PRICING IN MULTIPRODUCT FIRMS

Thesis submitted for the Degree of Doctor of Philosophy

MARK ARMSTRONG

ST. JOHN'S COLLEGE

OXFORD

SEPTEMBER 1992 [i.e. 1993]

PRICING IN MULTIPRODUCT FIRMS

Mark Armstrong
St. John's College
Oxford

D.Phil. Thesis
Trinity Term 1992

Abstract

This thesis is a theoretical analysis of optimal pricing by firms when consumer demands are uncertain. The purpose is to extend the familiar literature on single–product nonlinear pricing in two directions: to cases where the firm is regulated and to the case where the firm produces several products. Chapter 1 embeds these problems into the general setting of models of asymmetric information and, as well as covering existing work on the pricing decisions of firms facing adverse selection, discusses other areas including repeated contracts, auctions, signalling and the uses of what is known as the 'first–order approach'.

Chapter 2 analyzes nonlinear pricing by a regulated single–product firm. As an alternative to requiring the firm to offer a given linear tariff two different types of regulation which allow nonlinear pricing are considered, namely, average revenue regulation and optional tariff regulation.

Chapter 3 introduces the topic of multiproduct pricing when consumers have differing tastes for the various goods. The important simplifying assumption is that consumers wish to buy either one unit of a good or none at all. There are three main results: if consumers' taste parameters are continuously distributed then the firm will not offer all goods to all consumers; in the symmetric two–good case it is shown that (subject to a kind of 'hazard rate' condition) the firm will offer the bundle of two goods at a discount compared with the charge for the two goods separately; and the pricing strategy when the number of goods becomes large is solved approximately.

Chapter 4 relaxes the assumption of unit demands and uses differential methods to analyze the multiproduct nonlinear pricing problem. In the symmetric case when taste parameters are continuously distributed the firm will choose to exclude some low–demand consumers from the market. It is shown that when parameters are independently distributed the firm will wish to introduce a degree of cross–dependence into its tariff. Sufficient conditions for a tariff to be optimal are derived and any tariff which satisfies these conditions necessarily will induce 'pure bundling', so that once a consumer decides to participate in the market at all she will choose to buy all goods. A class of cases is solved explicitly using these sufficient conditions. Since other solutions may be hard to solve analytically, a procedure for numerically generating solutions for the two–good case is described and two more solutions are described.

Acknowledgements

I would like to thank, first of all, my supervisor Jim Mirrlees for his encouragement, insights and general expertise throughout the progress of this work.

I have also benefited from working with John Vickers for the past two years on the ESRC project *The Regulation of Firms with Market Power.* The many discussions I have had with him have helped to shape my thoughts on industrial economics.

Robert Wilson in Stanford has been working on similar matters to those covered in Chapter 4 below, and he has very kindly kept me informed of his progress in several helpful letters.

Finally, I wish to thank seminar participants in Oxford — including Simon Cowan, Paul Klemperer, Dan Maldoom, Meg Meyer and Hunter Monroe — for their helpful comments on some of this work.

PRICING IN MULTIPRODUCT FIRMS

Contents

Notation

$\boldsymbol{\alpha}$, $\mathbf{q}$	vectors in $\mathbb{R}^n$ (always in bold)
∂A	the boundary of a set A in $\mathbb{R}^n$
$f'(\alpha)$	the derivative of the scalar–valued function f evaluated at the scalar α
$\mathbf{q}(\boldsymbol{\alpha})$	a vector–valued function defined on a set in $\mathbb{R}^n$: $\mathbf{q}(\boldsymbol{\alpha}) = (q_1(\boldsymbol{\alpha}), \dots, q_n(\boldsymbol{\alpha}))$ is a 'vector field'.
$\boldsymbol{\alpha}'\mathbf{q}$	scalar product of two n–dimensional vectors $\boldsymbol{\alpha}$ and $\mathbf{q}$ $= \sum_1^n \alpha_i q_i$
$\lVert\boldsymbol{\alpha}\rVert$	Euclidian norm of an n–dimensional vector $= \left(\sum_1^n \alpha_i^2 \right)^{\frac{1}{2}}$
$\nabla s(\boldsymbol{\alpha})$	the 'gradient', or vector of derivatives, of the scalar–valued function s defined on a set in $\mathbb{R}^n$ at $\boldsymbol{\alpha}$ $= \left[\frac{\partial s}{\partial \alpha_i}(\boldsymbol{\alpha}), \dots, \frac{\partial s}{\partial \alpha_n}(\boldsymbol{\alpha}) \right]$
$\mathrm{div}(\mathbf{q})(\boldsymbol{\alpha})$	the 'divergence' of the vector field $\mathbf{q}(\cdot)$ at $\boldsymbol{\alpha}$ $= \sum_1^n \frac{\partial q_i}{\partial \alpha_i}(\boldsymbol{\alpha})$

CHAPTER 1

MODELS WITH ASYMMETRIC INFORMATION:

A SELECTIVE SURVEY

1. Introduction and plan of the thesis

During the past two decades or so much effort from economic theorists has been directed towards the study of markets in which some people are better informed than others, that is to say of markets with asymmetric information. It has become conventional to distinguish two broad classes of such models, those concerning moral hazard (or hidden action) and those concerning adverse selection (or hidden information). Examples of the former are found in insurance markets (where the insurance firm is only imperfectly able to monitor the care its customers take with their insured objects), in polluting industries (where monitoring authorities may be unable accurately to measure an undesirable output of a firm), in markets for credit, in the agricultural practice of sharecropping, in the relationship between a patient and doctor, between litigant and lawyer, and between a claim–holder and a manager of a firm (in all latter cases, the principal is unable to fully monitor or control the diligence of her agent). Typically in models with moral hazard the hidden action is the effort of the agent. This effort has a cost for the agent but a benefit to the principal; if effort is not directly observable by the latter then she must somehow encourage the agent to make sufficient effort by other means.

Examples of adverse selection occur in insurance markets again (since only the higher risk customers tend to become insured), in markets in which product quality is an issue (where the seller may have a superior idea of the quality or reliability of the object she sells), the regulation of monopolies (when the firm typically will have more detailed information concerning its costs than the regulator), income taxation (where tax–payers have greater information on their earning potential than tax–collectors), signalling (for instance in employment, where potential employees have better information on their productivity than would–be employers, and may attempt to signal this productivity by means of job qualifications), and when firms sell to potential consumers whose preferences for its good(s) are only imperfectly known (in which case the firm may offer a nonlinear tariff or, alternatively, sell the goods by auction).[1]

This thesis examines some problems involving adverse selection, or, more precisely,

---

[1] For a good discussion of the range of topics involving markets with asymmetric information see Arrow (1986).

it analyzes the optimal pricing strategies of firms facing consumers with hidden information concerning their preferences. The purpose of this thesis is to extend the familiar literature on single–product nonlinear pricing in two directions: to cases where the firm is regulated and to cases where the firm produces several products. This first chapter will introduce some of the underlying ideas — both economic and analytic — of models with asymmetric information and to discuss some of the high–points of the substantial literature this subject has generated. As well as nonlinear pricing, the topic which forms the bulk of the thesis, I survey the areas of optimal taxation, monopoly regulation, repeated contracts, auctions, signalling and the uses of what is called the 'first–order approach'. Almost all of the literature concerns problems where only a 'small' amount of information is hidden: in the nonlinear pricing problem, for instance, consumers' private information consists of a scalar parameter which simply shifts utility functions up or down, and so once a consumer's utility from consuming some given quantity is known her entire utility function is determined. Most models of income taxation, regulation, auctions and signalling also employ this assumption of having only small informational asymmetries, and this way of modeling private information has been used largely to to make the analysis more tractable rather than because it is particularly natural. This assumption seems to me to be a serious limitation of the analysis up until now and it would be useful to see how robust past results are to the introduction of larger quantities of private information.

Chapter 2 analyzes the pricing decisions of a regulated single–product firm continuing with the assumption that consumers are distinguished by a scalar parameter. As an alternative to requiring the firm to offer a simple linear tariff two forms of regulation which allow nonlinear tariffs are considered: under average revenue regulation the firm is shown to maximize output subject to the regulatory constraint and to choose the nonlinear tariff which maximizes the revenue from this output; when the firm is required to continue to offer consumers the option of using the old linear tariff in addition to its chosen nonlinear tariff the result is a Pareto improvement over the linear tariff. The welfare comparison between these two forms of regulation is ambiguous.

Chapter 3 introduces the analysis of multiproduct monopoly. Here, consumers have multidimensional parameters to represent their preferences over the various goods offered by the firm and, as such, have a larger quantity of private information than with the scalar case. The important simplifying assumption here is that consumers wish to buy either one unit of a good or none at all, and the parameters represent a consumer's willingness–to–pay for the goods. There are three main results: if parameters are continuously distributed then the firm will not offer all goods to all consumers; in the symmetric two–good case it is shown that (subject to a kind of 'hazard rate' condition) the firm will offer the bundle of both goods at a discount compared with the charge for the two goods separately; and the pricing strategy when the number of goods becomes large is solved approximately.

Chapter 4 relaxes the assumption of unit demands and can therefore make use differential techniques to analyze the multiproduct nonlinear pricing problem. Unfortunately, this analysis is significantly harder than is the case in the single–product/scalar parameter model set out in Chapters 1 and 2, and repeated use is made of both duality theory and multivariate integral and differential calculus. In the symmetric case when taste parameters are continuously distributed the firm chooses to exclude some low–demand consumers from the market. When parameters are independently distributed the firm will wish to introduce a degree of cross–dependence into its tariff. Sufficient conditions for a tariff to be optimal are derived and it is shown that any tariff which satisfies these conditions will induce 'pure bundling', so that once a consumer decides to participate in the market at all she will buy all goods. A class of cases is solved explicitly using these conditions by means of a change–of–variables technique. Since I believe that other solutions may be very hard to obtain in closed form, a procedure for numerically generating approximate solutions for the two–good case is described and used to find two more solutions.

2. Three models of adverse selection

While the list of examples of adverse selection given at the start of this thesis groups together some very disparate areas of economics, the methods of analysis involved in many of these — namely income taxation, monopoly regulation and firms facing uncertain consumers — formally are quite similar.[2] The key paper as far as introducing analytic techniques is concerned was Mirrlees' (1971) study of optimal income taxation. In his model workers differ in the hourly wage w they can command from employers. This wage cannot be altered by the workers and, for a given worker, cannot be observed by the tax authorities. It is this which is the source of the asymmetric information in the model. The distribution of w amongst the population is known, and is described by the distribution function $F(\cdot)$ (so that the fraction of workers commanding a wage less than w is $F(w)$). Mirrlees uses a differential approach to solve the problem, and in order for this to make any sense we require that there is a continuum of consumers (so that F is continuously differentiable). While this is not 'realistic', it should provide an adequate approximation to the solution if the number of agents is large. What the authorities can observe is the income (i.e. wage multiplied by total hours) of any worker, and can levy a tax as a function of this income. All workers have the same preferences over the quantity of labour q and level of consumption c given by $U(c, q)$, and for simplicity we assume that this function takes the separable form $U(c, q) = u_1(c) - u_2(q)$, where u_1 is concave increasing and u_2 is convex increasing. Given the tax schedule $T(\cdot)$ (which implies that a worker's consumption is $y - T(y)$ when her income is y), a worker with wage w can attain a surplus of

$$(1) \qquad s(w) \equiv \max: \{ u_1(wq - T(wq)) - u_2(q) \mid q \geq 0 \} .$$

This optimal surplus is obtained by the response function $q(\cdot)$, say, where $q(w)$ gives the labour supply of the type w worker faced by the tax schedule T.

We say that a given labour supply function is *implementable* if it can be induced by some tax schedule $T(\cdot)$. With the utility function $u_1(c) - u_2(q)$ it turns out that the response function $q(w)$ is implementable if and only if gross income $wq(w)$ is non–decreasing in w (we discuss this in more detail in Section 3). If the objective of

---

2 The other areas mentioned above — including problems in insurance markets (see Rothschild and Stiglitz, 1976) and of product quality (see Akerlof, 1970) — involve different techniques and tend to be less complex as far as the analysis goes. We discuss signalling in Section 8 of this chapter.

taxation is to maximize the aggregate surplus of workers subject to a government revenue requirement of r_0, the government will choose the tax schedule $T(\cdot)$ in order to

(P) maximize: $E[s(w)]$ subject to

$$E[T(wq(w))] \geq r_0$$

and $s(\cdot)$ and $q(\cdot)$ being determined from $T(\cdot)$ as given by (1)

where E is the expectation operator with respect to the distribution function $F(\cdot)$. The functions $s(\cdot)$ and $q(\cdot)$ are related in two ways. First, we can write (1) as

$$s(w) = u_1(wq(w) - T(wq(w))) - u_2(q(w)) \tag{2}$$

but, in addition, an envelope argument implies that if the surplus function $s(\cdot)$ is differentiable at w its slope is given by

$$s'(w) = q(w)u_2'(q(w))/w\,. \tag{3}$$

Thus, using the terminology of control theory we can think of surplus s as the 'state' variable and q as the 'control' variable, the equation of motion being given by (3). The tax schedule can be recovered from (2) once $s(\cdot)$ and $q(\cdot)$ are known. A useful insight was to consider the response function $q(\cdot)$ rather than the tax schedule $T(\cdot)$ as the control, and this change in emphasis has been continued in all subsequent work in these areas. Thus we can restate problem (P) as

(P_1) maximize: $E[s(w)]$ subject to

(i) $s'(w) = q(w)u_2'(q(w))/w$ (and $q \geq 0$)

(ii) $wq(w)$ is non–decreasing in w

(iii) $s(w) \geq u_1(0) - u_2(0)$, and

(iv) $E[T(wq(w))] \geq r_0$ where $T(wq(w))$ is given implicitly in (2).

Constraint (iii) requires that taxes must be feasible, i.e. that $T(0) \leq 0$. (It may be reasonable to suppose that $u_1(0) = -\infty$, in which case the constraint is redundant.) Since (i) implies that s is increasing in w, we can simply write (iii) as $s(\underline{w}) \geq u_1(0) - u_2(0)$, where $\underline{w}$ is the lowest possible wage. With this formulation the problem becomes a fairly standard constrained optimization problem, and several solution methods are applicable including the Pontryagin Maximum Principle and the more direct calculus of variations approach (although this is not to say that it will be easy to solve particular cases). Constraint (ii) is the hardest of the four to incorporate, and the most straightforward method of solution is to solve (P_1) by ignoring this constraint and then to check *ex post* that it is in fact satisfied (although there will be some cases in which this simple–minded approach will not be valid). We will see how to do this in Section 4.1.

This analysis extends readily to the other two problems, that of regulating a monopoly with unknown costs and the optimal pricing strategy for a firm facing unknown consumers. The main reference for the former problem is Baron and Myerson (1982). In that paper there is a regulated single–product monopoly with a cost function given by $\theta c(q)$, where q is the output of the firm. The parameter θ is unknown to the regulator and plays the role of the wage w in the Mirrlees paper.[3] The regulator has prior beliefs concerning the likelihood of the various realizations of θ summarized by the distribution function $F(\theta)$. The regulator can observe output q and levies a tax $T(q)$ on the firm as a function of this output. Given the inverse demand function $p(q)$ and the tax schedule $T(\cdot)$, the firm with cost parameter θ can obtain a maximum producer surplus of

3 However, in constrast to that paper there is a *single* agent (whose type is uncertain). Analytically this makes little difference.

(4) $$s(\theta) \equiv \max\colon \{\ qp(q) - \theta c(q) - T(q)\ |\ q \geq 0\ \}$$

and say that this surplus is attained by the type θ firm using the response function $q(\theta)$. This time, $q(\cdot)$ is implementable if and only if it is non–increasing. Suppose that the regulator wishes to maximize a weighted sum of consumer surplus and profits subject to the firm being willing to participate. With the schedule $T(\cdot)$ consumers have an total surplus of $v(p(q(\theta))) + T(q(\theta))$ if the firm has cost parameter θ (where $v(p)$ is the standard consumer surplus function when price is p). From (4) we can express the tax revenue in terms of $s(\cdot)$ and $q(\cdot)$ as

(5) $$T(q(\theta)) = q(\theta)p(q(c)) - \theta c(q(\theta)) - s(\theta)$$

so that total consumer surplus in state θ is

$$v(p(q(\theta))) + q(\theta)p(q(c)) - \theta c(q(\theta)) - s(\theta).$$

From (4) again, the envelope condition is $s'(\theta) = -c(q(\theta))$, and so the regulator's problem is to

(P_2) maximize: $E[v(p(q(\theta))) + q(c)p(q(c)) - \theta c(q(\theta)) - (1-\lambda)s(\theta)]$ subject to

(i) $s'(\theta) = -c(q(\theta))$ (and $q \geq 0$)

(ii) $q(\theta)$ non–increasing, and

(iii) $s(\bar{\theta}) \geq 0$.

(where $0 \leq \lambda \leq 1$ is the weight the regulator places upon profit, and $\bar{\theta}$ is the maximum possible cost parameter). Again, procedures for solving this problem are well known and typically involve ignoring constraint (ii). As before, the transfer schedule $T(\cdot)$ has no explicit part to play in the problem as stated above, and is recovered from (5) after the problem is solved.

The third problem is that of nonlinear pricing and has been analyzed by several authors, including Spence (1977), Mussa and Rosen (1978), Roberts (1979), Goldman *et al.* (1984) and Maskin and Riley (1984), and the whole topic including several real–world examples has been surveyed in the book by Wilson (1992b). Here, a (single–product) firm faces consumers with varying preferences for its output. For instance, suppose that a consumer of type α has a utility function given by $\alpha u(q) + y$, where q is the quantity of the good consumed and y is income. The function $u(\cdot)$ is increasing, concave and satisfies $u(0) = 0$. The parameter α reflects the value a consumer places on the good, and this information is private to the individual. The firm believes that α is distributed throughout the population by the distribution function $F(\alpha)$. One thing the firm can do is to offer consumers its output at a constant (marginal) price p say. However, typically the firm will do better if it offers a nonlinear tariff T, where $T(q)$ is the charge if a consumer purchases quantity q. Faced with such a tariff, the type α consumer can enjoy a surplus of

$$(6) \qquad s(\alpha) \equiv \max\text{:} \{ \alpha u(q) - T(q) \mid q \geq 0 \}$$

which is attained using the demand function $q(\alpha)$ say. In this setting $q(\cdot)$ is implementable if and only if it is non–decreasing, and the envelope condition is $s'(\alpha) = u(q(\alpha))$. If the firm has a constant unit cost of c in making the good, given a tariff $T(\cdot)$ it makes a profit of $T(q(\alpha)) - cq(\alpha)$ from a type α consumer. But (6) implies that $T(q(\alpha)) = \alpha u(q(\alpha)) - s(\alpha)$, and so the profit–maximizing firm will seek to

(P_3) maximize: $E[\alpha u(q(\alpha)) - s(\alpha) - cq(\alpha)]$ subject to

(i) $s'(\alpha) = u(q(\alpha))$ (and $q \geq 0$)

(ii) $q(\alpha)$ non–decreasing, and

(iii) $s(\underline{\alpha}) \geq 0$

(where $\underline{\alpha}$ is the smallest consumer type, the final constraint following from the assumption that consumers cannot be compelled to consume).

Thus we see that all three problems are very similar in terms of the analysis. In each case there is a principal facing a population of agents about which she is imperfectly informed. In order to maximize her objective she offers agents a payment schedule T and the various agents respond according to the response function $q(\cdot)$ and enjoy a surplus of $s(\cdot)$. In all three problems we can write the principal's objective in terms of these response and surplus functions rather than the payment schedule, and a constrained optimization problem thereby results. In each problem there is the equation of motion constraint (i), the implementability constraint (ii) and the participation constraint (iii). (In the tax problem there is the additional constraint (iv) on revenue.) Constraint (ii) is always the hardest to handle and in many cases it turns out that if the problem (P_i) is solved without the constraint that $q(\cdot)$ be implementable, the resulting solution does anyway satisfy this constraint; we investigate this possibility more fully in the next section.

3. Implementability and the first–order approach

It is instructive to examine the constraints in the above problems in a little more detail. For concreteness we will use the setting of problem (P_3) concerning nonlinear pricing. Recall that a demand function $q(\cdot)$ is said to be implementable if it can be induced by means of some nonlinear tariff $T(\cdot)$. We now establish the claim made in Section 2 that $q(\cdot)$ is implementable if and only if it is non–decreasing (and, of course, non–negative). In the following let α take values in the interval $[\underline{\alpha}, \bar{\alpha}]$.

LEMMA 1. *The response function* $q(\cdot)$ *is implementable if and only if it is non–decreasing.*

PROOF. *Necessity*: Let $q(\cdot)$ be implemented by the tariff $T(\cdot)$ and let $\alpha > \tilde{\alpha}$. Then by revealed preference we must have

$$\alpha u(q(\alpha)) - T(q(\alpha)) \geq \alpha u(q(\tilde{\alpha})) - T(q(\tilde{\alpha}))\ , \text{ and}$$
$$\tilde{\alpha} u(q(\tilde{\alpha})) - T(q(\tilde{\alpha})) \geq \tilde{\alpha} u(q(\alpha)) - T(q(\alpha))\ .$$

Subtracting one inequality from the other implies that $(\alpha-\tilde{\alpha})(q(\alpha)-q(\tilde{\alpha})) \geq 0$, i.e. that $q(\alpha) \geq q(\tilde{\alpha})$.

Sufficiency: Let $q(\cdot)$ be some given non–negative, non–decreasing function defined on $[\underline{\alpha}, \bar{\alpha}]$. We shall construct a tariff T which will induce the response function q. Let $s(\cdot)$ be a function defined by constraint (i) of (P_3), e.g.

$$s(\alpha) = \int_{\underline{\alpha}}^{\alpha} u(q(\tilde{\alpha}))\, d\tilde{\alpha} . \tag{7}$$

Define a tariff $T(\cdot)$ defined on the range of $q(\cdot)$ by the formula

$$T(q(\alpha)) = \alpha u(q(\alpha)) - s(\alpha) \tag{8}$$

(and $T(q) = +\infty$ for q outside the range of $q(\cdot)$). Then we can show that faced with the tariff $T(\cdot)$ a consumer of type α would choose the quantity $q(\alpha)$. Firstly, it is immediate that the participation constraint is satisfied since the surplus of a type α consumer faced with the tariff $T(\cdot)$ in (8) is at least $\alpha u(q(\alpha)) - T(q(\alpha)) \equiv s(\alpha)$ and $s(\cdot) \geq 0$ from (7). Therefore, all consumers make a non–negative surplus from the tariff $T(\cdot)$. Second, we need to show that:

$$\alpha \ \text{maximizes:}_{\tilde{\alpha}} \ \alpha u(q(\tilde{\alpha})) - T(q(\tilde{\alpha})) .$$

But q non–decreasing implies that the function $s(\alpha)$ defined in (7) is convex, and since convex functions lie above their tangents it satisfies

$$\begin{aligned} s(\alpha) &\geq s(\tilde{\alpha}) + s'(\tilde{\alpha})(\alpha-\tilde{\alpha}) \\ &= s(\tilde{\alpha}) + u(q(\tilde{\alpha}))(\alpha-\tilde{\alpha}) \quad \text{from (7).} \end{aligned}$$

Writing this in terms of the tariff $T(\cdot)$ in (8) we see that

$$\alpha u(q(\alpha)) - T(q(\alpha)) \geq \alpha u(q(\tilde{\alpha})) - T(q(\tilde{\alpha})),$$

i.e. that a type α consumer if faced with the tariff $T(\cdot)$ defined in (8) will choose to purchase the quantity $q(\alpha)$. □

As mentioned above, the usual approach in actually solving these problems is to ignore the implementability constraint (ii) in (P_i), to solve the less constrained problem and then to check that the response function $q(\cdot)$ which solves the problem is indeed implementable (i.e. non–decreasing in the nonlinear pricing context). Clearly, if this method works we will have found the solution to the fully constrained problem. We show next that this approach is closely related to the 'first–order approach' to solving principal–agent problems. This approach involves choosing the tariff $T(q)$ in order to maximize profits subject to the first–order condition on consumers' incentive compatibility holding (together with the participation constraint), but otherwise ignoring the constraint that consumers will choose their optimal quantity. Since the first–order condition for $q > 0$ to be the output choice of a type α consumer is $\alpha u'(q) = T'(q)$, the first–order approach to solving the firm's problem is to choose the demand function $q(\cdot)$ and tariff $T(\cdot)$ in order to:

(P_3^{fo}) maximize: $E[T(q(\alpha)) - cq(\alpha)]$ subject to

(i) $\alpha u'(q(\alpha)) - T'(q(\alpha)) \equiv 0$ and

(ii) $\alpha u(q(\underline{\alpha})) - T(q(\underline{\alpha})) \geq 0$.

Unfortunately this problem is not very well–behaved and there may be no solution to (P_3^{fo}) as stated.[4] However, in many cases the problem does have a solution. We can reformulate

4 The problem is that the first–order approach involves the implicit assumption that the tariff is differentiable, and whether or not the optimal tariff is differentiable is endogenous to the problem. If the optimal has kinks (so that the response function involves bunching) then the first–order approach to the problem has no solution.

(P_3^{fo}) as follows: constraint (i) of problem (P_3^{fo}) implies that

$$\frac{d}{d\alpha}[T(q(\alpha)) - \alpha u(q(\alpha)) + \int_{\underline{\alpha}}^{\alpha} u(q(\tilde{\alpha}))\, d\tilde{\alpha}\,] = 0$$

and so

$$T(q(\alpha)) = T(q(\underline{\alpha})) + \alpha u(q(\alpha)) - \int_{\underline{\alpha}}^{\alpha} u(q(\tilde{\alpha}))\, d\tilde{\alpha} - \underline{\alpha} u(q(\underline{\alpha}))\,. \tag{10}$$

Therefore, firm profit can be written as

$$E[\alpha u(q(\alpha)) - \int_{\underline{\alpha}}^{\alpha} u(q(\tilde{\alpha}))\, d\tilde{\alpha} - cq(\alpha)] + T(q(\underline{\alpha})) - \underline{\alpha} u(q(\underline{\alpha}))$$

and from constraint (ii) of (P_3^{fo}) the solution to (P_3^{fo}) will involve the firm setting $T(q(\underline{\alpha})) - \underline{\alpha}u(q(\underline{\alpha})) = 0$ and choosing $q(\alpha)$ in order to maximize

$$E[\alpha u(q(\alpha)) - \int_{\underline{\alpha}}^{\alpha} u(q(\tilde{\alpha}))\, d\tilde{\alpha} - cq(\alpha)]. \tag{11}$$

Since constraint (i) of problem (P_3) implies that

$$s(\alpha) = \int_{\underline{\alpha}}^{\alpha} u(q(\tilde{\alpha}))\, d\tilde{\alpha} + s(\underline{\alpha})$$

and constraint (iii) implies that $s(\underline{\alpha}) \geq 0$, it is clear that the maximization problem (P_3) without constraint (ii) will involve setting $s(\underline{\alpha}) = 0$ and choosing $q(\cdot)$ to maximize (11).

This argument demonstrates that the first–order approach to the problem coincides in many cases with problem (P_3) without constraint (ii). However, the correspondence is not quite exact. For instance, take an extreme example and suppose that the response function $q(\alpha)$ that maximized (11) was completely flat: $q(\alpha) \equiv q^*$ say. For such a case there is no tariff T which satisfies the first–order condition for q* to be an optimal choice for all types of consumer and (P_3^{fo}) has no solution. For all cases for which the solution to (11) is never flat, though, the two approaches do coincide and for convenience we therefore term the strategy of solving any of the problems (P_i) by ignoring the constraint that the response function be implementable (and then checking *ex post* that the solution is implementable) to be the 'first–order approach'.

Solving these problems using the first–order approach is easily done. In the next section we describe how to do this and also see under which conditions the approach is valid.

4.1 Using the first–order approach: the case of nonlinear pricing

We now find the solution to (P_3) ignoring constraint (ii). If α is distributed by the density function $f(\alpha)$, then the objective function of (P_3) is

$$\int_{\underline{\alpha}}^{\bar{\alpha}} [\alpha u(q(\alpha)) - s(\alpha) - cq(\alpha)]f(\alpha)\, d\alpha$$

where the support of α is the interval $[\underline{\alpha}, \bar{\alpha}]$. Since constraint (i) of (P_3) requires that $s'(\alpha) = u(q(\alpha))$, we can integrate by parts the term in $s(\cdot)$ above and write profits as

$$\int_{\underline{\alpha}}^{\bar{\alpha}} \{[\alpha u(q(\alpha)) - cq(\alpha)]f(\alpha) - (1 - F(\alpha))u(q(\alpha))\}\, d\alpha\,,$$

where we have used the fact that $F(\bar{\alpha}) = 1$ and $s(\underline{\alpha}) = 0$ (as is optimal). Therefore, if we ignore the constraint that $q(\cdot)$ be increasing, we can simply maximize the above integrand pointwise (subject to $q \geq 0$), i.e. we can set the optimal response function, denoted $q^*(\cdot)$, to satisfy:

$$\text{(12)} \qquad q^*(\alpha) \ \text{maximizes}_{q \geq 0}\text{:} \ u(q)[\alpha - \frac{1 - F(\alpha)}{f(\alpha)}] - cq\,.$$

Provided that (12) results in $q^*(\cdot)$ being non–decreasing then all constraints to problem (P_3) have been satisfied and the problem is solved. The missing link is the condition which determines whether q^* in (12) is non–decreasing. From (12) this condition is given by

(13) the function $\alpha \mapsto \alpha - \frac{1 - F(\alpha)}{f(\alpha)}$ is non–decreasing.[5]

Condition (13) is satisfied for a rather wide variety of well–known distributions for α, including all of those for which the hazard rate $\frac{f(\alpha)}{1 - F(\alpha)}$ is non–decreasing in α. Distributions which have increasing hazard rates include the normal, uniform, and exponential, and so for these distributions the first–order approach will yield meaningful answers.[6] If, however, (13) does not hold then $q^*(\cdot)$ in (12) will not be non–decreasing and from Lemma 1 it cannot possibly be implemented by any tariff. In such cases more delicate analysis will be needed. Mussa and Rosen (1978) discuss the issues involved here, and show that when the first–order approach breaks down it will be optimal for the firm to bunch some consumers together, so that there will be a non–trivial set of consumers all of whom buy the same (positive) quantity. Indeed, except for those borderline cases where (13) holds only weakly, there is a one–to–one relationship between the cases where the first–order approach is valid and the cases where it is optimal for the firm to completely separate those consumers who make a purchase.

We should not be too negative, however. Just because the first–order approach does not always yield the correct answer does not rule it out as a useful method of solution; indeed, the method continues to provide the most fruitful means of actually solving particular problems, and we shall use it extensively in Chapters 2 and 4 below.[7]

---

[5] Strictly, we do not need $\alpha - \frac{1-F(\alpha)}{f(\alpha)}$ to be non–decreasing for those values of α small enough that $q^*(\alpha) = 0$ in (12).

[6] More generally, any distribution for which the density function $f(\alpha)$ is log–concave has an increasing hazard rate — see Fudenberg and Tirole (1991, Chapter 7, footnote 21).

[7] The first–order approach has been proposed also for the moral hazard paradigm, for instance by Mirrlees (1974), Holmstrom (1979) and Jewitt (1988). Mirrlees (1975) and Rogerson (1985) find sufficient conditions for the first–order approach to be valid, although these conditions seem significantly stronger than the condition (12) which is required for the approach to be valid in the case of adverse selection. Grossman and Hart (1983) take a more fully optimizing approach, albeit in a very special model. For discussions of these

Returning to the analysis, what can we say about the shape of the optimal tariff T^*? The first–order condition for $q(\alpha) > 0$ to be the optimal choice of quantity for consumer α is $T'(q(\alpha)) = \alpha u'(q(\alpha))$, and so writing $p(q) \equiv T'(q)$ for the marginal price schedule for any tariff T, from (12) we can obtain the following characterization for the profit–maximizing marginal price schedule $p^*(q)$:

$$(14) \qquad p^*(q^*(\alpha)) = c/[1 - \frac{1 - F(\alpha)}{\alpha f(\alpha)}] \, .$$

As many have observed, equation (14) has two corollaries:

— Provided that $\bar{\alpha}$ is finite then $p^*(q^*(\bar{\alpha})) = c$, and so the consumers with the highest demand parameter $\bar{\alpha}$ are served efficiently; there are no mutually beneficial trades which could take place between the firm and such consumers.

— If $\frac{1 - F(\alpha)}{\alpha f(\alpha)}$ is decreasing then $p^*(q^*(\alpha))$ is decreasing in α. Since $q^*(\cdot)$ is increasing in α under the same assumption this means that $p^*(q)$ is decreasing in q, i.e. that the profit–maximizing tariff T^* is concave.

This second point, made by Maskin and Riley (1984), is worth stating as a result:

---

points see Rees (1985, pp. 16–22) and Laffont (1989, section 11.2).

RESULT 1 *(Maskin and Riley). The firm will find it optimal to offer quantity discounts, i.e. a tariff* T* *for which average cost* T*(q)/q *decreases with* q, *if*

(15) $$\frac{1 - F(\alpha)}{\alpha f(\alpha)} \text{ is decreasing.}$$[8]

This result is interesting. Quantity discounts are widely practised by actual firms: an airline may offer free flights if a certain number of miles are flown with it; banks offer higher rates of interest when larger amounts of money are deposited with it; and BT is now offering discounts of around 10% on a quarterly telephone bill of more than about £100. Whilst it may be possible that this it due in part to the fact that it is proportionally cheaper to serve consumers higher than lower quantities, the above result provides an additional motivation for the practice which is purely to do with the effects of asymmetric information.

4.2 Another view of nonlinear pricing: repeat purchases

The above model was very simple: consumers made their purchases at only one point in time and there was no uncertainty about the quality of (or their taste for) the good. In Section 5 we will consider general models of repeated contracts, but here we look at the special topic of repeat purchases. The phenomena we are interested in here include new magazine subscribers being offered a discount compared with those who are already subscribing, and firms distributing free trial samples of a new product in a supermarket or through our mailbox. These 'introductory offers' are forms of quantity premia, in the sense

---

[8] Like condition (12), condition (15) holds whenever α has an increasing hazard rate. Also, (15) implies (12). Therefore, the various strengths of assumptions on α may be ordered as:

$$\text{increasing hazard rate} \Rightarrow (15) \Rightarrow (12).$$

We argue in in Chapter 2 below that quantity discounts are not so much a result of profit–maximizing behaviour but of revenue–maximizing behaviour. There we show that the tariff that maximizes revenue for some given level of total output will be concave proved that condition (15) holds. An implication of this is that the profit–maximizing tariff is concave.

that subsequent units of the good are more expensive than the first. Alternatively, firms may include coupons on their product (e.g. coffee or breakfast cereal) which entitle the consumer to a discount on any subsequent purchase. Such examples, which include the airline example above, are forms of intertemporal quantity discounts.

The distinctive feature of models of repeat purchases is that consumers learn about the quality of the product, or how much they like the product, through time (otherwise the time dimension adds nothing to the analysis). Here we will focus on two models where *a priori* consumers do not know whether they will like the good (which may be a magazine or a breakfast cereal), and only discover their preferences *ex post.*[9] More specifically, let there be two periods with consumers buying at most one unit of the good in each period. If a consumer has consumed the good in the first period she discovers her taste for the good in the second. In principle the firm chooses three prices: p_1 (the price of the good in the first period), p_2 (the price of the good in the second period given that the consumer has purchased the good in the first period) and $\tilde{p}_2$ (the price of the good in the second period if no purchase is made earlier).

In one model consumers differ by the scalar parameter α with distribution function $F(\alpha)$.[10] The von Neumann/Morgenstern utility in each period of a consumer of type α buying a unit of the good at price p is

$$u = \begin{cases} \alpha - p & \text{if she likes the good} \\ -p & \text{if she does not like the good.} \end{cases}$$

(Her utility when no purchase is made is zero.) The proportion x $(0 < x < 1)$ of consumers of all types who like the good is exogenous and common knowledge. Prior to making her first purchase a consumer does not know whether she will like the good, this being revealed at the end of the first period. It is assumed that the firm is unable to

[9] Thus we do not consider the case where the quality of the good, which is controlled by the firm, initially is unknown to consumers and is discovered later. This would be the 'moral hazard' version of Section 4.2.

[10] This model is (loosely) based on Tirole (1988, section 2.6.1), which in turn is (loosely) based on Farrell (1986) and Milgrom and Roberts (1986).

distinguish first–time from second–time buyers in the second period, so that $\tilde{p}_2 = p_2$. If a consumer buys in the first period and prices are p_1 and p_2 her total utility is

$$u_1(p_1, p_2) = x\alpha - p_1 + \delta x[\alpha - p_2]_+$$

(where the discount rate is δ and $[\alpha - p_2]_+$ is the positive part of $\alpha - p_2$). This is because she will obtain a total utility of $\alpha - p_1 + \delta[\alpha - p_2]_+$ with probability x, and $-p_1$ with probability $1 - x$. If she buys only in the second period her total utility is

$$u_2(p_1, p_2) = \delta(x\alpha - p_2).$$

After tedious calculation, profits are found to be:

(i) $$(1 - F(p_1/x))(p_1 - c) + \delta x(1 - F(p_2))(p_2 - c) \quad \text{if } p_2 \geq p_1/x$$

(in this case p_2 is so high that nobody buys only in the second period, and only a fraction of those who buy in the first period buy in the second);

(ii) $$(1 - F(\frac{p_1 + \delta x p_2}{x(1 + \delta)}))(p_1 + \delta x p_2 - (1 + \delta x)c) \quad \text{if } p_1/x \geq p_2 \geq p_1/(1 + \delta(1 - x))$$

(in this case p_2 is sufficiently high for nobody to consume only in the second period, but low enough for all who consume in the first period to consume in the second);

(iii) $$(1 - F((p_1 - \delta(1 - x))/x))(p_1 + \delta x p_2 - (1 + \delta x)c) \; + \\ \{(F((p_1 - \delta(1 - x))/x) - F(p_2/x))(p_2 - c)\} \quad \text{if } p_2 \leq p_1/(1 + \delta(1 - x))$$

(here p_2 is so low compared to p_1 that some low type consumers buy only in the second period, and profits from this group is $\{\cdots\}$ in the above).

Tirole argues that it is the middle regime that will be optimal for the firm, so that the firm does not try to reach a new clientele in the second period but also does not wish to lose any of its existing consumers (its 'goodwill' clientele). In particular, $p_2 \leq p_1/x$. Tirole defines an 'introductory offer' to mean that $p_2 > p_1/x$, so that the first period price after discounting the possibility that a consumer will not like the good is less than the second period price. Therefore, with this definition the firm makes no introductory offer. It is, however, likely that $p_2 > p_1$, and so in this sense we see that quantity premia are a frequent outcome of this model.

Crémer (1984) proposes an alternative way of modelling this problem in which quantity discounts appear. In contrast to the first model consumers *ex ante* are identical. As with Section 4.1, take utility of a type α consumer to be $\alpha u(q) - T$. In the first period consumers do not know their own α but only the distribution from which it is drawn. They buy the good if their total expected utility (in money terms) is greater than their total expected outlay. Since all consumers are identical to start with, either all will buy in the first period or none will. Crémer shows that it cannot be optimal to make everyone buy only in the second period and, therefore, the distinction between p_2 and $\tilde{p}_2$ again vanishes. After consuming the good in the first period they discover their own particular taste parameter α and then make a second purchase if α is sufficiently high. Crémer shows that in this model the firm's profit–maximizing strategy involves setting $p_2 \leq p_1$, so that there are discounts for repeat buying. Thus we see that whether repeat purchases should be cheaper or more expensive than the initial purchase is sensitive to the way in which the uncertainty is modeled.[11]

[11] Repeat purchase discounts also make sense in a competitive framework, in that they can create switching costs between firms in the sense that once a consumer has been with one firm (e.g. an airline) for some time she may well face a lower marginal price from the firm than from other companies — see Caminal and Matutes (1990), the survey article by Klemperer (1990) and the references therein. Alternatively, first–time purchase discounts might be a means by which potential entrants to a market can pay the switching cost of a consumer changing brand from an incumbent, thereby easing their entry into the market.

For an exposition of the theory of nonlinear pricing (rather than of repeat purchases) in a competitive framework see Wilson (1992b).

5. The effect of asymmetric information and possibilities for the first–best

It is natural to consider what would happen if the principal *could* observe agents' hidden information and to contrast this with the second–best outcome. For the case of the monopolist facing uncertain consumers the first–best, full information outcome is simple to describe: since α is observable the firm can base the quantity and payment from a type α consumer directly upon α and for each type α the firm will choose the output $q_{fb}(\alpha)$ and lump–sum tax $T_{fb}(\alpha)$ in order to:

(P_3^{fb}) maximize $T - cq$ subject to

$$\alpha u(q) - T \geq 0,$$

i.e. $q_{fb}(\alpha)$ will maximize $\alpha u(q) - cq$, and $T_{fb}(\alpha) = \alpha u(q_{fb}(\alpha))$. The firm should make a take–it–or–leave–it offer to a consumer of type α of the quantity $q_{fb}(\alpha)$ in exchange for a fee of $T_{fb}(\alpha)$, an offer the consumer will (just) accept. Thus, this 'first–degree' price discrimination will be efficient and will leave all consumers with none of the gains from trade. Why does the firm fail to obtain this level of profit when it is unable to observe consumers' types? The function $q_{fb}(\alpha)$ is certainly implementable (since it is increasing in α), and would be the response function if the firm offered consumers the uniform tariff $T(q) = cq$. However, the firm does not get revenue of $T_{fb}(\alpha)$ from the type α consumer using this tariff. Suppose that the firm offered all consumers a menu of choices $\{ q_{fb}(\tilde{\alpha}), T_{fb}(\tilde{\alpha}) \}$, i.e. a consumer could choose the quantity $q_{fb}(\tilde{\alpha})$ for a charge $T_{fb}(\tilde{\alpha})$ (for any $\tilde{\alpha}$). Faced with this choice, the type α consumer would choose $\tilde{\alpha}$ in order to maximize

$$\alpha u(q_{fb}(\tilde{\alpha})) - T_{fb}(\tilde{\alpha}) = (\alpha - \tilde{\alpha})u(q_{fb}(\tilde{\alpha})) ,$$

(the equality arising from the definition of $T_{fb}(\tilde{\alpha})$). This must occur for some $\tilde{\alpha} < \alpha$. Therefore, a consumer of type α will 'pretend' to be of some other lower type $\tilde{\alpha}$. Consumption will be lower than is efficient but consumers will be left with some surplus.

As we would expect, then, introducing information which is private to consumers will reduce the monopolist's profit but increase every consumer's surplus relative to the first–best.

