delta_1 + \Delta_2 \geq P^A - P^B$ . If  $\bar{f}(\bar{\Delta})$  is the induced density for the average brand preference  $\bar{\Delta} = \frac{1}{2}(\Delta_1 + \Delta_2)$ , then the equilibrium bundle price satisfies  $P/2 = c + 1/(2\bar{f}(0))$ . Thus the per-product bundle price,  $P/2$ , is lower than the price without bundling if  $\bar{f}(0) > f(0)$ . In regular cases the density for the average of i.i.d. variables is more concentrated about the mean than is the underlying density. In these cases the per-product price is lower when sellers compete in bundles than when they compete product-by-product, due to the homogenizing impact of bundling on consumer valuations discussed in the introduction.<sup>55</sup>

Armstrong and Vickers (2010) extend this unit-demand framework so that consumers buy continuous quantities of the two products. Similar to Armstrong and Vickers (2001) and Rochet and Stole (2002) in the one-stop shopping context, in symmetric cases with full coverage and stochastic independence between horizontal and vertical taste parameters, an equilibrium exists in which sellers offer two-part tariffs where the marginal price for each product is equal to marginal cost and a seller offers a discount on its fixed charges when it supplies a consumer with both products. As before, the equilibrium is derived by supposing that sellers observe a consumer's vertical preferences but not her brand preferences, and

---

Matutes and Regibeau (1992), Anderson and Leruth (1993), and Thanassoulis (2007). Armstrong and Vickers (2010, Proposition 4) find conditions which ensure that when sellers engage in mixed bundling their profit is lower, and consumer surplus higher, relative to the situation with linear pricing.

<sup>54</sup>Early papers which study this comparison are Matutes and Regibeau (1988) and Economides (1989).

<sup>55</sup>Zhou (2015) suggests that this result only applies when there are few sellers. With more than two sellers, a firm is competing against the best offer from several rivals and so the relevant competitive margin is not at  $\Delta = 0$  as with duopoly but in the right-hand part of the distribution. While the average of i.i.d. variables is typically more concentrated about the mean, it is *less* concentrated in the tails, and as a result the per-product price can rise with bundling when there are more than two rivals.

showing that the resulting tariff does not depend on brand preferences. This framework allows for a more nuanced analysis of the welfare impact of nonlinear pricing, since relative to linear pricing marginal prices are lower (which is good for welfare) while the bundle discount induces excessive one-stop shopping (which harms welfare).

*Inter-firm bundling:* The previous discussion considered bundle discounts offered when consumers buy several products from the same multiproduct seller. We next focus on situations where nonlinear pricing is implemented across sellers, either unilaterally by a single seller or in a coordinated fashion.

A single-product seller might, if feasible, have an incentive to offer consumers a discount for its product if they also purchase a product supplied by another seller. For example, one magazine might offer an advertiser a discounted price to place an advert if that advertiser has already placed an advert in another magazine. To study this situation, suppose there are two products, labelled 1 and 2, each of which is supplied by one seller, and consumers have valuations  $(v_1, v_2, v_b)$  as described in section 3.1. If the price for product 2 is fixed at  $p_2$ , when seller 1 offers its product at price  $p_1$  but offers a discount  $\delta$  on its product if the consumer also purchases product 2, using the notation from section 3.1 its profit is

$$\pi = (p_1 - c_1)Q_1 - \delta Q_b .$$

In this scenario, the supplier of product 2 continues to receive its price  $p_2$  if the consumer buys the bundle, and so the bundle discount is funded entirely by seller 1. Similar to condition (20), one can show that it is profitable for the seller to introduce a unilateral discount of this form when  $N_b/N_1$  strictly decreases with  $p_1$ , so that bundle demand is more elastic than total demand for the seller’s product with respect to its price. In the “doubly independent” case where valuations are additive and stochastically independent, the seller has no incentive to offer a discount of this form, since whether or not a consumer has purchased product 2 has no bearing on her demand for product 1. However, if the two products are partial substitutes so that  $v_b < v_1 + v_2$  (which is plausibly the case in the advertising context, due to overlapping readership), or if valuations are additive but there is negative correlation between  $v_1$  and  $v_2$ , then the fact that a consumer has purchased product 2 is “bad news” for her propensity to buy product 1. In such cases the seller often has an incentive to offer a discount to these consumers, to reflect their more elastic demand.<sup>56</sup>

---

<sup>56</sup>For further details on this topic, see Schmalensee (1982), Lewbel (1985), and Armstrong (2013).

