



Munich Personal RePEc Archive

## **Nonlinear Pricing**

Armstrong, Mark

Department of Economics, University of Oxford

July 2015

Online at <https://mpra.ub.uni-muenchen.de/65779/>  
MPRA Paper No. 65779, posted 28 Jul 2015 20:11 UTC

# Nonlinear Pricing

Mark Armstrong\*

July 2015

## 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.

## 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 stroll 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. A newspaper consists of a bundle of distinct articles, and may itself be cheaper with a subscription than when casually purchased from a store. You may get a discount on a car rental when you book a flight, and the flight may be cheaper if you’ve 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.

---

\*Department of Economics and All Souls College, University of Oxford. I am grateful to David Myatt, David Sappington, John Vickers, Dan Vincent, and Jidong Zhou for helpful comments on this draft.

It is useful to distinguish various cases along two dimensions: whether there is a single product or multiple products, and whether there is a single seller or multiple sellers. Table 1 presents the situations 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 add-on pricing [section 4.1]	mix-and-match shopping inter-firm bundling [section 4.2]

Table 1: The plan of the paper

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 distinct products. In economic terms, though, the crucial distinction between the two scenarios is that with a single product a consumer’s utility depends only on the *number* of units they buy, while with multiple products she cares about *which* units they buy.

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 (at price  $p_1$ , say) followed by the “second” unit (at 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 order 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  $n^{\text{th}}$  unit in the consumer population, one can

calculate the most profitable price for this unit in isolation. This approach does not always generate the optimal tariff—some 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>1</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 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 cannot buy the basic product from one seller and the upgrade (or similar) from another.

The multiproduct pricing problem has a different structure, without this natural order over consumer decisions. With unit demand for two underlying products, 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 does not usually seem to be useful—or even well defined—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 usually 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. When competing sellers offer bundle discounts, however, artificial shopping costs are introduced which encourages more consumers to become one-stop shoppers.

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 optimal to introduce a bundle discount. A related insight is that bundling acts to reduce the seller’s uncertainty about a consumer’s surplus, just as a broad investment portfolio allows for more predictable returns. (This effect is most marked when *many* items are in the bundle, when the law of large numbers

---

<sup>1</sup>Wilson (1993) demonstrates the widespread usefulness of the demand profile approach.

can apply.) With a single seller, the decrease in consumer heterogeneity due to bundling often leads to higher profits and reduced consumer surplus. With multiple sellers, though, it can sharpen competition, so sellers are harmed and consumers benefit from this form of price discrimination.

## 2 Single-Seller, Single-Product Analysis

### 2.1 Nonlinear pricing with discrete choices

It is useful initially to discuss nonlinear (and related) pricing in a setting in which consumers have a binary choice from the monopoly seller, and can either buy a “basic” or a “premium” product. (Analysis with continuous choices is presented in the next section.) Here, the premium product might consist of greater quantity of the underlying product, relative to the basic product. That is, the seller supplies either a “small” or a “large” quantity, as with a café which sells coffee in two sizes. Alternatively, the premium product is of higher quality than the basic product (such as a car with a better engine), or includes an “add-on” product alongside the basic product (such as an enhanced warranty to go with a new television).

Consider the following discrete choice framework. A consumer is willing to pay  $v_1$  for the basic product and an extra amount  $v_2$  for the premium product relative to the basic product. The seller charges price  $p_1$  for the basic product and a total price  $p_1 + p_2$  for the premium product, so that  $p_2$  is the incremental price for the “upgrade”.<sup>2</sup> Likewise, let  $c_1$  be the seller’s cost of supplying the basic product, and  $c_2$  be its incremental cost of supplying an upgrade (so its total 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>3</sup>

The type- $(v_1, v_2)$  consumer buys the premium product if her surplus from doing so,  $v_1 + v_2 - (p_1 + p_2)$ , is above her surplus from the basic product,  $v_1 - p_1$  and from not buying anything (which is zero). She buys the basic product if the corresponding pair of inequalities is satisfied. The three cases partition the set of consumer types as shown

---

<sup>2</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>3</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).

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—either the basic or the premium product—when the tariff is  $(p_1, p_2)$ , and let  $Q_2(p_1, p_2)$  be the fraction who buy the premium product. In smooth cases we have Slutsky symmetry of demand functions, so that  $\partial Q_1/\partial p_2 \equiv \partial Q_2/\partial p_1 \leq 0$ . (Roughly speaking, this negative cross-price effect is equal to the measure of consumers on the diagonal line on Figure 1.) When these cross-price effects are negative the basic product and the upgrade are complementary products. The seller’s profit with tariff  $(p_1, p_2)$  is

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

For instance, suppose that  $(v_1, v_2)$  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  $(p_1, p_2) = (\frac{2}{3}, \frac{1}{6})$ .

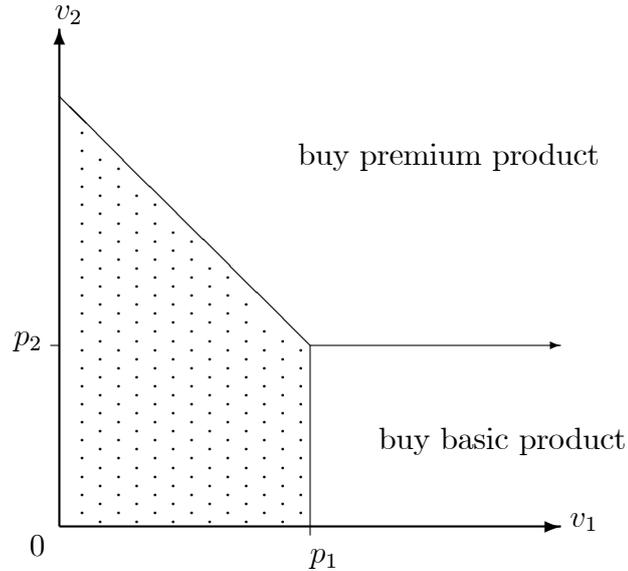


Figure 1: Pattern of demand with tariff  $(p_1, p_2)$

For  $i = 1, 2$  write  $N_i(p_i) \equiv \Pr\{v_i \geq p_i\}$ . Thus,  $N_1(p_1)$  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 given they have the basic product. From Figure 1 we see that  $N_1(p_1) \leq Q_1(p_1, p_2)$  and  $N_2(p_2) \geq Q_2(p_1, p_2)$ . 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 more general context of upgrades and add-ons.