The incentive incompatibility of the first–best solution is more acute in the case of the income taxation problem (P_1).[12] Here, with full observability of workers' productivity the problem reduces to that of determining optimal lump–sum taxation subject to a budget constraint. i.e. to choose consumption levels $c_{fb}(w)$ and labour supply $q_{fb}(w)$ in order to:

(P_1^{fb}) maximize $E[u_1(c(w)) - u_2(q(w))]$ subject to

$$E[wq(w) - c(w)] \geq r_0 .$$

Because of the separability of the utility function, we can decompose this problem: first consider how to share out a given level of total consumption C say. The function u_1 being concave implies that at the optimum all workers should enjoy the same level of consumption C (where we have normalized the number of consumers to equal one). Next, what is the optimal way to share out labour for a given level of production P, i.e. what is the solution to the problem of minimizing $E[u_2(q(w))]$ subject to $E[wq(w)] = P$? This is solved by setting $u_2'(q(w)) = \gamma w$ for some constant $\gamma > 0$ which, because of the assumed convexity of u_2 implies that at the optimum $q_{fb}(w)$ is increasing with w. Finally, it is necessary to choose aggregate consumption C and production P optimally. Therefore, whatever the result of this final optimization is, we have shown that the utility of the type w worker is $u_1(C_{fb}) - u_2(q_{fb}(w))$ which is decreasing in the worker's productivity w. This certainly cannot be incentive compatible when productivity is the private information of workers since all workers will then pretend to be of the lowest possible type $\underline{w}$.

Finally, we look at the problem of regulating a monopoly with unknown costs. Whenever the regulator places less weight on profit than on consumer surplus (i.e.

12 This paragraph is based upon Mirrlees (1974, section 9.3).

whenever $\lambda < 1$), the effect of asymmetric information is similar to that of the case of nonlinear pricing. However, if the objective is simply to maximize total surplus (i.e. $\lambda = 1$) then the first–best outcome becomes attainable. The mechanism is to offer the firm a subsidy equal to the entire consumer surplus, $v(p(q))$, if it produces an output q. With this method of regulation, which was proposed by Loeb and Magat (1979), the firm faces exactly the correct incentives to produce the efficient level of output no matter what its costs are. Of course, the result of this regulation is that consumers are left with none of the gains from trade and so the policy is unlikely to meet with universal approval.

The possibilities of the first–best have tended to receive more discussion in the moral hazard setting. For instance, Shavell (1979) demonstrates that if the agent is risk–neutral in the standard principal–agent problem, then the first–best, full information solution is possible. Mirrlees (1974, section 9.2) has an example with risk–averse agents where the first–best is (almost) attainable if the principal offers a payoff schedule which gives almost all agents the first–best payoff but punishes a few agents with very bad outcomes in a particularly severe manner. (In order for this draconian scheme to work it is necessary that agents' utility is unbounded from below.) Most usually, though, the optimal contract for the single–period, single–agent moral hazard problem will show deviations from the first–best allocation.

Typically, the presence of asymmetric information will lead to optimal second–best contracts which are inefficient relative to the first–best and which will leave agents with some rent on their private information. The principal tries to induce agents to reveal their type (or their effort level) by means of some incentive scheme and this will usually require distortions away from the first–best allocation. We have seen, however, that this is not invariably so and we see will more instances of first–best contracts below in the sections on repeated contracts and on competition between agents. Also, in Section 9 of Chapter 3 below we shall see another example of an adverse selection model for which the first–best is attained; in that case the principal is able to eliminate her uncertainty about consumer types due to a 'large numbers' effect.

6. Repeated contracts in principal–agent models

Except for the analysis of repeat purchases in Section 4.2 the three adverse selection models considered up to this point have been static: the principal and agent(s) meet once and then never again. In reality, workers are active for many years and their income is taxed on an annual basis; firms have long–standing relationships with their regulators. Is it possible that agents' responses in one period could be used to update the principal's information on types, and thereby improve her ability to reduce her losses due to the informational asymmetry in subsequent periods?

For instance, if the owner of a car has several years' accident–free driving then this may cause the insurance firm to favorably update its prior on the probability of that driver having an accident in the future and therefore to reduce its insurance premium (i.e. to offer a 'no–claims bonus'). If a footballer signs a contract to play for five years at a salary of £500,000 per year he has effectively revealed his earning potential for these years, and it could well be that the utilitarian tax authority would find it optimal to design a personal tax schedule for him, perhaps consisting simply of a large lump–sum tax. Alternatively, any savings a worker has in one period (either from past earnings or from bequests) may from that point on be regarded as lump–sum wealth and its removal by the tax authorities will not have any undesirable disincentive effects on the worker in the future. Therefore, such wealth taxes may well be optimal taking past decisions as given. Of course, the possibility of such 'expropriation' of wealth in the future could well affect present incentives of forward–looking workers and so such wealth taxes might not be optimal when considered in a dynamic context. More difficulties arise if the present government cannot commit that future governments will not expropriate the wealth generated by present workers. If a firm reveals that it has low costs now, then, provided that future costs were strongly correlated with present costs, the regulator would from that point onwards tighten regulation to the point where the firm makes only normal profits. Of course, the strategically–inclined firm would anticipate this behavior on the part of the regulator and might pretend to have a higher cost now in order to reap super–normal profits later. (This possibility is known as the 'ratchet effect'.) And if (for some reason) the regulator was able to commit to a particular dynamic regulatory regime, how would this change the equilibrium?

For concreteness, we shall use the framework of the problem of regulating monopoly with unknown costs.[13] Suppose that a firm exists for two periods and has the same cost parameter θ in each of these periods. There is no discounting. If the regulator was able to commit to a particular incentive scheme at the start of the two periods — say, by offering a subsidy to the firm of $T(q_1, q_2)$ if it produces output q_i in period i — then by symmetry she would want the firm to produce the same output in each of the two periods. Indeed, the regulator will induce the firm in each period to do exactly what it would have induced it to do at the optimum of the static model (P_2) above, which resulted in the response function $q^*(\theta)$ say. Therefore, with commitment power dynamic regulation is no different from the static equilibrium replicated in time. If, however, the regulator did not have such power, this symmetric incentive scheme no longer remains viable. To see this, remember that in many cases the static optimal response function $q^*(\cdot)$ completely separates firms of different type.[14] When this is so the regulator can deduce the type θ of the firm from its response $q^*(\theta)$ in the first period and can impose the first–best, full information solution in the second period (when the firm would obtain zero surplus). Anticipating this, the type θ type will find it optimal to deviate from its response $q^*(\theta)$ in order to obtain a positive surplus in the second period. For instance, if the type θ firm pretends to be the slightly higher cost firm $\theta + d\theta$ in the first period, from the optimality of $q^*(\theta)$ it makes only a second order loss then but will increase its profit by order $d\theta$ in the second period. Therefore, the two period equilibrium with commitment is not incentive compatible when the regulator cannot commit not to use the information obtained in the first period in order to squeeze profits later on. In fact, this

---

13 This paragraph draws upon section 10.4 of Laffont (1989) which in turn is a specialization to pure adverse selection of Laffont and Tirole (1988), a paper which included moral hazard as well as adverse selection.

14 It may be calculated using the method of Section 4.1 that the optimal response function q^* in the case where $c(q) \equiv q$ in problem (P_2) of Section 2 satisfies

$$p(q^*(\theta)) = \theta + (1 - \lambda)\frac{F(\theta)}{f(\theta)},$$

which completely separates firms of different type if the function $\theta + (1 - \lambda)\frac{F(\theta)}{f(\theta)}$ is strictly increasing.

argument demonstrates that in the no–commitment equilibrium there must be some degree of pooling by different firms in the first period — if firms were completely separated in the first period then they would subsequently all have profits squeezed to zero and firms would therefore maximize first period profit, so the only possible separating equilibrium is that using the optimal single–period response function $q^*(\cdot)$ (which we have shown to be sub–optimal).[15]

We might imagine that increasing the number of periods over which the principal and agents meet would increase this commitment problem since firms will have an added incentive to keep hold of information. Roberts (1984) shows heuristically that this can be true in a striking form in a dynamic taxation framework. He assumes that agents are infinitely long lived and do not discount (and so they wish to maximize their long–run average utility). The tax a worker must pay is allowed to depend upon the worker's past history of earnings as well as present earnings. Any information which the government obtains in one period will be used in future periods to maximize social welfare. In the two–period model above it was shown that there is necessarily some degree of pooling in the first period. Roberts argues that in his model there is *complete* pooling, so that all workers of all types have the same income in all periods. This is a dramatically inefficient outcome and shows the effect that lack of commitment can have in multi–period models of adverse selection.[16]

Repeated contracts have also been studied in the context of moral hazard. When there is no adverse selection (so that the principal is quite certain what the agent will do under any incentive contract) it is possible for a repeated contract to attain the first–best level. Rubinstein and Yaari (1983) consider a model in which there is a single

---

[15] All of this discussion assumes that the firm's costs are perfectly correlated from one period to the next. At the other extreme we can consider what happens if the cost parameters were independently distributed. In this case, it is simple to see that the optimal policy without commitment is just to repeat the static equilibrium q^* in each period, the reason being that a firm's response in the first period gives no new information about the firm's likely cost level in the second. Thus the no–commitment equilibrium when parameters are independent across time is precisely the same as the commitment equilibrium when parameters are perfectly correlated. For a thorough treatment of optimal dynamic regulation see Baron (1988, section 6).

[16] For more on dynamic taxation see Harris (1986).

infinitely–lived agent who does not discount the future (just as in the Roberts paper). This agent controls a random payoff y_t each period, the distribution of which depends upon her effort e_t that period (and not upon past history). She seeks to contract with a (competitive) principal to insure against this shock. We imagine that over time the principal will be able to deduce the effort level due to a large numbers effect. More precisely, let $e*$ be the first–best effort level with corresponding average payoff $y* = E[y \mid e*]$. When the agent chooses effort level $e*$, the long–run average of the observed payoffs $\frac{1}{T} \sum_1^T y_t$ will converge to $y*$ (with probability one) and so the principal can design a 'trigger' strategy whereby she gives the agent the first–best optimal insurance contract in any period whenever the average of the past payoffs does not deviate 'too far' from $y*$ in that period. In any period in which the observed average is judged to be too far from $y*$ the principal offers some less desirable contract. Provided that the definition of 'too far' is accurately chosen, the agent will receive optimal insurance in all but a finite number of periods if she chooses effort level $e*$, whereas if she chooses effort $e < e*$ she will miss out on the first–best an infinite number of times. Therefore, with no discounting she will choose the first–best effort level of $e*$ in each period. This is a subgame perfect equilibrium and therefore the principal needs no commitment power to enforce this first–best contract.

7. Competition amongst agents

Nonlinear pricing as discussed in Section 4 involved a firm selling a large number of units of a good to a large number of consumers. We restricted attention to nonlinear tariffs where the payment a consumer made for a given quantity did not depend on the demands of the other agents. In theory the firm may be able to increase its profits by relaxing this constraint.[17] However, when the number of agents is large then a 'large numbers' argument implies that in equilibrium the effect of introducing cross–dependence in the tariffs must

[17] In the taxation context, the possibility was considered by Roberts (1984, section 2). He demonstrates that these more sophisticated schemes — provided that they are anonymous — do no better than standard tax schedules with a random component which do *not* make a worker's tax liability depend on other workers' incomes. We show below in Result 3 that profits would not be increased if the firm introduced interdependencies between consumers.

vanish in the limit (provided that the scheme remains anonymous). We would not anticipate, therefore, that a large firm would benefit significantly from introducing cross–dependent tariffs.

At the other extreme, consider a 'firm' selling a single unit of a good to a few consumers. In practice this is often done using an 'ascending–bid' auction, where the bidding continues upwards at a steady rate until there is only one bidder willing to continue. In this case the seller forces agents to compete against each other in an extreme way.

In fact the theory of auctions can simply be incorporated into the adverse selection model of Section 4.1 above. Say that there are n potential buyers, $i = 1, \dots, n$. Consumer i's willingness–to–pay for the good is α_i, where *a priori* α_i is independently distributed across consumers with support $[\underline{\alpha}, \bar{\alpha}]$ and has the distribution function $F(\alpha)$ and density $f(\alpha)$. Agents are risk–neutral, so that if an agent of type α makes an expected payment of T and in return obtains the object with probability π, her utility is $\pi\alpha - T$. In principle, the class of mechanisms for allocating this object amongst the bidders is extremely large. However, Myerson (1981) demonstrates that without loss of generality the seller can restrict her attention to 'direct revelation mechanisms', where the bidders simply announce a type $\hat{\alpha}_i$ which they claim to be their own. Using this result, the most general auction mechanism is described by the functions $\pi_i(\boldsymbol{\alpha})$ and $t_i(\boldsymbol{\alpha})$, for $i = 1, \dots, n$, where $\boldsymbol{\alpha} \equiv (\alpha_1, \dots, \alpha_n)$ is the vector of reported values by the n consumers. Here, $\pi_i(\boldsymbol{\alpha})$ is the probability that i receives the good and $t_i(\boldsymbol{\alpha})$ is the price he must pay when the vector of announcements is $\boldsymbol{\alpha}$. Therefore

$$(16) \qquad \sum_1^n \pi_i(\boldsymbol{\alpha}) \leq 1, \quad \pi_i(\boldsymbol{\alpha}) \geq 0 \quad \text{for all } \boldsymbol{\alpha}.$$

We will restrict attention to symmetric auctions, so that π_i and t_i are all symmetric in the arguments of α_j, $j \neq i$ and π_i and π_j are the same functions except with the parameters permuted (and likewise for t_i and t_j). The seller herself places a value of c on the object and wishes to maximize expected profits $E[\sum (t_i(\boldsymbol{\alpha}) - c\pi_i(\boldsymbol{\alpha}))]$. There are constraints upon the functions π_i and t_i if they are to induce truthful announcements by consumers. Firstly, for i to wish to truthfully reveal her value, given that the other consumers do, it must be the case that

(17) α_i maximizes: $\{ \alpha_i \Pi(\hat{\alpha}_i) - T(\hat{\alpha}_i) \mid \underline{\alpha} \leq \hat{\alpha}_i \leq \bar{\alpha} \}$

where

$$\Pi(\hat{\alpha}_i) \equiv \int \cdot \cdot \int \pi_i(\alpha_1, .., \hat{\alpha}_i, ..., \alpha_n) f(\alpha_1) \cdot \cdot f(\alpha_{i-1}) f(\alpha_{i+1}) \cdot \cdot f(\alpha_n) \, d\alpha_1 \cdot \cdot d\alpha_{i-1} d\alpha_{i+1} \cdot \cdot d\alpha_n$$

and

$$T(\hat{\alpha}_i) \equiv \int \cdot \cdot \int t_i(\alpha_1, .., \hat{\alpha}_i, ..., \alpha_n) f(\alpha_1) \cdot \cdot f(\alpha_{i-1}) f(\alpha_{i+1}) \cdot \cdot f(\alpha_n) \, d\alpha_1 \cdot \cdot d\alpha_{i-1} d\alpha_{i+1} \cdot \cdot d\alpha_n$$

are the expected probability of obtaining the good and expected payment respectively, given i's announcement $\hat{\alpha}_i$ and the assumption that the other $n-1$ consumers report truthfully (by symmetry, these functions do not depend upon i). Finally, if consumers cannot be compelled to participate the seller must ensure that

(18) $\alpha \Pi(\alpha) - T(\alpha) \geq 0$ for all α.

Define $s(\alpha) = \alpha \Pi(\alpha) - T(\alpha)$ to be the expected surplus of the type α consumer. From (17) s is increasing, convex and satisfies $s'(\alpha) = \Pi(\alpha)$. As with Lemma 1 above, the surplus function $s(\cdot)$ is implementable if and only if it is increasing, non–negative and convex. Since $T(\alpha) = \alpha \Pi(\alpha) - s(\alpha)$ the seller's expected profit is

$$\Sigma E[\alpha_i \Pi(\alpha_i) - s(\alpha_i) - c\pi_i(\boldsymbol{\alpha})]$$

and she must maximize this subject to the constraint that $s'(\alpha) = \Pi(\alpha)$, $\Pi(\alpha)$ is non–decreasing, $s(0) \geq 0$ and (16). Writing this out explicitly, expected profit is

$$\int \cdots \int \{\sum_i^n [\alpha_i \pi_i(\boldsymbol{\alpha}) - s(\alpha_i) - c\pi_i(\boldsymbol{\alpha})]\} f(\alpha_1) \cdots f(\alpha_n)\, d\alpha_1 \cdots d\alpha_n$$

or

(19) $$\text{expected profit} = \int \cdots \int \{\sum_i^n [(\alpha_i - \frac{1 - F(\alpha_i)}{f(\alpha_i)} - c)\pi_i(\boldsymbol{\alpha})]\} f(\alpha_1) \cdots f(\alpha_n)\, d\alpha_1 \cdots d\alpha_n$$

where we have integrated the term in $s(\alpha_i)f(\alpha_i)$ by parts as in Section 4.1 using the fact that $s(\underline{\alpha}) = 0$ (as is optimal). Therefore, using the first–order approach we simply maximize the integrand of (19) pointwise with respect to each π_i subject to the constraint (16), so that the optimal probability functions $\pi_i{}^*$ satisfy

$$\pi_i{}^*(\boldsymbol{\alpha}) = \begin{cases} 1 \text{ if } \alpha_i - \frac{1 - F(\alpha_i)}{f(\alpha_i)} \geq \alpha_j - \frac{1 - F(\alpha_j)}{f(\alpha_j)} \;\forall j, \text{ and } \alpha_i - \frac{1 - F(\alpha_i)}{f(\alpha_i)} \geq c \\ 0 \text{ otherwise.} \end{cases}$$

When is such a scheme implementable? Suppose that the familiar condition (13) holds, and let α_{res} be the value of α that satisfies

(20) $$\alpha_{res} - \frac{1 - F(\alpha_{res})}{f(\alpha_{res})} = c\,.$$

In this case the above pointwise optimization reduces to setting

$$\pi_i{}^*(\boldsymbol{\alpha}) = \begin{cases} 1 \text{ if } \alpha_i \geq \alpha_j \;\forall j \text{ and } \alpha_i \geq \alpha_{res} \\ 0 \text{ otherwise,} \end{cases}$$

so that the object is awarded to the consumer with the highest valuation with probability one, provided that this valuation is above the reservation price α_{res}. In this case

$$\Pi(\alpha_i) = \text{prob}\{\alpha_i \geq \alpha_j \;\forall j\} = (F(\alpha_i))^{n-1}$$

which is certainly increasing α_i and so the first–order approach is valid. Finally, all that remains to do is to find a payment function $t_i^*(\boldsymbol{\alpha})$ which implements this scheme. Many such functions can be found, but the most obvious is the payment function corresponding to the Vickrey second–price (or, equivalently, the ascending–bid) auction, so that

$$(21) \qquad t_i^*(\boldsymbol{\alpha}) = \begin{cases} \max\{\alpha^{(2)}, \alpha_{res}\} & \text{if } \alpha_i \geq \alpha_j \ \forall j \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha^{(2)}$ is the second–highest valuation. Faced with this payment schedule it is certainly the dominant strategy of a consumer to participate in the auction if her valuation is greater than α_{res}, and in which case to announce her correct valuation. We sum up this discussion in the following result:

RESULT 2. *Suppose that consumers are risk–neutral, that valuations are identically and independently distributed by a density that satisfies (13), and let* α_{res} *be defined by (20). Then the Vickrey second–price (or ascending–bid) auction defined by (21) maximizes expected revenue for the seller over all auctions.*[18]

The implications of this result are:

(i) Since $\alpha_{res} > c$ in (20) there will cases when the optimal auction is inefficient *ex post* and there is too little trade (just as is the case with the traditional textbook monopolist);

(ii) that the optimal auction (including the reservation price) does not vary with the number of bidders (although the expected revenue for the seller rises with the number of bidders.) In particular, the optimal reservation price is the price the seller would offer a single buyer with uncertain willingness–to–pay;

(iii) that the bidder with the highest valuation gets the good (if the good is sold);

(iv) the seller is unable to extract all the surplus, and *ex ante* all bidders expect

18 For a heuristic explanation of this result using the language of third–degree price discrimination, see Bulow and Roberts (1989).

to obtain a positive surplus.

What is most interesting, however, is that the simple ascending–bid auction is optimal at the most rigorous theoretical level in such a wide range of circumstances.

This result depends upon some some special assumptions, namely: condition (13) must hold, and we need symmetry, risk–neutrality and independence of consumers values. In addition, there is only one unit of one good to be sold and there is only one seller. Attempts to relax these various assumptions has generated a voluminous literature — see the survey by McAfee and McMillan (1987) for instance. When (13) does not hold then the seller should choose the winner according to a random process (Myerson, 1981). Symmetry is quite easily dispensed with, whilst the case when agents are risk–averse is more problematic to deal with.[19] The single–good/single–seller assumption is crucial and extensions to the analysis beyond this are still at an early stage.[20] We now consider in turn what happens when two of these assumptions, that values are independent and that the (single) seller sells only one unit of the good, are relaxed.

Relaxing the assumption that values are independent results in many cases in a surprise: the seller can obtain the first–best allocation. To see that this can be so consider the following symmetric two–type/two–agent case. Each bidder has a value on the object which is either $\bar{\alpha}$ or $\underline{\alpha}$, where $\bar{\alpha} > \underline{\alpha}$, while the seller places no value on the good. The probability that one bidder has a low value given that the other has a low value is $\underline{p}$, and the probability that one has a low value given that the other has a high value is $\bar{p}$. The bidders' valuations are independent if and only if $\underline{p} = \bar{p}$. Let $t(\alpha_1, \alpha_2)$ be the payment from bidder 1 when she announces α_1 and her opponent announces α_2 (the payment from bidder 2 is just the same with subscripts swapped over). The mechanism we have in mind is that agents decide whether or not to participate in the auction prior to bidding, but once bidding has occurred agents are not permitted to leave if they make an *ex post* loss (i.e. the

[19] See the references in McAfee and McMillan (1987, section IX).

[20] For an analysis of the case where a single seller is auctioning more than one kind of good see McAfee and McMillan (1988), a paper which uses a similar kind of analysis to that used in Chapter 4 below, and Palfrey (1983), a paper discussed in Chapter 3 below. Wilson (1985a) examines double auctions (i.e. auctions where there are multiple sellers and buyers, all trading the same good), and this analysis is extended in Wilson (1992a), a paper which makes use of a technique proposed in an earlier draft of Chapter 4 below.

participation constraint is an *ex ante* one). The object is sold to the highest bidder and if there is a draw the object is allocated by flipping a coin. In order to elicit truthful responses and for the seller's revenue always to equal max: $\{\alpha_1, \alpha_2\}$ (as is required for efficiency), we should find a function t which satisfies:

$$0 = \underline{p}[\tfrac{1}{2}\underline{\alpha} - t(\underline{\alpha}, \underline{\alpha})] + (1 - \underline{p})[0 - t(\underline{\alpha}, \bar{\alpha})] \geq \underline{p}[\underline{\alpha} - t(\bar{\alpha}, \underline{\alpha})] + (1 - \underline{p})[\tfrac{1}{2}\underline{\alpha} - t(\bar{\alpha}, \bar{\alpha})]$$

and

$$0 = \bar{p}[\bar{\alpha} - t(\bar{\alpha}, \underline{\alpha})] + (1 - \bar{p})[\tfrac{1}{2}\bar{\alpha} - t(\bar{\alpha}, \bar{\alpha})] \geq \bar{p}[\tfrac{1}{2}\bar{\alpha} - t(\underline{\alpha}, \underline{\alpha})] + (1 - \bar{p})[0 - t(\underline{\alpha}, \bar{\alpha})].$$

The equalities in the above mean that the bidders *ex ante* have no surplus if they tell the truth, whereas if they do not tell the truth they do even worse. Substituting out $t(\underline{\alpha}, \bar{\alpha})$ and $t(\bar{\alpha}, \underline{\alpha})$ implies that we require

$$\underline{p}[\underline{\alpha} - ((1 - \bar{p})/\bar{p})(\tfrac{1}{2}\bar{\alpha} - t(\bar{\alpha}, \bar{\alpha})) - \bar{\alpha}] + (1 - \underline{p})(\tfrac{1}{2}\underline{\alpha} - t(\bar{\alpha}, \bar{\alpha})) \leq 0$$

and

$$\bar{p}[\tfrac{1}{2}\bar{\alpha} - t(\underline{\alpha}, \underline{\alpha})] - (1 - \bar{p})(\underline{p}/(1 - \underline{p}))(\tfrac{1}{2}\underline{\alpha} - t(\underline{\alpha}, \underline{\alpha})) \leq 0 .$$

It is clear that we can find numbers $t(\underline{\alpha}, \underline{\alpha})$ and $t(\bar{\alpha}, \bar{\alpha})$ that satisfy these inequalities if and only if $\underline{p} \neq \bar{p}$, i.e. if and only if the bidders' valuations are not independent.[21] The fact that the first–best auction is feasible when values are not independent is an interesting result. It does, however, depend quite crucially upon the assumptions of the risk–neutrality of the agents and an *ex ante* participation constraint, since when the values

[21] This argument appears originally to have been made by Demski and Sappington (1984) in the context of a regulated duopoly, and was introduced to the auction setting by Crémer and McLean (1985, 1988). The latter extend the argument to the case where agents have finitely many — rather than just two — possible valuations, although a stronger condition than the agents' valuations not being independent is then needed. McAfee and Reny (1992) extend further to the case where values are continuously distributed. Dana (1992) uses the framework of Demski and Sappington but assumes that the participation constraint is of the stronger *ex post* kind. He poses the question of whether a two–good industry should be regulated as a single firm or as a duopoly when industry costs are correlated. The former structure has superior risk–pooling properties whereas the latter enables there to be competition among the firms as with Demski and Sappington. The latter effect will outweigh the former whenever the degree of positive correlation is sufficiently large.

are almost independent (i.e. $\underline{p} \approx \bar{p}$) then the payment function t becomes very large in absolute value and the agent who loses may pay a large amount without receiving anything in return.

Turning next to the case where the firm can sell many units of the same good, we write the utility of bidders as in Section 2, and so if quantity q is consumed for an expected payment of T, the utility of a type α bidder is $\alpha u(q) - T$. For each unit of the good sold the seller incurs a cost c (and there are no capacity constraints). The seller's mechanism is to allocate a quantity $q_i(\boldsymbol{\alpha})$ and demand a payment $t_i(\boldsymbol{\alpha})$ from bidder i when the vector of announced types is $\boldsymbol{\alpha}$ (so we restrict attention to deterministic strategies although, as before, this is not a serious restriction as long as (13) holds). The analysis above for the case when the seller had only a single unit to sell all works in this multi–unit framework, and we can write the seller's total expected profit as:

$$(22) \qquad \int \cdot \cdot \int \{\sum_i^n [(\alpha_i - \frac{1 - F(\alpha_i)}{f(\alpha_i)})u(q_i(\boldsymbol{\alpha})) - cq_i(\boldsymbol{\alpha})]\} f(\alpha_1) \cdot \cdot f(\alpha_n)\, d\alpha_1 \cdot \cdot d\alpha_n .$$

As before, a response function $q_i(\boldsymbol{\alpha})$ is implementable if and only if

$$(23) \qquad \int \cdot \cdot \int q_i(\alpha_1, .. , \alpha_i, ... , \alpha_n)\, f(\alpha_1) \cdot \cdot f(\alpha_{i-1}) f(\alpha_{i+1}) \cdot \cdot f(\alpha_n)\, d\alpha_1 \cdot \cdot d\alpha_{i-1} d\alpha_{i+1} \cdot \cdot d\alpha_n$$

is increasing in α_i.

Using the first–order approach to maximize (22) implies that the optimal response function $q_i^*(\boldsymbol{\alpha})$ satisfies (12) above. Provided that (13) holds, q_i^* satisfies (23) and so this scheme is implementable. The response function which solves this auction problem is precisely the same as the one which solves the problem of finding the optimal nonlinear tariff in Section 4.1, a problem where we made the implicit assumption that the payment by an agent should only depend upon the action of that agent. We sum this discussion up as a result:

RESULT 3. *Suppose there are* n *consumers with utility of the form* $\alpha_i u(q) - t_i$, *where* α_i *is an independent draw from the distribution function* $F(\alpha)$ *which satisfies (13). The firm has unit cost* c. *Then the profit–maximizing response functions* $q_i^*(\boldsymbol{\alpha})$, $i = 1, \ldots, n$, *are functions only of* α_i: *making one agent's payoff depend upon another's action cannot increase profits.*

Thus we see that the firm has no incentive to make agents compete for its output when it faces no capacity constraint on the number of units it has to sell, even if the number of agents is small. Competitive auctions are optimal when the firm faces a capacity constraint (for instance, it has one unit to sell) rather than when it faces only a small number of consumers.[22]

8. Introducing multidimensional parameters: the case of signalling

In all three of the models of adverse selection described in Section 2 we made the implicit assumption that agents' type parameters (respectively w, θ and α) were scalars. In effect we assumed that agents were not really very different. In the third problem, for instance, since an agent's utility function took the form $\alpha u(q) + y$, once a particular agent's utility from consuming a given quantity is known, her entire utility function could be deduced. Similarly, in the second problem, once we know how much it costs the firm to produce a given quantity we know its entire cost function. Whilst this assumption is a natural first step towards an understanding of the effect of asymmetric information, it would be desirable to see how robust the results obtained in the scalar case are in more general frameworks.

One way in which we might generalize these models would be to allow the type

[22] The analysis of a principal facing competing agents has also been fruitful in the context of moral hazard. The simplest model is Shleifer (1985) where there are n identical firms acting in a non–stochastic environment and can reduce their costs by expending effort (which is unobservable by the regulator). By making the allowed price for one depend upon the observed costs of the other firms the first–best outcome is obtained. Hart (1983) has a model where managers' remuneration should be made partly on the basis of relative performance compared with other managers in order to reduce slack. More generally, Green and Stokey (1983) have a model in which 'tournaments', where an agent's payment is based entirely upon their performance ranking compared with other agents, are approximately optimal when the number of agents becomes large.

parameters to be multidimensional. In the income tax model it may be that there are several different sources of income a worker may have, for instance from different jobs or from savings, and workers will have different productivities in generating these incomes. In the setting of monopoly regulation, if the firm sold several products it seems unreasonable to subsume all uncertainty about the firm's cost function into a single parameter. Using the terminology of that section, a natural way to model the firm's costs might be to set

$$C(\boldsymbol{\theta}, \mathbf{q}) = \sum_{1}^{n} \theta_i c_i(q_i)$$

where $\mathbf{q} = (q_1, .. \, q_n)$ is the vector of outputs of the firm and $\boldsymbol{\theta} = (\theta_1, .. \, \theta_n)$ is the vector of parameters describing the private information of the firm, where the uncertainty of the i^{th} production process is described by θ_i. Similarly, in the third model if the firm was a multiproduct monopoly offering products to consumers with uncertain demands, we could set the utility function of a consumer of type $\boldsymbol{\alpha} = (\alpha_1, .. \, \alpha_n)$ to be

$$U(\boldsymbol{\alpha}, \mathbf{q}) = \sum_{1}^{n} \alpha_i u_i(q_i) - T$$

where α_i now represents a consumer's strength of feeling towards the i^{th} product.

When we describe the three models above it does not make much difference if the parameters are scalars or vectors. We can represent the principal's payoff in terms of the agents' surplus and response functions as before, there is a corresponding envelope condition (i) and the implementability constraint (ii) has a simple generalization. For instance, the statement of problem (P_3) when consumer types are vectors and the firm is a multiproduct monopoly is that the firm should choose the vector of optimal response functions $q_i{}^*(\boldsymbol{\alpha})$, $i = 1, \ldots, n$, in order to:

(P_3^n) maximize: $E[\sum_1^n \alpha_i u_i(q_i(\boldsymbol{\alpha})) - s(\boldsymbol{\alpha}) - \sum_1^n c_i q_i(\boldsymbol{\alpha})]$ subject to

(i) $\frac{\partial s}{\partial \alpha_i}(\boldsymbol{\alpha}) = u_i(q_i(\boldsymbol{\alpha}))$ for each i (and $q_i \geq 0$)

(ii) $s(\boldsymbol{\alpha})$ is convex, and

(iii) $s(\boldsymbol{\alpha}) \geq 0$

where E denotes expectation with respect to the joint distribution of types given by the distribution function $F(\boldsymbol{\alpha})$ (see Chapter 4 for more detail). What distinguishes the multidimensional case from the scalar case most strongly is that problem (P_3^n) is simply much more difficult to actually *solve* than (P_3) above. We shall see in Chapter 4 that certain rather striking results for the multidimensional case emerge if the first–order approach is valid (the first–order approach in this context involves solving the problem (P_3^n) ignoring constraint (ii)); however, in contrast to the scalar case (where the simple condition (13) determined whether the first–order approach worked) there is so far no way of telling whether this approach will be valid *a priori* in the multiproduct case.

As a means by which to introduce the special features of multidimensional adverse selection models we discuss the signalling model introduced by Spence (1973) with a scalar parameter, and generalized to the case of several parameters in an elegant paper by Quinzii and Rochet. Workers are heterogeneous and are indexed initially by the scalar parameter α. A worker of type α has a productivity of $w(\alpha)$ (an increasing function). Employers are unable to observe α, and so workers attempt to signal their productivity by obtaining the signal (or level of qualification) q, also a scalar initially. (In the simplest model this signal has no impact upon a workers productivity.) Workers of higher type find it less costly to obtain a given signal q and so employers might expect that a high signal will be correlated with high productivity. Specifically, let the cost of obtaining signal q given α be $c(\alpha, q)$. Since employers cannot directly observe α they can offer a wage schedule dependent only upon q, say $W(q)$. Faced with such a schedule, workers wish to maximize their net income and so the type α worker will choose her signal $q(\alpha)$ in order to

(24) $$\underset{q}{\text{maximize}} \quad W(q) - c(\alpha, q) .$$

One strong candidate for an equilibrium is where workers are completely sorted so that the map $\alpha \mapsto q(\alpha)$ sends different types to different signals. In this case employers can deduce the productivity $w(\alpha)$ of a worker with signal $q(\alpha)$, and if employers are competitive (so that equilibrium profits are zero) then $W(q(\alpha)) = w(\alpha)$: a worker's wage is equal to her productivity. This fact, in combination with (24) determines a signalling equilibrium. Therefore, $q(\cdot)$ is the equilibrium level of signalling if

(25) $$\alpha \ \underset{\tilde{\alpha}}{\text{maximizes}} \ w(\tilde{\alpha}) - c(\alpha, q(\tilde{\alpha})).$$

Spence showed that, subject to some technical assumptions, such a separating equilibrium will always exist. The vital condition for existence is that the following 'single–crossing' property of the cost function c holds:

(26) $$\frac{\partial^2}{\partial\alpha\partial q} c(\alpha, q) < 0 \quad \text{for all } (\alpha, q).$$

This condition ensures that, given any wage schedule $W(q)$, the optimal response function $q(\alpha)$ will be non–decreasing in α. For instance, if $c(\alpha, q) = q/\alpha$ and $w(\alpha) \equiv \alpha$ then a separating equilibrium exists and signalling schedules are of the form $q(\alpha) = \frac{1}{2}\alpha^2 +$ constant.[23]

A multidimensional version of this model can be set up with no great difficulty. Here workers have multiple characteristics $\boldsymbol{\alpha} = (\alpha_1, .. \alpha_n)$, the productivity of a type $\boldsymbol{\alpha}$ worker is $w(\boldsymbol{\alpha})$ and there are n signals $\mathbf{q} = (q_1, .. q_n)$. Quinzii and Rochet (1985) have examined this model for the special case where the cost of gaining the signal $\mathbf{q}$ for the

[23] This 'constant' remains undetermined within Spence's model (when the distribution of types has compact support) and so there are multiple signalling equilibria. Since higher values of this constant make all workers worse off and leave employers indifferent, these equilibria can be Pareto ranked.

type α worker is $\sum_1^n q_i/\alpha_i$ or, making a convenient change of variables to $\beta_i \equiv 1/\alpha_i$:

$$(27) \qquad c(\beta, \mathbf{q}) = \sum_1^n \beta_i q_i .$$

The productivity of a worker of type β is $w(\beta)$, where $w(\cdot)$ is now decreasing in β (after making the change of variables from α to β) and is assumed to be bounded. Again, we look for a separating equilibrium with zero profits. If employers offer the wage schedule $W(\mathbf{q})$, the type β worker will obtain a surplus of

$$(28) \qquad s(\beta) \equiv \max\colon \{ W(\mathbf{q}) - \beta'\mathbf{q} \mid \mathbf{q} \geq 0 \} .$$

If workers obtain a utility of zero if they withdraw from the labour market then we can suppose that $W(0) = 0$. In this case the surplus function $s(\cdot)$ in (28) is non–negative, non–increasing, convex and (where differentiable) satisfies the envelope condition $\frac{\partial s}{\partial \beta_i} = -q_i(\beta)$, where $\mathbf{q}(\beta)$ is the vector response function of the type β worker when faced with the wage schedule W. Since at competitive equilibrium we require that $W(\mathbf{q}(\beta)) = w(\beta)$, and from the identity $s(\beta) \equiv W(\mathbf{q}(\beta)) - \beta'\mathbf{q}(\beta)$ we see that the surplus function $s(\cdot)$ must satisfy the partial differential equation

$$(29) \qquad s(\beta) = w(\beta) + \beta'\nabla s(\beta)$$

(where ∇s is the vector of derivatives of s). That is to say, at a competitive equilibrium the surplus function of workers $s(\beta)$ is necessarily non–negative, convex, non–increasing and satisfies equation (29) (almost everywhere). Quinzii and Rochet show that these conditions are also necessary for equilibrium, and hence that there is a one–to–one correspondence between signalling equilibria and convex, non–negative, non–increasing solutions of (29). They also show that the *only* non–negative and non–increasing (but not necessarily convex) solution to (29) is given by:

$$(30) \qquad s(\beta) = \int_1^\infty w(t\beta)/t^2\, dt$$

(where, importantly, they assume that any $\beta > 0$ is possible). Therefore, there exists a signalling equilibrium if and only if s as defined in (30) is convex, and such an equilibrium is then unique.[24] Whether s is convex or not in (30) is dependent upon the productivity function w. So far all of the above reasoning is valid for both the single and multi–dimensional cases. Where the two cases diverge is the circumstances when s in (30) is convex:

— in the single–dimensional case considered by Spence s is always convex (provided of course that w is decreasing), so that therefore there always exists a signalling equilibrium if the cost function is as given in (27)

— when $n \geq 2$, on the other hand, this is not true. A sufficient condition for s in (30) to be convex is that the productivity function $w(\beta)$ be convex, and in such cases a signalling equilibrium will exist. However, taking $w(\beta) = 2/\|\beta\|$ (where $\|.\|$ is the standard Euclidean norm) means that s in (30) is not convex, and in this case no separating signalling equilibrium exists.

Thus, we see that there is a qualitative difference in the scalar and multidimensional signalling models, and in particular that existence problems seem likely to be more pronounced in the latter case.[25]

Quinzii and Rochet have managed an effective extension of Spence's model to a more general setting of multiple signals and types. While this has been done at the cost of a slight increase in mathematical difficulty the additional complexity is by no means insurmountable. Can we expect, then, that the transition to multiple types will be

---

[24] This uniqueness result contrasts with the earlier footnote describing the non–uniqueness of equilibria in the Spence model. However, Spence assumed that the distribution of types had compact support. When types can take any positive value (as was the case when expression (30) was derived) the equilibrium in his model is also unique.

[25] For more on multidimensional signalling see Wilson (1985b), Milgrom and Roberts (1986) and Engers (1987).

managed as smoothly in the other models of adverse selection? The answer is an unfortunate 'No'. The signalling model is significantly simpler than that of income taxation, regulation or monopoly pricing in that in order to solve the model we merely have to find a (convex) solution to a particular differential equation, not an optimal solution. In the latter three models there is a full two–stage optimization, the principal maximizing her payoff allowing for the fact the agents are maximizing their's. The effect of perfect competition among employers in the signalling model is to bypass the first optimization problem. In particular, the distribution of characteristics amongst the workforce plays no role in these signalling models. We shall see in Chapter 3 and especially Chapter 4 below that obtaining solutions in these other multidimensional adverse selection models can be quite problematic.

9. Conclusions

During the 1950s much work by economic theorists was concerned with models of 'general equilibrium' in a world where there were a complete set of Arrow–Debreu markets, or, what is the same thing, perfect information. Elegant results concerning the existence, uniqueness, stability and welfare properties of a perfectly competitive equilibrium were rigorously proved. The two main objections to this paradigm were, firstly, the prevalence of imperfect competition in actual markets, and secondly, the prevalence of imperfect information in actual markets. One way to tackle the problem of imperfect information has been to continue to use the language of general equilibrium theorists but allow some of the Arrow–Debreu markets to be missing. Another approach has been to focus more directly on the problems posed by imperfect information in a partial equilibrium framework, and this has been the topic of this survey chapter. The literature in this area has become vast in the past two decades, and I have completely ignored several vital areas, including bargaining, most settings involving moral hazard and many applications. I hope, though, it has been shown that the explicit modeling of the economic effects of incomplete information is necessary for a full understanding of many markets, and that the constraints imposed by imperfect information should be treated by economists with the same seriousness as the traditional resource constraints considered by the general equilibrium theorists.

CHAPTER 2

SINGLE–PRODUCT NONLINEAR PRICING AND REGULATION

This chapter is my account of joint work with Simon Cowan and John Vickers. My main contributions were the introduction of the 'revenue function' into the analysis and the proofs of all Lemmas and Results given here. The worked example of Section 5 was suggested by John Vickers and is included here for completeness. In addition, the section on average revenue regulation is parallel to Section 3 in Armstrong and Vickers (1991), and many of the ideas in this chapter stem from work on that earlier paper.

1. Introduction

Many aspects of single–product nonlinear pricing are now well understood by economists. As the previous chapter tried to make clear, nonlinear pricing is best seen in the context of a family of adverse selection problems in which the principal faces a variety of different kinds of agents but is unable to distinguish these agents from one another. Because of this, the principal must offer each agent the same contract which, in the present context, is a nonlinear tariff specifying the charge any consumer must pay in order to purchase a given quantity of the good. The two problems of designing the profit–maximizing nonlinear tariff (i.e. when the firm is the principal) and the welfare–maximizing nonlinear tariff (i.e. when a benevolent government is the principal) are much the same analytically and have been studied by various authors, including Spence (1977), Mussa and Rosen (1978), Roberts (1979), Goldman *et al.* (1984), Maskin and Riley (1984) and Wilson (1992b). The framework used is broadly similar to that described in Section 4.1 of the previous chapter and involves consumers varying by scalar parameter to capture the effect of asymmetric information. The main results have been

- discovering ways in which the problem may be solved (which is typically using the first–order approach)
- showing that in many cases the firm will wish to completely separate its customers (so that different customers buy different quantities)
- that the highest demand customers are served efficiently
- that in many cases it is optimal to offer a tariff with quantity discounts.

The purpose of this chapter is twofold: first, it will be useful to build on the preliminary discussion of nonlinear pricing given in the previous introductory chapter before moving onto the chapters below involving more than one product; and second, many of the leading examples of firms which practice nonlinear pricing — telecommunications and electricity supply companies for instance — are firms that are also regulated, and it would therefore be interesting to see in what way the previous analysis carries over to such cases.