More common than this kind of unilateral bundling is for sellers to combine to form an alliance which coordinates on a joint bundling strategy. For instance, tourist attractions in a city might coordinate to offer a “tourist pass” at a discount over the sum of their individual entry fees, where they agree on the price for the pass and how the revenue from the pass is allocated between them. In roughly symmetric cases, it may be straightforward for sellers to agree on a joint bundling arrangement. For instance, they could coordinate on a bundle discount, which they fund equally, but remain free to determine their standalone prices.<sup>57</sup> Since bundle discounts can convert what would otherwise be substitute products into complements, rivals may be able to use a coordinated inter-firm bundling scheme as a means to relax competition. To illustrate, consider the example from section 3.1 where all consumers have the subadditive valuation vector  $(v_1, v_2, v_b) = (2, 2, 3)$ , and there is costless production and a separate seller for each product. If sellers compete using linear prices the equilibrium price is  $p_1 = p_2 = 1$  and consumers enjoy positive surplus. If instead sellers agree to offer a bundle discount  $\delta = 1$  which they fund equally, and choose their standalone prices non-cooperatively, the equilibrium price rise to  $p_1 = p_2 = 2$ , and all surplus is extracted in profit.<sup>58</sup>

## 5 Conclusions

This paper has surveyed a number of forms of nonlinear pricing and discussed the techniques used to derive and understand optimal tariffs. In the simplest scenario, where a single seller supplies a single product (or product line), one can make much progress by using market-level demand functions which depend only on linear prices, rather than the disaggregated parameterized utility functions often employed in textbook models of nonlinear pricing. In particular, the demand profile method, where each incremental item was treated as a separate product, was an elegant and economically intuitive way to solve

---

Lucarelli, Nicholson, and Song (2014) investigate a market for pharmaceutical cocktails, where drugs from separate companies are combined for medical treatment as well as used on a stand-alone basis. They present simulations based on their empirical estimates which indicate that a seller would set a higher price for its drug when combined with another seller’s drug, for the reason that the combined drugs are complementary products.

<sup>57</sup>See Gans and King (2006) for a model along these lines.

<sup>58</sup>See Armstrong (2013, section 5) for further discussion. Closely related issues arise in the context of patent pools, where separate patent holders coordinate on licensing arrangements for their collection of patents. Lerner and Tirole (2004) analyze this issue, and show that when patents are partial substitutes the coordinated scheme can relax competition relative to the situation where each patent holder licenses its patent separately.

many single-product nonlinear pricing problems. Even when this approach does not generate the optimal tariff itself, standard aggregate demand functions can be used to determine whether offering a quantity discount, or in the multiproduct context a bundle discount, was more profitable than linear pricing.

Outside the single-product, single-seller context, it is natural to suppose that consumers differ along several dimensions. They have taste parameters for each of a number of products, say, or in competitive situations they may differ both in their “horizontal” and “vertical” preferences. Here, an often useful trick for solving the nonlinear pricing problem was to suppose the seller could observe a subset of a consumer’s private information, in which case it can screen over the remaining dimension(s). With stochastic independence of the appropriate kind the seller’s tariff does not actually depend on the hypothetically observed parameters, in which case the optimal tariff when there is multi-dimensional private information has been found. Another useful trick was to consider markets with many products, when the law of large numbers can eliminate most of the seller’s uncertainty about a consumer’s willingness-to-pay for its collection of products.

In the interests of space and focus I have neglected several important aspects of nonlinear pricing. For instance, I have said little about dynamic nonlinear pricing. If competing suppliers sell over time to consumers who purchase repeatedly, a seller might choose its price contingent on whether a consumer is an existing customer or not. When they cannot commit to future prices a seller will often set a higher price to a past customer (which is a kind of quantity premium), and at the same attempt to “poach” its rival’s customers with a low price. The resulting high rates of switching between sellers can harm welfare, i.e., there is excessive “two-stop shopping”, in contrast to the static model in section 4.2 where there was too much one-stop shopping.<sup>59</sup>

Finally, and remaining with the dynamic theme, I have said nothing about the use of nonlinear tariffs to deter entry or induce exit by rivals. For instance, an incumbent manufacturer might offer its retailers a lower wholesale price if they do not stock a rival manufacturer’s products, a reversal of the inter-firm bundling arrangements discussed in section 4.2, which might harm a rival’s ability to compete. Alternatively, a multiproduct incumbent which faces a potential rival for one product may choose to engage in (pure) bundling in order to commit itself to compete fiercely should the rival decide to enter.<sup>60</sup> As

---

<sup>59</sup>For instance, see Chen (1997) and Fudenberg and Tirole (2000).