It would be useful if the seller could calculate its optimal tariff merely from knowing its demand profile  $\{N_1(p_1), N_2(p_2)\}$  instead of the unwieldy and hard-to-estimate demand functions  $\{Q_1(p_1, p_2), Q_2(p_1, p_2)\}$ . That is, when can the seller treat consumer demand for the basic product and for the upgrade as separate, rather than complementary, markets? To understand this, suppose that the distribution of tastes is such that no consumers have taste vectors  $(v_1, v_2)$  to the “north-west” of the price vector  $(p_1, p_2)$  on Figure 1. This implies that all consumers buy as if they were myopic: they buy the basic product whenever  $v_1 \geq p_1$  and go on to buy the upgrade if  $v_2 \geq p_2$ . In this situation we have  $N_i(p_i) = Q_i(p_1, p_2)$ , there are no cross-price effects in consumer demand for the basic product and the upgrade, and the two products can (locally) be considered as separate markets. Therefore, 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 myopically—i.e.,  $v_2 \geq p_2$  implies  $v_1 \geq p_1$ —this tariff satisfies the first-order conditions for maximizing true profit (1).

The optimal tariff in the previous “uniform square” example cannot be derived using this demand profile method. (There, 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 studies of second-degree price discrimination assume there is scalar consumer heterogeneity, in the sense that  $v_2 \equiv kv_1$  and valuation vectors on Figure 1 all lie on a ray of slope  $k$  from the origin.<sup>4</sup> Following Johnson and Myatt (2003), suppose that the CDF for  $v_1$  is  $F(v_1)$ , which we assume to have an increasing hazard rate so that the profit functions in (2) are single-peaked. The demand profile approach 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  $c_2 \geq kc_1$ , so that cost of supply rises proportionately more steeply than consumer utility (i.e., there are “decreasing returns” to quality in the sense of Johnson and Myatt, 2003), this procedure implies that  $p_2 \geq kp_1$ . This in turn implies that all consumers purchase myopically and the method yields the optimal tariff. By contrast, if there are increasing returns to supplying

---

<sup>4</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 just shifts a consumer’s valuation vector equiproportionately. Similar analysis applies if the valuation vectors lie on a monotonically increasing curve, rather than a ray, as discussed more systematically in the next section.

quality, so that  $c_2 < kc_1$ , it is optimal to supply *only* the premium product to consumers. (The demand profile method does not always work in this case.)

Whenever the demand profile approach can be used, the second-degree pricing problem coincides with the more straightforward theory of single-product monopoly pricing, and the optimal tariff is determined by standard demand elasticities. The Lerner index for the basic product,  $(p_1 - c_1)/p_1$ , is equal to the reciprocal of the elasticity of demand for the basic product,  $N_1(\cdot)$ , while the Lerner index for the upgrade is equal to the reciprocal of the elasticity of its demand,  $N_2(\cdot)$ . It also follows that the upgrade price,  $p_2$ , is above its cost,  $c_2$ .<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  $p_2/c_2 < p_1/c_1$ , which is often the case. (For instance, this is true in the previous Mussa-Rosen example with  $c_2 > kc_1$ , and also in the next model with continuous choices we see that the optimal nonlinear tariff is often concave in quantity.)

An important question in the context of product lines is when it is optimal for the seller to supply just a single product or a menu of price/quality options.<sup>6</sup> Deneckere and McAfee (1996) provide analysis and several instances of “damaged goods”, where the firm deliberately introduces an inferior variant of its product (such as slower computer chip), which might even cost more than its standard product. (Having the inferior version cost more corresponds to  $c_2 < 0$  in the current notation.) As discussed, this situation cannot occur with “Mussa-Rosen” preferences. More generally, though, it is straightforward to find situations where it is profitable for the firm to introduce an inferior, but more costly, product variant. 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.

*Nonlinear pricing:* Further issues arise in the more specific context of nonlinear pricing. Suppose that the seller incurs the constant unit cost  $c$  to supply a unit of its product to

---

<sup>5</sup>This is not necessarily the case more generally, when, as discussed,  $Q_1$  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>For further discussion of this point, see Myatt and Johnson (2003, section III.B) and Anderson and Dana (2009).

a consumer. A consumer is willing to pay  $v_1$  for a single unit the product and  $v_2$  for a second unit. It is natural to assume diminishing returns, so that  $v_2 \leq v_1$  for all consumers. (By contrast, in the more general formulation, a buyer might put little value on the basic product but a lot on the premium variant.) 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>7</sup>

Given there are diminishing returns, the demand profile,  $N_i(p) = \Pr\{v_i \geq p\}$ , coincides with demand with the linear tariff  $p_1 = p_2 = p$ , which is  $Q_i(p, p)$ . Even when the demand profile approach cannot be used to generate the optimal nonlinear tariff, the demand profile can always be used to determine when it is profitable for the seller to offer a quantity discount. 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 introduces a small discount for buying the second unit is

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

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  $N_1$ . Therefore, the seller obtains greater profit by using 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>8</sup> The seller can determine whether a quantity discount will boost its profit with knowledge only of its demand at linear prices. (However, it does need to know its average demand per consumer, not merely aggregate demand  $N_1 + N_2$ .)

To illustrate the nonlinear pricing problem with an example, suppose that  $(v_1, v_2)$  is uniformly distributed with density 2 on the 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

<sup>7</sup>This discrete choice formulation of the nonlinear pricing problem is mentioned in Adams and Yellen (1976, pp. 488-489) and in Wilson (1993, section 4.3).

<sup>8</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).

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

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

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 rises.