This chapter is not the first attempt to do this, and aspects of nonlinear pricing by regulated firms have already been analyzed by several authors. Since there are various possible methods of regulating a firm, different authors have analyzed different kinds of regulation. Katz (1983) uses a discrete model to examine how minimum output regulation (i.e. regulation which requires the firm to produce at least a certain level of total output but leaves the choice of tariff otherwise unconstrained) could be used to improve welfare when a monopoly uses nonlinear tariffs. He shows that in many cases (though not all), forcing the firm to increase its output, starting from the position of the unregulated profit–maximizing nonlinear tariff, will increase overall welfare. One of the two forms of regulation considered in this chapter is average revenue regulation, where the firm can offer a nonlinear tariff only if the average revenue per unit of output is not greater than some price cap. It turns out that the kind of minimum output regulation proposed by Katz is precisely equivalent (in most cases) to average revenue regulation and so Katz in effect shows that average revenue regulation of a firm offering nonlinear tariffs will usually increase welfare compared to the situation with no regulation of any kind.

Srinagesh (1986) examines the effects of nonlinear pricing in the context of rate of return regulation, i.e. when the firm's profit per unit of capital employed is constrained to lie below some reference rate of return. He shows that high demand consumers are served at a price below marginal cost, a result we shall see is repeated under average revenue regulation. In addition there is the Averch–Johnson (1962) over–capitalization effect. Srinagesh constructs an example to show that nonlinear pricing can be superior in welfare terms to uniform pricing with the same rate of return.

Another form of regulation is considered by Willig (1978). He supposes that the firm is permitted to offer any tariff it chooses but is required also to offer some given uniform tariff as an option to consumers. Provided that the firm faces no competition and that buyers are final consumers such regulation is Pareto improving. (Ordover and Panzar (1980) show that if the firm's output is an intermediate good then this need not hold.) The crucial difference between Willig's optional nonlinear tariffs and nonlinear tariffs subject to average revenue (or minimum output) regulation is that in the former the firm is not permitted to trade off a higher charge for low users against a lower charge for high users — the firm must offer *all* consumers a (weakly) lower tariff.

Another closely related work is the analysis of Sappington and Sibley (1992). They use a dynamic model, restrict attention to two–part tariffs and consider regulation which caps the firm's average revenue using the previous year's output weights (this is a kind of average revenue regulation). This encourages strategic behaviour by the firm which may or may not cause discounted consumer surplus to fall over time. Price may be below marginal cost in some circumstances. They also examine the possibility of the firm offering optional two–part tariffs (as in Willig's paper), and demonstrate that with two types of consumer the firm will choose to offer the high–demand type a marginal price equal to marginal cost, whilst the low–demand type's situation remains unchanged. Finally, Armstrong and Vickers (1991, Section 3) considered the parallel problem of the welfare effects of third–degree price discrimination when there is average revenue regulation. Similarities between that paper and the section on average revenue regulation below will be much in evidence.

The model, which is set out in Section 2, is static. There is a continuum of consumers, who differ by a single parameter. The revenue function — i.e. the function which gives the maximum revenue available from a given level of total output — is analyzed (variants of this function were also used in the papers by Katz and Srinagesh), and it is shown that, subject to a hazard rate condition, the revenue–maximizing tariff will offer quantity discounts. In Section 3 the familiar unregulated monopoly case is briefly surveyed. In Section 4 two kinds of regulation are considered: average revenue regulation (Section 4.1) and optional tariffs (Section 4.2). In the former case, the constrained profit–maximizing tariff is found to be a revenue–maximizing tariff and so, in particular, it will often involve quantity discounts. Also, for most plausible cases, average revenue regulation is found to be equivalent to Katz's minimum output regulation. Consumers as a whole will certainly lose as a result of the move from a uniform tariff to a nonlinear tariff with the same average revenue. Total welfare will also decrease if the uniform tariff is close to the first–best tariff. More consumers will be excluded from the market when nonlinear tariffs are introduced. Total output will, however, rise. Indeed, all sufficiently high demand consumers will be served with higher quantities than is efficient (as was the case in Srinagesh's paper). The overall welfare effect is ambiguous. In Section 4.2, we find the profit–maximizing optional tariff which typically will also offer quantity discounts. It will not be a tariff that maximizes revenue, although quantity discounts will continue to be offered. It has the property that its marginal price is the lower of (i) the price cap $\bar{p}$, and (ii) the optimal marginal price for the unregulated monopolist. In particular, the highest type consumers are efficiently served (as in the Sappington and Sibley paper). Clearly, such regulation represents a Pareto improvement over the single regulated uniform tariff. However, optional tariffs need not be superior to the nonlinear tariffs induced by average revenue regulation. In Section 5 there is a worked example that illustrates all of these points. Brief conclusions follow in Section 6.

2. The model

We consider a single–product industry which is monopolized by a single firm able to offer nonlinear tariffs. The firm's costs are a function only of its total output Q, and written as $C(Q)$. The firm chooses a tariff schedule $T(q)$, which specifies for each quantity q the total sum that a consumer purchasing q units must pay to the firm. (A special case is a uniform tariff, where $T(q) = pq$ for some price p.) As in Chapter 1, consumers are indexed by a one–dimensional type parameter α, where $\alpha \in [\underline{\alpha}, \bar{\alpha}]$: a consumer of type α has a utility function $\alpha u(q) - T$ where q is consumption and T is payment. The function $u(\cdot)$ is increasing, strictly concave and satisfies $u(0) = 0$. The firm cannot distinguish the type of a given individual, but it knows that the distribution of types is $F(\alpha)$, with density $f(\alpha)$. For technical reasons we assume that the support of α is compact, so that $\bar{\alpha}$ is finite.

If the firm offers a tariff schedule $T(q)$, a type α consumer will have surplus of $s(\alpha)$ given by

$$s(\alpha) \equiv \max_{q \geq 0}: \alpha u(q) - T(q) .$$

Since consumers always have the option to consume nothing at a cost of zero, without loss of generality we can assume that the firm is constrained to ensure that $T(0) \leq 0$. From the definition of $s(\alpha)$, then, it satisfies:

— s is non–negative, increasing and convex

— if s is differentiable at α, the optimal quantity choice for α — which is denoted by $q(\alpha)$ — is unique and given by $s'(\alpha) = u(q(\alpha))$.

In fact, because s is convex it will necessarily be differentiable almost everywhere. The demand function $q(\cdot)$ is then well–defined almost everywhere and is necessarily increasing in α from the convexity of s. This means that we can write the revenue $T(q(\alpha))$ from a type α consumer as $T(q(\alpha)) \equiv \alpha u(q(\alpha)) - s(\alpha)$, an identity which holds almost everywhere.

The revenue function, $R(Q)$, is defined for each level of total output Q to be the maximum total revenue obtainable by the firm when selling output Q by means of some nonlinear tariff $T(q)$. More formally, set

$$R(Q) = \max \int_{\underline{\alpha}}^{\bar{\alpha}} T(q(\alpha))f(\alpha)\,d\alpha \quad \text{subject to}$$

$$\text{(i)} \quad q(\alpha) \text{ maximizes } \alpha u(q) - T(q) \text{ and } \alpha u(q(\alpha)) - T(q(\alpha)) \geq 0$$

$$\text{(ii)} \quad \int_{\underline{\alpha}}^{\bar{\alpha}} q(\alpha)f(\alpha)\,d\alpha = Q\,.$$

Constraint (i) ensures that the response function is incentive compatible while (ii) requires that total output equals Q. Let $T^*(\cdot)$ be the tariff which solves the above problem (clearly, this depends upon Q). Then T^* is solved using the method of Section 4.1 of the previous chapter as follows. Let T be any tariff. Since $T(q(\alpha)) \equiv \alpha u(q(\alpha)) - s(\alpha)$ we can write

$$\int_{\underline{\alpha}}^{\bar{\alpha}} T(q(\alpha))f(\alpha)\,d\alpha = \int_{\underline{\alpha}}^{\bar{\alpha}} \{\alpha u(q(\alpha)) - s(\alpha)\}f(\alpha)\,d\alpha.$$

This expression holds however many consumers choose to participate. Since $s'(\alpha) = u(q(\alpha))$, $F(\bar{\alpha}) = 1$ and $s(\underline{\alpha}) = 0$ (as is clearly optimal), we can integrate the $s(\alpha)$ term in the above by parts to get

$$\int_{\underline{\alpha}}^{\bar{\alpha}} T(q(\alpha))f(\alpha)\, d\alpha = \int_{\underline{\alpha}}^{\bar{\alpha}} [\alpha u(q(\alpha))f(\alpha) - (1 - F(\alpha))u(q(\alpha))]\, d\alpha.$$

Writing total revenue in this form we can define the revenue function in terms of the response function $q(\alpha)$ as

$$R(Q) = \max \int_{\underline{\alpha}}^{\bar{\alpha}} [\alpha f(\alpha) - (1 - F(\alpha))]u(q(\alpha))\, d\alpha \quad \text{subject to}$$

(i) $q(\alpha)$ non–decreasing, and

(ii) $\int_{\underline{\alpha}}^{\bar{\alpha}} q(\alpha)f(\alpha)\, d\alpha = Q$.

Constraint (i) is the incentive compatibility constraint: in order for $q(\cdot)$ to be the result of optimizing behaviour by consumers it is both necessary and sufficient that $q(\cdot)$ be non–decreasing — see Lemma 1 of the last chapter. In most cases we will ignore this constraint initially but verify that it is satisfied *ex post*, i.e. we will use the first–order approach.

Introducing $\lambda(Q)$ for the multiplier for constraint (ii), then if we do ignore (i), we simply maximize the integrand of

$$\int_{\underline{\alpha}}^{\bar{\alpha}} \{[\alpha f(\alpha) - (1 - F(\alpha))]u(q(\alpha)) - \lambda(Q)f(\alpha)[q(\alpha) - Q]\}\, d\alpha$$

pointwise with respect to q so that the optimal response function, denoted $q^*(\alpha)$, satisfies:

(2) $$q^*(\alpha) \text{ maximizes: } \max_{q \geq 0} u(q)[\alpha - \frac{1 - F(\alpha)}{f(\alpha)}] - \lambda(Q)q$$

where $\lambda(Q)$ is then given implicitly by the constraint that q^* in (2) satisfies $\int q^*(\alpha)f(\alpha)\, d\alpha = Q$.[1]

The condition which ensures that the maximand of (2) is non–decreasing as required is the following assumption on the distribution of types

(3) $$\alpha - \frac{1 - F(\alpha)}{f(\alpha)} \text{ is non–decreasing in } \alpha.$$

LEMMA 1. *Under assumption (3) the function* $R(Q)$ *is increasing and strictly concave.*

PROOF. If condition (3) holds then the optimal response function q^* is as given in (2) for each Q. In particular, if $\alpha - \frac{1 - F(\alpha)}{f(\alpha)}$ is negative then $q^*(\alpha) = 0$ for any Q. Let Q_1 and Q_2 be two output levels and $0 \leq t \leq 1$. Let $q_i^*(\cdot)$ be the optimal response function for output level Q_i, so that each $q_i^*(\cdot)$ is non–decreasing and $\int q_i^*(\alpha)f(\alpha)\, d\alpha = Q_i$. Then $tq_1^* + (1-t)q_2^*$ is a possible choice of response function for the intermediate output level $tQ_1 + (1-t)Q_2$,[2] and therefore

$$R(tQ_1 + (1-t)Q_2) \geq \int_{\underline{\alpha}}^{\bar{\alpha}} [\alpha f(\alpha) - (1 - F(\alpha))]u(tq_1^*(\alpha) + (1-t)q_2^*(\alpha))\, d\alpha$$

$$\geq tR(Q_1) + (1-t)R(Q_2) \; ,$$

the second inequality resulting from the assumption that $u(\cdot)$ is concave and the fact that $\alpha f(\alpha) - (1 - F(\alpha))$ being negative implies that each $q_i^*(\alpha)$ is zero. Since u is strictly concave, the above inequality is strict unless $q_1^* = q_2^*$ almost everywhere. However, $Q_1 \neq Q_2$ implies that $\int q_1^*(\alpha)f(\alpha)\, d\alpha \neq \int q_2^*(\alpha)f(\alpha)\, d\alpha$ and in particular that the two

---

[1] This expression just like (12) in Chapter 1 except that unit cost c is replaced the multiplier $\lambda(Q)$.

[2] Since each q_i^* is non–decreasing and satisfies $\int q_i^*(\theta)f(\theta)\, d\theta$ then $tq_1^* + (1-t)q_2^*$ is also non–decreasing and satisfies $\int [tq_1^* + (1-t)q_2^*]f(\theta)\, d\theta = tQ_1 + (1-t)Q_2$.

response functions q_i^* cannot coincide. Therefore, R is strictly concave. A similar argument demonstrates that R is also increasing. □

Since the Envelope Theorem implies that $R'(Q) \equiv \lambda(Q)$, the above lemma implies that the multiplier $\lambda(Q)$ is positive and strictly decreasing in output.

The following technical lemma will be useful in Section 4.1:

LEMMA 2. *Under assumption (3)*

(i) $u'(0) = +\infty \implies R'(0) = +\infty$

(ii) $u'(+\infty) = 0 \implies R'(+\infty) = 0.$

PROOF (sketch).

(i) We prove the equivalent claim that $R'(0) < +\infty \Longrightarrow u'(0) < +\infty$. Since R is concave, $\lim_{Q \to 0} R'(Q)$ exists and suppose that this limit is finite so that $\lambda(Q) = R'(Q) \leq M < \infty$ for all $Q > 0$. Then under assumption (3) the optimal response function q^* is given by (2). For all α close to $\bar{\alpha}$, $\alpha - \frac{1 - F(\alpha)}{f(\alpha)}$ is also close to $\bar{\alpha}$. Therefore, $q^*(\alpha)$ is close to the maximizing value of $\bar{\alpha}u(q) - \lambda(Q)q$ which is greater than the maximizing value of $\bar{\alpha}u(q) - Mq = \hat{q}$ say. More formally, we can find some $\delta > 0$ such that $q^*(\alpha) \geq \frac{1}{2}\hat{q}$ for $\alpha > \bar{\alpha} - \delta$ and for all $Q > 0$. Therefore, $Q = \int q^* f \, d\alpha \geq \frac{1}{2}[1 - F(\bar{\alpha} - \delta)]\hat{q}$ for all $Q > 0$, and so in particular we must have $\hat{q} = 0$. From the definition of $\hat{q}$ this implies that $u'(0) \leq M/\bar{\alpha} < \infty$ as claimed.

(ii) Similar. □

The example in Section 5 illustrates these two lemmas. There, $u(q) = \frac{1}{2}\sqrt{q}$, which satisfies the conditions of Lemma 2, and the parameter α is exponentially distributed. It is shown that the revenue function is $R(Q) = \sqrt{20Q}$, a function which is strictly concave and has a slope which decreases from infinity to zero.

What can we deduce about the shape of the nonlinear tariff $T^*(q)$ which solves the revenue maximization problem? Just as with (14) of Chapter 1 we have:

RESULT 1. *Under assumption (3), given any output level* Q *the revenue–maximizing marginal price schedule* $p^*(\cdot)$ *satisfies*

(4) $$p^*(q^*(\alpha)) = \lambda(Q)/[1 - \frac{1 - F(\alpha)}{\alpha f(\alpha)}].$$

COROLLARY 1. *If*

(5) $$\frac{1 - F(\alpha)}{\alpha f(\alpha)} \text{ is decreasing}$$

then for each level of total output Q *the revenue–maximizing tariff* $T^*(q)$ *is strictly concave.*

A comparison of interest between uniform pricing and tariffs with quantity discounts is the individual comparison of who wins and who loses in the move from one regime to another. This is described in the following lemma:

LEMMA 3. *Let* $T(q)$ *be some strictly concave tariff and* $\bar{T}(q) \equiv \bar{p}q$ *be some uniform tariff such that* $T'(0) > \bar{p}$. *Then there exists some* $\hat{\alpha}$ *such that all consumers of type* $\alpha < \hat{\alpha}$ *are (weakly) worse off and all types* $\alpha > \hat{\alpha}$ *are better off with the uniform tariff.*

PROOF. If $T(q) > \bar{T}(q)$ for all $q > 0$ then the claim holds trivially, so suppose that the two tariffs do cross at $\hat{q} > 0$ say. Since T is strictly concave $\hat{q}$ is the unique positive crossing point and $T(q) \geq \bar{p}q$ if and only if $q \leq \hat{q}$. Let s and $\bar{s}$ be the corresponding surplus functions with demand functions q and $\bar{q}$. By revealed preference the following hold:

$$\bar{q}(\alpha) \geq \hat{q} \implies \bar{s}(\alpha) < s(\alpha)$$
$$q(\alpha) \leq \hat{q} \implies \bar{s}(\alpha) > s(\alpha).$$

Let $\hat{\alpha}$ be any parameter type such that $s(\hat{\alpha}) = \bar{s}(\hat{\alpha}) > 0$. Then the above implications show that $\bar{q}(\hat{\alpha}) < \hat{q}$ and $q(\hat{\alpha}) > \hat{q}$. In particular, we see that $\bar{q}(\hat{\alpha}) < q(\hat{\alpha})$ which in turn implies that $s'(\hat{\alpha}) > \bar{s}'(\hat{\alpha})$. Therefore, the two surplus functions cross only once above the horizontal axis, and $s(\alpha) \leq \bar{s}(\alpha)$ if and only if $\alpha \lesssim \hat{\alpha}$. □

This lemma demonstrates that the utility functions s and $\bar{s}$ corresponding to the two tariffs typically are as shown in Figure 1, and formalizes the intuitive view that quantity discounts when compared to a uniform tariff will favour users with a strong preference for the good at the expense of other users.

3. Unregulated monopoly

Having analyzed the revenue function R(Q) we are in a position to solve the problem of the unregulated profit–maximizing firm. If the firm is free to set any tariff then its profit–maximizing level of output Q^* will satisfy

$$(6) \qquad Q^* \ \underset{Q \geq 0}{\text{maximizes:}} \ R(Q) - C(Q)$$

(recall that C(Q) is the firm's cost of producing total output Q). The profit–maximizing tariff T^* will then be that tariff which maximizes revenue at output level Q^*. We can then apply Corollary 1 to this particular case to deduce that the profit–maximizing tariff is concave.[3] The first–order condition for (6) will involve setting the profit–maximizing level of output Q^* to satisfy $C'(Q^*) = \lambda(Q^*)$. An examination of (4) then shows that the highest demand consumers are efficiently served:

$$p^*(q^*(\bar{\alpha})) = C'(q^*(\bar{\alpha})) \ ,$$

where we have used the fact that $1 - F(\bar{\alpha}) = 0$.

As in the case of third–degree price discrimination, nonlinear pricing leads to

---

[3] This is Maskin and Riley's (1984) quantity discount result, extended to the case of a more general cost function.

unequal marginal utilities across different consumers. For instance, in the present model marginal utility of a type α consumer is $p^*(q^*(\alpha))$ as given in (4), and this will vary with α. For a given level of total output this is inefficient and welfare would be increased by a redistribution of units of output from high to low users. Indeed, for a given level of output the tariff which would maximize welfare is uniform. On the other hand, the total level of output might well be increased with nonlinear pricing and this effect might in some cases outweigh the unequal marginal utilities effect. The example in Section 5 provides a case where nonlinear pricing is socially superior to uniform pricing in the unregulated case. As the foregoing discussion indicates, for total welfare to increase with the introduction of nonlinear pricing it is necessary (but not sufficient) that total output increases. The overall welfare effect of the move from uniform to nonlinear tariffs will be ambiguous.

4. Regulated monopoly

In practice there are only rare instances of pure unregulated monopolies, and it would be desirable to extend the analysis to encompass a wider class of cases. One way to do this might be to try to introduce competition.[4] Alternatively, if the firm is unlikely to be facing a significant degree of competition the possibilities for regulation should be considered. Several forms of regulation are feasible, and we shall analyze two: average revenue regulation, which is used in the regulation of some utilities in the UK, and optional tariffs, which have recently been introduced into the regulatory regime of BT.

4.1 Average revenue regulation

Suppose now that the monopoly faces an average revenue constraint, and so it may offer the nonlinear tariff T only if it satisfies

$$(7) \qquad \int_{\underline{\alpha}}^{\bar{\alpha}} T(q(\alpha))f(\alpha)\,d\alpha \leq \int_{\underline{\alpha}}^{\bar{\alpha}} \bar{p}q(\alpha)f(\alpha)\,d\alpha$$

where $q(\alpha)$ is the response function induced by $T(\cdot)$ and $\bar{p}$ is the permitted level of

[4] For instance, see Oren *et al.* (1983) and Wilson (1992b).

average revenue per unit of output (i.e. $\bar{p}$ is the 'price cap').

We shall suppose in the following that the average revenue constraint (7) is binding at the tariff chosen by the firm. In this case the firm's problem is to choose revenue R and output Q in order to maximize

$$R - C(Q) \quad \text{subject to}$$
$$\text{(i)} \quad R = \bar{p}Q \text{ , and}$$
$$\text{(ii)} \quad R \leq R(Q) \text{ .}$$

Constraint (i) is the binding average revenue constraint and (ii) requires that revenue be attainable at output level Q. Rewriting, the above problem reduces to

$$(P_{ar}) \qquad \underset{Q \geq 0}{\text{maximize}} \quad \bar{p}Q - C(Q) \quad \text{subject to}$$

$$R(Q) \geq \bar{p}Q \text{ .}$$

If (3) holds then by Lemma 1 $R(\cdot)$ is strictly concave and if $u'(0) = \infty$ and $u'(\infty) = 0$ then by Lemma 2 the same holds true for $R(\cdot)$. In such cases there is some unique positive level of output $\hat{Q}$ satisfying $R(\hat{Q}) = \bar{p}\hat{Q}$, and such that $R(Q) \geq \bar{p}Q$ if and only if $Q \leq \hat{Q}$ — see Figure 2. The firm's problem then becomes

$$\underset{Q}{\text{maximize}} \quad \bar{p}Q - C(Q) \quad \text{subject to} \quad 0 \leq Q \leq \hat{Q}.$$

Suppose that $\bar{p} \geq C'(Q)$ for $Q \leq \hat{Q}$ so that $\bar{p}Q - C(Q)$ is increasing in Q up to $\hat{Q}$. In this case the firm will simply maximize output subject to $Q \leq \hat{Q}$, i.e. it will set $Q = \hat{Q}$. Since revenue equals $R(\hat{Q})$ at this level of output, the firm will offer the nonlinear tariff that maximizes the revenue from output level $\hat{Q}$. We summarize this discussion in the following Result:

RESULT 2. *Suppose that the conditions of Lemma 2 hold and let* $\hat{Q}$ *be the unique level of output satisfying* $R(\hat{Q}) = \bar{p}\hat{Q}$. *Suppose further that* $\bar{p} \geq C'(Q)$ *for* $Q \leq \hat{Q}$. *Then if the average revenue constraint (7) binds at the optimum the firm will maximize output* Q *subject to the binding constraint (7). This maximum output is* $\hat{Q}$, *and the firm will offer the tariff* $\hat{T}$ *which maximizes revenue at output level* $\hat{Q}$.

Result 1 then implies that the optimal marginal price schedule is given by

$$(8) \qquad \hat{p}(\hat{q}(\alpha)) = \lambda(\hat{Q})/[1 - \frac{1 - F(\alpha)}{\alpha f(\alpha)}].$$

Corollary 1 implies that, provided that (5) holds, the regulated firm will choose to offer quantity discounts.

Recall that $Q*$ is the unregulated profit–maximizing level of output. Then it is clear that whenever the regulatory constraint binds we must have $Q* < \hat{Q}$ so that regulation increases total output. (To see this suppose on the contrary that $Q* \geq \hat{Q}$. Then from the concavity of R and the definition of $\hat{Q}$ it follows that $R(Q*) \leq \bar{p}Q*$, i.e. the unconstrained optimum is allowed by the constraint and so regulation has no effect.) Therefore, since $\lambda(\cdot)$ is decreasing in Q and $\lambda(Q*) = C'(Q*)$, if C is (weakly) convex then $\lambda(\hat{Q}) < C'(\hat{Q})$. Equation (8) then implies

COROLLARY 2. *If the firm has (weakly) convex costs and is permitted to engage in nonlinear pricing, then all sufficiently high–type consumers will be served with higher quantities than is efficient, so that* $\hat{p}(\hat{q}(\alpha)) < C'(\hat{Q})$ *for all* α *close to* $\bar{\alpha}$.

Another possible kind of regulation might be to require the firm to produce a certain minimum level of output (and to leave it otherwise unconstrained). In particular, suppose that the firm was required to produce at least output level $\hat{Q}$, where $\hat{Q}$ is greater than the unregulated profit–maximizing level of output $Q*$, but is otherwise unconstrained in the tariffs it may offer. Then, provided that profit $R(Q) - C(Q)$ is single–peaked in output Q, the firm will choose to let the output constraint bind and produce $\hat{Q}$, and will offer the tariff which maximizes its revenue from this output. In other words, this form of minimum output regulation coincides precisely with average revenue regulation for most plausible cases. We sum this up as a corollary:

COROLLARY 3. *Let the firm be required to supply a total output of at least* $\hat{Q}$ *but be otherwise unconstrained. Suppose that* $C'(Q) < R(\hat{Q})/\hat{Q}$ *for all* $Q \leq \hat{Q}$ *and let the profit function* $R(Q) - C(Q)$ *be single–peaked in* Q. *Then minimum output regulation is precisely equivalent to average revenue regulation with a price cap given by* $\bar{p} = R(\hat{Q})/\hat{Q}$.

There are two natural welfare comparisons we can make in the context of average revenue regulation: that between the unregulated profit–maximizing nonlinear tariff and some regulated nonlinear tariff; and that between some regulated uniform tariff and the constrained profit–maximizing nonlinear tariff which has the same average revenue. We consider each in turn.

As demonstrated above, for most cases average revenue regulation is the same as minimum output regulation, and this was the form of regulation considered by Katz (1983). More precisely, he examined the welfare change when the firm was required to increase its level of output from the unconstrained profit–maximizing level Q^* in (6) to some other level $\hat{Q} > Q^*$, but otherwise letting the firm offer whatever tariff it chooses. Katz argued that this kind of regulation was attractive because it had few informational requirements (e.g. on costs or rate–of–return), and required only that the regulator be able to observe the firm's output. It is not hard to see that requiring a small increase in output must in this model be unambiguously good for welfare since it will have a first–order benefit to consumers but, because Q^* was optimally chosen, only a second–order loss in profits for the firm. In other words, just as with uniform pricing, the unregulated monopolist typically chooses to restrict output.[5]

Turning to the other comparison, suppose that the firm was initially regulated and required to offer the uniform tariff $\bar{T}(q) = \bar{p}q$ and that regulation is relaxed so that the

[5] However, Katz shows that with utility functions other than of the form we use here (where $U = \alpha u(q) - T$) it need not be the case that the monopolist restricts output.

firm is permitted to offer any tariff subject to constraint (7). Then we have three results which parallel Propositions 2, 3 and 5 of Armstrong and Vickers (1991). First, consumers as a whole will be made worse off as a result of the move to a nonlinear tariff:

RESULT 3. *If* T *is such that the average revenue constraint (7) binds, then total consumer surplus with the nonlinear tariff is lower than total consumer surplus with the uniform tariff* $\bar{T}(q) = \bar{p}q$.

PROOF. Let $q(\alpha)$ be the demand function induced by the nonlinear tariff T. Let S(T) be consumer surplus given T, so that

$$S(T) \equiv \int_{\underline{\alpha}}^{\bar{\alpha}} s(\alpha)f(\alpha)\, d\alpha$$

(where $s(\alpha)$ is as given in Section 2). In addition, the following inequality holds:

$$S(\bar{T}) \geq S(T) + \int_{\underline{\alpha}}^{\bar{\alpha}} [T(q(\alpha)) - \bar{T}(q(\alpha))]f(\alpha)\, d\alpha, \tag{9}$$

where $q(\alpha)$ is the demand function induced by the tariff T. That this is true follows from the fact that, since $q(\alpha)$ is a possible choice for type α facing the uniform tariff $\bar{T}$, $\bar{s}(\alpha) \geq \alpha u(q(\alpha)) - \bar{T}(q(\alpha))$. Therefore

$$S(\bar{T}) \geq \int_{\underline{\alpha}}^{\bar{\alpha}} [\alpha u(q(\alpha)) - \bar{T}(q(\alpha))]f(\alpha)\, d\alpha = \int_{\underline{\alpha}}^{\bar{\alpha}} [s(\alpha) + T(q(\alpha)) - \bar{T}(q(\alpha))]f(\alpha)\, d\alpha,$$

which is equivalent to (9).

Using inequality (9) we see that

$$S(\bar{T}) \geq S(T) + \int_{\underline{\alpha}}^{\bar{\alpha}} [T(q(\alpha)) - \bar{p}q(\alpha))]f(\alpha)\, d\alpha$$

$$= S(T) \text{ from the binding constraint (7).} \quad \square$$

In contrast, the firm will clearly benefit from the freedom to offer nonlinear tariffs, and so Result 2 does not imply that the overall welfare effect (measured as the simple sum of consumer and producer surplus, for instance) will be negative. Recall that we can decompose the total welfare effect into an 'unequal marginal utilities' effect and an 'output' effect, and in order for total welfare to rise as a result of introducing a nonlinear tariff it is necessary for the output effect effect to be positive. The next result shows that this output effect of allowing the firm to offer nonlinear tariffs is (usually) positive:

RESULT 4. *If the conditions in the statement of Result 2 hold then total output will increase in the move to the profit–maximizing nonlinear tariff subject to average revenue regulation.*

PROOF. Result 2 shows that the firm will maximize output subject to the binding regulatory constraint. Since the uniform tariff $\bar{T}(q)$ satisfies the constraint, the result follows. □

The next result gives a sufficient condition for the overall welfare effect to be negative, and the intuition is as follows. The first–best allocation in terms of overall welfare is attained by offering a uniform tariff with price equal to marginal cost, and say that this occurs at the output level Q_{fb}, so that the first–best tariff is the uniform tariff with price $p_{fb} = C'(Q_{fb})$. Suppose that C is (weakly) convex so that the firm does not make a loss with this tariff. Suppose that the *ex ante* regulatory regime involved requiring the firm to offer the uniform tariff $\bar{p}q$, where $\bar{p}$ is close to p_{fb}. Under this regime the first–best allocation will have nearly been attained. On the other hand, if the firm were now permitted to offer a nonlinear tariff subject to (7), then it will in general offer a tariff which results in a rather different allocation to that of the first–best, and this allocation will have a lower level of welfare than that of the approximately optimal uniform tariff $\bar{p}q$.

RESULT 5. *Let* p_{fb} *be the first–best uniform price. Then for* $\bar{p}$ *sufficiently close to* p_{fb} *welfare is higher under the uniform tariff* $\bar{p}q$ *than if nonlinear tariffs subject to the average revenue constraint (7) are allowed.*

PROOF. Analogous to the proof of Proposition 5 in Armstrong and Vickers (1991).

Provided that (5) holds, the more detailed analysis of Lemma 3 may also be applied to the case of the concave tariff arising from average revenue regulation to obtain:

RESULT 6. *Suppose that the conditions of Result 2 and that condition (5) hold. Then there exists some* $\hat{\alpha}$ *such that all consumers of type* $\alpha < \hat{\alpha}$ *are worse off and all types* $\alpha > \hat{\alpha}$ *are better off when nonlinear pricing subject to average revenue regulation is permitted.*

Finally, having seen two negative results about allowing nonlinear tariffs subject to an average revenue constraint (Results 3 and 5), it is important to note that overall welfare could increase as a result of this change. The example in Section 5 is such that, provided that the price cap $\bar{p}$ is not too near the level of unit cost (in which case Result 5 would ensure the superiority of the uniform tariff), the move to nonlinear tariffs is good for welfare. Therefore, just as was the case in the context of third–degree price discrimination in Armstrong and Vickers (1991), the welfare effect of allowing nonlinear pricing subject to an average revenue constraint is ambiguous.

4.2 Optional nonlinear pricing

The last section considered one way in which the firm can be given freedom to offer a nonlinear tariff subject to price cap regulation. Here we look at a second possibility, namely, that the firm may offer any tariff it wishes subject to the constraint that it must continue to offer consumers the option of using the old uniform tariff $\bar{T}(q) = \bar{p}q$. Equivalently, the firm is constrained to offer tariffs $T(q)$ such that

$$(10) \qquad T(q) \le \bar{p}q \quad \text{for all } q.$$

Thus the firm must not increase the average revenue generated from any consumer, rather than simply not increase the average revenue from all consumers taken together (or put another way, constraint (10) says that (7) should hold pointwise rather than just on average). The firm is definitely worse–off under this form of regulation than under average revenue regulation (since any tariff which satisfies (9) will satisfy (7)). On the other hand, the firm and all consumers are (weakly) better off as a result of this change when compared to the uniform tariff $\bar{p}q$, and we will see that in most cases at least some consumers and the firm will be strictly better–off.

In this section we will assume that the firm has a constant unit cost of c: $C(Q) = cQ$. In this case the unregulated profit–maximizing tariff is given by the simple formula

$$(11) \qquad p^*(q^*(\alpha)) = c/[1 - \frac{1 - F(\alpha)}{\alpha f(\alpha)}]$$

with the optimal response function $q^*(\alpha)$ given by

$$(12) \qquad q^*(\alpha) \ \underset{q \ge 0}{\text{maximizes:}} \ u(q)[\alpha - \frac{1 - F(\alpha)}{f(\alpha)}] - cq .$$

(These were equations (14) and (12) in Chapter 1.)

It turns out to be simple to construct the profit–maximizing tariff subject to the constraint that $T(q) \le \bar{p}q$ (where we assume $\bar{p} > c$):

Result 7. *Let* $T^*(\cdot)$ *be the optimal nonlinear tariff for the unregulated firm with cost level* c, *so that* $p^*(\cdot)$, *the optimal marginal price function, is given by (11). Let* $\tilde{q}$ *be defined by* $p^*(\tilde{q}) = \bar{p}$. *Assume that regulation has an impact, so that it is not the case that* $T^*(q) \leq \bar{p}q$ *for all* q. *Then under assumption (5)* $\tilde{q}$ *is unique and the profit–maximizing tariff* $T_0(\cdot)$ *given the constraint (10) is:*

$$T_0(q) = \begin{cases} \bar{p}q & \text{if } q \leq \tilde{q} \\ T^*(q) - T^*(\tilde{q}) + \bar{p}\tilde{q} & \text{if } q \geq \tilde{q} \ . \end{cases} \tag{13}$$

Remark: Figure 3 illustrates how this tariff is constructed. Because (5) implies that T^* is concave the tariff T_0 does indeed satisfy the constraint, and by construction, T_0 is both continuous and differentiable. Alternatively, we can describe the optimal tariff in terms of the marginal price schedule, p_0, which satisfies $p_0(q) = \min: \{ p^*(q), \bar{p} \}$.

Proof. See Appendix.

To summarize, under assumption (5) the move from the uniform tariff $\bar{T}(q) = \bar{p}q$ to the optimal nonlinear tariff subject to $T(q) \leq \bar{p}q$ has the following desirable properties:

- no consumers are made worse off and the firm and all consumers of sufficiently high type are made strictly better off;
- output increases (since all marginal prices are at least weakly lower);
- there is no sub–marginal cost pricing (in contrast to average revenue regulation);
- quantity discounts are offered, with the highest–type consumers being served efficiently.

Welfare with the tariff T_0 is higher than with the uniform tariff $\bar{T}$. Therefore, the only remaining comparison of interest is between welfare under T_0 and welfare under the nonlinear tariff induced by average revenue regulation, $\hat{T}$. The fact that T_0 has two advantages over $\hat{T}$ (it is a Pareto improvement over the uniform tariff, whereas the average revenue tariff made consumers as the whole worse off, and there is no sub–marginal cost pricing) might suggest that optional tariff regulation will be superior to average revenue regulation. However, this is not always the case. It might be that T_0 involves fairly minor modifications to the uniform tariff, albeit of a Pareto improving kind, whereas average revenue regulation might involve the introduction of a dramatically more efficient tariff $\hat{T}$, albeit a tariff that leaves consumers worse off. The following example, which is by no means pathological, has T_0 socially inferior to $\hat{T}$ for a wide variety of parameter values.[6]

5. A worked example

Let $u(q) = 2\sqrt{q}$ and suppose that α is distributed on $[2, \infty)$ by the density function $f(\alpha) = e^{-(2 - \alpha)}$, a distribution which satisfies condition (5). Useful facts about this density function are that $E(\alpha) = 3$ and $E(\alpha^2) = 10$. We first calculate the revenue function. Equation (2) implies that for a given level of total output Q the revenue–maximizing response function q^* is given by (for $\alpha \geq 2$)

$$q^*(\alpha) = [(\alpha - 1)/\lambda(Q)]^2 \tag{14}$$

and, in particular, that all consumers participate in the market at all output levels. (This feature greatly simplifies the analysis.) Since the expectation of $q^*(\alpha)$ must equal Q, we see that

$$[\lambda(Q)]^2 = E[(\alpha - 1)^2]/Q = 5/Q.$$

Therefore, since $R'(Q) = \lambda(Q)$ and $R(0) = 0$ we deduce that the revenue function is

[6] In its recent review of BT's prices, the regulatory body Oftel (1992) decided to permit BT to engage in nonlinear pricing but that any quantity discounts offered will *not* count towards the firm's compliance with its price cap. Thus it allowed optional nonlinear pricing but not unrestricted nonlinear pricing subject just to an average price cap.

given by

$$R(Q) = \sqrt{20Q}.$$

Given any response function $q(\alpha)$, total welfare (measured as the simple sum of consumer surplus and profit) is given by

(15) $$E[\alpha u(q(\alpha)) - cq(\alpha)]$$

where the firm is assumed to have constant unit costs of c.

Unregulated monopoly: The unregulated profit–maximizing level of output is given by maximizing $R(Q) - cQ$, which is done by setting total output to $Q^* = 5/c^2$. From (12), the unregulated profit–maximizing response function is

(16) $$q^*(\alpha) = (\alpha - 1)^2/c^2.$$

The profit–maximizing tariff is

$$T^*(q) = cq + 2\sqrt{q} + 1/c\ .$$

Uniform tariffs: Faced with the uniform tariff $\bar{p}q$, a type α consumer will demand the quantity $\bar{q}(\alpha) = (\alpha/\bar{p})^2$. From (15), total welfare with the uniform tariff $\bar{p}q$ is

(17) $$\bar{W} = E[2\alpha^2/\bar{p} - c(\alpha/\bar{p})^2] = 10(2 - c/\bar{p})/\bar{p}$$

which for fixed c is maximized at $\bar{p} = c$. Profit with the uniform tariff is

$$\bar{\pi} = E[(\bar{p} - c)(\alpha/\bar{p})^2] = 10(\bar{p} - c)/\bar{p}^2$$

which is maximized at $\bar{p} = 2c$. For a uniform price cap $\bar{p}$ to bind, it must be that $\bar{p} < 2c$, which is assumed henceforth.

Average revenue regulation: Provided that the regulatory constraint binds, the profit–maximizing level of output $\hat{Q}$ is given by $R(\hat{Q}) = \bar{p}\hat{Q}$, or $\hat{Q} = 20/\bar{p}^2$. Therefore, from (2) the optimal response function $\hat{q}$ is $\hat{q}(\alpha) = 4(\alpha - 1)^2/\bar{p}^2$, and so total welfare is

$$\hat{W} = E[4\alpha(\alpha - 1)/\bar{p} - 4c(\alpha - 1)^2/\bar{p}^2] = (28 - 20c/\bar{p})/\bar{p} \tag{18}$$

which for fixed c is maximized at $\bar{p} = 10c/7$. The condition for the constraint not to bind is that the unregulated nonlinear tariff does not satisfy the average revenue constraint, which, since the unregulated level of output is $Q^* = 5/c^2$, requires that $\bar{p} < 2c$ (just as with the uniform tariff case). The constrained profit–maximizing tariff is

$$\hat{T}(q) = \bar{p}q/2 + 2\sqrt{q} + 2/\bar{p}\ .$$

Optional tariffs: The profit–maximizing optional tariff is as described in the proof in the Appendix, so the optimal response function is given by $q_0(\alpha) = \max: \{\ \bar{q}(\alpha), q^*(\alpha)\ \}$. The consumer type for which $\bar{q}(\alpha) = q^*(\alpha)$ is $\alpha = \bar{p}/(\bar{p} - c)$, so that

$$q_0(\alpha) = \begin{cases} (\alpha/\bar{p})^2 & \text{if } \alpha < \bar{p}/(\bar{p} - c) \\ (\alpha - 1)^2/c^2 & \text{if } \alpha > \bar{p}/(\bar{p} - c)\ . \end{cases}$$

It follows that

$$W_0 = \bar{W} + (2/(c\bar{p}^2))(2\bar{p} - c)(\bar{p} - c)\exp\{\bar{p}/(\bar{p} - c) - 2) \tag{19}$$

where $\bar{W}$ is as given in (17). W_0 is maximized at $\bar{p} = c$ for given c. The profit–maximizing optional tariff is

$$T_0(q) = \begin{cases} \bar{p}q & \text{if } q < 1/(\bar{p}-c)^2 \\ cq + 2\sqrt{q} - 1/(\bar{p}-c) & \text{if } q > 1/(\bar{p}-c)^2 . \end{cases}$$

Welfare comparisons:

(i) Since the second term of W_0 in (19) is certainly positive (provided that $\bar{p} > c$) then $W_0 > \bar{W}$ (as is generally true).

(ii) $\bar{W} > \hat{W}$ if and only if $\bar{p} < 5c/4$. Thus there is a range of $\bar{p}$ for which welfare is higher when nonlinear tariffs subject to average revenue regulation are allowed than when they are not (recall that we have assumed only that $c < \bar{p} < 2c$). This example also illustrates Result 5 that uniform pricing is superior to nonlinear tariffs whenever the price cap is sufficiently close to the level of unit cost. Finally, since $\bar{W} < \hat{W}$ when $\bar{p} = 2c$ — i.e. when the firm is unregulated — this example provides a case where the move from uniform to nonlinear pricing is good for total welfare in the absence of regulation.

(iii) It is not possible to obtain explicitly the range of values of $\bar{p}$ for which $W_0 < \hat{W}$, but calculating W_0 and $\hat{W}$ numerically we find that this is true when $\bar{p}$ lies in the range approximately given by $1{\cdot}255 < \bar{p} < 2$. Therefore, optional pricing is inferior to nonlinear pricing subject to average revenue regulation for a large range for the price cap.

Figure 4 shows how total welfare varies with the price cap $\bar{p}$ taking $c = 1$ under the three regimes,[7] while Figure 5 depicts the various optimal marginal price schedules taking $c = 1$ and $\bar{p} = 1.25$.

[7] It is interesting to see that the maximum welfare level for nonlinear pricing subject to average revenue regulation (which occurs at $\bar{p} = 10/7$) is rather close the the first–best welfare level (which occurs at a uniform tariff with $\bar{p} = 1$).