<sup>60</sup>For instance, see Whinston (1990).

we saw in section 4.2, when sellers compete on a margin where they win or lose two products rather than one they tend to compete aggressively, and for this reason the potential rival may decide not to enter if the incumbent bundles its products. The potentially anti-competitive impact of nonlinear pricing deserves a survey to itself.

## References

- ADAMS, W., AND J. YELLEN (1976): “Commodity Bundling and the Burden of Monopoly,” *Quarterly Journal of Economics*, 90(3), 475–498.
- ANDERSON, E., AND J. DANA (2009): “When is Price Discrimination Profitable?,” *Management Science*, 55(6), 980–989.
- ANDERSON, S., AND L. LERUTH (1993): “Why Firms May Prefer Not to Price Discriminate via Mixed Bundling,” *International Journal of Industrial Organization*, 11(1), 49–61.
- ARMSTRONG, M. (1996): “Multiproduct Nonlinear Pricing,” *Econometrica*, 64(1), 51–76.
- (1999): “Price Discrimination by a Many-Product Firm,” *Review of Economic Studies*, 66(1), 151–168.
- (2013): “A More General Theory of Commodity Bundling,” *Journal of Economic Theory*, 148(2), 448–472.
- ARMSTRONG, M., AND J.-C. ROCHET (1999): “Multi-Dimensional Screening: A User’s Guide,” *European Economic Review*, 43(4), 959–979.
- ARMSTRONG, M., AND J. VICKERS (2001): “Competitive Price Discrimination,” *Rand Journal of Economics*, 32(4), 579–605.
- (2010): “Competitive Non-linear Pricing and Bundling,” *Review of Economic Studies*, 77(1), 30–60.
- ATTANASIO, O., AND E. PASTORINO (2015): “Nonlinear Pricing in Village Economies,” mimeo.
- BAKOS, Y., AND E. BRYNJOLFSSON (1999): “Bundling Information Goods: Pricing, Profits, and Efficiency,” *Management Science*, 45(12), 1613–1630.

- (2000): “Bundling and Competition on the Internet,” *Marketing Science*, 19(1), 63–82.
- BARELLI, P., S. BASOV, M. BUGARIN, AND I. KING (2014): “On the Optimality of Exclusion in Multi-Dimensional Screening,” *Journal of Mathematical Economics*, 54, 74–83.
- BONATTI, A. (2011): “Brand-Specific Tastes for Quality,” *International Journal of Industrial Organization*, 29(5), 562–575.
- CHAMPSAUR, P., AND J.-C. ROCHET (1989): “Multiproduct Duopolists,” *Econometrica*, 57(3), 533–557.
- CHEN, Y. (1997): “Paying Customers to Switch,” *Journal of Economics and Management Strategy*, 6(4), 877–897.
- CHEN, Y., AND M. RIORDAN (2013): “Profitability of Product Bundling,” *International Economic Review*, 54(1), 35–57.
- CHU, C. S., P. LESLIE, AND A. SORENSEN (2011): “Bundle-Size Pricing as an Approximation to Mixed Bundling,” *American Economic Review*, 101(1), 263–303.
- CLERIDES, S., AND P. COURTY (2015): “Sales, Quantity Surcharge, and Consumer Inattention,” *Review of Economics and Statistics*, forthcoming.
- DELLAVIGNA, S., AND U. MALMENDIER (2004): “Contract Design and Self-Control: Theory and Evidence,” *Quarterly Journal of Economics*, 119, 353–402.
- (2006): “Paying Not to go to the Gym,” *American Economic Review*, 96, 694–719.
- DENECKERE, R., AND P. MCAFEE (1996): “Damaged Goods,” *Journal of Economics and Management Science*, 5(2), 149–174.
- DENECKERE, R., AND S. SEVERINOV (2014): “Multi-dimensional Screening: A Solution to a Class of Cases,” mimeo.
- ECONOMIDES, N. (1989): “The Desirability of Compatibility in the Absence of Network Externalities,” *American Economic Review*, 71(5), 1165–1181.