As usual, the calculation of the optimal nonlinear tariff is simplified when the demand profile approach is valid, so that  $p_1$  maximizes  $(p - c)N_1(p)$  and  $p_2$  maximizes  $(p - c)N_2(p)$ . In this case, the (second-degree) nonlinear pricing problem becomes an instance of third-degree price discrimination. For instance, if the demand profile  $\{N_1, N_2\}$  happened to consist of linear demand functions, it is well known that price discrimination causes total output to remain unchanged, so that total welfare falls. The exception to this is when price discrimination enables a market to “open up”, in which case a Pareto improvement results.<sup>9</sup>

*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

---

<sup>9</sup>Following Varian (1985), welfare can only increase with nonlinear pricing relative to linear pricing if total output  $Q_1 + Q_2$  rises. Suppose that  $v_2 = kv_1$  for  $k \leq 1$ ,  $c = 0$  and  $v_1$  is uniformly distributed on  $[0, 1]$ . Then the optimal nonlinear tariff is  $p_1 = \frac{1}{2}$  and  $p_2 = \frac{1}{2}k$ . If  $k \leq \frac{1}{3}$  then with linear pricing it is more profitable to serve only the one-unit demand, and to set  $p_1 = p_2 = \frac{1}{2}$ . A move to nonlinear pricing then opens up the market for two units, resulting in a Pareto improvement.

<sup>10</sup>The following discussion is essentially taken from the illustrative example in Eliaz and Spiegel (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).

the contract before the latter knows whether she wants one or two units. (This is a stylized representation of a gym contract, say, where the consumer signs a contract which specifies a monthly fee and/or a per-visit charge.)

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  $(p_1, p_2)$  is

$$v - p_1 + \alpha(v - p_2) , \quad (5)$$

while the seller's anticipated profit is

$$p_1 - c + \beta(p_2 - c) . \quad (6)$$

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

$$p_1 = (1 + \alpha)v ; p_2 = 0 . \quad (7)$$

(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 per visit.

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 will need 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 this, 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

---

<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 the model with competition.

$(p_0, p_1, p_2)$ , 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

$$p_0 = v ; p_2 = 0 ; p_3 = v . \tag{10}$$

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 rentals, where the consumer pays a fixed charge for a specified quantity (or less), and pays a steep “overage” charge when she goes beyond her contracted limit. In 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, and at the time of actual consumption their preferences differ from those at the time they choose their contract. 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. A (naive) person who is initially unaware that she has weak will will behave in a similar manner to an overoptimistic consumer above, in that they put too much weight on the event they will have large demand. Opposite effects apply when a consumer is tempted to purchase too much of a bad product: a sophisticated shopper who

---

<sup>12</sup>The extreme assumption that the consumer foresees no uncertainty in her demand is not important here, and the some 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.

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 (2011) 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}.

## 2.2 Nonlinear pricing with continuous choices

In contrast to the discussion in 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 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 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 two different consumers do not cross.<sup>14</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 maximum surplus

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

Regardless of the shape of  $T(\cdot)$ , consumers with higher  $\theta$  will choose a weakly greater quantity. (And conversely, any weakly increasing function  $q(\theta)$  can be implemented by a suitable nonlinear tariff.) The taste parameter  $\theta$  is distributed in the consumer population

---

<sup>13</sup>A key early paper in this literature is DellaVigna and Malmendier (2004). See Spiegel (2011, chapters 2–4) for an overview of contract design with dynamically inconsistent consumers, and Koszegi (2015) for a survey of contract design in the presence of biased agents.

<sup>14</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) below) is due to Mirrlees (1971). A useful overview of this analysis is presented in Rochet and Stole (2003, section 2).

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-type 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 second equality 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  $s(\theta_1) = 0$ .

The candidate solution to the seller's profit 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 (14) has increasing differences in  $(q, \theta)$ , it yields the seller's most profitable strategy. If the seller could observe a consumer's type directly, it would offer the type- $\theta$  consumer a 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 any such consumer exists), and provided that  $(1 - F(\theta))/f(\theta) \rightarrow 0$  as  $\theta \rightarrow \theta_2$  the type- $\theta_2$  consumer is served with the efficient quantity and there is "no distortion at the top".

Consider the "Mussa-Rosen" specification  $u(q, \theta) = \theta u(q)$ , so that  $\theta$  shifts utility multiplicatively. In this case,  $q(\theta)$  in (14) is increasing if

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

With costless production ( $c = 0$ ), expression (14) implies that consumers are either served with the maximum quantity (i.e., the quantity that maximizes  $u(\cdot)$ ) or nothing, so that the seller offers an "all you can eat" contract. As seen in section 2.1, this is not necessarily true outside this specification with scalar heterogeneity.

Since the type- $\theta$  consumer chooses her quantity to maximize  $\theta u(q) - T(q)$ , the incremental price of the  $q$ th unit, denoted  $p(q) \equiv T'(q)$ , satisfies  $p(q(\theta)) = \theta u'(q(\theta))$ . The first-order condition in (14) then implies that the Lerner index satisfies

$$\frac{p(q(\theta)) - c}{p(q(\theta))} = \frac{1 - F(\theta)}{\theta f(\theta)}. \quad (16)$$

If the right-hand side of (16) decreases with  $\theta$ , which is a stronger condition than (15), we deduce that marginal price decreases with quantity.<sup>15</sup> Similarly, if the right-hand side of (16) is constant, i.e., when  $\theta$  comes from a Pareto distribution, marginal price does not depend on quantity and a simple two-part tariff is the seller's optimal nonlinear tariff.

Since the marginal price satisfies  $p(q(\theta)) \equiv \theta u'(q(\theta))$ , the type- $\theta$  consumer's problem of choosing quantity to maximize  $\theta u(q) - T(q)$  is single-peaked in  $q$ .<sup>16</sup> Thus a consumer purchases myopically, and each infra-marginal unit generates positive surplus. As in section 2.1, since consumers purchase myopically we expect the demand profile approach to work.<sup>17</sup> As in section 2.1, let  $N_q(p)$  denote the fraction of consumers who purchase at least  $q$  units when faced with a linear price  $p$ . With the Mussa-Rosen specification, a consumer buys at least  $q$  units with linear price  $p$  if  $\theta u'(q) \geq p$ , and so

$$N_q(p) = 1 - F\left(\frac{p}{u'(q)}\right). \quad (17)$$

The formula for the marginal price in (16) corresponds to the incremental price  $p(q)$  being chosen for each  $q$  so that

$$p(q) \text{ maximizes } (p - c)N_q(p), \quad (18)$$

and the price for the  $q$ th unit simply maximizes the profit from selling this unit. Thus, the shape of the nonlinear tariff is governed by how the elasticity of the demand function  $N_q(p)$  varies with  $q$ .