6. Conclusions

We have examined three ways of regulating a monopolist which is able to offer nonlinear tariffs. Firstly, the regulator could simply require the firm to offer some fixed uniform tariff and so forbid the firm to practice nonlinear pricing at all. Alternatively, the firm could be permitted to offer nonlinear tariffs subject to the constraint that it does not increase its average revenue. Under this average revenue regulation, the firm will offer a revenue–maximizing tariff which, subject to a hazard rate condition on consumer types, will involve quantity discounts (Corollary 1). Consumers of high type will be served larger quantities than is efficient, the firm having an incentive to make up for losses from these high users by high charges to low users (Corollary 2). Compared to the uniform pricing regime with the same average revenue, output is higher with nonlinear pricing (Result 4) but consumer surplus is lower (Result 3). If the price cap is close to the first–best tariff then welfare will decrease in the move to nonlinear pricing (Result 5), although in general the total welfare effect is ambiguous. This form of regulation is usually equivalent to a regime which specifies a minimum output requirement for the firm but leaves it otherwise unconstrained (Corollary 3).

Thirdly, we looked at effects of requiring the firm to offer consumers an option of the uniform tariff together with whatever other tariff it chooses to offer. Such regulation clearly represents a Pareto improvement over the uniform tariff. In many cases, the optimal tariff has an attractively simple form: marginal price $p_o(q)$ is equal to the minimum of $\bar{p}$, the uniform price, and $p^*(q)$, the unregulated profit–maximizing marginal price (Result 7). All low demand consumers are unaffected by the introduction of the nonlinear tariff, but high users are made strictly better off compared to the uniform tariff. In this case there is no incentive for the firm to practice sub–marginal cost pricing. The optional tariff regime appears to have a double advantage over that of average revenue regulation since it represents a Pareto improvement over the uniform tariff, whereas average revenue regulation makes consumers as a whole worse off, and there are no 'perverse' incentives for sub–marginal cost pricing. However, unless the price cap is particularly close to the first–best price level (so that Result 5 is not applicable) there is no reason to think that this form of regulation is likely to be preferable to average revenue regulation, and the welfare comparison between the two is ambiguous.

Appendix: Proof of Result 4

First we show that $\tilde{q}$ exists and is unique. Assumption (5) implies that T^* is strictly concave, and so if regulation has an impact it must be true that $p^*(0) > \bar{p}$. Also, from (11) $p^*(q^*(\bar{\alpha})) = c < \bar{p}$. Therefore, since p^* is strictly decreasing there is one and only one $\tilde{q}$ satisfying $p^*(\tilde{q}) = \bar{p}$.

Let $\bar{s}(\alpha)$ be the surplus function corresponding to the uniform tariff $\bar{p}q$. The problem we need to solve is to maximize profits subject to (10). However, we shall solve this in the equivalent formulation of solving the problem subject to the constraint that $s(\alpha) \geq \bar{s}(\alpha)$. The method we use is simply to guess the solution and then use sufficiency arguments to show that we have guessed correctly. Our guess is the surplus function $s_0(\cdot)$ corresponding to the tariff $T_0(\cdot)$ in (13). Therefore, $s_0(\cdot)$ is defined as follows. Let $s^*(\cdot)$ be the surplus function corresponding to the unconstrained optimal tariff T^* and let $\tilde{\alpha}$ be the point such that $q^*(\tilde{\alpha}) = \bar{q}(\tilde{\alpha}) = \tilde{q}$ (where $q^*(\cdot)$ and $\bar{q}(\cdot)$ are the respective response functions). Then

$$s_0(\alpha) = \begin{cases} \bar{s}(\alpha) & \text{if } \alpha \leq \tilde{\alpha} \\ s(\alpha) - s(\tilde{\alpha}) + \bar{s}(\tilde{\alpha}) & \text{if } \alpha \geq \tilde{\alpha} \ . \end{cases}$$

Alternatively, our guess for the optimal response function is $q_0(\alpha) = \max: \{ \bar{q}(\alpha), q^*(\alpha) \}$.

Now suppose that we can find a continuous and piecewise continuously differentiable function $\mu(\alpha)$ such that $\mu(\alpha) = 0$ for all $\alpha > \tilde{\alpha}$, $\mu'(\alpha) \geq 0$ for $\alpha \leq \tilde{\alpha}$, and such that $s_0(\alpha)$ maximizes

(A1) $$\int_{\underline{\alpha}}^{\bar{\alpha}} [\alpha u(q(\alpha)) - s(\alpha) - cq(\alpha)]f(\alpha) + \mu'(\alpha)[s(\alpha) - \bar{s}(\alpha)] \, d\alpha$$

over all increasing functions s (where $s'(\alpha) \equiv u(q(\alpha))$). In this case, profits for the firm are higher with s_0 than with any other utility function s such that $s \geq \bar{s}$ as required.

Integrating by parts, (A1) becomes

$$\int_{\underline{\alpha}}^{\bar{\alpha}} [\alpha u(q(\alpha)) - cq(\alpha)]f(\alpha) - [1 - F(\alpha)]u(q(\alpha)) - \mu(\alpha)u(q(\alpha)) - \mu'(\alpha)\bar{s}(\alpha)\, d\alpha,$$

and maximizing this pointwise with respect to $q(\alpha)$ means that the constrained–optimal response function q_o should satisfy

$$\text{(A2)} \qquad q_o(\alpha) \text{ maximizes: } \max_{q \geq 0} u(q)[\alpha - \frac{1 - F(\alpha) + \mu(\alpha)}{f(\alpha)}] - cq.$$

Clearly, if $\mu(\alpha) = 0$ for $\alpha \geq \tilde{\alpha}$ we have $q_o(\alpha) = q^*(\alpha)$ for $\alpha \geq \tilde{\alpha}$, and so we want to pick μ such that $q_o(\alpha) = \bar{q}(\alpha)$ when $\alpha < \tilde{\alpha}$ (where $\bar{q}(\alpha)$ is the response function induced by the uniform tariff $\bar{p}q$). But writing

$$\text{(A3)} \qquad \mu(\alpha) = \begin{cases} 0 & \text{if } \alpha > \tilde{\alpha} \\ (1 - c/\bar{p})\alpha f(\alpha) - (1 - F(\alpha)) & \text{if } \alpha < \tilde{\alpha} \end{cases}$$

means that (A2) becomes

$$q_o(\alpha) \text{ maximizes: } \max_{q \geq 0} (c/\bar{p})[\alpha u(q) - \bar{p}q]$$

for $\alpha < \tilde{\alpha}$ so that $q_o(\alpha)$ in (A2) does indeed correspond to $\bar{q}(\alpha)$ over this range. Therefore, s_o as given above does maximize (A1) when μ is as given in (A3).

Only one condition remains to be checked, namely, that μ is non–decreasing for $\alpha \leq \tilde{\alpha}$. However, from the definition of $\tilde{\alpha}$ equation (11) implies that

$$(1 - c/\bar{p}) = (1 - F(\tilde{\alpha}))/(\tilde{\alpha}f(\tilde{\alpha}))$$

and assumption (5) then implies that

$$\text{(A4)} \qquad \alpha \leq \tilde{\alpha} \Rightarrow 1 - c/\bar{p} \leq \frac{1 - F(\alpha)}{\alpha f(\alpha)}.$$

The derivative of μ in (A3) is $(1 - c/\bar{p})(\alpha f(\alpha))' + f(\alpha)$. If $(\alpha f(\alpha))' \geq 0$ then μ is indeed increasing. On the other hand, if $(\alpha f(\alpha))' < 0$ then (A4) implies that

$$\mu'(\alpha) \geq \frac{1 - F(\alpha)}{\alpha f(\alpha)}(\alpha f(\alpha))' + f(\alpha).$$

But the right–hand side of this is positive from assumption (5). Therefore, all the sufficient conditions for $s_0(\cdot)$ to be optimal have been satisfied. □

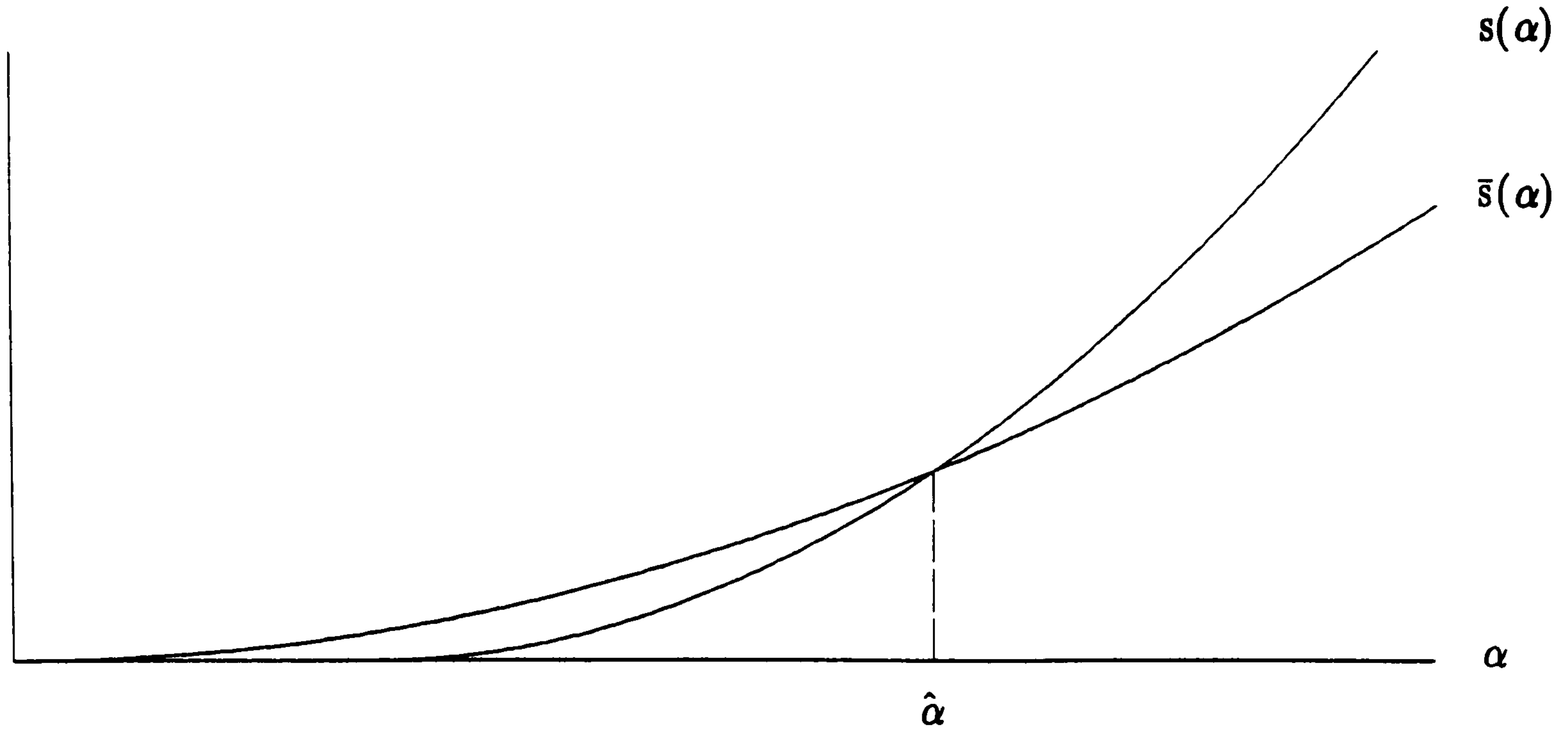

Figure 1: A comparison of the surplus functions induced by a uniform tariff and a tariff with quantity discounts

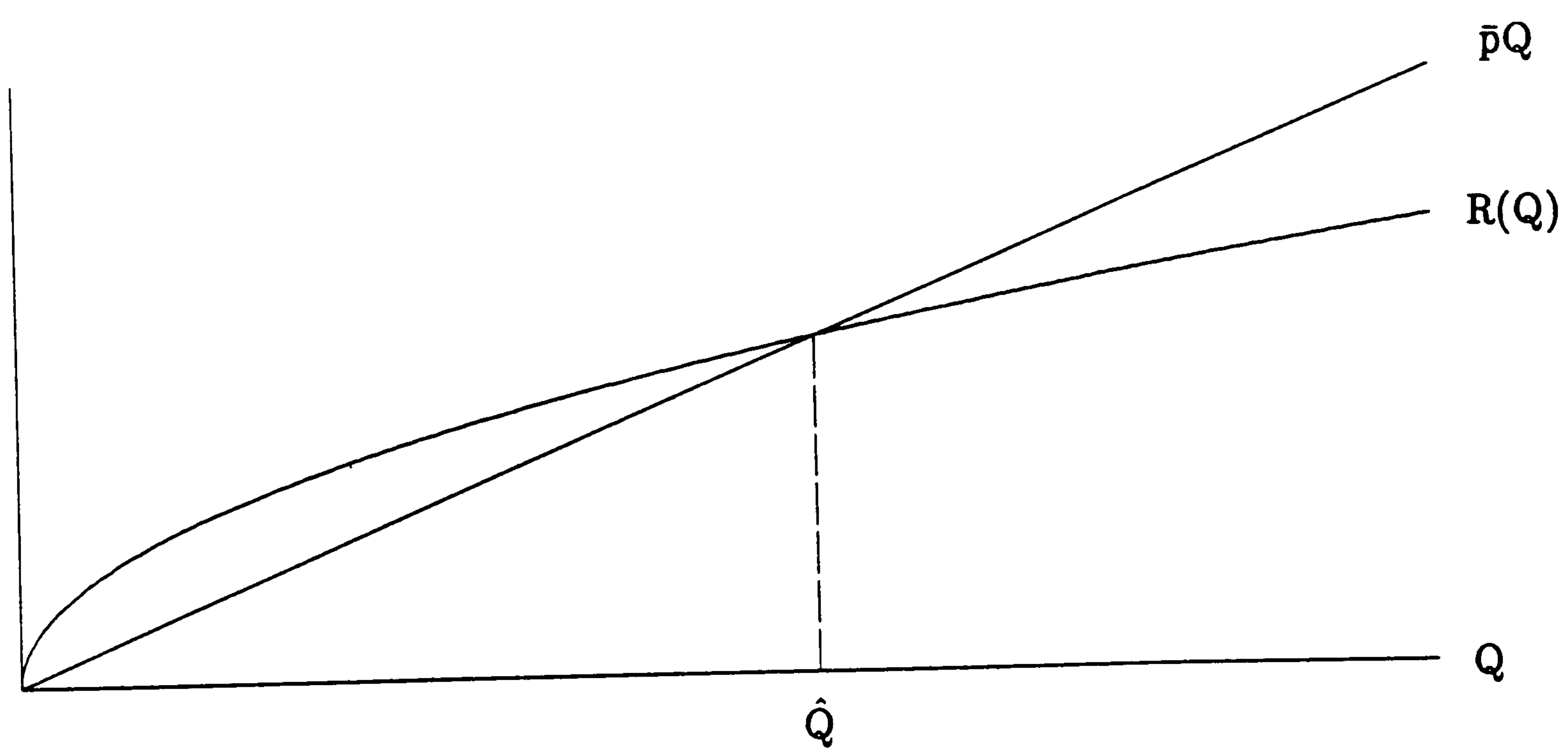

Figure 2: Definition of the output level $\hat{Q}$

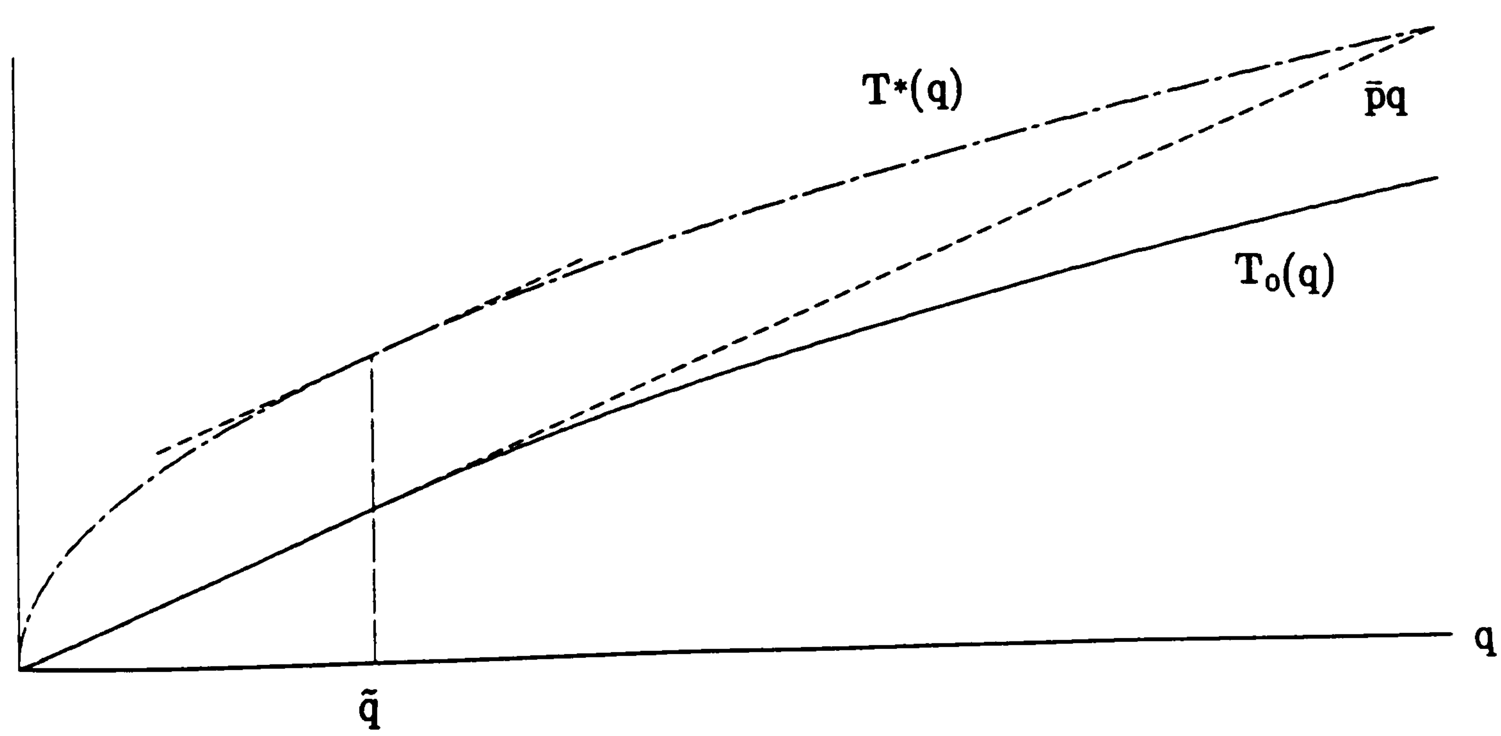

Figure 3: Construction of the profit–maximizing optional tariff

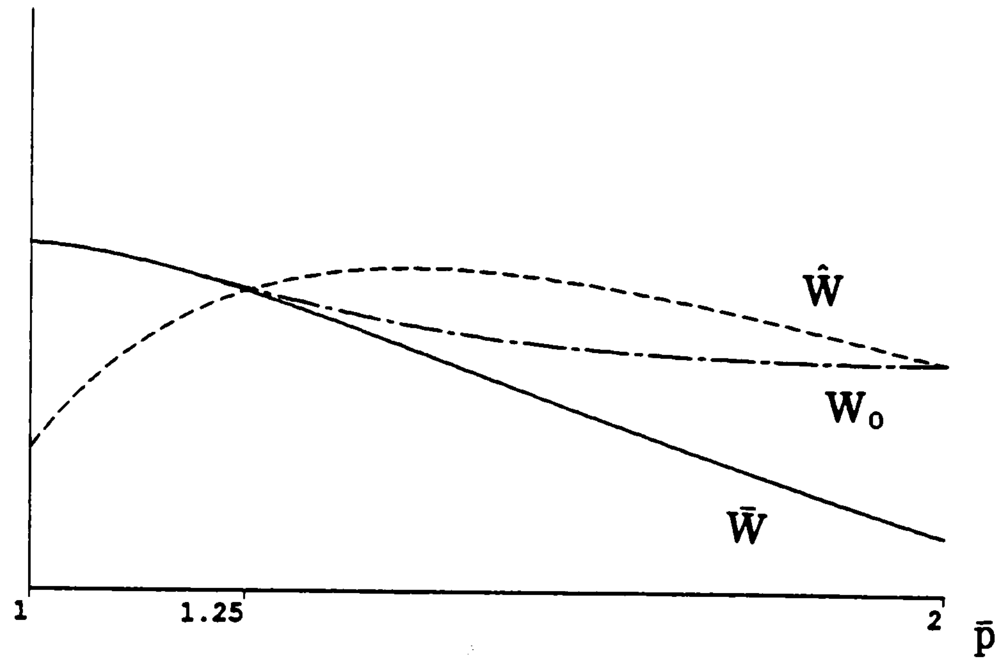

Figure 4: How welfare varies with $\bar{p}$ in the three regimes $(c = 1)$

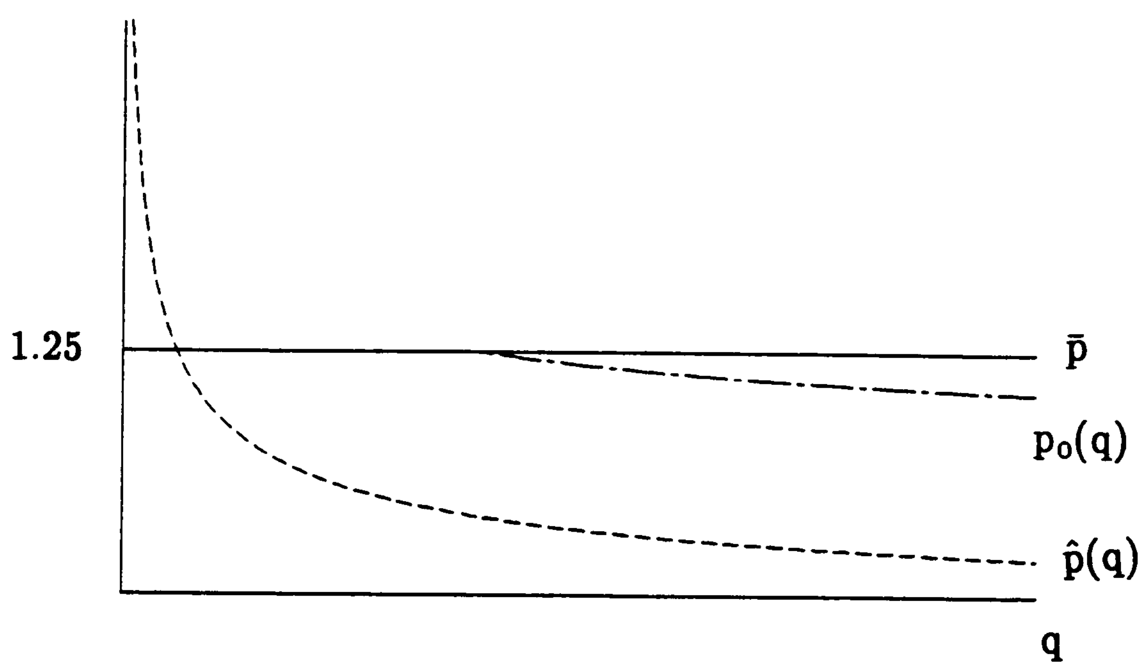

Figure 5: Marginal prices in the three regimes $(c = 1,\ \bar{p} = 1.25)$

CHAPTER 3

BUNDLING AND THE CORRELATION OF TASTES

1. Introduction

The form of price discrimination known as bundling occurs when a multiproduct firm makes the joint price of a collection — or *bundle* — of goods differ from the sum of the prices it offers for the goods individually. Instances of this widespread business practice include set meals versus *a la carte* menus in restaurants, season tickets (either for travel or entertainment), package holidays and computer retailers offering free software with computers. In these examples the firm gives a discount for joint purchases so that consumers are encouraged to buy a bundle of goods (we call this subadditive bundling). It nevertheless is a possibility for some firms to discourage joint consumption by requiring an extra payment for purchases of more than one product (superadditive bundling). It will be the major purpose of this chapter to determine when a firm will choose which of these strategies.

The crucial difference between bundling and the analysis of nonlinear pricing in Chapters 2 and 4 is that consumers here are assumed to want at most one unit of any particular good. Whilst this assumption of unit demands is broadly realistic for the examples given above — and so the analysis of this case has independent interest — it has the advantage of making the analysis rather more tractable than is the case when consumers choose continuous quantities. Hence this chapter, together with the wider bundling literature, can be viewed as a stepping stone towards a better understanding of the multiproduct nonlinear pricing problem of the next chapter, and indeed of multidimensional incentive problems in general.

The bundling literature briefly is as follows. Stigler (1963) provided an early and informal analysis of the problem illustrated by the example of the 'block booking' of films by distributors. He observed that when there is negative correlation in consumer tastes the seller can increase its profit by selling disparate goods (films in this instance) as a single block. The most widely cited paper is Adams and Yellen (1976). Through the use of (discrete) examples they construct a two–good model to examine the possible changes in profit and welfare in a move from independent pricing to bundling. They do not consider the possibility of superadditive bundling, and show that the welfare effects of introducing subadditive bundling are ambiguous. In common with other forms of price discrimination there are two opposing welfare effects: output may increase with bundling, which will tend to be good for welfare, but this output will be suboptimally distributed across consumers due to the differing marginal prices of the goods. They discuss loosely how correlation in consumers' tastes affects the optimal bundling strategy of the firm, and indicate that negative correlation may make subadditive bundling especially desirable. However, the paper provides no general conditions for subadditive bundling to be the dominant strategy for the firm, nor are there any principles suggested for an understanding of whether bundling is likely to be good for overall welfare as well as profits.

Schmalensee (1984) examines computationally the profitability of bundling in another two–good model on the assumption that reservation prices for the various goods are jointly Normally distributed. He shows that pure bundling (the special form of bundling where the firm offers only the bundle of both goods for sale) is more profitable than simple independent pricing if tastes are not too positively correlated, a result which formalizes (in his very special framework) Adams and Yellen's intuition that negative correlation in consumer tastes tends to encourage (pure) bundling. However, no optimal tariffs are found, nor is there any indication of what might happen with other distributions of consumer tastes. In a comment on the above paper, Long (1984) concisely provides many insights. He establishes what I call Result 3 in this chapter, and finds a useful equivalence between two–part tariffs and bundling. McAfee, McMillan and Whinston (1989) provide the result which I summarize as Result 2. Both of these papers look at the *marginal* incentives for bundling (starting from the position of the optimal non–bundled prices), and an implication of their analysis is that if reservation prices are independently distributed this marginal incentive is towards subadditive bundling. One missing link in this literature seems to be a discussion of what the *optimal* bundled prices look like.

The structure of this chapter is as follows. I discuss informally an aspect of bundling which is not covered by the model below, namely, the benefits of bundling in a competitive environment (Section 2). The formal model is introduced in Section 3 and it is shown that, provided consumers' tastes are continuously distributed (together with a technical assumption on the shape of the support of the distribution) the firm will choose to exclude some consumers from consuming some goods (Result 1). This has the corollary that the introduction of bundling could strictly increase *all* prices compared with non–bundling, a result which contrasts with most other kinds of price discrimination.

To gain more insights this model is specialized to the two–good case introduced by Adams and Yellen. Result 2 demonstrates that when types are continuously distributed it will almost always be the case that either sub– or superadditive bundling will dominate non–bundling as a pricing strategy. In the symmetric case a much stronger result is possible (Result 4): provided that the probability a consumer is prepared to pay a price p for one good given that she would pay p for the other good is non–increasing in p, the firm will engage in subadditive bundling (regardless of whether or not consumers are continuously distributed). The degree of correlation between customers' tastes for one good and their tastes for the other therefore plays a crucial role here and elsewhere in the chapter. This connection is examined informally in Section 5 using a rigorous concept of 'increasing correlation', and it is argued that there is no simple connection between increasing the degree of negative correlation in tastes and either increasing the firm's profits or increasing the incentives towards subadditive bundling. An application of Result 4 to the single–product case demonstrates that quantity discounts typically are good for a single–product firm's profits even when consumers have multidimensional preferences (Result 5).

The welfare implications of bundling are then discussed and an inequality is derived between the total surplus of any two bundling tariffs. In the symmetric case this reduces to the result that, in order for bundling to increase welfare it is necessary that total output rises. And as with these other forms of discrimination the overall welfare effect is certainly ambiguous, and there seem to be no clearcut principles governing when bundling will enhance welfare. Matters become easier to understand when the firm's price level is regulated (Section 8). There I consider three possible rules under which the regulated firm might be allowed to engage in bundling, two of which will unambiguously increase total welfare.

A second way in which the analysis can be simplified occurs when we let the number of goods become large (Section 9). When all consumers are *ex ante* identical (in the sense that their reservation prices are drawn from the same distribution) and reservation prices are independently distributed, it is shown that in the limit the firm is able to obtain the first–best profit level and consumers are left with no surplus. The optimal tariff may be approximated by a two–part tariff with marginal price equal to marginal cost (a form of subadditive bundling). Compared with the optimal non–bundling tariff welfare unambiguously rises with the introduction of bundling. On the other hand, when consumers differ in their income the problem reduces in the limit to the nonlinear pricing problem (where quantity is chosen continuously) considered, for instance, in Section 4.1 of Chapter 1 above. Subject to a hazard rate condition on consumers' incomes the firm will again offer a subadditive tariff. In this case consumers retain some of the gains from trade, and the overall welfare effect is ambiguous. Concluding remarks and connections with other literature follow in Section 10.

2. Bundling as a competitive strategy

It seems worth discussing briefly the connections between 'tie–in sales', product compatibility and bundling as I model it below. Tie–in sales — the (legally controversial) practice whereby firms make the purchase of some item conditional on the purchase of another — can be regarded as a special kind of bundling, so–called 'pure bundling' whereby the firm allows consumers to purchase only the composite good. Examples include film distributors requiring cinemas to purchase several films at once in a block booking and firms which lease capital equipment requiring their machines to be serviced by its staff. A well–documented case is that of Rank Xerox which used to tie the purchase of its brand of toner with the leasing of its photocopying equipment.[1] However, there are significant differences between some of these examples and the model below. In this chapter, bundling is used as a method by which to sort consumers out according to their differing preferences for the various goods. Tie–in sales seem almost inevitably to involve two goods which are perfect complements (e.g. photocopiers and toning fluid), and so customers will wish either to buy both goods or nothing. In effect there is only one good desired by consumers (in the

[1] The report into the affair (MMC, 1976) recommended that this was against the public interest. The Commission later made a general investigation into the desirability of tie–ins and found that unequivocal condemnation could not be justified — see MMC (1981).

above example it is the use of a photocopying machine).[2]

A second distinguishing characteristic of tie–in sales is that the firm attempts to link the purchase of one good (for instance, a photocopier) in which they have market power to the purchase of another good (for instance, toner or paper) which is available from other sources. There are at least two reasons why a firm might wish to do this.[3] Firstly, it is a method of sorting customers into high– and low–demand groups. If the company could control the price at which customers buy the complementary product then, by charging a higher price together with a lower price for the capital equipment it would, in effect, charge a higher price for the combined good to high users of the combined good. It is natural to think of this in terms of two–part tariffs. If, say, the supply of toner was competitive so that it may be purchased at cost, then, without tying, the photocopier firm is constrained to offer two–part tariffs with marginal price equal to marginal cost. With tying it would able to reduce the 'fixed charge' and increase the usage charge, something which will always increase profits provided that there is some degree of consumer diversity.

A second, more strategic motive for tie–in sales is analyzed by Whinston (1990). In his model there is an incumbent firm operating in two markets, one captive and one with an endogenous degree of competition. Prior to any entry decisions by other firms, the incumbent firm is able to commit to bundling its two goods together. The way that Whinston's model is constructed, if it does so this has the effect that the incumbent firm will price more aggressively in the second market than if it did not bundle; this will be anticipated by potential entrants and may deter entry in some circumstances. Thus bundling can be a means of extending market power from a monopolized market to other markets.

It is perhaps unreasonable to suppose that the incumbent could be able to commit

[2] The exception to this in the above list is that of films which fits the model described below well. Also, book clubs where members can buy certain books at a discount provided that they buy a certain number of books at full price seem to be a genuine example of bundling as I model it.

[3] Schmalensee (1982) considers the motives for tying a good produced by a monopolist with a second, competitively produced good when consumers' tastes are *not* perfectly correlated. This is just the model considered by Adams and Yellen except that the firm is constrained to offer one of the goods at price equal to marginal cost. He shows that, provided there is negative correlation between the two types, bundling remains a profit–increasing strategy.

to bundle its outputs unless it makes its monopolized good incompatible with other manufacturers' complementary goods. It is hard to imagine how this could be done with, say, photocopying paper, but examples such as cameras and lenses, video recorders and tapes, or computer hardware and software come to mind.[4] This is a complex area since the existence of network externalities is commonplace. A decade ago, a consumer considering whether to buy a 'Betamax' or a 'VHS' video recorder would wish to have the machine which would be compatible with the largest number of video tapes available then and in the future (all else being equal). In many cases this means that she would choose the machine of the firm which will have what she expects to be the larger market share. This is a game where starting first will often be an advantage.[5]

It is clear that tie–in sales and product compatibility can only be discussed in a competitive environment. In a monopoly setting, if all consumers view two goods as perfect complements the firm has no incentive to introduce bundling as people will buy both goods anyway (if they buy at all). The bulk of this chapter is about the use of bundling as a means by which to sort consumers out in the presence of asymmetric information, and specifically, it is about those cases where consumers' demands for the various goods are not perfectly correlated. This kind of multidimensional bundling by firms in a competitive environment has received little attention from economists so far, and this chapter is no exception.

3. The model

We shall use the formulation which has become standard in the bundling literature. Consumers have unit demands for each of the goods (i.e. a consumer will either buy one unit of any good or none at all) and have no complementarities or substitutabilities in

4 A more prosaic example is that of (old–fashioned) razors and blades. Razor manufacturers made their razors (and, they hoped, the blades) unique so that its customers would not buy other firms' blades. However, they would not mind those customers with other razors using their own blades too. Therefore, a firm would ideally like to have its blades compatible with all razors but no other blades compatible with its own razors. This accounts for the peculiar shape of these blades — see Adams (1978).

5 A special issue of the Journal of Industrial Economics is devoted to the topic of product compatibility — see Gilbert (1992). In that issue Katz and Shapiro (1992) have a model where starting first may be a disadvantage.

demand (so the utility of two goods is simply the sum of the separate utilities). More formally, let a monopolist manufacture n goods, label these goods 1, 2, ... , n, and the set of goods by $N = \{1, 2, \dots, n\}$. A *bundle* of goods is a subset B of N. A *tariff* is then a function $T(\cdot)$ giving the payment $T(B)$ required for each bundle B. A tariff T is *subadditive* if $T(B_1 \cup B_2) \leq T(B_1) + T(B_2)$ wherever B_1 and B_2 are disjoint bundles, and T is *superadditive* if the inequality is reversed. In order for a strictly subadditive tariff to be feasible for the firm we must assume that there are no effective markets for the resale of any individual goods which have been unbundled by arbitrageurs, and for a strictly superadditive tariff to be feasible the firm must be able to monitor its sales to particular consumers to avoid these consumers making up their own bundles by repeated single purchases. From now on we shall abstract from such considerations and assume that the firm is able to charge any kind of tariff. A special case of a tariff is when $T(B)$ is additive, so that $T(B) = \sum_{j \in B} p_j$ for some prices $\{p_i\}_{i=1}^n$ (i.e. when the tariff is both sub– and superadditive). We term this a *non–bundling* tariff since a consumer simply faces a price for a given good which is independent of her consumption of the other goods.

Consumers of these goods are distinguished by an n–dimensional parameter $\boldsymbol{\alpha} = (\alpha_1, .. \alpha_n)$ in the following manner: let a consumer of 'type $\boldsymbol{\alpha}$' obtain a utility (in monetary terms) of α_i if they consume a unit of the good i. The parameter α_i is the reservation price of the consumer for good i in the sense that the consumer will purchase the good given the non–bundling tariff $\{p_i\}$ if and only if $\alpha_i \geq p_i$. Therefore, if she consumes the bundle B and pays a fee T she enjoys a utility of $\sum_{j \in B} \alpha_j - T$. If consumers always have the option of not purchasing any good we may assume that $T(\emptyset) = 0$. Given a tariff schedule $T(\cdot)$, a type $\boldsymbol{\alpha}$ consumer will choose the bundle $B(\boldsymbol{\alpha}, T)$ in order to:

$$(1) \qquad \underset{B \subseteq N}{\text{maximize}} \qquad \sum_{j \in B} \alpha_j - T(B) \; .$$

Demands will always be an increasing function of $\boldsymbol{\alpha}$, i.e. if $\tilde{\boldsymbol{\alpha}} \geq \boldsymbol{\alpha}$ then $B(\tilde{\boldsymbol{\alpha}}, T) \supseteq B(\boldsymbol{\alpha}, T)$. Let $D(B, T)$ be the total number of consumers that choose to purchase bundle B given tariff T. If $s(\alpha)$ is the surplus of a type α consumer, i.e.

$$s(\alpha) = \max_{B}: \sum_{j \in B} \alpha_j - T(B))$$

then $s(\cdot)$ is convex, increasing, non–negative and, if it is differentiable at α, its derivative will be the type α consumer's demand function:

$$\text{(2)} \qquad \frac{\partial s}{\partial \alpha_i}(\alpha) = \begin{cases} 1 & \text{if } i \in B(\alpha, T) \\ 0 & \text{otherwise.} \end{cases}$$

The monopolist cannot observe a particular consumer's type but is assumed to believe that α is distributed according to some joint distribution function $F(\alpha)$. If the firm faces a cost of supplying any customer with a bundle B of $C(B)$ it will choose its tariff T^* in order to:

$$\text{(P)} \qquad \underset{T}{\text{maximize:}} \quad \pi(T) \equiv E\,[T(B(\alpha, T)) - C(B(\alpha, T))]$$

where E is expectation with respect to the underlying distribution of consumer types F and the demand function B is given by the solution to (1), given T. Writing this in terms of the aggregate demand function $D(B, T)$ the firm's problem is to:

$$\text{(P}'\text{)} \qquad \underset{T}{\text{maximize:}} \quad \pi(T) \equiv \sum_{B} D(B, T)[T(B) - C(B)]\,.$$

Thus, if there are n underlying goods the firm's problem in choosing the optimal bundling tariff may be viewed as choosing the optimal prices for the 2^n bundles, given the bundle demand functions $D(B, T)$. Under the assumptions of the model consumers view the various bundles as substitutes:

LEMMA 1. *For any tariff* T, *increasing the charge for some bundle* B *will increase the demand for all other bundles* $\tilde{B}$.

PROOF. Given tariff T, from (1) a consumer of type α will purchase bundle $\tilde{B}$ if and only if

$$\sum_{j \in \tilde{B}} \alpha_j - T(\tilde{B}) \geq \sum_{j \in B} \alpha_j - T(B) \quad \text{for all } B \neq \tilde{B}.$$

Increasing the charge on any of these other bundles B leaves this inequality being satisfied and the set of consumers choosing bundle $\tilde{B}$ will increase (at least weakly). □

COROLLARY 1. *The profit–maximizing tariff* T^* *satisfies* $T^*(B) \geq C(B)$ *for all* B.

PROOF. Suppose on the contrary that there is some non–empty subset of bundles $B_1, \ldots, B_k$ which are optimally priced strictly below marginal cost: $T^*(B_i) < C(B_i)$. Then increasing the charge for each of these bundles to equal cost whilst keeping the charge for other bundles fixed will (i) eliminate the loss the firm makes from the sale of these bundles and (ii), using Lemma 1, increase the profit the firm makes on all other bundles (which are priced above cost). Therefore profit increases and the original tariff could not be optimal. □

There are at least three broad questions we are interested in answering about problem (P):

— What determines when the optimal tariff T^* is sub– or superadditive, i.e. when will the firm choose to give 'quantity' discounts or 'quantity' premia?

— How does the degree of correlation in consumer tastes affect profits and the incentives toward subadditive bundling?

— How will the optimal bundling tariff compare with the optimal non–bundling tariff? In particular, is it possible that the introduction of bundling will either reduce or increase *all* prices? Will the number of customers served increase or decrease as a result of bundling? More generally, when will bundling result in an increase in total welfare?

So far there is only one result that holds at (nearly) this level of generality:

RESULT 1. *If there are* n *goods and consumer types are continuously distributed on a closed convex set* $A \subset \mathbb{R}^n_+$ *(n > 1) by a density function* $f(\boldsymbol{\alpha})$ *that is bounded, and* A *is such that the point* $\boldsymbol{a} = (a_1, \dots, a_n)$ *that minimizes* $\{ \sum_1^n \alpha_i \mid \boldsymbol{\alpha} \in A \}$ *is unique, then it will be optimal for the firm to exclude some customers from consuming some goods.*

PROOF. See Appendix.

The proof works by showing that if the tariff is increased by a small amount ϵ then the firm gains revenue of ϵ from those that continue to buy, and except in the single–product case the number who leave the market is of order strictly greater than ϵ. Therefore, the former effect will outweigh the latter and it cannot be optimal to have all consumers consume all goods. The only cases where this argument breaks down is when the lower boundary of the set of types A is of the form $\sum_1^n \alpha_i = \text{constant}$, in which case the number of consumers leaving as a result of the change will also be of order ϵ.

This result contrasts with the non–bundling equilibrium. For instance, suppose that $\boldsymbol{\alpha}$ is uniformly distributed on the cube with $\mathbf{a}$ as the south–west vertex and $\mathbf{a} + \mathbf{1}$ as the north–east vertex. Then, if it had no costs and did not engage in bundling, the firm would choose to set its price for good i equal to a_i provided that $a_i \geq 1$, so that if $\mathbf{a} \geq \mathbf{1}$ then all consumers would buy all goods. However, we have seen that this strategy would not remain optimal if bundling were allowed and hence, in this case, the introduction of bundling will increase all charges:

COROLLARY 2. *Allowing the firm to bundle could result in an increase in the charge for all bundles, which would in turn imply a reduction in welfare (as measured by consumer surplus plus profit).*[6]

Before leaving this exposition of the general model we examine the possible desirability of the firm offering a stochastic tariff. This question is analyzed in McAfee and McMillan (1988) when all agents are risk–neutral. For simplicity consider the two–good case. A deterministic tariff will then consist of three charges, one for each of the individual goods, and one for the bundle of both goods. A stochastic tariff gives a charge $T(\gamma_1, \gamma_2, \gamma_b)$ for the lottery consisting of the consumer getting good 1 only with probability γ_1, good 2 only with probability γ_2 and both goods with probability γ_b. These three probabilities must be non–negative and sum to less than one. In purchasing such a lottery, a type α consumer gains a surplus of

$$\alpha_1(\gamma_1 + \gamma_b) + \alpha_2(\gamma_2 + \gamma_b) - T(\gamma_1, \gamma_2, \gamma_b)$$

and will choose $(\gamma_1, \gamma_2, \gamma_b)$ in order to maximize this surplus given the nonlinear tariff T. Provided that the firm has linear unit costs, McAfee and McMillan show that, subject to a condition on the density of types, such stochastic tariffs can do no better than the simple determinstic tariffs we have been using up until now.[7] For the remainder of the chapter, then, we will continue to consider only deterministic strategies by the firm.