- ELIAZ, K., AND R. SPIEGLER (2008): “Consumer Optimism and Price Discrimination,” *Theoretical Economics*, 3(4), 459–497.
- ELLISON, G. (2005): “A Model of Add-on Pricing,” *Quarterly Journal of Economics*, 120(2), 585–637.
- FUDENBERG, D., AND J. TIROLE (2000): “Customer Poaching and Brand Switching,” *Rand Journal of Economics*, 31(4), 634–657.
- GABAIX, X., AND D. LAIBSON (2006): “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets,” *Quarterly Journal of Economics*, 121(2), 505–540.
- GANS, J., AND S. KING (2006): “Paying for Loyalty: Product Bundling in Oligopoly,” *Journal of Industrial Economics*, 54(1), 43–62.
- GENG, X., M. STINCHCOMBE, AND A. WHINSTON (2005): “Bundling Information Goods of Decreasing Value,” *Management Science*, 51(4), 662–667.
- GOLDMAN, B., H. LELAND, AND D. SIBLEY (1984): “Optimal Nonuniform Prices,” *Review of Economic Studies*, 51(2), 305–319.
- GRUBB, M. (2009): “Selling to Overconfident Consumers,” *American Economic Review*, 99(5), 1770–1807.
- (2015): “Overconfident Consumers in the Marketplace,” *Journal of Economic Perspectives*, 29(4), 9–35.
- HART, S., AND N. NISAN (2014): “How Good are Simple Mechanisms for Selling Multiple Goods?,” mimeo.
- ITOH, M. (1983): “Monopoly, Product Differentiation and Economic Welfare,” *Journal of Economic Theory*, 31(1), 88–104.
- JAHEDI, S. (2013): “A Taste for Bargains,” mimeo, University of Arkansas.
- JOHNSON, J., AND D. MYATT (2003): “Multiproduct Quality Competition: Fighting Brands and Product Line Pruning,” *American Economic Review*, 93(3), 748–774.

- (2006a): “Multiproduct Cournot Oligopoly,” *Rand Journal of Economics*, 37(3), 583–601.
- (2006b): “On the Simple Economics of Advertising, Marketing and Product Design,” *American Economic Review*, 96(3), 756–784.
- (2015): “The Properties of Product Line Prices,” *International Journal of Industrial Organization*, forthcoming.
- KOSZEGI, B. (2014): “Behavioral Contract Theory,” *Journal of Economic Literature*, 52(4), 1075–1118.
- LAFFONT, J.-J., E. MASKIN, AND J.-C. ROCHET (1987): “Optimal Nonlinear pricing with Two-Dimensional Characteristics,” in *Information, Incentives and Economic Mechanisms*, ed. by T. Groves, and S. Reiter, pp. 256–266. University of Minnesota Press, Minneapolis, MN.
- LERNER, J., AND J. TIROLE (2004): “Efficient Patent Pools,” *American Economic Review*, 94(1), 691–711.
- LEWBEL, A. (1985): “Bundling of Substitutes or Complements,” *International Journal of Industrial Organization*, 3(1), 101–107.
- LEWIS, T., AND D. SAPPINGTON (1989): “Countervailing Incentives in Agency Problems,” *Journal of Economic Theory*, 49(2), 294–313.
- LONG, J. (1984): “Comments on ‘Gaussian Demand and Commodity Bundling’,” *Journal of Business*, 57(1), S235–S246.
- LUCARELLI, C., S. NICHOLSON, AND M. SONG (2014): “Bundling Among Rivals: A Case of Pharmaceutical Cocktails,” mimeo.
- MANELLI, A., AND D. VINCENT (2006): “Bundling as an Optimal Selling Mechanism for a Multiple-good Monopolist,” *Journal of Economic Theory*, 127(1), 1–35.
- (2007): “Multidimensional Mechanism Design: Revenue Maximization and the Multiple-good Monopoly,” *Journal of Economic Theory*, 137(1), 153–185.