The nonlinear pricing problem is especially neat when (15) is linear in  $\theta$ , so that  $\theta - (1 - F(\theta))/f(\theta) = b\theta - a$  for constants  $a$  and  $b > 0$ .<sup>18</sup> (A variety of familiar distributions fall into this category, including the uniform, Pareto and exponential.) In this case, (18) implies that  $bp(q) = c + au'(q)$ . Provided that it is optimal for some consumers to be excluded,

<sup>15</sup>See Maskin and Riley (1984, Proposition 6).

<sup>16</sup>The derivative of this objective, evaluated at  $q(\tilde{\theta})$ , is  $\theta u'(q(\tilde{\theta})) - p(q(\tilde{\theta})) = (\theta - \tilde{\theta})u'(q(\tilde{\theta}))$ . When  $q(\tilde{\theta})$  is increasing, this implies that  $\theta u(q) - T(q)$  is increasing for  $q < q(\tilde{\theta})$  and decreasing for  $q > q(\tilde{\theta})$ .

<sup>17</sup>For further discussion see Wilson (1993, sections 4.1 and 6.5) and Rochet and Choné (1998, section 2.3.1).

<sup>18</sup>See Itoh (1983), which is extended in Johnson and Myatt (2015) including to Cournot situations.

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 of these cases is that this  $T(q)$  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 buy 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 prices each bundle of units 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) simply by supposing that that was the only product it supplied.

Another situation where the solution is simple is the linear case where gross surplus for the type- $\theta$  consumer is  $\theta q$  and  $q$  is constrained to lie in the interval  $[0, 1]$ . This corresponds to the situation where a risk-neutral consumer only wants a single indivisible unit of the product, which she values at  $\theta$ , and where the seller considers offering its product stochastically, so that a consumer can obtain the product with probability  $q$  in return for payment  $T(q)$ . Then (14) 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>19</sup> As discussed in section 3.1, this result is specific to the single-product context.

While the demand profile approach—where the price for the  $q$ th unit is chosen to maximize profit in (18)—works elegantly in this parameterized specification in which consumer demand curves never cross, it applies to single-product situations far more widely. For example, consumer heterogeneity might be multi-dimensional rather than scalar, in which case we expect demand curves from different consumers sometimes to cross. (Demand curves might be linear, say, but consumers differ both in the slope and the intercept of demand.) 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>20</sup> This single-product analysis suggests that it is not so much the

---

<sup>19</sup>In fact 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 more details.

<sup>20</sup>Wilson (1993, section 8.4) and Rochet and Stole (2003, section 5) solve 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,

lack of a natural order on “types” *per se* which makes solving nonlinear pricing problems hard. Rather, as discussed in the introduction, with multiple products there is a lack of a natural in *decisions*, and this makes it hard to derive 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 a known 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>21</sup> As above, with 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

$$\int_{\theta_1}^{\theta_2} G(s(\theta), \theta) [u(q(\theta), \theta) - s(\theta) - cq(\theta)] f(\theta) d\theta . \quad (19)$$

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 (19) means we can no longer use the trick of integrating by parts to eliminate

---

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 total surplus is bounded.) One can then calculate  $N_q(p)$ , the fraction of consumers who buy at least  $q$  units with linear price  $p$ . With costless production one can check that the price which maximizes profits from the  $q$ th unit,  $pN_q(p)$ , is  $p(q) = \frac{1}{3}(1 - q)$ , and the corresponding nonlinear tariff is  $T(q) = \frac{1}{3}q - \frac{1}{6}q^2$ . One can also check that every consumer has sufficiently concave utility that they buy myopically when faced with this tariff.

<sup>21</sup>See also Yang and Ye (2008).

$s(\theta)$  as we did in expression (13), and instead one must often resort to using optimal control methods or similar.

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 consumers even if it knew they had strong preferences for its product.<sup>22</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  $S(\theta) = \max_q u(q, \theta) - cq$  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  $x = S(\theta) + \epsilon$ , where  $\epsilon$  is an additive shock to  $S(\theta)$  which is independent of  $\theta$ . (While this seems a very particular specification with monopoly supply, it naturally emerges in the context of duopoly in section 4.1.) I assume that  $x = S(\theta) + \epsilon$  is the consumer’s sole outside option even if this is negative. Suppose, hypothetically, the seller can observe a consumer’s parameter  $\theta$  (but not her  $\epsilon$ ). 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  $S(\theta) - P \geq x = S(\theta) + \epsilon$ , 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>23</sup>

---

<sup>22</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>23</sup>Spence (1975) makes this point in the context of quality choice.

### 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 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. Thus a consumer's preferences are described by the vector  $(v_1, v_2, v_b)$ . 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 great majority of the articles written on bundling as a form of price discrimination assume valuations are additive, as much for tractability as for realism.

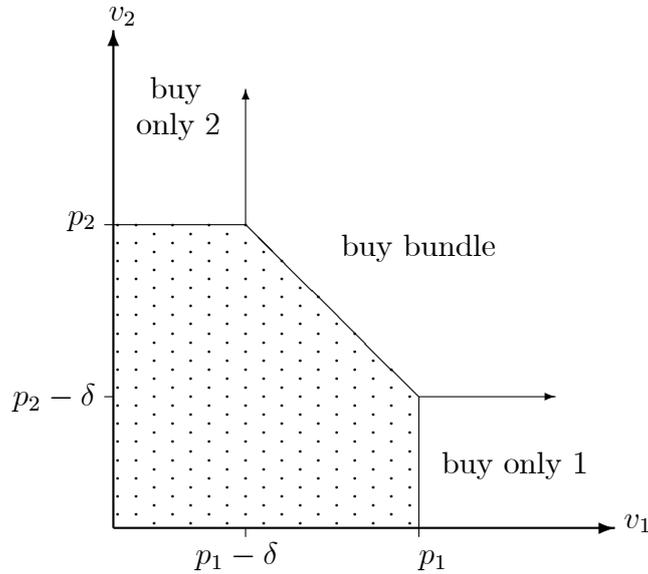


Figure 2: Pattern of demand with additive valuations and tariff  $(p_1, p_2, \delta)$

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 2 shows the pattern of demand in the case of additive valuations ( $v_b \equiv v_1 + v_2$ ) with a positive bundle discount.