Finally in this section we relax the assumption that there are many consumers, that there are many units of each good and that the firm offers a non–negotiable tariff, and instead suppose that there are comparatively few consumers, that the firm has only a single

---

[6] This corollary contrasts with third–degree price discrimination, where allowing the firm to discriminate usually results in some prices increasing and some decreasing (for instance, this is certainly true if the multiproduct profit function π may be written as $\pi(p) = \Sigma\pi_i(p_i)$ and each π_i is single peaked).

[7] Their conclusion is valid only in the case of two goods and when consumers are risk–neutral. If $f(\alpha_1, \alpha_2)$ is the density of types, their condition on f is $3f + \alpha_1 f_1 + \alpha_2 f_2 \geq 0$, where f_i denotes the partial derivative with respect to α_i. This also happens to be equation (23) in Chapter 4 below.

unit of each good and that these goods are sold by auction. The problem for the seller is to decide whether to sell her goods in separate auctions or in a single bundled auction. Palfrey (1983) examines this when consumers are risk–neutral and *ex ante* identical. There are two goods and two consumers. Consumer j has a willingness to pay for good i denoted by $\alpha_j{}^i$ and for the two goods together of $(\alpha_j{}^1 + \alpha_j{}^2)$ (so there are no substitutabilities or complementarities in consumption), where $i, j = 1, 2$. For each consumer j the two parameters $(\alpha_j{}^1, \alpha_j{}^2)$ are independently drawn from some joint distribution with distribution function $G(t_1, t_2)$ (although this is not to say that given j, the two parameters $(\alpha_j{}^1, \alpha_j{}^2)$ are themselves independently distributed). The problem for the seller, then, is whether to have two separate auctions or to sell the two goods together as a bundle. Once the seller has decided upon her strategy and given the fact that the joint valuations $(\alpha_j{}^1, \alpha_j{}^2)$ are independently distributed across the two consumers, we know from Section 7 of Chapter 1 above that the Vickrey second price auction is optimal. What are the expected payoffs to the seller under the two options? We consider the non–bundling and bundling strategies in turn:

Non–bundling: If the seller has two auctions, one for each of the goods, then she will expect to receive $E[\ \alpha^i_{min}\ ]$ from the auction for good i, where α^i_{min} is the lower draw of α^i across the two consumers: $\alpha^i_{min} = \min\{\ \alpha_1{}^i, \alpha_2{}^i\ \}$. Total expected revenue will be $E[\ \alpha^1_{min} + \alpha^2_{min}]$. Thus in this case the payoff will not depend on the joint distribution of (α^1, α^2), but only on the marginal distributions of α^1 and α^2.

Bundling: In this case, the expected revenue from the single auction is $E[\ (\alpha^1 + \alpha^2)_{min}\ ]$, where $(\alpha^1 + \alpha^2)_{min}$ is the lower draw of the single random variable $(\alpha^1 + \alpha^2)$ across the two consumers. Since the distribution of $(\alpha^1 + \alpha^2)$ does depend on the joint distribution of (α^1, α^2), the expected revenue will also depend on the joint distribution.

It is now simple to see what the seller should do. Since $\alpha^1_{min} + \alpha^2_{min} \leq (\alpha^1 + \alpha^2)_{min}$ for every possible draw we see that the variable $\alpha^1_{min} + \alpha^2_{min}$ is stochastically dominated by the variable $(\alpha^1 + \alpha^2)_{min}$ and so the seller must prefer to bundle the two goods together in a single auction (no matter what her attitude is towards risk).[8]

4. Optimal bundling with two goods and constant unit costs

Suppose now that there are just two goods, good 1 and good 2, and that the firm's cost in producing a unit of good i is c_i (this is exactly the framework introduced by Adams and Yellen). Let $F(t_1, t_2)$ be the joint distribution function for consumer types (α_1, α_2), and denote the marginal distribution functions by $G_i(t)$. In this case a choice of tariff involves the firm setting three prices, the price for a single purchase of good i, denoted t_i, and the bundled price of the two goods, t_b. A non–bundling tariff involves setting two prices $\{p_i\}$, where p_i is the price of good i (regardless of the other purchase decision). We first solve the problem of finding the optimal non–bundling tariff, then consider the more complex problem of finding the optimal bundling tariff.

Faced with the prices $\{p_i\}$ a consumer buys good i if and only if $\alpha_i \geq p_i$, and in particular, she makes her decision on good i independently of her taste for the other good. Since the number of consumers who have a reservation price for good i greater than p_i is $1 - G_i(p_i)$ (where we have normalized the total number of consumers to 1), with these prices the firm will make a profit of $\sum_1^2 (p_i - c_i)(1 - G_i(p_i))$, and so will choose the optimal non–bundling prices, p_i^*, to satisfy:

$$(3) \qquad p_i^* \text{ maximizes } \qquad (p - c_i)(1 - G_i(p)) \quad \text{over } p \geq 0.$$

In determining the optimal non–bundling tariff, then, the firm has no interest in the joint distribution of consumer types, F, and will base its prices only on each marginal distribution of types, G_1 and G_2.

If we now allow bundling the problem of how many consumers will buy which

---

8 This result would not be affected if there were more than two goods. When there are more than two buyers things become more complex, and Palfrey argues that as the number of consumers gets large it is less likely that bundling will continue to dominate non–bundling as a strategy.

bundles becomes more complex. If the tariff is given by t_1, t_2 and t_b, the pattern of demand is as given in Figures 1A and 1B. These diagrams look very different depending upon whether the bundled price t_b is greater than or less than the sum of the individual prices $t_1 + t_2$, i.e. upon whether a super– or subadditive bundling tariff is offered. The crucial difference is that in the former case demand for one good is a (weakly) decreasing function of the other taste parameter, whereas in the latter it is an increasing function. In either case, a consumer will no longer base her decision on whether to purchase a given good only on her taste parameter for that good, but on both of her taste parameters.[9]

Let R_b denote the set of consumers who buy both goods, R_i the set who buy only good i, and R_0 the set who buy nothing. Let $\mu(\cdot)$ denote the measure on the plane induced by the distribution function F, so that $\mu(E)$ is the number of consumers in the set E. Then the firm's profit is given by:

$$\pi(t_1, t_2, t_b) \equiv \sum_1^2 (t_i - c_i)\mu(R_i) + (t_b - c_1 - c_2)\mu(R_b) \tag{4}$$

where R_i and R_b depend on the choice of tariff as shown in Figure 1. We could try to solve this problem by direct calculation of the functions $\mu(R_i(t_1, t_2, t_b))$, but this turns out to be more difficult than might be expected (although well suited to computer solution for particular choices of F). Two solutions are possible, however:

<u>Example 1</u>: Let there be two goods with zero costs, and let consumer tastes be distributed uniformly on the unit square $[0, 1]^2$. In this case, symmetry implies that $t_1 = t_2$, so denote this common price by t, and $\mu(R_1) = \mu(R_2)$. There are then two situations to consider. If $t_b < 2t$ then $\mu(R_1) = \mu(R_2) = (1 - t)(t_b - t)$, and the number of people who

---

9 A useful way of thinking about bundling has been suggested by Long (1984). Given the bundling tariff t_1, t_2 and t_b, define p_i to be the marginal price of good i given that the other good is also being purchased, i.e. $p_i = t_b - t_j$ $(j \neq i)$. Let A be the discount given by the bundling tariff for joint purchase, i.e. $A = t_1 + t_2 - t_b$. Then the bundling tariff is precisely equivalent to the two–part tariff given by the fixed charge A for entry to the market and the marginal prices p_i. (We should, however, bear in mind that this equivalence holds only in the two–good case.) Thus the problem of whether the firm should offer a super– or subadditive tariff corresponds to the question of whether it should offer a negative or positive fixed entry–fee to the market.

buy both goods is $(1-t)(1+t-t_b)+\frac{1}{2}(2t-t_b)(2-t_b)$. Therefore, total profit is

$$2t(1-t)(t_b-t)+t_b\{(1-t)(1+t-t_b)+\tfrac{1}{2}(2t-t_b)(2-t_b)\}.$$

On the other hand, if $t_b > 2t$ profits are given by

$$t\{1-t^2-(1+t-t_b)^2\}+t_b(1+t-t_b)^2.$$

A check shows that it is best to take the former case, which is maximized at $t=\frac{2}{3}$ and $t_b=(4-\sqrt{2})/3 \approx 0.862$, obtaining a profit of about 0.55. Thus, the firm does best if the pattern of demand is as shown in Figure 1B. In comparison with the optimal non–bundling price of $p^*=\frac{1}{2}$ (which gives a level of profit of 0.5), the move to a bundling tariff results in an increase in the price of a single good, but a reduction in the price for both goods. □

To see that subadditive bundling may also be optimal consider the following discrete example:

Example 2: There are three consumers, one with type $\alpha=(1,0)$, a second with type $\alpha=(0,1)$ and the remaining having type $\alpha=(2,2)$. If unit costs are zero then the optimal tariff for the monopolist is to set $t^*=1$ and $t_b^*=3$. No other combination of t and t_b will do better and so in this case superadditive bundling is optimal. Of course, here φ is certainly not decreasing — it starts at $\varphi(0)=1$, immediately drops to a half and finally rises (at $t=1$) back to one again. □

One thing which is simple to do is to find which directions of movement from the optimal non–bundling prices increase profit, i.e. setting $t_i=p_i^*$ and $t_b=p_1^*+p_2^*$ with p_i^* given by (3), what are the partial derivatives of π in (4)? In order for this question to make any sense we must have π differentiable, i.e. that consumer types are distributed by some continuous density function $f(\alpha)$. In this case, given that it may make a small change, would the firm prefer to move to the regime of Figure 1A or Figure 1B? To understand this, introduce the notation

$$\varphi_1(p_1, p_2) = \text{prob}\{\alpha_1 \geq p_1 \mid \alpha_2 \geq p_2\}$$
$$\varphi_2(p_1, p_2) = \text{prob}\{\alpha_2 \geq p_2 \mid \alpha_1 \geq p_1\}.$$

Thus $\varphi_i(p_1, p_2)$ gives the probability that a consumer has a reservation price for good i of at least p_i, given that she would be prepared to pay at least p_j for a unit of the other good. Given our assumption that types are continuously distributed these functions are differentiable. Using these functions we can derive the following result:

RESULT 2 *(McAfee et al. (1989), restated). If* p_i^* *are the optimal non-bundling prices which are interior solutions to (3), then*

(5) $$\frac{\partial \pi}{\partial t_b}(p_1^*, p_2^*, p_1^* + p_2^*) = -(1 - G_1)\varphi_2 +$$
$$(p_2^* - c_2)(1 - G_2)\frac{\partial \varphi_1}{\partial p_2} + (p_1^* - c_1)(1 - G_1)\frac{\partial \varphi_2}{\partial p_1}$$

(where all functions are evaluated at (p_1^*, p_2^*)*).*

REMARK: For this result to hold it is important that p_i^* be interior solutions in (3). Result 1 demonstrates that when p_i^* is at the lower end of the support of α_i the marginal incentive is to increase the price of the bundled good regardless of the form of the functions φ_i.

PROOF. We may write the number of consumers who buy both goods in two ways:

$$\mu(R_b(p_1^*, p_2^*, p_1^* + p_2^*)) = (1 - G_2(p_2^*))\varphi_1(p_1^*, p_2^*)$$
$$= (1 - G_1(p_1^*))\varphi_2(p_1^*, p_2^*)\ ,$$

and

$$\mu(R_i(p_1^*, p_2^*, p_1^* + p_2^*)) = 1 - G_i(p_i^*) - \mu(R_b(p_1^*, p_2^*, p_1^* + p_2^*)).$$

Suppose the firm reduces its charge for the two goods by ϵ, so that $t_b = p_1^* + p_2^* - \epsilon$. In this case, the change in the number of consumers who used to buy only good 1 but now buy both goods is approximately

$$\{(1 - G_2)\frac{\partial \varphi_1}{\partial p_2} - g_2\varphi_1\}\epsilon$$

(where g_i is the marginal density of good i), and similarly for good 2. This expression holds whether ϵ is positive or negative. The firm makes an extra profit of approximately $(p_2 - c_2)$ from each of these consumers, $(p_1 - c_1)$ from each of the consumers who used to buy only good 2 and now buy both goods, but loses ϵ from each of the consumers who buy both goods in any case. The total change in profit has the sign of (after canceling ϵ):

$$-(1 - G_2)\varphi_1 - (p_2^* - c_2)\{(1 - G_2)\frac{\partial \varphi_1}{\partial p_2} - g_2\varphi_1\} - (p_1^* - c_1)\{(1 - G_1)\frac{\partial \varphi_2}{\partial p_1} - g_1\varphi_2\}$$

$$= (1 - G_1)\varphi_2 - (p_2^* - c_2)(1 - G_2)\frac{\partial \varphi_1}{\partial p_2} - (p_1^* - c_1)(1 - G_1)\frac{\partial \varphi_2}{\partial p_1}$$

where we have used the fact that $1 - G_i(p_i^*) = (p_i^* - c_i)g_i(p_i^*)$ since p_i^* is an interior solution to (3). Therefore, the marginal incentive is for the firm to move towards sub–additive bundling whenever the term in (5) is negative. □

COROLLARY 3. *It is sufficient for the marginal incentive of the firm to be towards sub–additive bundling if the functions* φ_1 *and* φ_2 *are non–increasing in* p_2 *and* p_1 *respectively. A special case occurs when* α_1 *and* α_2 *are independently distributed.*

If the functions φ_1 and φ_2 are decreasing in p_2 and p_1 respectively then the taste parameters are in a sense negatively correlated, and so we see that the intuition of the earlier literature that negative correlation in tastes is associated with subadditive bundling is confirmed here (but see Section 5 below).

Results are more clear cut if we make the additional assumption of symmetry: unit costs are the same for each good, say c, and the joint distribution function F is symmetric. In this case a tariff consists of two prices, the price of any single good t, and the price of two goods t_b. Write firm profits given this tariff as $\pi(t, t_b)$. Also, the optimal non–bundling prices are equal and denoted by p^*. Now define

$$\varphi(p) \equiv \text{prob}\{\alpha_1 \geq p \mid \alpha_2 \geq p\}$$

to be the probability that a consumer would buy one good at price p given that she buys the other at price p (and so φ is a kind of 'hazard rate' on the joint distribution of types). If α_1 and α_2 are independently distributed it is clear that φ is decreasing. Similarly, if types are negatively correlated in the sense that the function $\varphi_1(p_1, p_2)$ is decreasing in p_2, then φ is also decreasing. On the other hand, if φ is increasing over some range then, in some sense, tastes must be strongly positively correlated in certain regions: not only must $\varphi_1(p_1, p_2)$ now be increasing in p_2, it must be increasing in p_2 faster than it is decreasing in p_1 (on the diagonal). Whether φ is increasing or not is vital to the understanding of the problem as the following result demonstrates.

Result 3 *(Long (1984), restated). Starting from the non–bundling optimum of* $(t, t_b) = (p^*, 2p^*)$, *where* p^* *is an interior solution to (3), profits are increased by increasing (respectively decreasing)* t_b *if* φ *is increasing (respectively decreasing) at* p^*.

Proof. Using the fact that $\varphi(p) \equiv 1 - \frac{G(p) - F(p, p)}{1 - G(p)}$ it is simple to calculate that the change in profits obtained by increasing t_b (or equivalently by decreasing t) is given by

$$(6) \qquad \frac{\partial \pi}{\partial t_b}(p^*, 2p^*)\,) \qquad = (1 - G(p^*))\varphi(p^*) + (p^* - c)\frac{d}{dp}[\varphi(p^*)(1 - G(p^*))]$$

$$= \varphi'(p^*)(p^* - c)(1 - G(p^*)) \qquad \text{from (3).}$$

Thus, profits increase when t_b is decreased if and only φ is decreasing at p^*. □

While this condition is very simple, it would more satisfactory to be able to get beyond results about these marginal incentives — when will it be the case that the *optimal* tariff (t^*, t_b^*) results in super– or subadditive bundling? This is answered by the following central result of this chapter:

RESULT 4. *If* φ *is non–increasing it is optimal for the firm to set* $t_b^* \leq 2t^*$.

REMARK: We do not need any kind of differentiability for this proof, and, in contrast to the above two results about marginal changes (and Result 1), this result is valid both for discrete and continuous distributions of types.

PROOF. *By contradiction*: Suppose that φ is non–increasing and that optimally $t_b^* > 2t^*$, so that the pattern of demand is as shown in Figure 1A. In this case the number of consumers who buy both goods is $\varphi(t_b^* - t^*)(1 - G(t_b^* - t^*))$ and the number who buy at least one of any good is $2(1 - \varphi(t_b^* - t^*))(1 - G(t_b^* - t^*))$. Therefore, optimal profit π^* is given by:

$$\pi^* = (t_b^* - t^* - c)\varphi(t_b^* - t^*)(1 - G(t_b^* - t^*)) +$$
$$(t^* - c)[2(1 - G(t^*)) - \varphi(t^*)(1 - G(t^*))],$$

or

$$(7) \qquad \pi^* = \varphi(t_b^* - t^*)\bar{\pi}(t_b^* - t^*) + (2 - \varphi(t^*))\bar{\pi}(t^*)$$

where $\bar{\pi}(p) \equiv (p - c)(1 - G(p))$ is the single–product profit function. By revealed preference the following three inequalities must hold:

$$\varphi(t_b^* - t^*)\bar{\pi}(t_b^* - t^*) + (2 - \varphi(t^*))\bar{\pi}(t^*) \geq 2\bar{\pi}(p^*),$$

$$\bar{\pi}(p^*) \geq \bar{\pi}(t^*)$$

and

$$\bar{\pi}(p^*) \geq \bar{\pi}(t_b^* - t^*) .$$

The first inequality arises because profit with bundling must be at least as high as optimal profit without bundling, and the second two arise from the definition of p^* in (3). Substituting the second two inequalities into the first implies that

$$\varphi(t_b^* - t^*)\bar{\pi}(p^*) + (2 - \varphi(t^*))\bar{\pi}(p^*) \geq 2\bar{\pi}(p^*), \text{ or}$$

(8) $$\varphi(t_b^* - t^*) - \varphi(t^*) \geq 0 .$$

If the inequality is strict in (8) then the assumption that φ is non–increasing implies that $t_b^* - t^* < t^*$, i.e. that the firm practices subadditive bundling in contradiction to the initial assumption. If, on the other hand, we have equality in (8) then in particular we must have

$$\pi^* = \varphi(t_b^* - t^*)\bar{\pi}(t_b^* - t^*) + (2 - \varphi(t^*))\bar{\pi}(t^*) = 2\bar{\pi}(p^*)$$

so that the firm does no better by bundling than by not bundling, and therefore that it is (at least weakly) optimal to set the non–bundling price p^* for each good. In either case the result is proved. □

Very many joint distributions of taste parameters satisfy the condition of this result and for these distributions the firm will offer a subadditive tariff.[10] In order for superadditive bundling to be a feasible option for the firm it must be able to monitor sales to particular consumers to prevent consumers saving money and making up their own bundle by repeated purchases of single goods. For many firms such monitoring is not possible. However, for firms facing consumers who satisfy the condition of this result this inability to offer superadditive tariffs does not harm their profits.

5. The impact of increasing correlation in tastes

The importance of the degree of correlation in consumers' taste parameters in Section 4 was evident. Also, much of the previous literature has suggested that as consumers' tastes become 'more negatively correlated', either the firm's profits increase or (subadditive) bundling is more likely to be the optimal strategy.[11] Here we rigorously formulate this claim and test to see whether or not this is the case.

In order to discuss this issue we must first make precise what we mean by more or less correlation. Epstein and Tanny (1980) propose a measure of 'increasing correlation' in a pair of joint random variables analogous to the definition by Rothschild and Stiglitz (1970) of 'increasing risk' in a single random variable. In order to isolate the effect of correlation from other factors they keep the marginal distributions of the two random variables fixed. The equivalent concept of a 'mean–preserving spread' is the 'elementary correlation–increasing transformation' (CIT) which, roughly, involves removing mass from north–west and south–east corners of the support of the joint density function and replacing it in the north–east and south–west corners, in such a way as to preserve the marginal distributions. An example of such a transformation is depicted in Figure 2A. The reverse procedure defines a correlation–decreasing transformation (CDT). A pair of

---

10 It is not possible to consider the analogous result to this — that φ everywhere increasing implies that super–additive bundling is optimal — since (apart from the uninteresting case of perfect correlation) there are no joint distributions of tastes which have φ everywhere increasing: it is always the case that $\varphi(0) = 1$ and $\varphi(t) \leq 1$. Therefore φ increasing implies $\varphi(t) \equiv 1$ which in turn implies perfect correlation.

11 For instance, see Adams and Yellen (1976, page 485). In his model with tastes Normally distributed Schmalensee shows that pure bundling (a kind of subadditive bundling) is more likely to dominate independent pricing the more negatively correlated are tastes.

joint random variables is then said to be 'more positively correlated' than a second pair if it may be obtained from the latter by a (convergent) sequence of these CITs.[12]

What effect does increasing correlation in this sense have on the firm's profits? Suppose tastes are initially distributed in such a way that the firm chooses optimally to engage in superadditive bundling, so that the pattern of demand is as shown in Figure 1A (we assume throughout this section that there are two goods). Then a CIT as shown in Figure 2A will unambiguously increase the firm's profits. To see this suppose that the weight in each square is δ (which is positive in the north–east and south–west corners, negative elsewhere) and that the firm keeps its bundling tariff unchanged. The resulting change in profit is

$$\delta\{(t_b{}^* - 2c) - 2(t^* - c)\} > 0 ,$$

the inequality holding since $t_b{}^* > 2t^*$ by assumption. In many cases the firm will be able to even better than this by changing its tariff. Since this argument is not affected if a sequence of CITs is used we deduce that, starting from a distribution of tastes which induces the firm to offer a superadditive tariff, any other distribution with higher positive correlation unambiguously will increase the firm's profits. However, even though reducing the degree of correlation (i.e. putting $\delta < 0$ in the above) will reduce firm profit keeping its tariff fixed, we cannot deduce that making the tastes parameters less positively correlated will reduce firm profits because the firm may be able to change its tariff in such a way as to counteract the loss of weight from the north–east consumers. Indeed, if production is costless a move from perfect positive correlation to perfect negative correlation will increase firm profits to the first–best level. Using an envelope argument, though, we can

12 This defines a partial ordering of pairs of random variables (all of which have the same marginal distributions) which turns out to have a particularly simple characterization in terms of joint distribution functions: if (α_1, α_2) and (β_1, β_2) are two joint random variables with respective joint distribution functions $F_\alpha(t_1, t_2)$ and $F_\beta(t_1, t_2)$, then the former variables are more positively correlated than the latter if and only if $F_\alpha \geq F_\beta$. In addition, if (α_1, α_2) are more positively correlated than (β_1, β_2) then the scalar variable $(\alpha_1 + \alpha_2)$ is more risky (in the sense of Rothschild and Stiglitz) than $(\beta_1 + \beta_2)$.

show that a sufficiently small CDT will certainly reduce the firm's profits.

Suppose now that the initial distribution of tastes is such that the firm offers subadditive bundling, the case that Result 4 indicated might be the more typical of the two possibilities. Unfortunately the analysis now is less clear cut. If we impose a CDT in a manner which takes weight from Regions R_0 and R_b and replaces it in Regions R_1 and R_2 as given in Figure 1B, then the change in firm profit keeping the tariff unchanged is

$$\delta\{2(t^* - c) - (t_b^* - 2c)\} > 0 ,$$

where δ is the weight of each square of the CDT and the inequality follows from the assumption that $2t^* > t_b^*$. Thus, any such CDT and any sequence of such CDTs will increase the profits of the firm. But now suppose that the CDT has the configuration as shown in Figure 2B, so that the north–west and south–east squares no longer lie in Regions R_1 and R_2, but the north–east square remains in Region R_b (such a configuration was not possible in the superadditive case of Figure 1A). In this case, if it keeps its tariff fixed the firm will find its profits decrease, and by using an envelope argument, if δ is sufficiently small the firm will lose money even after taking into account its ability to change the tariff. Therefore, it is ambiguous whether increasing the degree of negative correlation will increase or decrease the profit of the firm practicing subadditive bundling.

Turning to the question of whether increasing the degree of positive correlation will make subadditive bundling more or less likely, matters become even less clear–cut. It has been argued in the earlier literature that increasing the degree of negative correlation in tastes makes subadditive bundling more likely. Using the formal definition of increasing negative correlation we could state this result as: if tastes are initially distributed so that the firm offers a subadditive tariff, then if the distribution is altered so that tastes become more negatively correlated then the firm will certainly continue to offer a subadditive tariff. If (α_1, α_2) is a pair of taste variables which are more negatively correlated than the pair (β_1, β_2) then

$$\varphi_\alpha(p) \equiv \text{prob}\{\ \alpha_1 \geq p \mid \alpha_2 \geq p\ \} \ \leq\ \varphi_\beta(p) \equiv \text{prob}\{\ \beta_1 \geq p \mid \beta_2 \geq p\ \}.$$

However, this is of no help when using Result 4 since the fact that φ_α is less than φ_β does not imply that φ_α is decreasing whenever φ_β is decreasing. Indeed, one can construct an example where initially the firm offers a subadditive tariff but when the degree of negative correlation is increased the firm switches to a superadditive bundling strategy:

Example 3: There are 11 consumers and zero costs. Initially there are three consumers with type $\boldsymbol{\alpha} = (0, 0)$, three of type $\boldsymbol{\alpha} = (1, 1)$, three of type $\boldsymbol{\alpha} = (2, 2)$, one of type $\boldsymbol{\alpha} = (\frac{1}{2}, 1\frac{1}{2})$ and one of type $\boldsymbol{\alpha} = (1\frac{1}{2}, \frac{1}{2})$. In this case it is optimal for the firm to offer the subadditive tariff consisting simply of the single charge $t_b^* = 2$ for the two goods. Now consider increasing the degree of negative correlation so that the distribution becomes three consumers with type $\boldsymbol{\alpha} = (0, 1)$, three of type $\boldsymbol{\alpha} = (1, 0)$, three of type $\boldsymbol{\alpha} = (2, 2)$, one of type $\boldsymbol{\alpha} = (\frac{1}{2}, 1\frac{1}{2})$ and one of type $\boldsymbol{\alpha} = (1\frac{1}{2}, \frac{1}{2})$ (this is certainly a CDT since three units are transferred from each of the points $(0, 0)$ and $(1, 1)$ and replaced at the points $(0, 1)$ and $(1, 0)$, all else remaining constant). With this new distribution, the situation is similar to Example 2 and it is now optimal for the firm to set the superadditive tariff $t^* = 1$ and $t_b^* = 3$. □

We deduce from this example that there is no simple connection between the degree of negative correlation in consumers' tastes and the incentive towards subadditive bundling.

6. An application to the nonlinear pricing of one good

Maskin and Riley (1984) demonstrate that it will often be optimal for the single–product firm to offer quantity discounts to consumers. They use a model where consumers differ by only a one–dimensional parameter: once a particular consumer's utility from consuming a given quantity is known, her complete utility schedule is known. It is useful to see how robust their conclusions are with respect to this assumption. We can use the above model to partially answer the question.

For a consumer of type $\boldsymbol{\alpha}$, let α_i be the utility obtained by her consumption of the i^{th} unit of a good, given that $i-1$ units have already been consumed. Let t_m be the firm's charge for m units of the good. By the principal of diminishing marginal utility it makes sense to restrict attention to those consumers with type $\boldsymbol{\alpha}$ such that $\alpha_1 \geq \alpha_2 \geq .. \geq \alpha_n$ where n is now the maximum number of units of the good that any consumer would wish to have. In the special case where $n = 2$ this amounts to requiring that the distribution of types lies below the line $\alpha_1 = \alpha_2$ — see Figure 3. We can now reflect the distribution of types along this 45 degree line to end up with the symmetric case considered in Result 4. If we do this then the set E_2, those consumers who buy two units, corresponds to half the set R_b in Figure 1B, and those who consume one unit, E_1, corresponds to R_1. Therefore, we can simply apply Result 4 to this case to obtain:

RESULT 5. *Suppose that consumers wish to consume at most two units of a single good and that they enjoy a diminishing marginal utility from the good. Then it is optimal for the firm to set* $t_2 \leq 2t_1$ *(i.e. to offer a quantity discount) if the function*

$$\varphi(p) = \text{prob}\{\alpha_2 \geq p \mid \alpha_1 \geq p\}$$

is non–increasing.

Thus we see that it is very likely that the firm will wish to offer quantity discounts (but by suitably modifying Example 2 it is certainly not inevitable that this be the case), and so this nonlinear pricing strategy continues to be desirable in a simple model where consumers differ by a parameter of more than one dimension.[13]

[13] This is also the case in the more sophisticated example of Laffont *et al.* (1985). There, a consumer's taste for a single product is determined by a two–dimensional parameter and quantity is continuous (rather than being restricted to 0, 1 or 2 units). They find that the optimal tariff is concave.

7. The welfare effects of bundling

The first observation to make is that it is (almost) always possible to use bundling to implement Pareto improvements over the more familiar non–bundling tariffs. For instance, Result 3 shows us that, if the function φ is decreasing the firm can increase profits by reducing its charge for the joint purchase of the goods by some small amount ϵ. Alternatively, if φ is increasing the firm can increase profits by decreasing its charge for the single good (keeping the joint charge fixed). In either case, charges for all bundles are lower and so both the firm and all consumers are better off. These Pareto improvements do not result from any demand–side or cost–side complementarities or substitutabilities, but rather from the enhanced ability to sort consumers which is afforded by bundling.

Of course, potential Pareto improvements do not imply that at the profit–maximizing tariff all consumers will be better off. For instance, Example 1 is a case where the introduction of bundling results in a decrease in the charge for the joint purchase and an increase in the charge for a single good. Therefore, all those consumers who continue to buy just one good will certainly be worse off because of the introduction of bundling. The question it is natural to ask is, how does *aggregate* consumer surplus and/or welfare change with the introduction of bundling? We can usefully think of there being two opposing effects at work. On the one hand, the above–mentioned superior targeting of consumers tends to reduce the welfare loss due to the asymmetric information. But there will also be an unequal marginal utilities effect (or 'allocative inefficiency), since the firm's output is sub–optimally distributed across consumers. Indeed, the socially optimal method of distributing some given level of firm outputs would always be to set some non–bundling tariff so that no gains from trade amongst consumers remain.

As is usually the case when analyzing price discrimination, the combination of these two effects will be ambiguous for welfare. The corollary to Result 1 shows that introducing bundling certainly can be bad for welfare as in that case it results in an increase of all charges. On the other hand, if production is costless and if consumers have perfectly negatively correlated types so that each consumer's set of types always summed to a constant, then it is the case that bundling will increase total surplus (in fact it will attain the first–best). Thus we have established that the welfare effect of bundling is indeed ambiguous.

When, then, should bundling be allowed? Following Varian (1985) we can obtain bounds on the welfare change by means of convexity. We go back to the more general case of Section 3 where there are n goods and arbitrary distribution of types. Let S(T) be total consumer surplus when the firm offers the bundling tariff $T(\cdot)$, i.e. S is the expectation of individual consumer surplus $s(\boldsymbol{\alpha})$ as given in Section 3. Then S is convex in T and differentiable, with partial derivatives obtained by taking expectations of equation (2):

$$\frac{\partial}{\partial T(B)} S(T) = -D(B, T)\,. \tag{9}$$

Since S is convex and convex functions always lie above their tangents, total consumer surplus given two tariffs T and $\tilde{T}$ satisfies the following inequality:

$$S(\tilde{T}) \geq S(T) + \sum_B D(B, T)[T(B) - \tilde{T}(B)]\,. \tag{10}$$

Measuring total welfare with tariff T, denoted W(T), by consumer surplus plus profit, (10) implies

$$W(\tilde{T}) \geq W(T) + \sum_B [\tilde{T}(B) - C(B)][D(B, \tilde{T}) - D(B, T)]. \tag{11}$$

As usual, matters are clearest in the symmetric case. Let $\tilde{T}$ be the uniform non–bundling tariff where the price of each good is p^*, and let the cost of making each good be c. Equation (11) implies that

$$W(T) \leq W(\tilde{T}) + (p^* - c)\sum_B [D(B, T) - D(B, \tilde{T})] \times [\text{number of goods in bundle } B],$$

so that a move from a uniform non–bundling tariff to some other bundling tariff can only increase welfare if it results in an increase in total output, measured as the total number of units of any good demanded.

This result is familiar from standard analyses of third–degree price discrimination — see Schmalensee (1981) and Varian (1985) — and from the single–product nonlinear pricing model — see for instance Chapter 2 above. However, it begs the question of when exactly output increases. Varian considers the special case of independent linear demands and he shows that, provided that no new markets are opened up, output remains constant and hence welfare must necessarily fall in the move from a uniform to a discriminatory set of prices. In the bundling setting, demands for bundles are never independent (Indeed, Lemma 1 shows that bundles are substitutes for each other). Varian's special case cannot, therefore, be modified so as to be suitable for the bundling case, and, so far there are no satisfactory sufficient conditions which determine whether the move from non–bundling to bundling is good or bad for welfare.

8. Bundling by a regulated firm[14]

Many of the firms most likely to practice bundling also enjoy substantial market power and the prices they may charge are therefore regulated. Already existing examples of bundling by utilities in the UK include BT offering telephone directories 'free of charge' (indeed, they previously offered directory enquiries at zero marginal charge as well) and water companies offering water supply for a charge that is not volume–related. For firms which are regulated it is natural to find desirable ways in which to promote additional bundling within the overall regulatory environment. Ideally, such regulatory schemes should not depend upon cost conditions for two reasons. First, the regulatory body is unlikely to have reliable information about costs, especially if this information comes from the regulated firm, and second, regulation which depends on costs is liable to undesirable incentive properties as regards cost reduction. Because of these and other reasons, in recent years regulation in the UK has tended to favour 'price–based' rather than 'cost–based' mechanisms. In this section we present three possible price–based regimes

---

14 The following section is an extension to the bundling setting of the analysis of third–degree price discrimination by a regulated monopoly in Armstrong and Vickers (1991).

which would allow the regulated firm to engage in bundling. We suppose that the firm is initially required to offer the non–bundling tariff denoted by $\bar{T}$ given by the prices $\{\bar{p}_i\}$, where $\bar{p}_i$ is the regulated price for service i.

Method 1. There is one way to regulate bundling which is Pareto improving: allow the firm to bundle, but require it to continue to offer the old non–bundling tariff $\bar{T}$ as an option to consumers (this is analogous to the regulation proposed in Section 4.2 of Chapter 2). More precisely, with this scheme the regulator should allow the firm to offer any bundling tariff T so long as it satisfies:

$$\text{(C1)} \qquad T(B) \leq \bar{T}(B) \equiv \sum_{j \in B} \bar{p}_j \quad \text{for each bundle } B.$$

The result of this regulation cannot leave any consumer worse off, and we have seen in Result 2 that in general, the firm will be able to find some tariff which will leave it and some consumers strictly better off.

With the above method the firm faced a separate cap on each of its bundles and was not permitted to offset high charges for some bundles against low charges on others. In contrast, the next two methods involve capping the firm's *average* charges for its bundles.

Method 2. In this case the firm is permitted to offer a tariff T if it satisfies:

$$\text{(C2)} \qquad \sum_B D(B, \bar{T})T(B) \leq \sum_B D(B, \bar{T})\bar{T}(B) \,.$$

Equation (10) implies that any tariff satisfying constraint (C2) will result in a higher total consumer surplus than the the non–bundling tariff $\bar{T}$. Clearly, the firm will also be better off since it continues to have the option of offering the old tariff $\bar{T}$. Therefore, welfare unambiguously must rise with the introduction of bundling under this scheme. However, the result will not be a Pareto improvement since the price of certain bundles will rise (the possibility which was forbidden in the previous method).

Method 3. The constraint (C2) weighted the tariff charges by the demands under the old non–bundling tariff. An alternative is to use the new weights, so that T must satisfy:

$$\text{(C3)} \qquad \sum_B D(B, T)T(B) \leq \sum_B D(B, T)\bar{T}(B)\,.$$[15]

(This is analogous to the average revenue regulation described in Section 4.1 of Chapter 2.) If the firm chooses to make constraint (C3) bind, then (10) implies that consumers in the aggregate are now *worse off* as a result of this change (although some particular consumers will prefer the new regime), while, as before, the firm is better off. The overall welfare change will certainly be ambiguous.

We cannot deduce, however, that Method 3 is worse for welfare than the other methods since we have have not compared the resulting total welfare in each of the three cases. Indeed, the example in Section 5 of Chapter 2 suggests that it would be possible to find cases where welfare under the third kind of regulation is superior to the Pareto improving first method. Moreover, since regulation in practice takes place over time and not just during the transition from non–bundling to bundling, the third method has informational advantages over Method 2 (but not Method 1) in so far as demands at the hypothetical tariff $\bar{T}$ need not be known in the future to implement the constraint.

9. Large numbers of goods

There are some situations where it may not be unreasonable to take the number of goods to be large and for a given consumer's preferences over these goods to be independently drawn from a common distribution. For instance, a supermarket will stock many items and it might not be too unrealistic to suppose that the information that a

[15] In Method 2, the constraint is equivalent to requiring that $(\Sigma\, \bar{D}T)/(\Sigma\, \bar{D}\bar{T}) \leq 1$, i.e. that according to a Laspeyres price index, average prices do not increase. Method 3, on the other hand, requires that a Paasche price index does not increase in the move to bundling.

particular shopper likes one item tells us little about the value they place upon others. Similarly, there are very many possible telephone routes, and whether a caller in Oxford makes many calls to Edinburgh may be unrelated to how many calls he makes to Liverpool, say. More importantly, it would be good to gain intuition about how the number of goods affects the optimal price schedule. We shall look first at this simple case where consumers' valuations are independently distributed; later in this section we will generalize this to encompass income differences which will introduce a degree of correlation.

To model these situations we shall return to an earlier assumption of symmetry but increase the number of goods to n. By symmetry, the firm will then set a charge t_m for a bundle of any m goods, where $m = 1, 2, .., n$. In the following we are going to let n be large, and so we will normalize utility and costs in order to keep the problem bounded. Therefore, a consumer of type $\boldsymbol{\alpha}$ has a utility function of

$$\frac{1}{n}\sum_{j \in B} \alpha_j - t$$

if she buys the bundle B for payment t. The cost per unit for the firm is now c/n. Suppose that for a given consumer each type parameter α_i is an independent draw from the distribution function, $G(\cdot)$. For now we will take G to be constant across consumers. We now find a price schedule which lies arbitrarily close to the optimal price schedule for sufficiently large n.

RESULT 6. *Fix* $\epsilon > 0$. *For each* $n = 1, 2, \ldots$ *there is an* n*-good industry where the firm has a cost of* c/n *for the production of a unit of any good and the utility of a type* $\boldsymbol{\alpha}$ *consumer from consuming bundle* B *is* $\frac{1}{n}\sum_{j\in B} \alpha_j - t$, *where* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. *Each* α_i *is independently drawn from some distribution function* $G(\alpha)$. *Then for all sufficiently large* n *a tariff of the form*

$$(13) \qquad t_m^* = A + cm/n$$

(where t_m^* *is the charge for a bundle of any* m *goods) obtains profit for the firm to within* ϵ *of the optimal tariff for the* n*-good firm. In addition, such a tariff obtains profit to within* ϵ *of the first-best profit level (i.e. the profit level when the firm can observe consumers' types).*

REMARK. This result is special case of Result 7 below and so, strictly, the following proof is unnecessary. However, because it is so much simpler than the proof of that result I believe intuition is better assisted by its inclusion.

PROOF. The first–best profit level provides an upper bound on profits which may feasibly be obtained when the firm cannot observe types. For any given consumer of type $\boldsymbol{\alpha}$ this maximum profit is obtained by charging a marginal price equal to marginal cost c/n and extracting her entire surplus by means of a fixed charge. For a type $\boldsymbol{\alpha}$ consumer this surplus is:

$$(14) \qquad \bar{s}_n(\boldsymbol{\alpha}) \qquad = \tfrac{1}{n}\Sigma\,[\alpha_i - c]_+$$

where $[x]_+$ is the positive part of x (so that $[x]_+ = x$ if $x \geq 0$, otherwise $[x]_+ = 0$). The random variable $\bar{s}_n(\boldsymbol{\alpha})$ is just the sample mean of n independent draws from the distribution $[\alpha - c]_+$, and so satisfies:

$$(15)\qquad \begin{aligned} E[\bar{s}_n] &= E[\alpha - c]_+ \\ \mathrm{Var}[\bar{s}_n] &= \tfrac{1}{n}\mathrm{Var}[\alpha - c]_+ . \end{aligned}$$

(We assume that these moments exist.) The total first–best profit is the sum of $\bar{s}_n(\boldsymbol{\alpha})$ over all $\boldsymbol{\alpha}$:

$$(16)\qquad \pi_{fb} = E[\bar{s}_n] = E[\alpha - c]_+$$

which is independent of n.

Suppose that the firm offers a two–part tariff of the form (13). Faced with this, a consumer of type $\boldsymbol{\alpha}$ will participate in the market if and only if $A \leq \bar{s}_n(\boldsymbol{\alpha})$. Using equations (15) and (16), the 'Weak Law of Large Numbers' implies that $\bar{s}_n \longrightarrow \pi_{fb}$ in probability as n tends to infinity. But this means that for sufficiently large n

$$(17)\qquad \mathrm{prob}\{\, |\bar{s}_n - \pi_{fb}| > \tfrac{\epsilon}{2} \,\} < \frac{\epsilon}{2\pi_{fb}} .$$

If the firm charges a tariff of the form (13) with $A = \pi_{fb} - \frac{\epsilon}{2}$ then, by construction, its profit will be exactly $\pi_{fb} - \frac{\epsilon}{2}$ times the number of participants in the market, which from (17) comes to at least $(\pi_{max} - \frac{\epsilon}{2})(1 - \frac{\epsilon}{2\pi_{max}})$ for sufficiently large n. Since this in turn is greater than $\pi_{fb} - \epsilon$ we have found a tariff which attains profits to within ϵ of the first–best tariff, and hence to within ϵ of the optimal tariff for the case when the firm is unable to observe consumers' types. □

Therefore, for large n we may approximate the optimal tariff by a two–part tariff with marginal price equal to marginal cost $(= c/n)$ and fixed charge equal to a little less than expected consumer surplus. In particular, the firm will engage in subadditive bundling. The tariff will be (almost) efficient and will result in the firm appropriating nearly all of the consumer surplus. In this very special case, then, the firm will choose a bundling tariff which asymptotically is the first–best.[16] This contrasts with the non–bundling optimum, where for a given n the firm chooses the single price p_n to maximize $(p - c/n)(1 - G(np))$, and so sets $np_n = \text{constant} = p^*$, say. Clearly, $p_n > c/n$. This means that the optimal non–bundling tariff is of the form $t_m = p^*m/n$. For this example, then, a move from non–bundling to bundling is just like the move from a linear price schedule (with price above marginal cost), to an efficient two–part tariff with marginal price equal to marginal cost. Moreover, this two–part tariff has the advantage that very few consumers will be driven from the market as a result of the fixed charge.