- MASKIN, E., AND J. RILEY (1984): “Monopoly with Incomplete Information,” *Rand Journal of Economics*, 15(2), 171–196.
- MATUTES, C., AND P. REGIBEAU (1988): “Mix and Match: Product Compatibility without Network Externalities,” *Rand Journal of Economics*, 19(2), 221–234.
- (1992): “Compatibility and Bundling of Complementary Goods in a Duopoly,” *Journal of Industrial Economics*, 40(1), 37–54.
- MCAFEE, R. P., J. MCMILLAN, AND M. WHINSTON (1989): “Multiproduct Monopoly, Commodity Bundling and Correlation of Values,” *Quarterly Journal of Economics*, 104(2), 371–384.
- MCMANUS, B. (2007): “Pricing in an Oligopoly Market: The Case of Speciality Coffee,” *Rand Journal of Economics*, 38(2), 512–532.
- MENICUCCI, D., S. HURKENS, AND D.-S. JEON (2015): “On the Optimality of Pure Bundling for a Monopolist,” *Journal of Mathematical Economics*, forthcoming.
- MIRRLEES, J. (1971): “An Exploration in the Theory of Optimum Taxation,” *Review of Economic Studies*, 38, 175–208.
- MUSSA, M., AND S. ROSEN (1978): “Monopoly and Product Quality,” *Journal of Economic Theory*, 18(2), 301–317.
- MYERSON, R. (1981): “Optimal Auction Design,” *Mathematics of Operations Research*, 6, 58–73.
- PAVLOV, G. (2011): “Optimal Mechanism for Selling Two Goods,” *The B.E. Journal of Theoretical Economics*, 11(1, Advances), Article 3.
- RILEY, J., AND R. ZECKHAUSER (1983): “Optimal Selling Strategies: When to Haggle, When to Hold Firm,” *Quarterly Journal of Economics*, 98(2), 267–289.
- ROCHET, J.-C., AND P. CHONÉ (1998): “Ironing, Sweeping, and Multidimensional Screening,” *Econometrica*, 66(4), 783–826.
- ROCHET, J.-C., AND L. STOLE (2002): “Nonlinear Pricing with Random Participation,” *Review of Economic Studies*, 69(1), 277–311.

- (2003): “The Economics of Multidimensional Screening,” in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, ed. by M. Dewatripont, L. P. Hansen, and S. Turnovsky. Cambridge University Press, Cambridge.
- SALANT, S. (1989): “When is Inducing Self-selection Suboptimal for a Monopolist?,” *Quarterly Journal of Economics*, 104(2), 391–397.
- SCHMALENSEE, R. (1982): “Commodity Bundling by Single-Product Monopolies,” *Journal of Law and Economics*, 25(1), 67–71.
- SHI, T. (2016): “Nearly Optimal Pricing Using Shipping Fees,” mimeo.
- SHILLER, B., AND J. WALDFOGEL (2011): “Music for a Song: An Empirical Look at Uniform Song Pricing and its Alternatives,” *Journal of Industrial Economics*, 59(4), 630–660.
- SPENCE, M. (1975): “Monopoly, Quality, and Regulation,” *Bell Journal of Economics*, 6(2), 417–429.
- (1977): “Nonlinear Pricing and Welfare,” *Journal of Public Economics*, 8(1), 1–18.
- SPIEGLER, R. (2011): *Bounded Rationality and Industrial Organization*. Oxford University Press, Oxford, UK.
- STIGLER, G. (1963): “United States v. Loews Inc.: A Note on Block Booking,” *Supreme Court Review*, pp. 152–157.
- STOKEY, N. (1979): “Intertemporal Price Discrimination,” *Quarterly Journal of Economics*, 93(3), 355–371.
- THANASSOULIS, J. (2007): “Competitive Mixed Bundling and Consumer Surplus,” *Journal of Economics and Management Strategy*, 16(2), 437–467.
- VERBOVEN, F. (1999): “Product Line Rivalry and Market Segmentation - With an Application to Automobile Optional Engine Pricing,” *Journal of Industrial Economics*, 47(4), 399–425.

- (2002): “Testing for ‘Monopoly’ Market Power when Products are Differentiated in Qualities,” *Manchester School*, 70(1), 115–133.
- WHINSTON, M. (1990): “Tying, Foreclosure and Exclusion,” *American Economic Review*, 80(4), 837–859.
- WILSON, R. (1993): *Nonlinear Pricing*. Oxford University Press, New York.
- YANG, H., AND L. YE (2008): “Nonlinear Pricing, Market Coverage, and Competition,” *Theoretical Economics*, 3(1), 123–153.
- ZHOU, J. (2015): “Competitive Bundling,” mimeo.