As functions of the tariff parameters  $(p_1, p_2, \delta)$ , denote by  $Q_i$  the proportion of consumers who buy product  $i = 1, 2$  (either on its own or as part of the bundle), and denote by  $Q_b$  the proportion who buy both products. As can be checked from Figure 2 in the special case with additive valuations, in general we have Slutsky symmetry of cross-price effects, so that  $\partial Q_i / \partial \delta \equiv -\partial Q_b / \partial p_i$  for  $i = 1, 2$ . As before, it is useful to define 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.<sup>24</sup> 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>25</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

$$\pi = (p_1 - c_1)Q_1 + (p_2 - c_2)Q_2 - \delta Q_b . \quad (20)$$

One can calculate the most profitable linear prices, denoted  $(\hat{p}_1, \hat{p}_2)$  say, and then derive the impact on profit (20) 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

$$\left. \frac{d N_b(\hat{p}_1 + t[\hat{p}_1 - c_1], \hat{p}_2 + t[\hat{p}_2 - c_2])}{dt N_i(\hat{p}_1 + t[\hat{p}_1 - c_1], \hat{p}_2 + t[\hat{p}_2 - c_2])} \right|_{t=0} < 0 , \quad (21)$$

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 bundling is profitable the seller needs knowledge only of its demand system with linear prices.

---

<sup>24</sup>From Figure 2, which involves additive valuations, when  $\delta > 0$  an increase in  $p_1$  reduces the demand for product 2.

<sup>25</sup>For details see Long (1984) and Armstrong (2013).

Consider the case where consumer valuations are additive, i.e.,  $v_b \equiv v_1 + v_2$ , in which case  $N_1$  is a function only of  $p_1$  and  $N_2$  only depends on  $p_2$ . 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 (21) 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>26</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 offers the same price for either product on its own. This example is essentially the same as the “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  $\frac{2}{3}$  for either single product and the incremental price  $\frac{2}{3} - \frac{1}{3}\sqrt{2}$  for the second item, so that the bundle discount is  $\delta = \sqrt{2}/3 \approx 0.47$ .<sup>27</sup> For comparison, the most profitable linear prices are  $p_1 = p_2 = \frac{1}{2}$ , and so the use of bundling causes the price for a single product to rise, but the price for the bundle to fall.

When valuations are additive, the revenue from selling two products using a bundling tariff is more 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 bundling is profitable. Negative correlation puts less weight in the “north-east” part of the distribution on Figure

---

<sup>26</sup>Long (1984) and McAfee, McMillan, and Whinston (1989) independently derived this result. Menicucci, Hurkens, and Jeon (2014) 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 exceed cost for each product.

<sup>27</sup>Manelli and Vincent (2006, Theorem 5) show that the optimal tariff in the corresponding 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. Unfortunately, in Armstrong (2010, section III.A) I tried to solve the same example (in ignorance of the earlier paper) and did so incorrectly.

2, which tends to make bundle demand  $N_b$  more elastic.<sup>28</sup> Likewise, making products more substitutable also reduces the scale of bundle demand, which will often increase its elasticity.<sup>29</sup> Subadditive preferences give a direct reason, in addition to screening motives, to engage in bundling. For example if the 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 with a bundling tariff.

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 lay 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 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 made worse off when a second product is introduced, since that allows 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, when the number of products is relatively *large*. (Examples where buyers have demand for many products include online retailing of music 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 marginal cost of supply  $c$ . 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 where she can purchase any product for price  $p = c$  by paying a fixed charge  $P$  equal to her resulting surplus, which is

$$P = \sum_{i=1}^n \max\{v_i - c, 0\} . \quad (22)$$

---

<sup>28</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>29</sup>See the discussion in Armstrong (2013, pages 457-8).

The 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 (22) 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 ability 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 gains from trade.) When  $c = 0$ , as with information goods such as music downloads or electronic journals, the seller simply offers the grand bundle of all products in return for a fixed charge.<sup>30</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. (Some consumers like music more than others, for instance.) 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 approximated by the underlying distribution,  $G(v, \theta)$ . Since products are symmetric, suppose the seller chooses a tariff which depends only on the number of products purchased. Since there are many

---

<sup>30</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.

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

$$u(q, \theta) = \int_{V(q, \theta)}^{\infty} vg(v, \theta)dv , \quad (23)$$

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 (23) satisfies the requirements for the classical analysis of single-product nonlinear pricing presented in section 2.2, and the seller's optimal strategy is given in expression (14).<sup>31</sup> In particular, with information goods ( $c = 0$ ) only the grand bundle is offered, but now the bundle price may exclude some fraction of consumers. In sum, when consumers have unit demands for many products, an approximately optimal bundling tariff can often be calculated using the standard approach to single-product nonlinear pricing.

This discussion of situations with many products is an instance of an approach to price discrimination which tries to approximate the optimal tariff by means of a simple tariff, or to gauge the loss incurred when simple tariffs are used. For instance, Wilson (1993, section 8.3) argues in the context of nonlinear pricing that the loss of profit involved in using a nonlinear tariff consisting of  $N$  linear segments relative to the optimal tariff is of order  $1/N^2$ . In the bundling context, Hart and Nisan (2014) show with two products and consumer valuations which are identically and independently distributed, selling the items separately (i.e., with a linear tariff) can achieve at least 73% of the possible revenue from any mechanism (including the stochastic schemes discussed shortly). 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 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

---

<sup>31</sup>A version of this discussion was contained in Armstrong (1999, section 3).

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.