We now modify the above model to include a degree of correlation in consumers' willingness–to–pay for the various goods. In the above case, all consumers' willingness–to–pay parameters were drawn from the same distribution. Suppose now that these parameters for a given consumer continue to be independent, but the distributions from which they are drawn now vary from consumer to consumer. For instance, differences in income across consumers could mean that a poorer person would most likely be prepared to pay less than a better–off person for all items in a supermarket, whilst keeping their particular valuations independent across goods. Rich people may, *ex ante*, be prepared to pay more for all telephone routes. To capture this idea we introduce a one–dimensional parameter θ which is distributed in the consumer population with distribution function $H(\cdot)$. As before consumers also have taste parameters α_i independently from the distribution function $G(\cdot)$. We shall also assume that these parameters are distributed independent of θ. If θ measures the reciprocal of an agent's marginal utility of income, this means that a consumer of prior type θ with type α_i for good i gains utility of $\theta\alpha_i/n$ from consuming good i when measured in money terms. More precisely, a consumer with income parameter θ and taste parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, .. , \alpha_n)$ obtains a utility from consuming the bundle B of

$$(18) \qquad \frac{\theta}{n} \sum_{j \in B} \alpha_j - t .$$

[16] For other examples of adverse selection models where the principal is able to obtain the first–best outcome see Chapter 1 above.

This means that the reservation prices for goods i and j, $\theta\alpha_i/n$ and $\theta\alpha_j/n$, are now positively correlated.

In order to find an approximately optimal tariff for the firm for this case, we solve a slightly different problem first. Suppose that there are, not just a large number, but a continuum of goods. In this problem each consumer has a distribution of taste parameters actually equal to the theoretical distribution $G(\alpha)$, so that the density of taste parameters equal to α is $g(\alpha)$. Assume that the firm incurs a cost of cq in supplying a fraction q of all possible goods to any consumer and so by symmetry the firm is going to offer a tariff $T(q)$ where q is the fraction of goods in the bundle. What is a type θ consumer's utility from consuming a fraction q of all possible goods, where $0 \leq q \leq 1$? Clearly, she will choose to buy the fraction q of her most valued goods, and so obtain a utility of:

$$\theta u(q) - T(q), \text{ where } u(q) \equiv \int_{\alpha(q)}^{\bar{\alpha}} \alpha g(\alpha)\, d\alpha \tag{19}$$

and $\alpha(q)$ is given implicitly by $1 - G(\alpha(q)) \equiv q$. The utility function $u(\cdot)$ will necessarily be strictly concave and increasing in q. The firm chooses the nonlinear tariff $T^*(q)$ in order to maximize its profits given that consumers have utility given in (19). We solved this problem using the first–order approach in Section 4.1 of Chapter 1, where it was shown that, provided

$$\theta - \frac{1 - H(\theta)}{h(\theta)} \text{ is non–decreasing} \tag{20}$$

the optimal response function $q^*(\theta)$ satisfies

$$q^*(\theta) \underset{1 \geq q \geq 0}{\text{maximizes:}} \; u(q)[\theta - \frac{1 - H(\theta)}{h(\theta)}] - cq\,. \tag{21}$$

The optimal tariff T^* which gives rise to this response function is given by the marginal price schedule:

$$(22) \qquad p^*(q^*(\theta)) = c/[1 - \frac{1 - H(\theta)}{\theta h(\theta)}] .$$

As we might expect, the optimal tariff T^* in the above problem is a good approximation to the optimal bundling tariff when there are a large but finite number of goods. The proof of this is technical, and we leave it in the Appendix to this chapter.

RESULT 7. *Let* $\epsilon > 0$ *be given. For each* $n = 1, 2 \ ...$ *there is an* n*-good industry where the firm has a cost of* c/n *for the production of a unit of any good and the utility of a type* $(\theta, \boldsymbol{\alpha})$ *consumer from consuming bundle* B *is* $\frac{\theta}{n} \sum_{j \in B} \alpha_j - t$, *where* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, .. , \alpha_n)$. *All of the parameters* $\{ \theta, \alpha_i \}$ *are independent,* α_i *being drawn from the smooth distribution function* $G(\alpha)$ *which has compact support and* θ *being drawn from the distribution function* $H(\theta)$ *which satisfies (20). Let* $u(q)$ *be given by (19) and* $T^*(q)$ *be the tariff given by (21) and (22). Then for all sufficiently large* n *a tariff of the form*

$$(23) \qquad t_m^* = T^*(m/n)$$

obtains profit for the firm to within ϵ *of the optimal tariff for the* n*-good firm.*

PROOF. See Appendix.

Just as was the case in Result 1 of Chapter 1, we have the following corollary:

COROLLARY 4. *The (approximately) optimal tariff* t_m^* *will be subadditive if*

$$(24) \qquad \frac{1 - H(\theta)}{\theta h(\theta)} \text{ is decreasing.}$$

PROOF. It is sufficient (although not necessary) for T^* in (22) to be subadditive in q that it be concave, and T^* is concave if (24) holds. The tariff $t_m{}^*$ in (23) is subadditive whenever T^* is. □

When there were no differences in income (i.e. when θ was constant), the move to bundling was unambiguously good for welfare (although bad for consumers). In effect the firm is able to diversify away all risk. When income differences are introduced, in order to judge the welfare effect of bundling we must compare welfare under the nonlinear tariff T^* with the profit–maximizing linear tariff $T(q) = p^*q$, say. From Chapter 2 this will certainly be ambiguous. Here, consumers retain some private information, θ, from which the firm is unable to diversify and so retain some of the gains from trade.

9. Conclusions

In the model considered above there were n underlying goods, the demands for which were independent. The move from non–bundling to bundling involved changing the focus of analysis from these n goods to the 2^n bundles, 'goods' which were always mutual substitutes (Lemma 1). For a wide variety of distributions of consumer tastes we know two things about the profit–maximizing tariff: the firm will choose not to offer all goods to all consumers (Result 1), and in the symmetric case of two goods it will choose to offer the bundle of two goods at a discount compared to the sum of the prices for the two goods separately. The vital ingredient needed for this last prediction was that consumers' tastes were not 'too' positively correlated in any range so that the probability that a consumer buys one good at price p given she buys the other at that price is non–increasing (Result 4). To this extent we have confirmed the intuition of some of the earlier literature that negative correlation encourages subadditive bundling. Result 4 had the corollary that when consumers wish to buy at most two units of a single good (and they exhibit diminishing marginal utility) the firm will offer the second unit at a discount.

It is clear that the welfare effects of bundling are ambiguous. A move from a separable tariff to a bundling tariff could strictly increase all prices (see Result 1), which will certainly decrease total surplus. But there are circumstances in which bundling will result in the first–best welfare level, for instance if consumer types were perfectly negatively correlated and production was costless. In Section 7 an inequality was derived using the fact that consumer surplus was a convex function of the bundling tariff. Again, the symmetric case was clearest, where the inequality implied that the introduction of bundling could increase welfare only if total output increases. However, no sufficient conditions have yet been found which guarantee any particular direction of output change, and more work needs to be done in this area. The situation is less ambiguous if the monopoly is regulated as to its price level. In Section 8 three kinds of price cap regulation were examined, two of which resulted in bundling which simultaneously benefits the firm and consumers.

In the last section of this chapter we saw that increasing the number of goods could well work to the advantage of the firm. When consumers did not differ in their income the firm could appropriate almost all consumer surplus, an outcome which is more efficient than the non–bundling equilibrium. The optimal tariff was approximated by a two–part tariff with marginal price equal to marginal cost, a tariff which is subadditive. When consumers differed in their income they were able to retain some of the gains from trade. In that case the firm offered a tariff derived from the familiar single–product nonlinear pricing problem and, provided that a hazard rate condition on incomes held, a subadditive tariff was again (approximately) optimal.

In all of this analysis the role of symmetry has been large. In the proof of Result 4, it enabled us to deduce the firm's profit given only that we know the number of consumers who buy both goods and the number who buy neither; the number buying only a given single good then being half the remaining consumers. In the section on large numbers of goods, symmetry was necessary in order to be able to invoke 'large number' arguments. I doubt whether this symmetry assumption can easily be dropped, and this is unfortunate because there is no reason to think of symmetry as being in any sense natural for many multiproduct firms.

The importance of the degree of correlation in tastes will have been apparent throughout. In the two–good case with perfect positive correlation there is no motive to bundle, whereas if tastes were perfectly negatively correlated and production was costless the firm could obtain the first–best profit level by pure subadditive bundling. In the case of many goods, without income differences consumer tastes were independent and the firm obtained first–best profits in the limit, but when income was not constant tastes become positively correlated and the firm was forced to share some surplus with consumers. An informal discussion in Section 5, using the notion of 'increasing correlation' in tastes, proposed the rule of thumb that increasing the degree of positive correlation (all else being equal) increases the profit of the superadditive bundling firm, whereas increasing the degree of negative correlation *tended* to improve the position of the subadditive bundling firm. Increasing correlation increases the degree of uncertainty faced by the firm which cannot be overcome by means of subadditive bundling, and so worsens the position of the subadditive bundling firm. In the language of finance, with independent demands — or better still, with negatively correlated demands — by subadditively bundling its outputs the firm is able to diversify away from most of the risk associated with its uncertain knowledge of consumer tastes. There is, however, no rigorous connection between increasing the degree of negative correlation and either increasing the firm's profits or increasing the incentives towards subadditive bundling (as shown by Example 3).

In the model, the motive for bundling comes entirely from the improved sorting of consumers according to their ability to pay for the various bundles of goods. For instance, there are no economies of scope either in consumption or the manufacture of two or more goods. If these effects were present it seems likely that the incentive to bundle would be all the more. A good example of a firm which, because of cost reasons, would prefer consumers to consume two goods is any firm with 'peak load' problems such as electricity or telecommunications suppliers. The French electricity supply industry, for instance, offers a complex tariff to its industrial users where it is cheaper to do two tasks involving power consecutively rather than concurrently. Joint consumption of peak and off–peak services is thereby encouraged. This is a clear use of bundling to spread the load and reduce average costs.

Another instance of bundling in an entirely different area occurs when a firm issues a stock of equity. Typically, a firm's share offers its holder both rights to future profits (via dividends and capital gains) and rights of control (via voting). An interesting question to ask, then, is why firms choose to bundle these two 'goods' rather than sell them separately (although this does sometimes happen when a firm offers non–voting shares). Is it again a desire on the firm's part for buyers to have both goods? It may be that the present value of the firm is increased when those with the means to change firm policy will also have to face the outcome of these changes in terms of their receipt of future profits. Harris and Raviv (1989) have investigated precisely this question of security design. For their model, they find that a single voting security is indeed optimal, and the present value of the firm is decreased by 'unbundling' control from rights to profits.

Bundling is important also in some models of moral hazard. For instance, in most penal systems first offenders are treated more leniently than recidivists, i.e. the 'charge' for two crimes is more than twice that for a single crime. Perhaps the rationale for this is that conviction is only a noisy signal of a person's innate 'criminality' — a single offense may be the result of being 'led astray', or indeed of being convicted falsely — and that increased punishment for repeat offenders is then a means by which to punish the true 'criminals' more severely. Polinsky and Rubinfeld (1991) construct a formal model to try to answer this question, and show that at the optimum the penalty could be either increasing or decreasing for repeat offenders.

However, the purpose of this paper has not simply been to understand the phenomenon of bundling. Economists have for some time had difficulty in solving 'multidimensional' problems. In more complex models than this (for instance, see the next chapter) a more sophisticated kind of mathematics is required than is familiar to a mainstream economist — any gain in intuition which has been the result of this chapter and other work will be valuable in making progress in solving such models.

Appendix.

Proof of Result 1. *By contradiction:* Let $T^*(\cdot)$ be the optimal tariff and suppose that with this tariff all consumers buy the bundle of all goods N. Then in particular $T^*(N) \leq \sum_1^n a_i$. If T^* is optimal then we must have equality here, so that $T^*(N) = \sum_1^n a_i$. Let

$$A(\epsilon) = \{ \alpha \mid \alpha \in A \text{ and } \sum_1^n \alpha_i \leq T^*(N) + \epsilon \}.$$

Then each $A(\epsilon)$ is compact and convex, and $A(0) = \{\mathbf{a}\}$ by assumption. In addition, $A(\cdot)$ is a continuous family of sets, so that $A(\epsilon) \longrightarrow A(0) = \{\mathbf{a}\}$ as $\epsilon \longrightarrow 0$.[17] Let $S(A(\epsilon))$ be the surface area of the set $A(\epsilon)$. Let $g(t)$ and $G(t)$ be respectively the density and distribution function of the scalar variable $\sum_1^n \alpha_i$ induced by the density function $f(\alpha)$ on consumer types. If B is a bound on f we must have

$$\text{(A1)} \qquad g(T^*(N) + \epsilon) \leq B \times S(A(\epsilon)) .$$

(This is because $g(T^*(N) + \epsilon)$ is the integral of f over the set

$$\{ \alpha \mid \alpha \in A \text{ and } \sum_1^n \alpha_i = T^*(N) + \epsilon \} ,$$

---

[17] Compact convex subsets of a bounded set in $\mathbb{R}^n$ are very well–behaved and the space of such sets can be given a notion of distance so that the space is a compact metric space — see Eggleston (1958). In particular it makes sense to speak of a continuous family of sets. The facts about this space I use in this proof are that if C is a compact convex set then the volume $V(C)$ and surface area $S(C)$ of such a set are both well–defined and both continuous in C under this metric. We define the surface area of the set C as a kind of differential:

$$S(C) = \lim_{\epsilon \to 0} [V(C + \epsilon B) - V(C)]/\epsilon$$

where B is the solid n–dimensional unit ball. In particular, the surface area of a single point is zero unless $n = 1$. Moreover, $S(C)$ is decreasing in C so that

$$C \subset C' \Rightarrow S(C) \leq S(C')$$

(this is not always true for non–convex sets).

a set which is a subset of the boundary of $A(\epsilon)$, and therefore $g(T^*(N) + \epsilon)$ is less than the integral of f over the boundary of $A(\epsilon)$ which in turn is less than $B \times S(A(\epsilon))$.) Therefore

$$G(T^*(N) + \epsilon) = \int_0^{\epsilon} g(T^*(N) + \tilde{\epsilon})\, d\tilde{\epsilon}$$

$$\leq \int_0^{\epsilon} B \times S(A(\tilde{\epsilon}))\, d\tilde{\epsilon} \quad \text{from (A1)}$$

$$\text{(A2)} \qquad \leq \epsilon B \times S(A(\epsilon))$$

(where in the final inequality we used the fact that $\tilde{\epsilon} < \epsilon$ implies that $A(\tilde{\epsilon}) \subset A(\epsilon)$, which in turn implies $S(A(\tilde{\epsilon})) \leq S(A(\epsilon))$). Now consider replacing the tariff T^* with the pure bundling tariff where the charge for all goods is $T^*(N) + \epsilon$ (and consumers must buy either all goods or none at all). The number of consumers who leave the market as a result of this change is $G(T^*(N) + \epsilon)$ and all consumers who remain now pay an extra ϵ. The total change in profits is

$$\Delta\pi = \epsilon(1 - G(T^*(N) + \epsilon)) - (T^*(N) - C(N))G(T^*(N) + \epsilon)$$

$$\text{(A3)} \qquad \geq \epsilon(1 - G(T^*(N) + \epsilon)) - (T^*(N) - C(N))\epsilon B \times S(A(\epsilon))$$

(since $(T^*(N) - C(N)) \geq 0$ by Corollary 1). But since $S(\cdot)$ is a continuous function the fact that $A(\epsilon) \longrightarrow A(0) = \{\mathbf{a}\}$ as $\epsilon \longrightarrow 0$ implies that $S(A(\epsilon)) \longrightarrow S(\{\mathbf{a}\})$ as $\epsilon \longrightarrow 0$. Therefore, unless $n = 1$

$$\text{(A4)} \qquad S(A(\epsilon)) \longrightarrow 0 \text{ as } \epsilon \longrightarrow 0 .$$

Finally, provided that $n > 1$ the combination of (A3) and (A4) implies that $\Delta\pi > 0$ for all sufficiently small $\epsilon > 0$. As this is a contradiction, the proof is complete. □

Proof of Result 7. For each n, given α and t define $\tilde{G}_n(t)$ to be the sample distribution function for the type α consumers, i.e. $\tilde{G}_n(t)$ is the proportion of the sample α which take values less or equal to t. Therefore, for each t $\tilde{G}_n(t)$ is a random variable. It is intuitive that for large sample sizes this observed distribution function $\tilde{G}_n(t)$ will approximate the underlying distribution function $G(t)$. In fact, the following lemma shows that convergence occurs in the following 'weak' sense:

LEMMA 2. *$\tilde{G}_n \longrightarrow G$ uniformly in probability as $n \longrightarrow \infty$, i.e. given $\delta > 0$ then*

$$prob\{\ |\tilde{G}_n(t) - G(t)| \geq \delta \text{ for some } t\ \} \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

PROOF. For each t the random variable $\tilde{G}_n(t)$ is the fraction of successes in an n–shot Binomial distribution where the probability of a success is $G(t)$. Therefore, $\tilde{G}_n(t)$ has mean $G(t)$ and variance $\frac{1}{n}G(t)(1 - G(t))$. For a given $\delta_1 > 0$ Chebychev's Inequality implies

$$\text{prob}\{\ |\tilde{G}_n(t) - G(t)| \geq \delta_1\ \} < \frac{G(t)(1 - G(t))}{n\delta_1{}^2}$$

$$\text{(A5)} \qquad \leq \frac{1}{n\delta_1{}^2} \text{ for all } t\ .$$

Let α have support $[\underline{\alpha}, \bar{\alpha}]$ and divide up this support into K equal sub–intervals, each of length $(\underline{\alpha} - \bar{\alpha})/K$. The k^{th} endpoint of these sub–intervals is $t_k = \underline{\alpha} + k(\underline{\alpha} - \bar{\alpha})/K$, for $k = 0, 1, \ldots K$. From (A5),

$$\text{(A6)} \qquad \text{prob}\{\ |\tilde{G}_n(t_k) - G(t_k)| > \delta_1 \text{ for any } t_k\ \} < \frac{K - 1}{n\delta_1{}^2}$$

(there are $K + 1$ points, but we know that the two functions are precisely equal at $t = \underline{\alpha}$ and $t = \bar{\alpha}$). Therefore, $\tilde{G}_n$ is close to G at each of the points $\{t_k\}$ with high probability for large n. We must now show that the two functions are also close between the points $\{t_k\}$. Since g is continuous, let m be the maximum value it takes on $[\underline{\alpha}, \bar{\alpha}]$, which means that the slope of G is everywhere less than m. Both the functions $\tilde{G}_n$ and G are (weakly) increasing, and so the 'worst case' deviation which could happen in the sub–interval $[\alpha_{k-1}, \alpha_k]$ is $2\delta_1 + m(\underline{\alpha} - \bar{\alpha})/K$.[18] Writing $\delta = 2\delta_1 + m(\underline{\alpha} - \bar{\alpha})/K$, (A6) implies that

$$\text{(A7)} \qquad \text{prob}\{\ |\tilde{G}_n(t) - G(t)| \geq \delta \text{ for some } t\ \} < 4(K-1)/(n[\delta - m(\underline{\alpha} - \bar{\alpha})/K]^2),$$

provided that $\delta > m(\underline{\alpha} - \bar{\alpha})/K$. Therefore, for a given δ pick a fixed K large enough so that $\delta > m(\underline{\alpha} - \bar{\alpha})/K$, and then let n get large in (A7). □

We are going to suppose that the firm offers a continuous tariff schedule $T(q)$ where $0 \leq q \leq 1$ is a real number. In the n–good case this means that the charge for $m \leq n$ goods is $T(m/n)$. Define $\tilde{\alpha}_n(q)$ to be the type α consumer's $[nq]^{th}$ best draw in her sample (where $[x]$ is the integer part of x). Then a type (θ, α) consumer's utility from consuming her most valued fraction q of the possible n goods is $\theta\tilde{u}_n(q)$, where

$$\text{(A8)} \qquad \tilde{u}_n(q) \equiv \tfrac{1}{n}\Sigma\{\ \alpha_i \mid \alpha_i \geq \tilde{\alpha}_n(q)\ \}.$$

(Note that the function $\tilde{u}_n(q)$ is neither continuous nor concave as it is constant on the intervals $(\frac{m}{n}, \frac{m+1}{n})$.) The above lemma suggests that $\tilde{u}_n(q)$ should approximate $u(q)$ for large n, where u is given in (19) and this is given a sketch proof in the following lemma.

LEMMA 3. $\tilde{u}_n(q) \longrightarrow u(q)$ *uniformly in probability as* $n \longrightarrow \infty$.

---

[18] That is to say, if the two functions $\tilde{G}_n$ and G agree to within δ_1 at each end of a sub–interval of length $(\underline{\alpha} - \bar{\alpha})/K$, each is increasing and one always has a slope less than m, then they must agree to within $2\delta_1 + m(\underline{\alpha} - \bar{\alpha})/K$ throughout the sub–interval.

PROOF. We can write $\tilde{u}_n(q)$ in (A8) using a Riemann–Stieltjes integral:

$$\text{(A9)} \qquad \tilde{u}_n(q) = \int_{\tilde{\alpha}_n(q)}^{\bar{\alpha}} \alpha \, d\tilde{G}_n(\alpha).$$

Now, using similar arguments to those used in Lemma 1, it is possible to show that $\tilde{\alpha}_n(q) \to \alpha(q)$ uniformly in probability. Also, it is a fact that $\tilde{G}_n$ uniformly close to G implies that integrating with respect to $\tilde{G}_n$ gives a close answer to integrating with respect to G — see section 6.2 of Burkill and Birkill (1970) for more details. Therefore, Lemma 2 and the uniform convergence of $\tilde{\alpha}_n(q) \to \alpha(q)$ implies that

$$\tilde{u}_n(q) = \int_{\tilde{\alpha}_n(q)}^{\bar{\alpha}} \alpha \, d\tilde{G}_n(\alpha) \longrightarrow \int_{\alpha(q)}^{\bar{\alpha}} \alpha \, dG(\alpha) = u(q) \quad \text{as } n \longrightarrow \infty$$

where the convergence is uniform in probability. □

Let $\pi_n{}^*$ be the maximum profit for the n–good firm and let π^* be the maximum profit for the continuum–of–goods firm. Then Lemma 3 implies the following:

COROLLARY 4. $\limsup_{n \to \infty} \pi_n{}^* \leq \pi^*$.

PROOF. Let $\epsilon > 0$ be given. From Section 4.1 of Chapter 1 we can write π^* as

$$\text{(A10)} \qquad \int_{\underline{\theta}}^{\bar{\theta}} \max_{q \geq 0} \{ [\theta u(q) - cq]h(\theta) - (1 - H(\theta))u(q) \} \, d\theta .$$

Profit $\pi_n{}^*$ can be no greater than the 'second–best' profit obtainable when the firm is able to observe the taste parameters $\boldsymbol{\alpha}$ (but not the income parameter θ) which we write as $\pi_n{}^{**}$. Fix n and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ and let $\tilde{u}_n(q)$ be as given in (A8). If $\boldsymbol{\alpha}$ is known by the firm then $\tilde{u}_n(q)$ is known, and so optimal expected profit from this consumer is obtained using the same procedure as described in Section 4.1 of Chapter 1 and is given by

$$\int_{\underline{\theta}}^{\overline{\theta}} \max_{q \geq 0} \{ [\theta \tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \} \, d\theta$$

and so that total second–best profits are

$$\pi_n{}^{**} = E\,[\int_{\underline{\theta}}^{\overline{\theta}} \max_{q \geq 0} \{ [\theta \tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \} \, d\theta\,]$$

where E denotes expectation with respect to the parameter α. Since α and θ are independent we may write this as

$$\pi_n{}^{**} = \int_{\underline{\theta}}^{\overline{\theta}} E[\max_{q \geq 0} \{ [\theta \tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \}] \, d\theta \,.$$

If $|\tilde{u}_n(q) - u(q)| \leq \delta$ for all q then

$$\max_{q \geq 0} \{ [\theta \tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \} \leq$$

$$\max_{q \geq 0} \{ [\theta u(q) - cq]h(\theta) - (1 - H(\theta))u(q) \} + \delta[\theta h(\theta) - (1 - H(\theta))] \,.$$

Therefore, letting $\delta = \epsilon / \max_{\theta}[\theta h(\theta) - (1 - H(\theta))]$ we see that if $|\tilde{u}_n(q) - u(q)| \leq \delta$ for all q then

$$\max_{q \geq 0} \{ [\theta\tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \} \leq$$

$$\max_{q \geq 0} \{ [\theta u(q) - cq]h(\theta) - (1 - H(\theta))u(q) \} + \epsilon \quad \text{for all } \theta.$$

But Lemma 3 implies that $\text{prob}\{ \, |\tilde{u}_n(q) - u(q)| \leq \delta \text{ for all } q \} \longrightarrow 0$ as $n \longrightarrow \infty$ and so letting n tend to infinity we see that

$$\limsup_{n \to \infty} E[\max_{q \geq 0} \{ [\theta\tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \}]$$

$$\leq \max_{q \geq 0} \{ [\theta u(q) - cq]h(\theta) - (1 - H(\theta))u(q) \} + \epsilon \quad \text{for all } \theta.$$

Therefore

$$\limsup_{n \to \infty} \pi_n^{**} = \limsup_{n \to \infty} \int_{\underline{\theta}}^{\overline{\theta}} E[\max_{q \geq 0} \{ [\theta\tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \}] \, d\theta$$

$$\leq \int_{\underline{\theta}}^{\overline{\theta}} \limsup E[\max_{q \geq 0} \{ [\theta\tilde{u}_n(q) - cq]h(\theta) - (1 - H(\theta))\tilde{u}_n(q) \}] \, d\theta$$

(by Fatou's Lemma)

$$\leq \int_{\underline{\theta}}^{\overline{\theta}} \max_{q \geq 0} \{ [\theta u(q) - cq]h(\theta) - (1 - H(\theta))u(q) \} \, d\theta + \epsilon(\overline{\theta} - \underline{\theta})$$

$$= \pi^* + \epsilon(\overline{\theta} - \underline{\theta}) .$$

Since ϵ is arbitrary we deduce that

$$\limsup_{n \to \infty} \pi_n^* \leq \limsup_{n \to \infty} \pi_n^{**} \leq \pi^* . \quad \square$$

We next need a lemma of pure analysis:

LEMMA 4. *Let* $z(q, \theta)$ *be a continuous function defined on the domain* $[0, 1] \times [\underline{\theta}, \overline{\theta}]$ *and suppose that* $q(\theta)$ *is the unique maximand of* $z(q, \theta)$ *over* $0 \leq q \leq 1$. *For each* n *define*

$$B_i = \{\, y(q, \theta) \text{ such that } |\, y(q, \theta) - z(q, \theta) \,| < \tfrac{1}{i} \text{ for all } (q, \theta) \in [0, 1] \times [\underline{\theta}, \overline{\theta}] \,\}$$

so that B_i *contains all functions (not necessarily continuous) which uniformly approximate* z *to within* $1/i$. *For each* θ *let* $q_i(\theta)$ *be a maximand (not necessarily unique) of* $y(\cdot, \theta)$ *over* $0 \leq q \leq 1$ *for some* $y \in B_i$. *Then* $q_i(\theta)$ *uniformly converges to* $q(\theta)$ *as* i *tends to infinity.*

PROOF. *By contradiction*: Suppose not, so that there exists $\delta > 0$, a sequence θ_i and a subsequence of $\mathbb{N}$ (which without loss of generality we take to be $\mathbb{N}$) such that

$$|\, q_i(\theta_i) - q(\theta_i) \,| \geq \delta \quad \text{for all } i$$

and $q_i(\theta_i)$ maximizes $y_i(\cdot, \theta_i)$ over $0 \leq q \leq 1$ for some $y_i \in B_i$. Define

$$\epsilon = \min: \{\, z(q(\theta), \theta) - z(q', \theta) \mid \theta \in [\underline{\theta}, \overline{\theta}] \text{ and } |q(\theta) - q'| \geq \delta \,\}. \tag{A11}$$

Then the continuity of z and the compactness of $[\underline{\theta}, \overline{\theta}] \times [0, 1]$ implies that this minimum exists, and the fact that $q(\theta)$ is the unique maximand of $z(\theta, \cdot)$ implies that $\epsilon > 0$. Then

$$y_i(q(\theta_i), \theta_i) - y_i(q_i(\theta_i), \theta_i) = [\, y_i(q(\theta_i), \theta_i) - z(q(\theta_i), \theta_i) \,] + [\, z(q(\theta_i), \theta_i) - z(q_i(\theta_i), \theta_i) \,] + [\, z(q_i(\theta_i), \theta_i) - y_i(q_i(\theta_i), \theta_i) \,]$$

$$\text{(A12)} \qquad \geq [\, y_i(q(\theta_i), \theta_i) - z(q(\theta_i), \theta_i) \,] + \epsilon + [\, z(q_i(\theta_i), \theta_i) - y_i(q_i(\theta_i), \theta_i) \,]$$

from (A11). But the fact that $y_i \in B_i$ implies that for $i > 2/\epsilon$ and for such n (A12) implies that $y_i(q(\theta_i), \theta_i) - y_i(q_i(\theta_i), \theta_i) > 0$, contradicting the assumption that $q_i(\theta_i)$ maximizes $y_i(\cdot, \theta_i)$. □

We are finally in a position to prove the result:

Proof of Result 7: Let $\epsilon > 0$ be given. Let condition (20) hold and let the firm offer the tariff $T^*(q)$ given by (21) and (22) in the text. Then the response function which maximizes $\theta u(q) - T^*(q)$ over $0 \leq q \leq 1$ is given by $q^*(\theta)$ in (21), and in particular $q^*(\theta)$ is unique for each θ. Let $\tilde{u}_n(q)$ be given by (A8) and let $\tilde{q}_n(\theta)$ be the demand function for the type (θ, α) consumer facing the tariff T^*. In the continuum case total profits for the firm is

$$\text{(A13)} \qquad \pi^* = \int_{\underline{\theta}}^{\bar{\theta}} [T^*(q^*(\theta)) - cq^*(\theta)]\, h(\theta)\, d\theta$$

and in the the finite n case total profit is

$$\text{(A14)} \qquad \pi_n = \int_{\underline{\theta}}^{\bar{\theta}} E[T^*(\tilde{q}_n(\theta)) - c\tilde{q}_n(\theta)]\, h(\theta)\, d\theta$$

where E takes expectations with respect to the parameter α (since θ and α are independent variables, this expectation does not depend upon θ). Regarding π^* in (A13) as a functional of $q(\cdot)$ the continuity of T implies that π^* is a continuous functional of $q(\cdot)$ using the 'sup' norm, i.e. there exists $\delta > 0$ such that

$$| \tilde{q}_n(\theta) - q^*(\theta) | < \delta \quad \text{for all } \theta \Longrightarrow$$

$$\left| \int_{\underline{\theta}}^{\bar{\theta}} [T^*(\tilde{q}_n(\theta)) - c\tilde{q}_n(\theta)]h(\theta)d\theta - \int_{\underline{\theta}}^{\bar{\theta}} [T^*(q^*(\theta)) - cq^*(\theta)]h(\theta)d\theta \right| < \tfrac{1}{2}\epsilon .$$

From Lemma 4 we know that there exists $\delta_1 > 0$ such that

$$| \tilde{u}_n(q) - u(q) | < \delta_1 \text{ for all } q \Longrightarrow | \tilde{q}_n(\theta) - q^*(\theta) | < \delta \text{ for all } \theta .$$

Let M be some upper bound on the profit the firm can make on any type of consumer (e.g. the first–best profit from $\bar{\theta}$). Lemma 3 implies that there exists some n_0 such that

$$\text{prob}\{ \, | \tilde{u}_n(q) - u(q) | < \delta_1 \text{ for all } q \,\} > 1 - \epsilon/2M \quad \text{for all } n > n_0.$$

By construction, if π_n is as given in (A14) then

(A15) $\qquad |\pi_n - \pi^*| < \epsilon \quad \text{for all } n > n_0.$

Therefore, for sufficiently large n the profit obtained by offering the tariff T^* in the n–good industry approximate π^* to within ϵ. However, Corollary 4 implies that for all sufficiently large n the maximum profit from offering any tariff in the n–good industry is less than $\pi^* + \epsilon$. Therefore, for all sufficiently large n if the n–good firm offers the tariff T^* it will achieve profits to within 2ϵ of its optimal profits. □

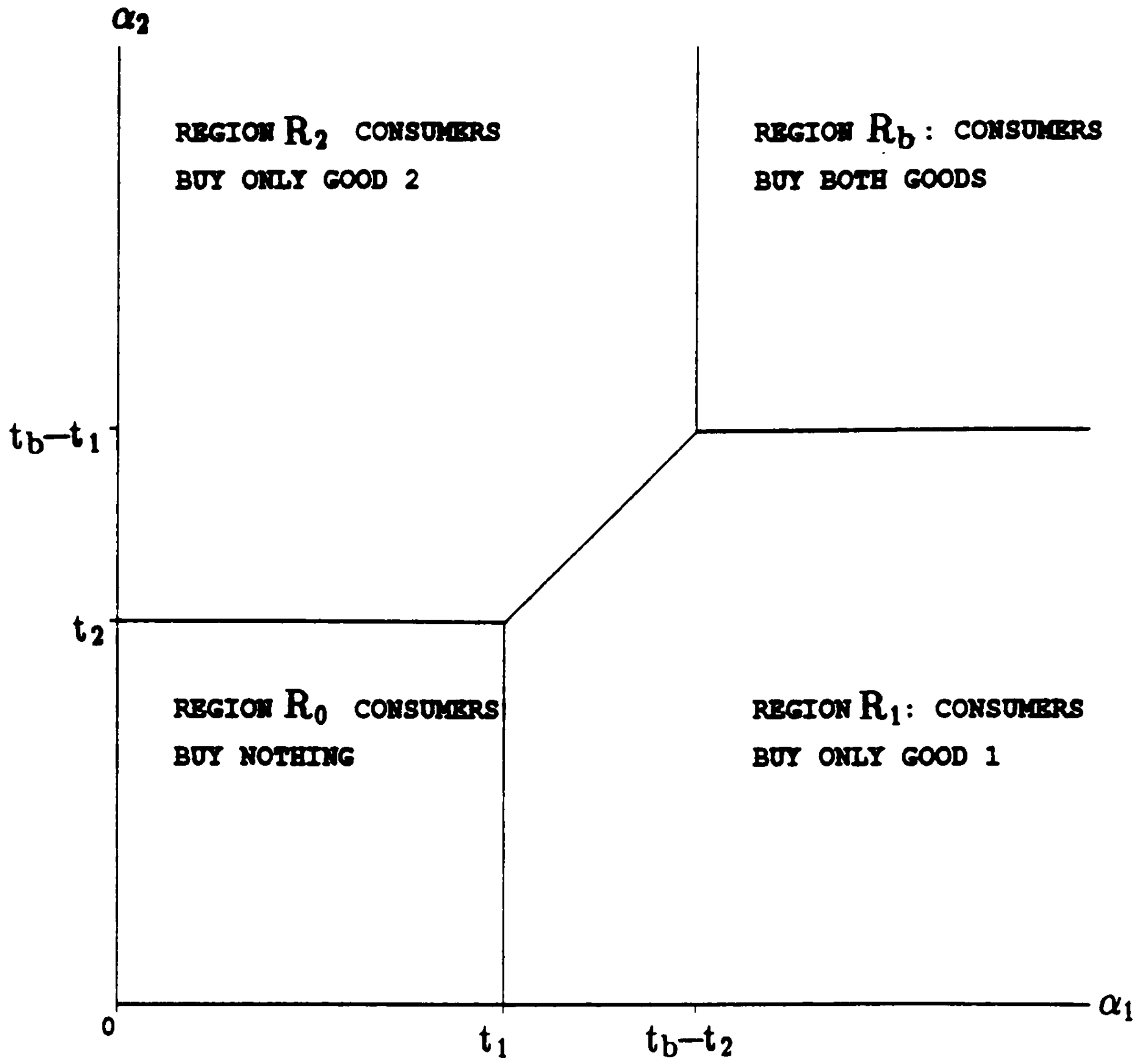

Figure 1A: The distribution of demand when $t_b > t_1 + t_2$

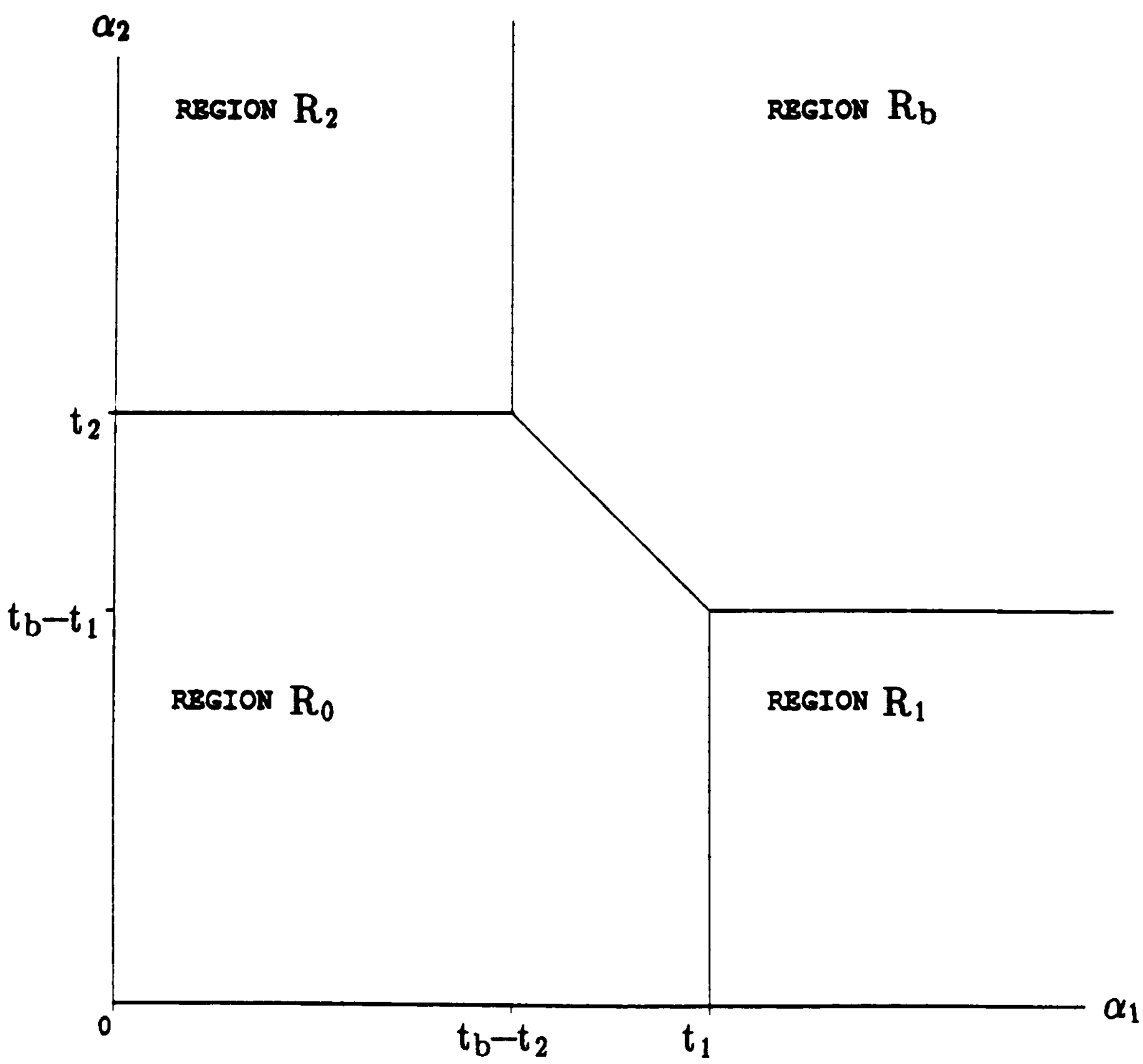

Figure 1B: The distribution of demand when $t_b < t_1 + t_2$

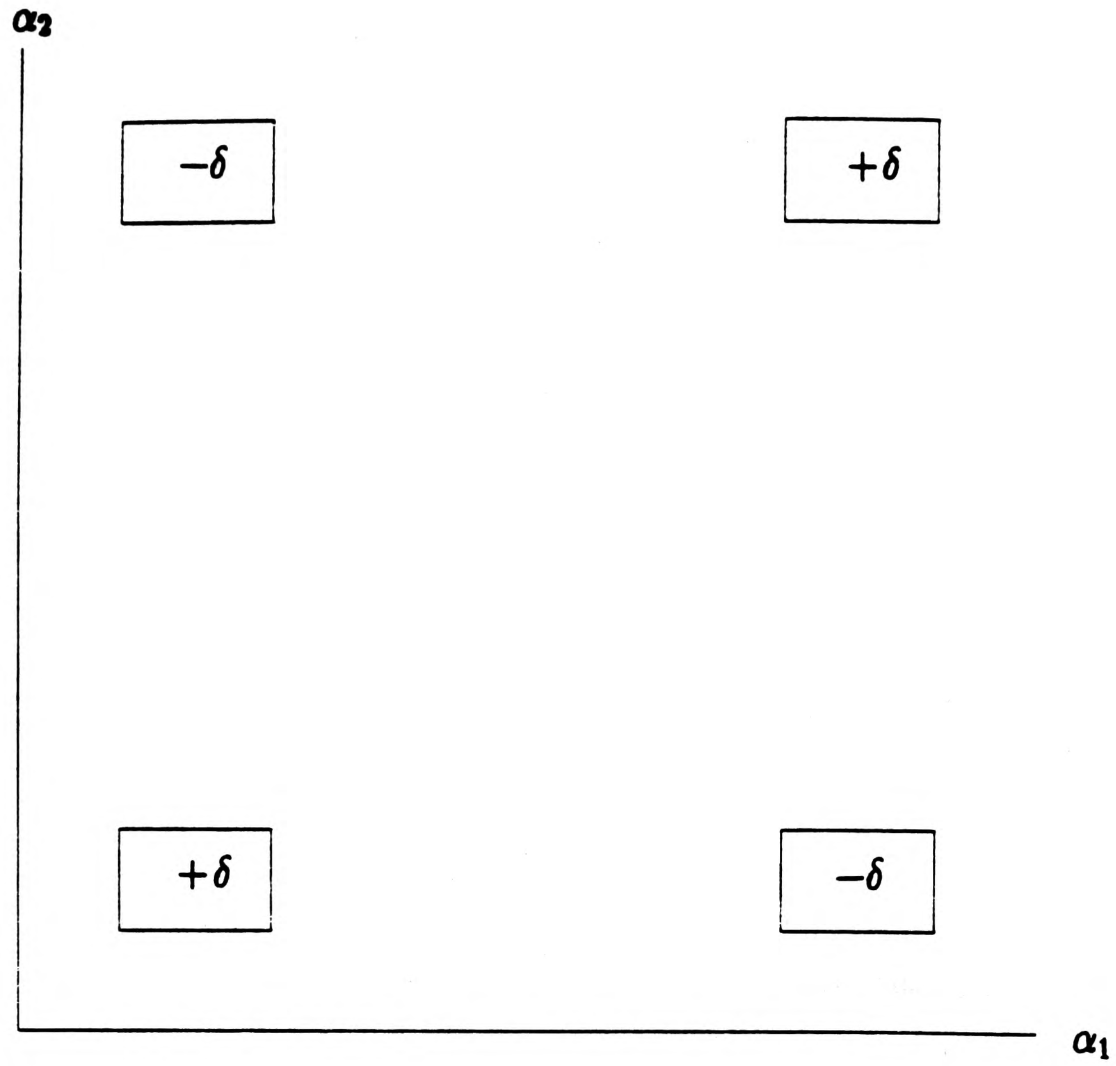

Figure 2A: A correlation–increasing transformation

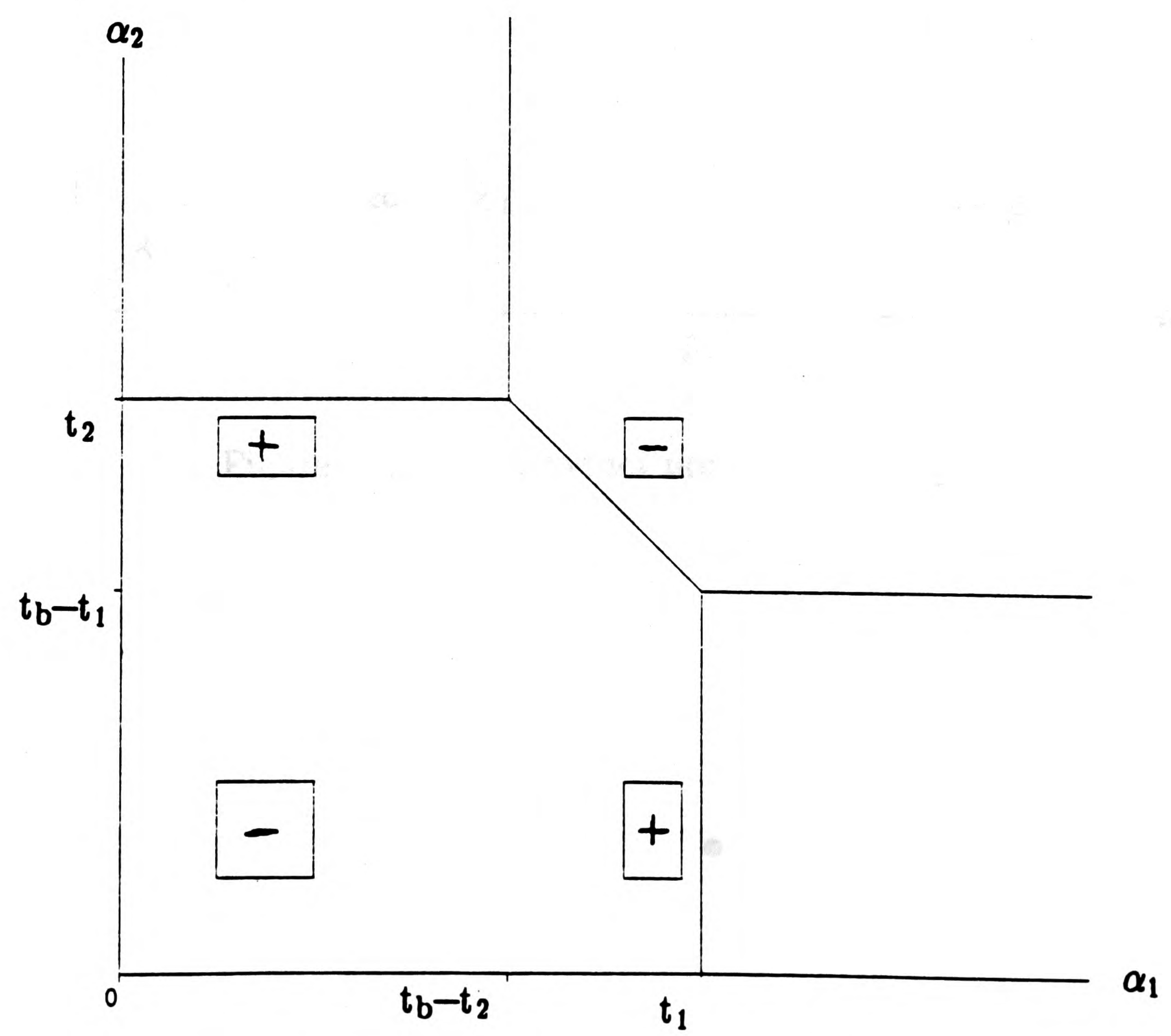

Figure 2B: How increasing negative correlation can lower profits when $t_b < t_1 + t_2$