*Stochastic selling schemes:* So far in this section I have only 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, in which 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 3 has  $(v_1, v_2) = (9, 9)$ . Here, 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 incentive compatibility constraints are slack). However, the seller can extract more revenue by offering consumer  $i = 1, 2$  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 (precisely, 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 “small” surplus-increasing changes which do not interact with the slack incentive compatibility 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 which 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 can be 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 employs nonlinear pricing or bundling, 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 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- $w$  consumer given by

$$w_1 U\left(\frac{q_1}{w_1}\right) + w_2 U\left(\frac{q_2}{w_2}\right), \quad (24)$$

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

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

Here,  $\theta$  represents the average scale of the consumer's demands and  $r$  represents her relative taste for product 1 compared to product 2. A property of the utility specification (24) 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, one can check that

$$\max \left\{ w_1 U\left(\frac{q_1}{w_1}\right) + w_2 U\left(\frac{q_2}{w_2}\right) \text{ subject to } q_1 + q_2 = q \right\} = \theta U\left(\frac{q}{\theta}\right), \quad (25)$$

where the individual quantities which solve this problem are  $q_1 = rq$  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 total quantity, 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  $T(q, \omega)$ . The function  $u(q, \theta) \equiv \theta U(q/\theta)$  in expression (25) has increasing differences in  $(q, \theta)$ , and we can therefore apply the 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

$$u(q, \theta) - \frac{1 - F(\theta, r)}{f(\theta, r)} u_\theta(q, \theta) - cq . \quad (26)$$

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 (26) 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 (26) represents the optimal strategy even when the seller cannot observe  $r$ . In such cases, we have derived the optimal multiproduct nonlinear tariff, which depends only on the total quantity demanded by the consumer.

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 the consumer's scale parameter  $\theta$ , which gives a consumer an incentive to mis-report her parameter  $r$  as well as her  $\theta$ . I do not know how to solve any such cases analytically.<sup>32</sup>

To illustrate the method, suppose that  $w_1$  and  $w_2$  are independent exponential variables with parameter  $\lambda$ . A property of i.i.d. exponential variables is that their sum and ratio

---

<sup>32</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.

are also independent random variables, where  $\theta \equiv w_1 + w_2$  is a Gamma variable with CDF  $F(\theta) = 1 - (1 + \lambda\theta)e^{-\lambda\theta}$  and density  $f(\theta) = \lambda^2\theta e^{-\lambda\theta}$  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 production is costless,  $\lambda = 1$ , 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 (26) 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 allocation of product  $i = 1, 2$  to the type- $w$  consumer is

$$q_i(w) = w_i \frac{(w_1 + w_2)^2}{1 + (w_1 + w_2) + (w_1 + w_2)^2}. \quad (27)$$

The quantities the seller would allocate if it knew each consumer's tastes perfectly are  $q_i = w_i$ , and expression (27) 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 (27) 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 (24) 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>33</sup> Indeed, in (27) we see that all consumers buy something in this optimal scheme. Rochet and Choné (1998, section 7.2) analyzed 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 (24) 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  $(w_1 + w_2)$  are excluded, consumers with intermediate  $(w_1 + w_2)$  are “bunched” in the sense that a consumer in this region buys the same quantity of each product, while consumers with large  $(w_1 + w_2)$  each buy a distinct bundle.

---

<sup>33</sup>The argument in Armstrong (1996) assumed that gross utility was convex and homogeneous degree 1 in taste parameters, while in (24) neither of these requirements are satisfied.

## 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 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 buying all items from one 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 only make sense to suppose that the upgrade (or similar) must be 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 cost the  $c_1$  to supply the basic product and an extra cost  $c_2$  to supply the upgrade.

As well as discussing how to derive equilibrium tariffs in a competitive setting, a focus in this section concerns the relationship between the price and the cost for the upgrade. Empirical evidence on this point is mixed. 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 markups. Verboven (1999) suggests that prices for a better car engine and other upgrades are priced above cost. And it is well documented that many kinds of “add-ons”, the price of which might be revealed to (or noticed by) consumers only after they have chosen their basic product and are locked-in to their supplier, can be expensive.

Perhaps the most straightforward situation is when sellers offer homogenous product lines and compete in Cournot fashion.<sup>34</sup> (This approach might apply well to situations where capacity constraints are important, such as with air travel.) Here, the type- $(v_1, v_2)$

---

<sup>34</sup>For further discussion, see Wilson (1993, section 12.3) and Johnson and Myatt (2006a).

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>35</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>36</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. This scenario is potentially more complicated 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) switch to the best option available from a rival, or (iii) exit the market altogether. 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 (iii) does not apply. From an individual seller’s perspective, perhaps the main difference when (ii) rather than (iii) 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.

In more detail, suppose there are two sellers, denoted  $A$  and  $B$ , and that consumers are heterogeneous in two dimensions: their brand preference for seller  $A$  over seller  $B$ , represented by parameter  $\Delta$ , and their 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  $\Delta \equiv V^A - V^B$  matters for consumer choices.<sup>37</sup> Suppose that seller  $i$  charges  $p_1^i$  for its basic product and

<sup>35</sup>For instance, for the “uniform square” example from section 2.1, somewhat laborious calculations reveal that the Cournot duopoly equilibrium tariff is  $(p_1, p_2) = (\frac{3}{5}, \frac{1}{30})$ .

<sup>36</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>37</sup>This set-up is similar to that studied by Verboven (1999). He assumes that  $V^A$  and  $V^B$  are extreme

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 3, where the parameter  $\Delta$  represents “horizontal” preferences while  $\theta$  captures the “vertical” preferences of consumers. 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.

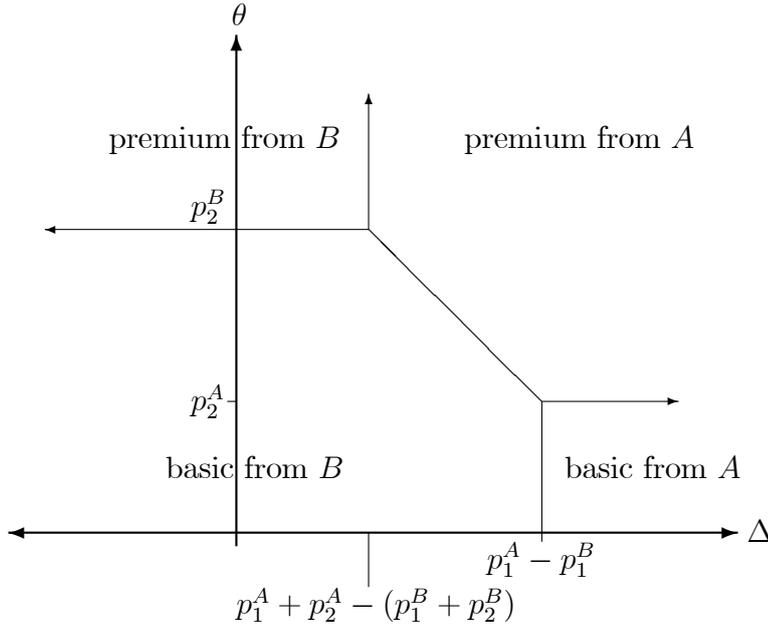


Figure 3: Pattern of demand with one-stop shopping ( $p_2^A < p_2^B$ )

One can calculate the proportion of consumers who choose each of the four options in Figure 3 and so derive the equilibrium tariffs offered by the two sellers. 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>38</sup> The reason, as with monopoly, is that the preferences of a seller’s marginal customers do not differ systematically from the value random variables, so that the difference  $\Delta$  has a logistic distribution.

<sup>38</sup>With distributional assumptions for  $\Delta$  and  $\theta$ , this result is shown in Verboven (1999, Proposition 1).

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 in this framework its price is forced down to cost.

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 3. With nonlinear pricing, we know that in equilibrium  $p_2 = c$ , and one can check that  $p_1 = c + 1/(2f(0))$  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 3 shows that a price cut by a seller 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 formally 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>39</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>40</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 one aspect of the consumer's private information, solve for the

<sup>39</sup>For a proof of this claim in a richer model, see Armstrong and Vickers (2010, Proposition 5).

<sup>40</sup>See Armstrong and Vickers (2001, Proposition 5) and Rochet and Stole (2002, Propositions 2 and 6).

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>41</sup>

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 weight. (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, suppose sellers do not reveal their upgrade price  $p_2$  until a consumer has committed to purchase at least the basic product from a seller, while consumers are rational and anticipate their seller’s incentives to choose this price.<sup>42</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

---

<sup>41</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 consumers have brand preferences over the supplier of the upgrade.

<sup>42</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 seller only by comparing  $p_1^A$  ad  $p_1^B$ .

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.

This discussion of hidden upgrade charges is taken from Verboven (1999) and Ellison (2005). Verboven assumes independence between  $\theta$  and  $\Delta$ , which implies that equilibrium profit in the game in which sellers shroud their upgrade prices is the same as the game when consumers see both prices from the start. In this case, hiding upgrade prices is not a profitable strategy. Ellison assumes positive correlation between  $\theta$  and  $\Delta$ , which implies that profit is higher in the “add-on pricing game” in which sellers coordinate to conceal upgrade prices. Ellison (1995, 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 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). They show how it can be an equilibrium strategy 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 publicise its tariff.

## 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. There are a number of situations where, 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, as with a frequent flier program, and such tactics encourage more 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. (In the airline context, these might be two routes, say.) Each consumer wants at most one unit of each product. 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 total value for the bundle where product 1 is purchased from seller  $j$  and product 2 is purchased from seller  $k$  is just  $v_1^j + v_2^k$ . (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.

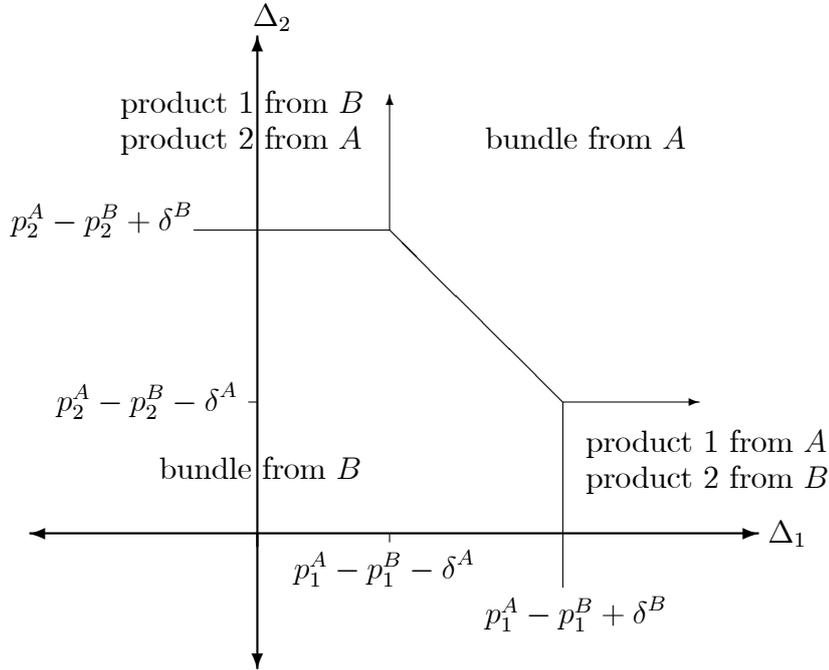


Figure 4: Pattern of demand when duopolists engage in bundling

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 4. 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 in the sense that the density of 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 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,  $\delta$ .

There is an intuitive formula for the size of the equilibrium discount,  $\delta$ . 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 4, 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

$$-\frac{\delta\Phi'(\delta)}{\Phi(\delta)} = 2 \tag{28}$$

and the elasticity of two-stop shopping is equal to 2.<sup>43</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 (21) was satisfied. Expression (28) 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

---

<sup>43</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.