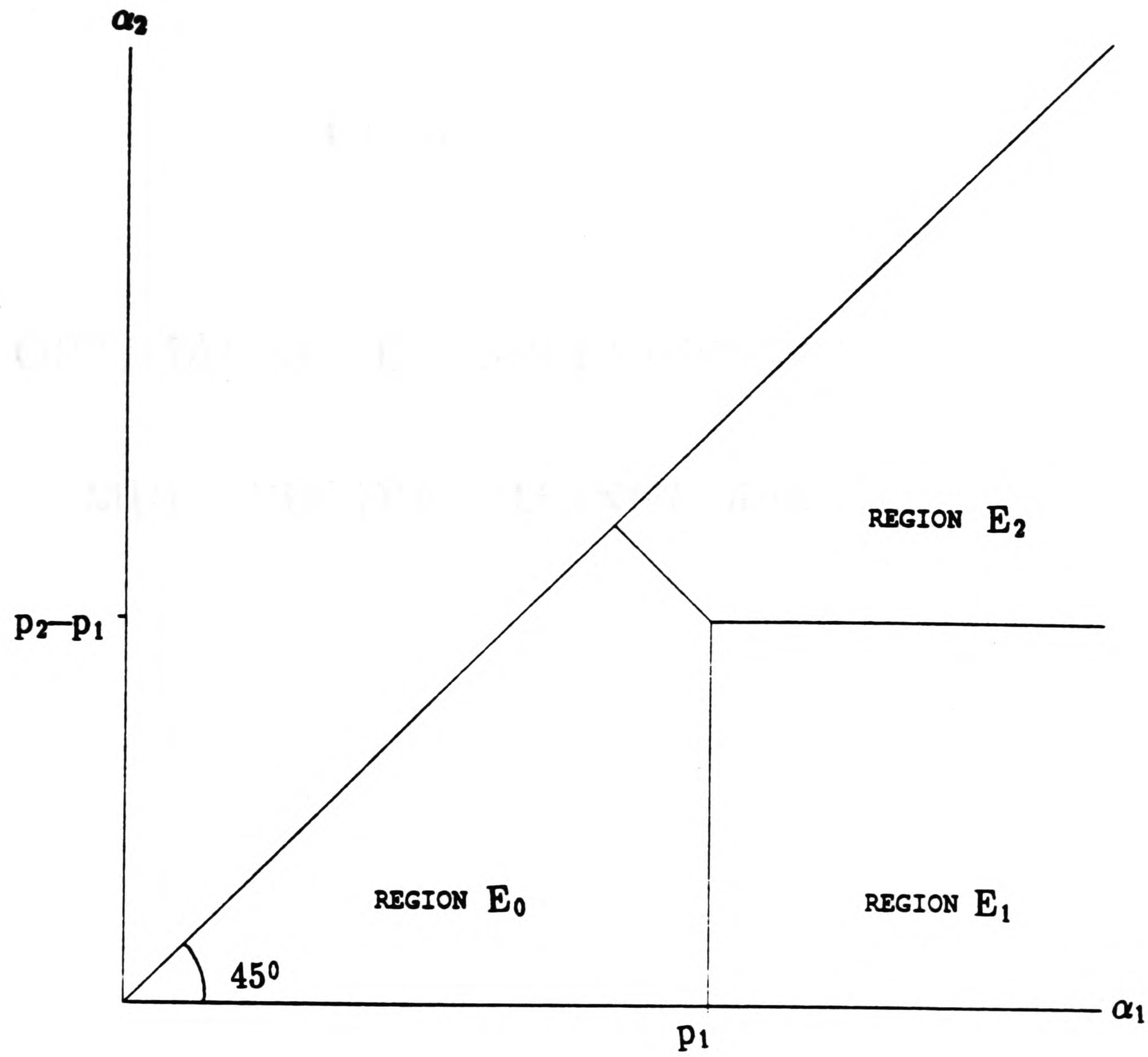

Figure 3: Single–product nonlinear pricing when $p_2 < 2p_1$

CHAPTER 4

OPTIMAL NONLINEAR PRICING BY A MULTIPRODUCT MONOPOLIST

1. Introduction

The first two chapters of this thesis have described and analyzed nonlinear pricing by a single–product firm in an environment with a small amount of hidden consumer information. Chapter 3 analyzed pricing strategies by a multiproduct firm under the assumption that consumers had several dimensions of hidden information and had unit demands. This chapter will bring together these two settings in a model where consumers have differing preferences for the output of a multiproduct monopolist and where quantities are chosen continuously.

An interesting question is how far any of the single–product results extend to multiproduct monopoly, or, more generally, what exactly will a profit–maximizing multiproduct nonlinear tariff look like? In the present context, questions we might like to answer include:

— Is the optimal multiproduct tariff concave? (This would be a generalization of Maskin and Riley's quantity discount result.)

— Are the 'cross–terms' negative, i.e. does the marginal price of one good decrease with the consumption of other goods? (This would be analogous to subadditive bundling in the last chapter.)

— What does the pattern of demand look like? Are consumers encouraged to buy all goods if they make any purchase, or will there be some consumers who buy only a strict subset of the goods on offer?

While it would be desirable to be able to answer these questions, the analysis is hampered by the fact that this multiproduct problem is simply much more difficult to solve than the corresponding problem for the single–good firm. The reward would be, however, solutions to a host of incentive problems which have so far been intractable.

Papers in this area are sparse. Most often the framework is simplified to that of 'unit demands', where consumers buy either one unit of a good or none at all. This was the topic of the previous chapter on bundling. Moving beyond the assumption of unit demands the literature is as follows. Mirman and Sibley (1980) analyze the multiproduct nonlinear pricing problem under the assumption that consumers vary only by a scalar parameter (although in their conclusion they pose the problem of multidimensional types). Such a problem may be solved using similar methods to the standard single–product case, but I would argue that this model sheds little light on the design of good multiproduct tariffs. Since consumers differ only by a scalar, the bundles which are chosen (given any tariff) will all lie on some one–dimensional curve in the product space, and so it will not be possible to address the question of whether consumers are encouraged or discouraged to buy two products rather than one. Moreover, it is rarely the case that knowledge of a given consumer's taste for one product determines exactly what her tastes are for any other (although this is not to deny that tastes could well be correlated).

Laffont, Maskin and Rochet (1985) consider the opposite problem to that of Mirman and Sibley: consumers buy only one product, but are differentiated by a two–dimensional parameter. As a result there will necessarily be 'bunching', so that some consumers with different parameters will buy the same quantity. It is interesting to see that in their example the optimal tariff has quantity discounts and the consumer with the highest parameters is efficiently served. These properties, then, are not purely the result of an assumption that consumers differ only be a scalar parameter. The analysis required to solve their example is significantly more complex than that used in the scalar problem, and indicates that mathematical difficulties occur when it is the type space rather than the decision space which is multi–dimensional. Again, though, their model is not capable of answering the question of what makes a good multiproduct tariff.

McAfee and McMillan (1988) is chiefly concerned with extending the paper by Laffont *et al.* to a more general framework. They describe a 'generalized single–crossing property' which guarantees that necessary conditions for a (differentiable) incentive scheme to be implementable are also sufficient (see footnote 1 below). They use this condition to solve the single–good/multidimensional–type problem for a class of cases (Laffont *et al.* simply found an example). This involves the rather restrictive (but crucial) assumption that there is one type parameter which must be positive for the consumer to gain any utility from consuming the good (this is their condition (21)).

A paper which has both multiple products and multiple types is Spence (1980).

Indeed, consumer i in his model simply has a utility function $u_i(\mathbf{x})$, and so consumers are not parameterized at all (other than by i). He models consumers discretely and solves only a sub–problem: given that the various consumers are each going to be assigned a certain bundle, what is the best way to make them pay for these bundles (either for profit or for social welfare). He has, however, no results concerning the optimal method of assigning these bundles in the first place. Nevertheless, he does highlight the central problem in solving the multidimensional case: in the single–good/single–parameter case it is possible to give a natural ordering of consumers since higher types choose higher quantities and, under a wide variety of conditions, only the 'downward' incentive compatibility constraints will bind. In contrast, in Spence's model it is not possible to predict in advance which of the many incentive constraints will prove to be binding, and this makes the problem rather intractable.

The starting points for the present chapter is Section 4 of Mirrlees (1976) and Section 7 of Mirrlees (1986). In the context of optimal tax theory, he sets out the framework of continuously distributed multidimensional types which I use, and derives some first–order conditions for an incentive scheme to be optimal. However, no examples are solved, nor are any hints given as to what an optimal scheme might look like.

In a soon–to–be–published book on nonlinear pricing, Wilson (1992b, Chapters 12–14) analyzes a similar model to that presented below.[1] He derives first–order conditions for an incentive scheme to be optimal although he has no results concerning the shape of the optimal tariff nor of the pattern of demand it induces. In addition to the parametric approach employed here, he also analyzes optimal nonlinear pricing when information comes in the form of a 'demand profile', by which is meant the function describing the number of consumers choosing a particular bundle of goods given a particular two–part tariff (although it seems likely that this approach is better suited to the single– rather than multi–product setting). This may be of greater practical use than the present approach which relies upon consumers being characterized by various unobserved parameters. Wilson also uses numerical simulations in order to obtain

[1] For the record, none of the results in this chapter appear in Wilson's book.

particular solutions to the problem, and his technique is to construct an algorithm to find solutions to the first–order conditions. Finally, Wilson has a chapter on real–world examples of multiproduct nonlinear tariffs, which include advertising rates in American periodicals, electricity tariffs in France, telephone tariffs in America and express delivery firms.

The structure of the chapter is as follows. In Section 2 I set out the model and use some ideas from duality to derive some facts which be useful later on. In Section 3 I find the conditions for a given tariff to be optimal and show that this optimum is unique. In Section 4 it is proved that, in contrast to the single–good case, the multi–product firm will almost always wish to exclude potential consumers from the market. In Section 5 I analyze which directions of movement in the optimal additively separable tariff will increase profits under the assumption that parameters are independently distributed. A family of profit–enhancing changes is found and so, in particular, an additively separable tariff cannot be optimal with independent parameters. Many of these movements will also benefit all consumers and so Pareto improvements are possible to separable tariffs.

In Section 6, which contains most of the substance of this chapter, I set out a strategy for solving the problem which involves ignoring a necessary convexity constraint and checking afterwards that this constraint has indeed been satisfied. This 'first–order' approach is a direct extension of the method so successfully used in the standard single–product case. Whenever this first–order approach is valid not only are the high demand consumers efficiently served (as is the case in the single–product problem), but also those with low demands. This has the corollary that, subject to a condition on costs, at the optimum there is pure bundling, so that once a consumer decides to buy any goods she will buy all goods. This is a striking result which contrasts with the previous chapter.

In Section 7 a general solution is found for a simple class of cases. These include the multivariate Normal, t and Weibull distributions. In the solved examples, the form of the optimum depends in a surprising way upon whether there are an even or odd number of goods (Section 8). I believe that in most cases which are not covered by the solved class mentioned above an explicit solution will not be obtainable, and so in Section 9 I describe how numerical simulations may be used to approximate the true optimum. The procedure is expensive in computing time but reliable, and uses a direct approach rather than an algorithm for solving some set of first-order conditions (as used by Wilson). All the examples generated by this means also have pure bundling, and it therefore remains an attractive conjecture that pure bundling will turn out to be a characteristic feature of these multidimensional models. Finally, Section 10 offers some concluding remarks on this exploratory work.

2. The model

We shall consider a firm with a monopoly on n goods. Consumers of these goods have a variety of preferences over these goods, parameterized by the vector $\boldsymbol{\alpha}$. A general form for a type $\boldsymbol{\alpha}$ consumer's utility would be $U(\boldsymbol{\alpha}, \mathbf{x}, y)$, where $\mathbf{x}$ is the bundle of goods consumed and y is income. However, in this first attempt to understand this problem we shall immediately specialize this function to be of the form

$$U(\boldsymbol{\alpha}, \mathbf{x}, y) = \sum_1^n \alpha_i q_i(\mathbf{x}) + y \tag{1}$$

where all the $q_i(\cdot)$ functions are increasing and concave. Therefore, utility is linear in $\boldsymbol{\alpha}$ and there are no income effects. In contrast to the papers of Mirman and Sibley and Laffont *et al.*, then, I take the dimension of the product space to be equal to the dimension of the type space. We assume that the firm has constant unit cost of c_i in producing good i. In this case, for the purposes of analysis it will be much more convenient to consider $\mathbf{q}$ rather than $\mathbf{x}$ as the 'output' of the firm, and because of the constant unit cost assumption we can write its cost of supplying any agent with utility $\mathbf{q}$ as $C(\mathbf{q})$ regardless of its supplies to other consumers. (For example, suppose that $c_i \equiv 1$ and that $q_i(\mathbf{x}) = \sqrt{x_i}$, then $C(\mathbf{q}) = \|q\|^2$.) The cost-of-utility function $C(\cdot)$ is necessarily convex given that the $q_i(\cdot)$ are concave. By an abuse of terminology, we will refer to the vector $\mathbf{q}$ as the vector of 'quantities'. In sum, a consumer of type $\boldsymbol{\alpha}$'s utility given consumption of $\mathbf{q}$ and income y is

$$U(\boldsymbol{\alpha}, \mathbf{q}, y) \equiv \boldsymbol{\alpha}'\mathbf{q} + y$$

while the firm's cost in supplying any consumer with quantities $\mathbf{q}$ is $C(\mathbf{q})$. The reason I choose this very special linear framework is that we can use some useful results from the theory of duality and convex functions to help us solve the problem.

We have to make a judgment about whether to model consumers discretely or continuously. The former has the advantage of being more intuitive and 'realistic', but notationally is messier, for instance in its use of inequalities to describe the various incentive compatibility constraints. The paper by Spence takes this approach. If, on the other hand, we choose to have a continuous distribution of types then, firstly we will be able to use calculus with more powerful effect, and secondly there will then be a continuous distribution of different bundles chosen — this means that we will get a tariff schedule instead of merely an assignment of prices to a finite number of bundles. This will probably be more illuminating. I therefore choose the latter.

The firm is not able to observe a given consumer's type, but has prior beliefs over the distribution of types, described by the density function $f(\boldsymbol{\alpha})$ which has support $A \subset \mathbb{R}^n_+$. For technical reasons, we suppose that f is smooth and that A is closed and convex. It aims to maximize its profit by offering a suitable tariff $T(\mathbf{q})$, where $T(\mathbf{q})$ is the charge for the bundle of quantities $\mathbf{q}$. We suppose that there are no possibilities of arbitrage amongst consumers and that the firm is able to monitor the sales to each consumer and therefore that the tariff T may take any form (so it may have either increasing or decreasing returns).

Since we assume consumers cannot be forced to consume, we suppose that $T(0) = 0$. Faced with a particular tariff T, a consumer of type $\boldsymbol{\alpha}$ obtains a surplus of:

$$(2) \qquad s(\boldsymbol{\alpha}) \equiv \max: \{\boldsymbol{\alpha}'\mathbf{q} - T(\mathbf{q}) \mid \mathbf{q} \geq 0\}.$$

Since $T(0) = 0$ this function will always be non–negative. In fact s is the convex (or 'Fenchel') dual of the firm's tariff schedule $T(\cdot)$, and, as such, it is also necessarily convex, continuous and increasing.

DEFINITION: The surplus function $s(\cdot)$ is said to be *implementable* by the firm if it may be induced from (2) by means of some tariff T such that $T(0) = 0$.

The following lemma is analogous to Lemma 1 of Chapter 1:

LEMMA 1. *The function* $s(\cdot)$ *is implementable if and only if it is convex, increasing, non–negative and continuous on the set* A.[2]

PROOF. (Sketch) Necessity follows from above discussion. The argument for sufficiency goes as follows. For such a surplus function, the dual of the dual brings us back to the original function again. More precisely, let $T(\cdot)$ be given by

$$(3) \qquad T(q) = \max: \{ \boldsymbol{\alpha}' q - s(\boldsymbol{\alpha}) \mid \boldsymbol{\alpha} \in A \},$$

and define $s^{**}(\boldsymbol{\alpha})$ as the surplus function which results from the tariff T given by (2).

---

[2] It is this simple characterization of the implementable surplus functions which is the main advantage of using the representation of utility u in (1) which is linear in $\boldsymbol{\alpha}$. When the parameter and product spaces are only of a single dimension, more general utility functions $u(q, \alpha)$ may be considered: provided that u satisfies the 'Spence–Mirrlees' single–crossing property

$$\frac{\partial^2 u}{\partial \alpha \partial q} \geq 0 \quad \text{for all } (\alpha, q),$$

then the response function $q(\alpha)$ may be implemented by some tariff $T(q)$ if and only if $q(\alpha)$ is non–decreasing. For a general survey of this scalar case see section 7.3.1 in Fudenberg and Tirole (1991). McAfee and McMillan (1988) propose an extension of the single–crossing property to the multidimensional case which they define to hold for a general utility function $u(\boldsymbol{\alpha}, q)$ if, given q, $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$ there always exists $\lambda > 0$ such that the following vector equality holds:

$$u_q(\boldsymbol{\alpha}, q) - u_q(\hat{\boldsymbol{\alpha}}, q) = \lambda u_{\alpha q}(\hat{\boldsymbol{\alpha}}, q)'(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}).$$

This is a rather restrictive assumption since it implies that, for those types which participate, the set of types $\boldsymbol{\alpha}$ who buy the same bundle q is some kind of hyperplane.
Rochet (1987) provides a slight generalization of the formulation considered here, but considers only utility functions that are linear in the parameters. Indeed, he concludes that "when u is not linear w.r.t. [$\boldsymbol{\alpha}$], it is very difficult to get sufficient conditions for [implementability]." Other papers which rely on utility functions which are linear in types, and for a quite different reason, are Caplin and Nalebuff (1988, 1991).

Then duality implies that $s = s^{**}$ if and only if s satisfies the conditions of the lemma. Therefore, s is implemented by the tariff T in (3). □

This demonstrates that, given a surplus function which satisfies the conditions of the lemma, we can recover the tariff which would induce this function by means of (3).[3]

If $s(\cdot)$ is differentiable at $\boldsymbol{\alpha}$ it will satisfy the envelope condition

$$(4) \qquad \nabla s(\boldsymbol{\alpha}) = \mathbf{q}(\boldsymbol{\alpha})$$

where $\mathbf{q}(\boldsymbol{\alpha})$ is the (then necessarily unique) set of outputs which maximize utility for type $\boldsymbol{\alpha}$ consumer given T. The fact that s is convex means that it is differentiable almost everywhere (in the sense of Lebesgue measure) on A. This implies that there will be a unique optimal choice of outputs for almost all consumers (given that consumers are continuously distributed on A). Therefore, we will use the notation $\mathbf{q}(\boldsymbol{\alpha})$ for the demand function even when this function is not necessarily everywhere well defined.

3. The optimal tariff

Since it is almost always the case that $T(\mathbf{q}(\boldsymbol{\alpha})) = \boldsymbol{\alpha}'\mathbf{q}(\boldsymbol{\alpha}) - s(\boldsymbol{\alpha})$, the profit obtained by the firm from a type $\boldsymbol{\alpha}$ consumer is $\boldsymbol{\alpha}'\mathbf{q}(\boldsymbol{\alpha}) - s(\boldsymbol{\alpha}) - C(\mathbf{q}(\boldsymbol{\alpha}))$, where $\mathbf{q}(\boldsymbol{\alpha})$ is given in (4). Its total profit is then the expectation of all of these individual profit contributions. Therefore, the central problem we shall be concerned with is to:

$$\text{(P)} \qquad \begin{aligned} &\text{maximize:} && \pi(s) \equiv \int_A [\boldsymbol{\alpha}'\mathbf{q}(\boldsymbol{\alpha}) - s(\boldsymbol{\alpha}) - C(\mathbf{q}(\boldsymbol{\alpha}))]\, f(\boldsymbol{\alpha})\, dV \\ &\text{subject to:} && \text{(i)} \ \nabla s(\boldsymbol{\alpha}) \equiv \mathbf{q}(\boldsymbol{\alpha}) \text{ almost everywhere on } A \text{ and} \\ & && \text{(ii)} \ s(\cdot) \text{ is continuous, convex, increasing and non-negative.} \end{aligned}$$

---

[3] Since any tariff generated from (3) is necessarily continuous and convex we have the corollary that the firm can restrict attention to continuous and convex tariffs such that $T(0) = 0$. In particular, a tariff with a 'fixed charge' F (so that $T(0) = F$) will always be equivalent to some other tariff which tends to zero for low quantities.

This problem is of the standard 'calculus of variations' type, except for the fact the integral is maximized subject to the constraint that s be convex. This is a non–standard and difficult sort of constraint to handle.

Write $s^*(\cdot)$ for the solution to problem (P). Then $\pi(s^*)$ (where π is given in (P)) must be greater than $\pi(s)$ for all other implementable functions $s(\cdot)$. In particular, it must be greater than $\pi(s)$ for all $s(\cdot)$ which are near to s^* (in some sense). With this in mind we introduce the concept of an admissible deformation of the surplus function s. It is tempting to say that a function η is admissible given s, if $s + \eta$ is also implementable. However, it turns out that we will only be interested in which directions of movement in $s(\cdot)$ are allowed. We therefore make the following definition:

DEFINITION: Call η (a function defined on A) an *admissible deformation* of s if $s + t\eta$ is implementable for some $t > 0$.

Since the set of implementable functions on A is itself convex, it follows that, if $s + t\eta$ is implementable, then so is $s + \epsilon\eta$ for all $0 < \epsilon < t$. This tells us that, given $s(\cdot)$, the set of admissible deformations is a cone. In addition, it is convex and contains all implementable functions.

We can now now derive the first order conditions for $s^*(\cdot)$ to be optimal for the firm. If η is admissible and s^* is optimal then $\pi(s^*) \geq \pi(s^* + t\eta)$ for all sufficiently small t. This implies that $\frac{d}{dt}\pi(s^* + t\eta)\Big|_{t=0} \leq 0$. If A is compact and provided that s^* and η are continuously differentiable,[4] we can use the Leibniz Rule to reverse the order of differentiation and integration to give:

$$(5) \qquad \frac{d}{dt}\pi(s^* + t\eta)\Big|_{t=0} = \int_A [(\boldsymbol{\alpha} - \nabla_q C(q^*))'\nabla\eta - \eta] f \, dV$$

[4] In fact, it is generally true of convex functions that if they are differentiable they are also continuously differentiable.

(where $\mathbf{q}^* = \nabla s^*$ is the optimal demand function). Therefore, the first–order condition for s^* to be optimal is that:

$$(6) \qquad \int_A [(\boldsymbol{\alpha} - \nabla_q C(\mathbf{q}^*))' \nabla \eta - \eta] f \, dV \ \leq \ 0$$

for all admissible deformations η of s^*.

In fact, since profit $\pi(s)$ in (P) is concave in s, the first–order condition (6) is also sufficient. Condition (6) is not particularly easy to work with, largely because it is hard to characterize which deformations of s^* are allowed (although it is useful to show that certain surplus functions are *not* optimal — see Results 2 and 3 below). However, we can use it to obtain one result:

RESULT 1. *The optimal surplus function* s^* *is unique.*

REMARK: It does not follow that the optimal tariff $T(\cdot)$ is unique since many different tariffs will induce the surplus function s^*.

PROOF. Let s^* be optimal. Therefore, the function $t \longmapsto \pi(ts^*)$ must be maximized at $t = 1$. But this means that:

$$(7) \qquad 0 \ = \ \frac{d}{dt} \pi(ts^*) \Big|_{t=1} \ = \ \int_A [\boldsymbol{\alpha}' \mathbf{q}^* - s^* - \nabla_q C(\mathbf{q}^*)' \mathbf{q}^*] f \, dV.$$

Now suppose that s is also optimal, and let $\eta = s - s^*$ be considered as a deformation of s^* (it is admissible). Then, if $\mathbf{q} = \nabla s$ then from (6) we must have:

$$0 \ \geq \ \int [\boldsymbol{\alpha}'(\mathbf{q} - \mathbf{q}^*) + s^* - s - \nabla_q C(\mathbf{q}^*)'(\mathbf{q} - \mathbf{q}^*)] f \, dV$$

$$= \ \int [\nabla_q C(\mathbf{q}) - \nabla_q C(\mathbf{q}^*)]' \mathbf{q} \, f \, dV \quad \text{from the optimality of } s^* \text{ and } s \text{ and (7).}$$

Similarly,

$$0 \geq \int [\nabla_q C(\mathbf{q}^*) - \nabla_q C(\mathbf{q})]'\mathbf{q}^* \, f \, dV.$$

But $C(\cdot)$ strictly convex means that for $\mathbf{q} \neq \mathbf{q}^*$,

$$(\nabla_q C(\mathbf{q}^*) - \nabla_q C(\mathbf{q}))'(\mathbf{q}^* - \mathbf{q}) > 0.$$

This contradicts the two above inequalities unless $\mathbf{q}^* = \mathbf{q}$ almost everywhere. Therefore $s^* = s + \text{constant}$, and since the firm will always choose to make the non–negativity constraint bind at an optimum, this constant must be zero. □

We might hope to find cases where the optimal s^* allows *any* deformation η — if this were so then (6) could be considerably simplified. Unfortunately, this can never be the case in the multidimensional setting, and one reason for this is given in the next section.

4. The inevitability of exclusion

When the firm assigns an output $\mathbf{q}$ to a type $\boldsymbol{\alpha}$ consumer it faces a dilemma: increasing this output means that typically it will increase the profits from this consumer; on the other hand, from the equation $\nabla s(\boldsymbol{\alpha}) = \mathbf{q}(\boldsymbol{\alpha})$ we see that it will pay the penalty of having to give all consumers of type higher than $\boldsymbol{\alpha}$ a higher surplus (and hence lower profits). An interesting question then is whether the latter effect will outweigh the former for low types, and so whether the firm will choose to exclude some low–demand potential customers.

In the single–product case the answer to this is ambiguous. With one product α is a scalar and A is an interval, say $[\underline{\alpha}, \bar{\alpha}]$. Using the same method as described in Section 4.1 of Chapter 1 provided that

(8) $$\alpha - \frac{1 - F(\alpha)}{f(\alpha)} \text{ is increasing}$$

then then the profit–maximizing demand function $q^*(\alpha)$ is given by

$$q^*(\alpha) \ \underset{q \geq 0}{\text{maximizes:}} \ q[\alpha - \frac{1 - F(\alpha)}{f(\alpha)}] - C(q) .$$

Using this formula we see that the firm chooses to exclude no consumers if and only if

$$C'(0) < \underline{\alpha} - (1 - F(\underline{\alpha}))/f(\underline{\alpha}),$$

where $\underline{\alpha}$ is the lowest type. Even for the lowest type consumers, then, the profits gained from their inclusion will often outweigh the lost profit this policy will entail for the higher types. Just as in the bundling chapter, it is striking that this is (almost) never the case for the multiproduct firm.

RESULT 2. *If the economy is symmetric (i.e. if* f *and* C *are symmetric functions),* f *is bounded and* $A \subset \mathbb{R}^n_+$ $(n > 1)$ *is closed, convex and is such that the point* $\boldsymbol{a}$ *that minimizes* $\{\sum_1^n \alpha_i \mid \boldsymbol{\alpha} \in A\}$ *is unique, then the firm will choose to exclude a non–trivial set of consumers from the market.*

PROOF. See Appendix.[5]

Whilst the proof of this result is a little technical, the intuition is I think clear and is exactly as was the case with Result 1 of Chapter 3. If a tariff is such that all potential consumers do join the market then consider what happens if some fixed charge ϵ is added to this tariff. The firm will gain ϵ in profits from every consumer who remains in the market, but will pay the penalty of causing some low demand consumers to exit. The proof of the result demonstrates that for the multi–product case for small ϵ the number who exit is of order strictly greater than ϵ. Therefore, the former effect will necessarily

[5] But see also the remark that follows Result 4 below.

dominate the latter, and exclusion will be optimal.[6] Even if the demand parameters are very high in relation to cost, the firm will not wish to serve all the market. This is not a result of rapacious profit–maximization — if the objective was utilitarian welfare maximization the result would still hold. What is crucial, on the other hand, is the assumption of a continuous distribution — if consumers were modeled discretely this result would not necessarily hold.

5. The benefit of introducing cross–dependence in the tariff

In this section we ask the question of whether multiproduct firms should be allowed or encouraged to offer more complex tariff structures than is usually the case at present. For instance, should a telephone company make the marginal price of long–distance calls depend upon the number of local calls a user makes? And if so, should the dependence be positive or negative?

In Result 2 of the bundling chapter we posed the question of which directions of movement in the tariff would increase profits, starting from the optimal non–bundling tariff. We saw that when the parameters were independently distributed the marginal incentive was to offer the two goods at a discount compared with the charge for the two goods separately. In the present context, the equivalent of a non–bundling tariff is an additively separable tariff, i.e. in the two–good case a tariff T which can be written in the form $T(q_1, q_2) = T_1(q_1) + T_2(q_2)$. Faced with a separable tariff, then, a consumer makes her purchase decision for good i independently from her decisions concerning other goods. Analogously to Result 2 in that chapter, it is natural to discover which directions of movement increase profits, starting from the position of the optimal separable tariff. What kinds of cross–dependence in the tariff will prove to be beneficial to the firm at the margin?

In order to answer this, we must first find the optimal separable tariff. Suppose that the cost function is additively separable: $C(q_1, q_2) = C_1(q_1) + C_2(q_2)$. Let the two

[6] Just as in the bundling case it is not quite true that it is never the case that the multiproduct firm will wish to exclude potential consumers, since if consumers were distributed on a set like $A = \{ \alpha \mid 1 \leq \Sigma\, \alpha_i \leq 2 \}$, then the proof would not be valid.

parameters be independently distributed, with α_i having distribution function F_i and density function f_i. Then the optimal separable tariff is given by $T_1^*(q_1) + T_2^*(q_2)$, where $T_i^*(q_i)$ is the optimal single–product tariff for good i. Therefore, the optimal demand function for good i, $q_i^*(\alpha_i)$ is given below (8) above. Let α_i^{min} be the minimal type for which $q_i^*(\alpha_i) > 0$. Then the first–order condition for $q_i^*(\alpha_i)$ to be optimal is

$$(9) \qquad (\alpha_i - C_i'(q_i^*(\alpha_i)))f_i(\alpha_i) = 1 - F_i(\alpha) \qquad i = 1, 2 \text{ and } \alpha_i > \alpha_i^{min}.$$

If this response function is induced by the single–product tariff $T_i^*(\cdot)$ then the optimal additively separable tariff is $T_1^*(q_1) + T_2^*(q_2)$. In the following result we describe a family of profit–enhancing changes to this optimal separable tariff.

RESULT 3. *In the two–good case with separable costs, suppose that demand parameters are independently distributed with distribution functions given by* F_1 *and* F_2 *which each strictly satisfy (8). If* $T_1^*(q_1) + T_2^*(q_2)$ *is the optimal separable tariff then profits will increase with any deformation to this tariff of the form:*

$$T(q_1, q_2) = T_1^*(q_1) + T_2^*(q_2) - \epsilon\delta_1(q_1)\delta_2(q_2)\ ,$$

where ϵ *is small and positive and* $\delta_1(\cdot)$ *and* $\delta_2(\cdot)$ *are any smooth functions which are either both positive or both negative and satisfy* $\delta_i(0) = 0$.

PROOF. Let $s_1^*(\alpha_1) + s_2^*(\alpha_2)$ be the optimal separable surplus function corresponding to the optimal separable tariff schedule $T_1^*(q_1) + T_2^*(q_2)$, with demand functions given by $q_i^*(\alpha_i)$ satisfying (9). Consider a deformation of the form $\eta(\alpha_1, \alpha_2) = \phi_1(\alpha_1)\phi_2(\alpha_2)$. If (8) holds strictly then $q_i^*(\alpha_i)$ is strictly increasing for $\alpha_i > \alpha_i^{min}$. In this case η is admissible provided that ϕ_i is smooth and $\phi_i(\alpha_i) \equiv 0$ for $\alpha_i < \alpha_i^{min}$. From (5), the marginal change in profit is given by

$$\int_{\alpha_2^{\min}}^{\bar{\alpha}_2}\int_{\alpha_1^{\min}}^{\bar{\alpha}_1} [\alpha_1\phi_1'(\alpha_1)\phi_2(\alpha_2) + \alpha_2\phi_2'(\alpha_2)\phi_1(\alpha_1) - \phi_1(\alpha_1)\phi_2(\alpha_2) - C_1'(q_1^*)\phi_1'(\alpha_1)\phi_2(\alpha_2) - C_2'(q_2^*)\phi_2'(\alpha_2)\phi_1(\alpha_1)]f_1(\alpha_1)f_2(\alpha_2)\, d\alpha_1\, d\alpha_2$$

$$= \int_{\alpha_2^{\min}}^{\bar{\alpha}_2}\int_{\alpha_1^{\min}}^{\bar{\alpha}_1} [(1-F_1(\alpha_1))\phi_1'(\alpha_1)\phi_2(\alpha_2)f_2(\alpha_2) + (1-F_2(\alpha_2))\phi_2'(\alpha_2)\phi_1(\alpha_1)f_1(\alpha_1) - \phi_1(\alpha_1)\phi_2(\alpha_2)f_1(\alpha_1)f_2(\alpha_2)]\, d\alpha_1\, d\alpha_2$$

(substituting from (9))

$$= \left(\int_{\alpha_1^{\min}}^{\bar{\alpha}_1} \phi_1(\alpha_1)f_1(\alpha_1)d\alpha_1\right)\left(\int_{\alpha_2^{\min}}^{\bar{\alpha}_2} \phi_2(\alpha_2)f_2(\alpha_2)d\alpha_2\right)$$

(using integration by parts and the fact that $\phi_i(\alpha_i^{\min}) = 0$). Clearly this is positive provided that the ϕ_i are either both positive or both negative, in which case any change in the surplus function in the direction of $\phi_1(\alpha_1)\phi_2(\alpha_2)$ will increase profits. As a particular case of this we can take $\phi_i(\alpha_i) = q_i^*(\alpha_i)$ (this satisfies the ϕ_i both being positive and $\phi_i(\alpha_i) = 0$ for $\alpha_i < \alpha_i^{\min}$). This implies that deforming the surplus function by $\epsilon q_1^*(\alpha_1)q_2^*(\alpha_2)$ for some small $\epsilon > 0$ will increase profits. By the envelope theorem, this is precisely the deformation induced by subtracting a term ϵq_1q_2 from the original separable tariff $T_1^*(q_1) + T_2^*(q_2)$. More generally, any deformation to $s_1^* + s_2^*$ of the form $\epsilon\delta_1(q_1^*(\alpha_1))\delta_2(q_2^*(\alpha_2))$ will again be implementable and profit–enhancing so long as the δ_i are either both positive or both negative and satisfy $\delta_i(0) = 0$. But such a deformation is induced by subtracting a term $\epsilon\delta_1(q_1)\delta_2(q_2)$ from the original tariff as claimed. □

If $\delta_1(\cdot)$ and $\delta_2(\cdot)$ are non–negative all consumers will be better off and so, as with the bundling case, we see that introducing many kinds of non–separability into the tariff will be Pareto improving. As a corollary to this we see that, if tastes are independently distributed, it is never optimal for the firm to use a separable tariff. These results mirror those obtained in the simpler bundling setting.

In Result 4 in the bundling chapter I showed that, not only is the marginal incentive to reduce the price for a bundle of two goods, but that at the optimum we will have the price of two goods less than the sum of the prices of the two goods separately (in many cases). We have shown in the above result that the marginal incentive is to introduce negative cross–terms into the tariff (at least along the diagonal), and this perhaps suggests that, under certain conditions, the optimal tariff will also have negative cross–derivatives. In the examples we calculate below this is indeed the case. Sadly, however, no equivalent general condition has yet been found.

6. A strategy for solving the problem

Up until now we have two results on the form of the optimal tariff: it must exclude some consumers and it cannot be separable if tastes are independent. What is clearly lacking is a method which could generate a solved example. That is the purpose of this section.

Let us try to mirror the technique used to solve the single–good case as described in Section 4.1 of Chapter 1. There, the method involved solving the problem (P) ignoring the constraint that s must be convex, and then checking *ex post* to see that this condition is indeed satisfied. This 'first–order' approach was valid if and only if (8) held. We can try to do exactly the same thing when the dimension of A is greater than one, and therefore try to solve:

$$(\mathrm{P}^{\mathrm{fo}})\qquad \begin{aligned} &\text{maximize} \quad \int_A [\boldsymbol{\alpha}'\mathbf{q}(\boldsymbol{\alpha}) - \mathrm{s}(\boldsymbol{\alpha}) - \mathrm{C}(\mathbf{q}(\boldsymbol{\alpha}))]\mathrm{f}(\boldsymbol{\alpha})\, \mathrm{dV} \\ &\text{subject to} \quad \mathrm{s}(\boldsymbol{\alpha}) \geq 0, \ \ \mathbf{q}(\boldsymbol{\alpha}) \geq \mathbf{0} \ \text{ and } \ \nabla \mathrm{s}(\boldsymbol{\alpha}) = \mathbf{q}(\boldsymbol{\alpha}) \end{aligned}$$

(i.e. we ignore the constraint that s be convex). Again, for convenience we refer to this approach as the first–order approach, even though it is not quite equivalent (see Section 3 of Chapter 1).

Suppose we write

$$f(\alpha) = \mathrm{div}(\mu) \tag{10}$$

where $\mu(\alpha)$ is some continuous and piecewise continuously differentiable vector field on A, and $\mathrm{div}(\mu) \equiv \sum_1^n \frac{\partial \mu_i}{\partial \alpha_i}(\alpha)$ is the *divergence* of the field μ. The field μ will play the analogous role to $-(1 - F)$ in the one–dimensional setting, and it is helpful to think of μ as the 'integral' of f. There are many fields μ which satisfy (10), and the major difficulty with this approach will be to find the one solution which will work best.

For this multiproduct problem we shall need to use a multidimensional version of 'integration by parts', and for this we use the Divergence Theorem — for instance, see section 10.6 in Loomis and Sternberg (1990) for an explanation of this and the general theory of integration over surfaces.