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 (28) implies that the equilibrium bundle discount is  $\delta = \frac{1}{2}$  and one in eight consumers chooses to mix and match. With costless production, one can then show using Figure 4 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 bundling *all* prices fall relative to the situation with linear pricing. It follows that all consumers are better off (even though they sometimes buy a less preferred product), while sellers obtain lower profit with bundling than with linear pricing. 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>44</sup>

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>45</sup> In more detail, suppose that the brand preference 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 there is costless production, the equilibrium price for each product is  $p_1 = p_2 = 1/(2f(0))$ . 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  $\hat{f}(\bar{\Delta})$  is the induced density for the average brand preference  $\bar{\Delta} = \frac{1}{2}(\Delta_1 + \Delta_2)$ , then the equilibrium bundle price satisfies  $P/2 = 1/(2\hat{f}(0))$ . Thus the per-product bundle price,  $P/2$ , is lower

---

<sup>44</sup>Papers which discuss this “competition intensifying” property of competitive mixed bundling include 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>45</sup>Important papers which study this comparison are Matutes and Regibeau (1988) and Economides (1989).

than the price without bundling if  $\hat{f}(0) > f(0)$ . 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 as discussed in the introduction.<sup>46</sup>

Armstrong and Vickers (2010) extend this unit-demand bundling 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 charge 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 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 a given multiproduct seller. We next focus on situations where bundling strategies are 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, a museum might offer a discount on its entry fee if a visitor shows she went to another museum or has a travel card for the city. 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

---

<sup>46</sup>Zhou (2015) suggests that this result might only apply when there are few sellers, however. With more than two sellers, a firm is competing against the best offer from its several rivals and so the relevant competitive margin is not  $\Delta = 0$  as with duopoly but in the right-hand part of the distribution. While the average of i.i.d. variables is often 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 enough rivals.

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 (21), 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  $v_b < v_1 + v_2$ , 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>47</sup>

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. Since bundle discounts convert substitute products into complements (see section 3.1), sellers may be able to use a bundle discount as a means with which to relax competition (Armstrong, 2013, section 5).

As discussed in section 3.1, the use of bundling often means that the maximum revenue from selling several products together is strictly greater than the revenue obtained from selling products separately, and this implies it is hard to reward each seller’s incremental contribution to revenue at the same time as balance the alliance’s budget. This problem of “super-additivity” can vanish in the limit when *many* separate sellers combine their products into a large bundle. (For example, academic journals or television channels might

---

<sup>47</sup>For further details on this topic, see Schmalensee (1982), Lewbel (1985), and Armstrong (2013).

combine to form a bundle, or musicians might offer their content on a music streaming service.) Using the framework discussed in section 3.1, suppose that seller  $i = 1, \dots, N$  has a product for which consumers have valuation  $v_i$  which has expected value  $\mu_i$  in the population of consumers and which costs nothing to supply. (Consumers have additive and stochastically independent valuations across products.) Using a large numbers argument, the supplier of a large bundle of products, say  $\mathcal{B} \subset \{1, \dots, N\}$ , can sell its bundle to approximately all consumers for bundle price approximately equal to  $\sum_{i \in \mathcal{B}} \mu_i$ . Thus, seller  $i$ 's contribution to bundle revenue is simply its expected valuation,  $\mu_i$ , regardless of which other products are included in the bundle. Sellers can combine to sell their products in a large bundle, and divide up the resulting revenue according to each product's expected valuation. Such an arrangement benefits each seller, relative to supplying its product with a stand-alone price, and boosts welfare, but harms consumers. An alternative arrangement would be for an independent intermediary to emerge, which buys content from individual sellers and retails the bundle to final consumers. If there were several such intermediaries, competition for content might force them to pay each seller their contribution,  $\mu_i$ , to the bundle revenue.<sup>48</sup> If for some reason there was a dominant intermediary, though, it might only have pay a seller its outside option (i.e., the profit it makes if it sells its product on a stand-alone basis). The size of payments from bundling intermediaries to their content providers (be they academic journals or musicians) is a frequent source of controversy.

## 5 Conclusions

This paper has surveyed a number of topics and techniques concerning nonlinear pricing. Where possible, I have tried to make use of familiar consumer demand functions rather than the more abstract “types” often emphasized in models of nonlinear pricing. In particular, the demand profile approach associated with Bob Wilson was shown to be a useful and economically intuitive way to solve many nonlinear pricing problems in the single-product context. Even when this approach cannot generate the optimal tariff itself, standard demand functions could be used to determine whether a quantity or a bundle discount was profitable.

In most situations where the demand profile approach cannot be used, consumers differ

---

<sup>48</sup>Armstrong (2010) discusses this point in the context of academic journals, when the intermediaries are commercial publishers.

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 on 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 cases with many products, when the law of large numbers often operates to wash out most of the seller’s uncertainty about a consumer’s willingness-to-pay for its products.

In the interests of space and focus I have neglected several important aspects of nonlinear pricing. For instance, I have not covered 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 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>49</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 the 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 hard should the rival decide to enter.<sup>50</sup> As 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. This important topic deserves a survey to itself.

---

<sup>49</sup>For instance, see Chen (1997) and Fudenberg and Tirole (2000).

<sup>50</sup>For instance, see Whinston (1990).

## 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.
- (2010): “Collection Sales: Good or Bad for Journals?,” *Economic Inquiry*, 48(1), 163–176.
- (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 Nonlinear Pricing and Bundling,” *Review of Economic Studies*, 77(1), 30–60.
- 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.
- BONATTI, A. (2011): “Brand-Specific Tastes for Quality,” *International Journal of Industrial Organization*, 29(5), 562–575.

- 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.
- 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.
- 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.
- 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.
- 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. (2011): “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. (2015): “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.
- 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.

- 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 (2014): “On the Optimality of Pure Bundling for a Monopolist,” mimeo.
- MIRRELES, 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.
- SCHMALENSEE, R. (1982): “Commodity Bundling by Single-Product Monopolies,” *Journal of Law and Economics*, 25(1), 67–71.
- 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.
- THANASSOULIS, J. (2007): “Competitive Mixed Bundling and Consumer Surplus,” *Journal of Economics and Management Strategy*, 16(2), 437–467.
- VARIAN, H. (1985): “Price Discrimination and Social Welfare,” *American Economic Review*, 75(4), 870–875.
- 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.
- 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.