THEOREM. *Let* A *be a compact convex set of full dimension in* $\mathbb{R}_n$, *and let* $\mathbf{m}(\alpha)$ *be a continuous and piecewise continuously differentiable vector field on* A. *Then*

$$\int_A \mathrm{div}(\mathbf{m})\, dV = \int_{\partial A} \mathbf{m}'\hat{\mathbf{n}}\, dS \tag{11}$$

where ∂A *is the boundary of* A, *and* $\hat{\mathbf{n}}$ *is the outward–pointing unit normal vector at a point on* ∂A.

For the examples we derive below we will need a refinement of this theorem to encompass non–compact domains. Therefore, suppose that A is merely closed (but not necessarily bounded). Then A may be written as the limit of some sequence of sets $\{ A_i \}_1^\infty$, each of which is compact. We can use the above theorem to deduce that:

$$\text{(12)} \qquad \int_A \operatorname{div}(\mathbf{m})\, dV = \lim_{i\to\infty} \int_{A_i} \operatorname{div}(\mathbf{m})\, dV = \lim_{i\to\infty} \int_{\partial A_i} \mathbf{m}'\hat{\mathbf{n}}\, dS .$$

For instance, if A is the whole of $\mathbb{R}_n$ then writing A_i for the set of points inside the sphere of radius i means that $A = \lim_{i\to\infty} A_i$ and each A_i is compact. Then (12) implies that the integral of the divergence of $\mathbf{m}$ over $\mathbb{R}_n$ is equal to the surface integral of $\mathbf{m}$ over the 'sphere at infinity', by which we mean the limit of the sequence of surface integrals over larger and larger spheres.

Using the identity $\operatorname{div}(s\boldsymbol{\mu}) \equiv \boldsymbol{\mu}'\nabla s + s\operatorname{div}(\boldsymbol{\mu})$ and the fact that $\nabla s \equiv \mathbf{q}$, equation (11) implies that

$$\text{(13)} \qquad \int_A s\operatorname{div}(\boldsymbol{\mu})\, dV = \int_{\partial A} s\boldsymbol{\mu}'\hat{\mathbf{n}}\, dS - \int_A \boldsymbol{\mu}'\mathbf{q}\, dV$$

and it is this formula which is analogous to simple integration by parts. When A is non–compact, (13) has the analogous version. Using (10) and substituting (13) into (P^{fo}) we see that profits may be written as

$$\text{(14)} \qquad \int_A [\boldsymbol{\alpha}'\mathbf{q}(\boldsymbol{\alpha}) - C(\mathbf{q}(\boldsymbol{\alpha}))]f(\boldsymbol{\alpha}) + \boldsymbol{\mu}(\boldsymbol{\alpha})'\mathbf{q}(\boldsymbol{\alpha})\, dV - \int_{\partial A} s(\boldsymbol{\alpha})\boldsymbol{\mu}(\boldsymbol{\alpha})'\hat{\mathbf{n}}\, dS.$$

Equation (14) represents firm profit no matter which choice of $\boldsymbol{\mu}$ we take in (10).

The strategy for finding the optimal demand function $\mathbf{q}^*(\boldsymbol{\alpha})$ goes as follows: let $\mathbf{q}^*(\boldsymbol{\alpha})$ be defined by

$$\mathbf{q}^*(\boldsymbol{\alpha}) \ \underset{\mathbf{q}\,\geq\,0}{\text{maximizes}} \ [\boldsymbol{\alpha}'\mathbf{q} - C(\mathbf{q})]f(\boldsymbol{\alpha}) + \boldsymbol{\mu}(\boldsymbol{\alpha})'\mathbf{q} .$$

Let $s^*(\boldsymbol{\alpha})$ be the resulting surplus function given the choice of integral $\boldsymbol{\mu}$ in the above expression. Suppose that we can find a choice of $\boldsymbol{\mu}$ which satisfies

$$\left.\begin{matrix} \mu(\alpha)'\hat{n} \geq 0 \\ v^*(\alpha) \geq 0 \end{matrix}\right\} \quad \text{comp. for } \alpha \text{ on } \partial A$$

(where 'comp.' stands for the standard complementary slackness condition, so that $\mu'\hat{n} \geq 0$, $s^* \geq 0$ and $s^*\mu'\hat{n} = 0$ on ∂A). If we can do this then is it clear we have found the optimum solution for the problem (P^{fo}). If, in addition, the resulting s^* is convex we have found the solution to the full problem (P). In fact, it will be convenient to modify this method slightly to allow $\text{div}(\mu) \leq f$ provided that we have equality whenever consumer surplus s^* is positive. To fix ideas, we write this method as a result:

RESULT 4. *Let* μ *and* s^* *be continuous and piecewise continuously differentiable functions satisfying:*

(15) $$\nabla s^*(\alpha) \ \text{maximizes}_{q \geq 0} \ [\alpha' q - C(q)]f(\alpha) + \mu(\alpha)' q \,,$$

(16) $$\left.\begin{matrix} \mu(\alpha)'\hat{n} \geq 0 \\ s^*(\alpha) \geq 0 \end{matrix}\right\} \quad \text{comp. for } \alpha \text{ on } \partial A, \text{ and}$$

(17) $$\left.\begin{matrix} \text{div}(\mu) \leq f \\ s^*(\alpha) \geq 0 \end{matrix}\right\} \quad \text{comp. for } \alpha \text{ in } A.$$

Then provided that s^* *is also convex it is the optimal choice of surplus function for the firm.*

PROOF. Let s be any other surplus function and let q and q^* be the respective demand functions corresponding to s and s^*. Then profit with s is:

$$\pi(s) = \int_A [\boldsymbol{\alpha}'\mathbf{q} - s - C(\mathbf{q})]f \, dV$$

$$\leq \int_A [\boldsymbol{\alpha}'\mathbf{q} - C(\mathbf{q})]f - s\,\mathrm{div}(\boldsymbol{\mu}) \, dV \quad \text{(since } s \geq 0 \text{ and } \mathrm{div}(\boldsymbol{\mu}) \leq f)$$

$$= \int_A [\boldsymbol{\alpha}'\mathbf{q} - C(\mathbf{q})]f + \boldsymbol{\mu}'\mathbf{q} \, dV - \int_{\partial A} s\boldsymbol{\mu}'\hat{\mathbf{n}} \, dS$$

(by the Divergence Theorem)

$$\leq \int_A [\boldsymbol{\alpha}'\mathbf{q} - C(\mathbf{q})]f + \boldsymbol{\mu}'\mathbf{q} \, dV \quad \text{(since } s \geq 0 \text{ and } \boldsymbol{\mu}'\hat{\mathbf{n}} \geq 0)$$

$$\leq \int_A [\boldsymbol{\alpha}'\mathbf{q}^* - C(\mathbf{q}^*)]f + \boldsymbol{\mu}'\mathbf{q}^* \, dV \quad \text{(from (15))}$$

$$= \int_A [\boldsymbol{\alpha}'\mathbf{q}^* - C(\mathbf{q}^*)]f + \boldsymbol{\mu}'\mathbf{q}^* \, dV - \int_{\partial A} s^*\boldsymbol{\mu}'\hat{\mathbf{n}} \, dS$$

(since $s^*\boldsymbol{\mu}'\hat{\mathbf{n}} \equiv 0$ on ∂A from (16))

$$= \int_A [\boldsymbol{\alpha}'\mathbf{q}^* - C(\mathbf{q}^*)]f - s^*\mathrm{div}(\boldsymbol{\mu}) \, dV$$

(by the Divergence Theorem)

$$= \int_A [\boldsymbol{\alpha}'\mathbf{q}^* - C(\mathbf{q}^*)]f - s^*f \, dV$$

(since $s^*(f - \mathrm{div}(\boldsymbol{\mu})) = 0$ in A by (17))

$$= \pi(s^*) .$$

Therefore, $s*$ is the solution to problem (P^{fo}). If $s*$ also happens to be convex it is the solution to the fully constrained problem (P). □

REMARK. This result provides corroboration for the earlier Result 2 which showed that there will be exclusion at the optimum. To see this, suppose that (almost) all consumers do have a positive surplus. In this case (17) implies that $\text{div}(\mu) = f$ throughout A and (16) implies that $\mu'\hat{n} = 0$ on the whole boundary of A. Using the Divergence Theorem:

$$0 = \int_{\partial A} \mu'\hat{n}\, dS = \int_{A} \text{div}(\mu)\, dV = \int_{A} f\, dV = 1$$

(if the number of consumers is normalized to unity). As this is a contradiction, any solution generated by the method of Result 4 will involve exclusion.

REMARK. We can extend Result 4 to the case where A is non–compact as follows. Let $A = \lim_{i\to\infty} A_i$ where each A_i is compact. Then, using the refinement of the Divergence Theorem in (12), the proof of Result 4 is valid provided that:

(i) equations (15) and (17) continue to hold throughout A

(ii) instead of (16) we have $\mu'\hat{n} \geq 0$ on ∂A and the complementary slackness condition is replaced with

(18) $$\int_{\partial A_i} s*\mu'\hat{n}\, dS \longrightarrow 0 \quad \text{as } i \longrightarrow \infty .$$

Suppose now that A is a rectangle with 'south–west' vertex **a** and 'north–east' vertex **b** (where **b** could be at infinity), and also that costs are additive: $C(q) = \Sigma\, C_i(q_i)$ say. Then one immediate fact is, if the problem (P) can be solved by the method of Result 4 then we have the result (familiar from the single–good case) that consumers with the highest demands are efficiently served. For suppose that a consumer has parameter b_i for good i. Then (provided she is served at all, i.e. $s* > 0$) from (15) and (16) we see that $q_i^*(\alpha)$ is chosen to maximize $b_i q - C_i(q)$, so that marginal utility from the i^{th} good is equal to marginal cost. This is perhaps not too surprising. However, what is more striking is that those consumers with the *lowest* type parameters are also served efficiently (again, provided that they are served at all). That this is true follows again from (16) which implies that $\mu_i(\alpha) = 0$ on the lower boundary $\alpha_i = a_i$ whenever $s^*(\alpha) > 0$. This has the following interesting corollary:

RESULT 5. *Suppose that types are distributed on a rectangle:* $A = \prod_1^n [a_i, b_i]$. *If the firm's problem can be solved by means of Result 4, if the cost function is separable and satisfies* $C_i'(0) > a_i$, *then (except for possible sets of measure zero) there will be pure bundling at the optimum: provided that a consumer participates at all she will (almost certainly) purchase all goods.*

PROOF. Let $s^*(\alpha)$ be the convex solution to (P) which was be solved by means of Result 4 using the field $\mu(\alpha)$. Suppose that $q_i^*(\alpha) = 0$ for some α and i. Then we will prove that (except for sets of measure zero) it follows that $q_j^*(\alpha) = 0$ for all j.

Let $\hat{\alpha}$ be such that $\hat{\alpha}_i = a_i$ and $\hat{\alpha}_k = \alpha_k$ for $k \neq i$. Thus, $\hat{\alpha}$ is the projection of α onto that part of the lower boundary normal to the i^{th} axis. Then, under the assumptions of this result it is necessarily the case that $s^*(\hat{\alpha}) = 0$. (To see this we can argue as follows: $s^*(\hat{\alpha}) > 0$ implies that $\mu_i(\hat{\alpha}) = 0$ from (16) and hence that good i is efficiently supplied to the type $\hat{\alpha}$ consumer. $C_i'(0) > a_i$ then implies that $q_i^*(\hat{\alpha}) > 0$. But s^* convex implies that $q_i^* \equiv \frac{\partial s^*}{\partial \alpha_i}$ is increasing in α_i, and hence that we must have $q_i^*(\alpha) > 0$ which is a contradiction.) This implies that $q^*(\hat{\alpha}) = 0$. Unless α happens to lie within the measure zero set of points which project onto a boundary point of the set of types with zero surplus, small changes in $\hat{\alpha}$ (within A) will also have zero demands. Set $\tilde{\alpha}$ to be any vector sufficiently near to $\hat{\alpha}$ (within A) such that $q^*(\tilde{\alpha}) = 0$. Then the fact that s^* is convex implies that $(q^*(\alpha) - q^*(\tilde{\alpha}))'(\alpha - \tilde{\alpha}) \geq 0$, which means that $q^*(\alpha)'(\hat{\alpha} - \tilde{\alpha}) \geq 0$ (since $q^*(\tilde{\alpha}) = 0$, $q_i^*(\alpha) = 0$ and $\hat{\alpha}_k = \alpha_k$ for $k \neq i$). Since we may freely vary $\tilde{\alpha}$ around $\hat{\alpha}$ in any way that does not decrease the i^{th} component, this in turn implies that $q^*(\alpha) = 0$. Therefore, except possibly for a measure zero set of consumers, there is pure bundling. □

Figure 1 shows the pattern of demand in the two–good case. Result 5 implies that there are at most two regions, R_0 and R_b, and Result 2 implies that R_0 is non–trivial.

Result 5 is in itself of limited value since it begs the question of when exactly the method of Result 4 will pick out the solution. Is there a way of finding out when this will be the case? In the single–product case, the first–order approach was valid if and only if a simple hazard rate condition was satisfied (condition (8) above). In the multiproduct case we need to know under which conditions on the joint density function $f(\alpha)$ we can find a field μ and a surplus function s^* which together satisfy equations (15), (16) and (17) of Result 4 *and* such that s^* is convex. Such an result remains to be derived, and in the examples below the strategy will be to ignore the convexity constraint and to check *ex post* that s^* is convex.

To show that the method of Result 4 is sometimes useful, and hence that Result 5 has some content, we next solve a particular example. This is followed by a complete solution for a class of cases.

Example 1A: Quadratic costs and taste parameters following the (truncated) Normal distribution (with two goods).

Suppose first that there are n goods and consumer types are distributed on $\mathbb{R}^n_+$ with density $f(\alpha) = \exp\{-\|\alpha\|^2/2\sigma^2\}$. Assume that the cost function is $C(q) = \frac{1}{2}\|q\|^2$. Because of spherical symmetry of both costs and types, we guess that the optimal surplus function s^* is also spherically symmetric.[7] Therefore, write $s^*(\alpha) = s^*(z)$ where $z \equiv \frac{1}{2}\|\alpha\|^2$. In this case $q^*(\alpha) = \alpha(s^*)'(z)$, and, for those consumers which participate, (15) implies that $\mu(\alpha) + \alpha(1-(s^*)'(z))f(z) = 0$. Equation (17) requires that $\text{div}(\mu) = f(z)$ if $s^*(z) > 0$,

[7] In fact, as the next result will make clear, it is not $C(q)$ being spherically symmetric which is relevant, but rather that the dual function to C is spherically symmetric. It so happens that in the case of quadratic costs the cost function is equal to its dual.

and these two together require that

$$f(z) + n(1-(s*)'(z))f(z) - 2z(s*)''(z)f(z) - \frac{2}{\sigma^2}z(1-(s*)'(z))f(z) = 0, \quad \text{or}$$

(19) $$1 + n(1-(s*)'(z)) - 2z(s*)''(z) - \frac{2}{\sigma^2}z(1-(s*)'(z)) = 0$$

for all $z > z_{min}$ (where a consumer participates only if $z > z_{min}$). Equation (19) is easiest to solve when $n = 2$, in which case it has the following solution (among many):

$$(s*)'(z) = 1 - \frac{\sigma^2}{2z},$$

which implies that the demand function is

(20) $$q^*(\alpha) = \alpha[\, 1 - \frac{\sigma^2}{\|\alpha\|^2}\,]$$

for all $\|\alpha\| > \sigma$. Since $s^*(z) = [\, z - \frac{\sigma^2}{2}\log(z) + \text{constant}\,]$ it is easy to verify that $s^*(z(\alpha))$ is convex in α as required. A consumer of type α participates in the market if and only if $\|\alpha\| > \sigma$.

We must next check that the boundary conditions (18) are satisfied. From (15), $\mu(\alpha) = -\alpha\frac{\sigma^2}{\|\alpha\|^2}f(z)$ for all $\|\alpha\| > \sigma$. We should extend μ to the whole of $\mathbb{R}^2_+$ in such a way that μ is continuous across the the curve $\|\alpha\| = \sigma$. The following definition of μ turns out to be satisfactory:

$$\mu(\alpha) = \begin{cases} -\alpha f(z) & \text{if } \|\alpha\| \leq \sigma \\ -\alpha\frac{\sigma^2}{\|\alpha\|^2}f(z) & \text{if } \|\alpha\| \geq \sigma\ . \end{cases}$$

With this definition μ is continuous and piecewise continuously differentiable on $\mathbb{R}^2_+$. By construction, $\text{div}(\mu) = f$ when $\|\alpha\| > \sigma$ and it may be calculated that $\text{div}(\mu) \leq f$ when

$\|\alpha\| < \sigma$.[8] Therefore, (17) is satisfied. Also, μ satisfies $\mu'\hat{\mathbf{n}} = 0$ on ∂A (here, ∂A consists simply of the non–negative axes). Taking A_i to be that part of the disk of radius i which lies in $\mathbb{R}^2_+$, then the boundary ∂A_i has two parts, the part made up of the axes and the circular arc. The integral of $\mu'\hat{\mathbf{n}}$ over any part of the axes is zero, and therefore the integral of $s^*\mu'\hat{\mathbf{n}}$ over ∂A_i is simply equal to its integral over the quarter–circle of radius i. This is:

$$\int_{\partial A_i} s^*\mu'\hat{\mathbf{n}}\, dS \qquad = \frac{\pi}{2} i[(i^2/2 - \sigma^2 \log(i) + \text{constant})\frac{\sigma^2}{i} \exp\{-i^2/2\sigma^2\}]$$

$$\longrightarrow 0 \quad \text{as } i \longrightarrow \infty \quad \text{as required.}$$

We may therefore invoke the refinement to non–compact domains of Result 4 and the optimality of the demand function given in (20) is proved. □

Figure 2 shows how demand for good 1 varies with the two parameter types. As expected, this example exhibits pure bundling.[9] In addition, demand for good 1 is an increasing function of the second type parameter, just as was the case in the usual solution to the bundling problem in the last chapter.

7. A full solution for a class of cases

The above example suggests a more general result. Suppose that the cost function C is homothetic, and so may be written as

$$\text{(21)} \qquad C(q) = c(H(q)) \text{ ,}$$

---

[8] $\text{div}(\mu) = -2f + \frac{\|\alpha\|^2}{\sigma^2} f$ if $\|\alpha\| \le \sigma$, and so $f - \text{div}(\mu) \ge 0$ if $\|\alpha\| \le \sigma$ as required.

[9] In fact this example does not quite satisfy the conditions of Result 5 since here we have $C_i'(0) = a_i = 0$, rather than $C_i'(0) > a_i$. However, a small perturbation of the example would satisfy the requirements of the result.

where $H(q)$ is an increasing convex function which has constant returns (i.e. is homogeneous of degree 1), and $c(H)$ is an increasing convex function (of a single variable). For instance, in the Normal example above $H(q) = \|q\|$ and $c(H) = \frac{1}{2}H^2$.

Let $h(\alpha)$ to be the dual function to $H(q)$:

$$(22) \qquad h(\alpha) = \max: \{ \alpha' q \mid H(q) \leq 1 , q \geq 0 \} .$$

Then h is also increasing, convex and exhibits constant returns. In the case where $H(q) = \|q\|$ the dual may be calculated to be $h(\alpha) = \|\alpha\|$.[10]

The density function $f(\alpha)$ may without loss of generality be written as $f(h(\alpha), \alpha/h(\alpha))$. This is simply a change of variables and gives the density of α given the ray α is on from the origin — described by the $(n-1)$–dimensional parameter $\alpha/h(\alpha)$ — and the distance of α from the origin as measured by the norm $h(\alpha)$. In the following result we will assume that under this change of variables f is multiplicatively separable:

$$(23) \qquad f(\alpha) = f_1(h(\alpha))f_2(\alpha/h(\alpha)).$$

This assumption, therefore, implies that the two random variables $h(\alpha)$ and $\alpha/h(\alpha)$ are independently distributed. For example, with two goods and quadratic costs — so that $h(\alpha) = \frac{1}{2}\|\alpha\|^2$ — a change to polar coordinates (r, θ) results in the two variables r and θ being independent.[11] In the Normal example above we could write f in the form $f(\alpha) = f_1(h(\alpha))$, and so (23) was trivially satisfied.

In the proof that follows we will also need to make use of another condition on the density function f (but see the remark following Result 6):

---

[10] More generally, if $C(q) = \Sigma q_i^\gamma$ for $\gamma > 1$, then $H(q) = \{ \Sigma q_i^\gamma \}^{1/\gamma}$, $c(H) = H^\gamma$ and $h(\alpha) \propto \{ \Sigma \alpha_i^\delta \}^{1/\delta}$, where δ is given by $\frac{1}{\delta} + \frac{1}{\gamma} = 1$.

[11] Polar coordinates are an alternative to the Cartesian system in two dimensions whereby the location of a point is measured by its distance from the origin, r, and the angle the joining the point to the origin makes with the 'x–axis', θ. Using the notation of (23), the variable $\alpha/h(\alpha) = \alpha/\|\alpha\|$ is in a 1–1 relationship with the coordinate θ, and therefore in two dimensions condition (23) may be re–expressed as requiring that the two variables $r(\alpha)$ and $\theta(\alpha)$ are independently distributed.

(24) $$(n+1)f(\alpha) + \alpha'\nabla f(\alpha) \geq 0 \quad \text{for all } \alpha \text{ in } A.$$[12]

Finally, in order to ensure that the optimal surplus function s^* is convex we will need an assumption on $f_1(h)$ analogous to the single–product condition (8):

(25) the function $h \mapsto h - \left(\int_h^\infty \tilde{h}^{n-1} f_1(\tilde{h})\, d\tilde{h} \right) / (h^{n-1} f_1(h))$ is non–decreasing

(where n is the number of goods).[13]

RESULT 6. *Suppose that the cost function may be written as in (21), that types are distributed on the whole of the non–negative orthant, that the density function may be written as in (23), that conditions (24) and (25) hold, that* $\frac{\partial H}{\partial q_i}(q) = 0$ *if* $q_i = 0$, *and that the first–best profit for the firm is finite. Then optimal surplus may be written simply as* $s^*(h(\alpha))$, *where* $(s^*)'(h)$ *is given by:*

(26) $$(s^*)'(h) \text{ maximizes}_{q \geq 0} \; (hq - c(q))h^{n-1}f_1(h) - q\int_h^\infty \tilde{h}^{n-1} f_1(\tilde{h})\, d\tilde{h}.$$

Optimal demands are then given by

(27) $$q^*(\alpha) = \nabla h(\alpha)(s^*)'(h(\alpha)) .$$

REMARK: Equation (26) is easily obtained once we have decided that s^* is going to be a function of $h(\alpha)$, in which case the problem reduces to a one–dimensional problem in h. Assumption (23) means that the density of h is simply $h^{n-1}f_1(h)$ (up to a multiplicative

[12] It is interesting that McAfee and McMillan (1988) also make use of this assumption on f in a related context — (24) is their condition (46).

[13] This is the version suitable for the case where f has support equal to the whole positive orthant. If f had compact support, we must change the upper limit on the integral to be h_{max} say. The whole of the following theorem may be modified to deal with this case in a similar fashion.

constant), so that (25) just the corresponding version of (8) above and (26) corresponds to equation (12) in Chapter 1. The content of the proof comes in showing that it *is* optimal to write s^* as a function of h.

PROOF. See Appendix.

REMARK. Condition (24) is very strong and in fact was required only to hold in the region $h(\alpha) \leq h_{min}$ in order for the proof to work (the condition was needed so that the extension of μ to the region $h(\alpha) \leq h_{min}$ was straightforward). For instance, in the two–good Normal example above (24) certainly did not hold everywhere but was satisfied for $||\alpha|| < \sigma$. In the other examples below a similar situation will occur.

Result 6 is most naturally used for the case where $A = \mathbb{R}^n_+$. If A is compact, being able to write the density of types as in (23) requires that the upper boundary of A must have the rather artificial form $h(\alpha) = \text{constant}$, and in particular, A being a rectangle will never satisfy the conditions of the Result.

We next list two corollaries to this result:

COROLLARY 1. Under the conditions of Result 6 the optimal tariff T^* is a function only of $H(q)$. In other words, the tariff is simply a function of the cost $C(\mathbf{q})$ of providing the bundle $\mathbf{q}$. This will not generally be true. T^* can be recovered from s^* by:

$$T^*(H(q)) = \max: \{ xH(q) - s^*(x) \mid x \geq 0 \}.$$

For instance, the tariff which gives rise to the optimal demands in Example 1A may be calculated to be

$$T^*(\mathbf{q}) = \tfrac{1}{4}(||\mathbf{q}|| + \sqrt{(||\mathbf{q}||^2 + 4\sigma^2)}) + \sigma^2\log(||\mathbf{q}|| + \sqrt{(||\mathbf{q}||^2 + 4\sigma^2)}) + \text{constant}$$

which is indeed just a function of $||\mathbf{q}||$. Returning to our original formulation of utility in terms of $\mathbf{x}$ rather than $\mathbf{q}$ in (1), the assumption of quadratic costs is equivalent to writing utility as $\Sigma\, \alpha_i\sqrt{x_i} + y$. In this case Corollary 1 has the attractive implication that the optimal tariff $T^*(\mathbf{x})$ is a function only of the sum of quantities, $\Sigma\, x_i$. Figure 3 shows optimal demand for good 1 for the Normal example when output is taken to be $\mathbf{x}$ rather than $\mathbf{q}$.

COROLLARY 2. Equation (27) implies that (for $i \neq j$)

$$\frac{\partial q_i^*}{\partial \alpha_j} = h_{ij}(s^*)'(h) + h_i h_j (s^*)''(h)$$

and so demand for one good certainly increases with the type parameters of the other goods if $h_{ij} \geq 0$, which in turn holds if $H_{ij} \leq 0$. Therefore, if there are weak cost complementarities in production, it will certainly be optimal to have demand being an increasing function of all type parameters.

8. Solved examples

In each of the following examples the cost function is taken to be quadratic: $C(\mathbf{q}) = \frac{1}{2}\|\mathbf{q}\|^2$. Therefore, $H(\mathbf{q}) = \|\mathbf{q}\|$ and $h(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|$. As required by Result 6, $\frac{\partial H}{\partial q_i}(\mathbf{q}) = 0$ if $q_i = 0$.

Example 1B: the (truncated) Normal distribution with n goods.

As with Example 1A, the density function is $f(\boldsymbol{\alpha}) = \exp\{-\|\boldsymbol{\alpha}\|^2/2\sigma^2\}$. For $n = 1, 2, \ldots$ let the function $I_n(h)$ be given by

$$I_n(h) = \int_h^\infty \tilde{h}^{n-1}\exp\{-\tilde{h}^2/2\sigma^2\}\, d\tilde{h}.$$

Then Result 6 implies that the optimal surplus s^* may be written as function of $h = \|\boldsymbol{\alpha}\|$ and satisfies (for $s^*(h) > 0$):

$$(s^*)'(h) = h - I_n(h)/(h^{n-1}\exp\{-h^2/2\sigma^2\}).$$

Integrating by parts gives (for $n > 2$) the difference equation:

$$I_n(h) = \sigma^2 h^{n-2}\exp\{-h^2/2\sigma^2\} + (n-2)\sigma^2 I_{n-2}(h).$$

Starting values are $I_2(h) = \sigma^2\exp\{-h^2/2\sigma^2\}$, and $I_1(h)$ which has no closed–form expression. The form of the optimal demands then depends crucially upon whether there is an even or an odd number of goods. If n is even then equation (27) means that

$$q^*(\alpha) = \alpha[\,1 - \frac{k_2}{\|\alpha\|^2} \cdots - \frac{k_n}{\|\alpha\|^n}\,]$$

where $k_2, k_4, .., k_n$ are some positive constants. On the other hand, if n is odd then

$$q^*(\alpha) = \alpha[1 - \frac{k_2}{\|\alpha\|^2} \cdots - \frac{k_{n-1}}{\|\alpha\|^{n-1}} - \frac{k_n I_1(\|\alpha\|)\exp\{+\|\alpha\|^2/2\sigma^2\}}{\|\alpha\|^n}\,].$$

When n is even then condition (25) certainly holds, whereas if n is odd this is less clear. Since it is hard to explicitly calculate h_{min} for each n for this example, it is also hard to verify that (24) holds for $h < h_{min}$. □

Example 2: the (truncated) t distribution with n goods.

The multivariate t distribution with d degrees of freedom is defined by taking the quotient of a multivariate Normal distribution and a Chi–squared distribution with d degrees of freedom. When the Normal variables are identical and independent with variance σ^2, the density function for this t distribution has the form:

$$f(\alpha) = [1 + \|\alpha\|^2/(d\sigma^2)]^{-\frac{1}{2}(n+d)}\,. \tag{28}$$

This again is spherically symmetric (and so satisfies (23) taking $f_2 \equiv 1$), and Result 6 applies. In order to ensure that the first–best profit is bounded we require that second moments exist, which in turn requires that $d \geq 3$.

For $n = 1, 2, \ldots$ define the function

(29) $$J_n(h) = \int_h^\infty \tilde{h}^{n-1}[1 + \tilde{h}^2/(d\sigma^2)]^{-\frac{1}{2}(n+d)}\, d\tilde{h} .$$

Then (26) implies that

$$(s^*)'(h) = h - J_n(h)/\{h^{n-1}[1 + h^2/(d\sigma^2)]^{-\frac{1}{2}(n+d)}\}.$$

Integrating by parts gives the following difference equation in n (for $n > 2$)

$$J_n(h) = \frac{d\sigma^2 h^{n-2}}{(d+n-2)}[1 + h^2/(d\sigma^2)]^{-\frac{1}{2}(n+d-2)} + \frac{d\sigma^2 (n-2)}{(d+n-2)} J_{n-2}(h) .$$

When n is even this sequence terminates with

$$J_2(h) = \sigma^2(1 + h^2/(d\sigma^2))^{-\frac{1}{2}d}$$

and so equation (27) implies that in this case demands are

$$q^*(\alpha) = \alpha[\, 1 - \frac{1}{d+n-2} - \frac{\tilde{k}_2}{\|\alpha\|^2} \cdots - \frac{\tilde{k}_n}{\|\alpha\|^n}\,]$$

for some constants $\tilde{k}_i$. In particular, when $n = 2$

(30) $$q^*(\alpha) = \alpha[\, 1 - \frac{1}{d} - \frac{\sigma^2}{\|\alpha\|^2}\,] .$$

A consumer participates if and only if $\|\alpha\|^2 \geq \sigma^2/(1 - 1/d)$. This demand function is superficially rather similar to that found for Example 1A in equation (20), but has the important difference that outputs are not asymptotically efficiently supplied as α gets large.[14] However, since

$$\lim_{d\to\infty}: [1 + \|\alpha\|^2/(d\sigma^2)]^{-\frac{1}{2}(n+d)} \equiv \exp\{-\|\alpha\|^2/2\sigma^2\}$$

we see that the t distribution approximates the Normal when the number of degrees of freedom is large, and so it is no surprise that demands in (30) approximate (20) for large d.

On the other hand, when there are an odd number of goods the solution will once again look very different. In order to solve this case, we need to find the function J_1. Making the substitution $x = \sqrt{(d\sigma^2)}\tan\theta$ in (26) gives

$$J_1(h) = \sqrt{(d\sigma^2)} \int_{\tan^{-1}(h/\sqrt{(d\sigma^2)})}^{\pi/2} \cos^{d-1}(\theta)\, d\theta .$$

For simplicity, set $d = 3$ and $\sigma^2 = 1/3$. Then it may be calculated that

$$J_1(h) = \tfrac{1}{2}\{ \tfrac{\pi}{2} - \frac{h}{1 + h^2} - \tan^{-1}(h) \}.$$

Therefore, equation (27) implies that when $n = 3$

---

14 When A is compact it is necessary that the high demand consumers are served efficiently. For instance, in one dimension if $\bar{\alpha}$ is the maximum type then equation (9) implies that $C'(q(\bar{\alpha})) = \bar{\alpha}$ and so there is efficient production for these types. But if α takes any non-negative value there is no reason why demands should eventually be efficient since $(1 - F)/f$ need not get small for large α (e.g. if $f(\alpha) = \exp(-x)$ then $(1 - F)/f \equiv 1$).

$$q^*(\alpha) = \alpha[\frac{\tilde{k}_{-2}}{\|\alpha\|^2} + \frac{\tilde{k}_{-1}}{\|\alpha\|} + \tilde{k}_0 + \tilde{k}_1\|\alpha\| + \tilde{k}_2\|\alpha\|^2 + \tilde{k}_3\|\alpha\|^3 + \tilde{k}_4\|\alpha\|^4$$

$$+ \frac{(1+\|\alpha\|^2)^3}{8\|\alpha\|^2}\tan^{-1}(\|\alpha\|)]$$

for some constants $\tilde{k}_i$.

Condition (25) is again satisfied when n is even, but it is hard to determine in general whether it is satisfied for all odd n. On the other hand, condition (24) is satisfied for all α when when f is as defined in (28) for each n and d. □

The final example uses a density function which satisfies condition (23) less trivially in the sense that the density of types depends not only upon the distance from the origin.

<u>Example 3</u>: The multivariate Weibull distribution.

The multivariate density function of n identical and independent Weibull variables with parameters $\sigma^2 > 0$ and 'b' = 2 is given by

(31) $$f(\alpha) = (\prod_1^n \alpha_i)\exp\{-\|\alpha\|^2/2\sigma^2\}\ .$$

This may be written in the form of (23) by taking $f_1(h) = h^n\exp\{-h^2/2\sigma^2\}$ and $f_2(\alpha/h(\alpha)) = \prod_1^n (\alpha_i/h)$. Therefore, Result 6 implies that

$$(s_*)'(h) = h - I_{2n}(h)/(h^{2n-1}\exp\{-h^2/2\sigma^2\})$$

where I_{2n} is as defined in Example 1B. Since 2n is always even, this means that demands always have a comparatively simple form:

$$q^*(\alpha) = \alpha[\ 1 - \frac{k_2}{\|\alpha\|^2} \cdots - \frac{k_{2n}}{\|\alpha\|^{2n}}\ ]$$

for some positive constants k_i. Thus the n–dimensional Weibull distribution results in a similar demand function to the Normal distribution with an even number of goods, but with twice as many terms. In particular, in the case of two goods optimal surplus is given by

$$(32) \qquad s^*(\alpha) = \tfrac{1}{2}\|\alpha\|^2 - \sigma^2\log(\|\alpha\|) + \frac{\sigma^4}{\|\alpha\|^2} + \sigma^2\log(\sigma\sqrt{2}) - \tfrac{3}{2}\sigma^2$$

with demands of

$$q^*(\alpha) = \alpha[\,1 - \frac{\sigma^2}{\|\alpha\|^2} - 2\frac{\sigma^4}{\|\alpha\|^4}\,]$$

and a consumer participates if and only if $\|\alpha\|^2 > 2\sigma^2$. For this example, condition (25) is certainly satisfied for each n (whether even or odd). Since it is hard to calculate h_{min} for each n, whether condition (24) is valid for $h < h_{min}$ for each n is unclear, although it is valid in the above two–good case.

Unlike the two previous examples, in this case we can explicitly compute the optimal additively separable tariff as discussed in Section 5 above. The optimal single–product surplus function s_1^* may be calculated to be

$$(33) \qquad s_1^*(\alpha) = \tfrac{1}{2}\alpha^2 - \sigma^2\log(\alpha) - \tfrac{1}{2}\sigma^2 + \sigma^2\log(\sigma)$$

for all $\alpha > \sigma$ (and $s_1^* = 0$ otherwise). Figure 4 shows which types of consumers are better off with the fully optimal tariff than with under the optimal separable tariff by plotting $s^*(\alpha) - s_1^*(\alpha_1) - s_1^*(\alpha_2)$, where s^* is given by (32) and s_1^* by (33). We see that those consumers with particularly asymmetric tastes (low α_i with high α_j, $i \neq j$) fare better with the separable tariff, whereas those with high preferences for both goods prefer the bundled tariff. This accords with intuition. It would be an interesting topic of future research to see how widespread is this phenomenon. □

Result 6 has provided only a partial solution to the problem of finding the optimal multiproduct tariff, and we can speculate as to how further progress might be made when the density of types could not be written in the separable form of (23). In such cases the surplus function given in (26) can no longer be optimal, since it no longer satisfies the combination of equations (15) and (17). It is likely that the change of coordinates from α to $(h(\alpha), \alpha/h(\alpha))$ will remain a good one even in a more general setting. For instance, in the two good case with quadratic costs, changing to polar coordinates (r, θ) means that given the general density function $f(r, \theta)$, the optimal surplus function $s^*(r, \theta)$ must satisfy the partial differential equation:

$$(34) \qquad (fs^*_\theta)_\theta + r(rfs^*_r)_r = (r^3f)_r$$

for those consumers of type (r, θ) who participate (we have used subscripts for partial derivatives). (This is obtained from equations (15) and (17).) Result 6 demonstrates that when f is multiplicatively separable in r and θ (34) has a particularly simple solution given by s^* being a function only of r. For more general functional forms of f this is harder to solve, but it may be possible for certain particular cases.

As before, this change of coordinates is only likely to be useful when type are distributed on the whole positive orthant. If, instead of this, types are distributed on a rectangle, I predict that solutions will invariably be particularly complex, largely because the upper boundary conditions in (16) are especially troubling. Moreover, many may feel the effort involved in finding the solution in such cases to be out of balance with the return, especially considering the ease with which approximate solutions can be obtained with the help of computers.

9. Numerical simulations

While we have managed to make some progress in solving the problem, for a general density function $f(\alpha)$ which does not satisfy the conditions of Result 6 we still have no solution. Indeed, I believe that a general solution to problem (P) will be hard to achieve. And if some more explicit solutions are found, then because of the rather complex manner in which demands and types interact, the solutions may be too complicated to be manageable. I have therefore resorted to numerical analysis to find some other, approximate, solutions. Using these solutions I will be able to test how general are the two main observations so far, namely, that pure bundling is a widespread phenomenon, and that demands are an increasing function of all type parameters. Moreover, we will see whether our strategy in Result 4 of solving the problem by ignoring the constraint that surplus be convex is usually valid.

The programming approach I use goes as follows. I consider distributions of types defined on the unit square (so that there are two goods) and suppose that the firm has quadratic costs: $C(q) = \frac{1}{2}\|q\|^2$. I approximated continuous distributions by a fine discrete distribution; in the examples below I used a regular 30×30 mesh, indexing types by the integers (i, j), $1 \leq i, j \leq 30$. A 'density' function was then a weight $w(i, j)$ given to each of these points. Surplus $s(i, j)$ was assigned to each of these points, subject to the proviso that surplus be (weakly) increasing in (i, j) but *not* so that s was necessarily convex. Therefore, I try to solve problem (P^{fo}) rather than (P). Given the array $s(i, j)$, demand for the first good was taken to be $30[s(i+1, j) - s(i, j)]$ in the neighborhood if (i, j) (and similarly for good 2), and then total profit was calculated by the formula (P). The only subtlety comes when ensuring that s is non–decreasing. The method by which this was done was to define an initial array $x(i, j)$ and then define s inductively in terms of x by the iterative formula:

$$(35) \qquad s(i, j) = x(i, j) + \max: \{ s(i-1, j), s(i, j-1) \}$$

for $i, j > 1$, with boundary conditions of $s(i, 1) = x(i, 1) + s(i-1, 1)$, $s(1, j) = x(1, j) + s(1, j-1)$ and finally, $s(1, 1) = x(1, 1)$.

Using this procedure, s is non–decreasing and non–negative if and only if x is non–negative. Thus, profit was defined as a function of the array x (via s), and a routine was employed to maximize this function subject to the condition that x be non–negative. This latter constraint is much easier to incorporate than a non–decreasing constraint. This method is very reliable but extravagant in computing time. When using a 30×30 grid, the problem reduces to maximizing a function of 900 variables (or about

450 variables if the problem is symmetric), and takes about an hour of processing time on Oxford University's CONVEX vectorizing computer to get reliable and accurate results.[15]

Two runs are included here, one where types are uniformly distributed and one where types have the density function $f(x, y) = xy$. Demand for good 1 in the uniform case is shown in Figures 5A and 5B. Figure 5A gives a three dimensional picture of how demand varies with the two type parameters, while 5B shows how demand varies with the first type parameter for some fixed values of the other parameter. As is required by efficiency, demand for good 1 tends to 1 as the first parameter tends to 1. We can contrast this with the optimal demand function in the single–product case which is given by $q(x) = 2x - 1$ for $x > \frac{1}{2}$, zero otherwise. Thus the full nonlinear tariff will induce demands which are higher than the separable tariff for consumers with high tastes for the second good, but lower than the separable tariff otherwise. We can also see that demand for good 1 increases with the second type parameter, that there is pure bundling and I also checked that the resulting surplus function was convex.

When the density was given by the function $f(x, y) = xy$, the results were broadly similar except for the fact that demands started later and were steeper. Figures 6A and 6B are the corresponding diagrams.[16] In this case, the optimal separable demand function is given by $q(x) = \frac{3}{2}x - 1/2x$ for $x^2 > 1/3$, and zero otherwise. Again, we have pure bundling, demands are increasing in all parameters and the surplus function is convex.

I tried a few other density functions, and all looked broadly the same. It seems that the picture of demand given in any of the three dimensional diagrams is the 'typical' one, that pure bundling appears to be widespread, that demand is usually an increasing function of all type parameters, and that the convexity constraint is not often binding.[17] The fact that demand increases in all parameters at the optimum corresponds to an optimal tariff in which marginal price of one good is a decreasing function of the quantities of the other

---

15 The method would be quite inappropriate for solving the problem with three or more goods due to the inordinate amount of time required.

16 The reason why the demand function is slightly badly behaved for some very low values of the taste parameters is because the density vanishes on the lower boundary.

17 If the convexity constraint does turn out to be a problem, it is possible to incorporate this constraint directly by means of a more complex 'second–order' version of equation (35).

goods. Thus, there will be subadditive bundling, in the sense that it will be cheaper to buy a package of goods rather than the goods separately.

10. Conclusions

This has been a long and rather difficult chapter. However, some useful steps have been made and our intuition has been helped along a little. I would single out the apparent frequency of pure bundling being a dominant strategy for the firm as the most striking result, and a result which represents a significant departure from the previous chapter on bundling. It demonstrates that this topic, for all its mathematical complexity, may be governed by some simple principles.

The assumption that utility in (1) was linear in types was absolutely crucial to the analysis. Without it, we would have had little idea which surplus functions would have been implementable by the firm. Once the linearity assumption was made we immediately changed variables so that the firm 'manufactured' utility $\mathbf{q}$ rather than the actual good $\mathbf{x}$. This meant that consumer utility was linear in both type and quantity, an assumption that facilitated much subsequent analysis. We formulated the firm's problem and showed that its solution was unique (Result 1). In a wide variety of situations the firm will find it optimal to exclude certain low–demand consumers from the market (Result 2), a result which is a distinctive feature of all of these multiproduct models. By analogy to Result 2 in the bundling chapter, I looked for what kinds of deformations to the optimal additively separable tariff would increase profits when the two taste parameters were independently distributed. It was found that very many deformations would increase profits, some of which would also be Pareto improving (Result 3). In particular, with independent types a separable tariff could never be optimal.

Section 6 was concerned with finding a method to actually solve the problem. The proposed method involved following that used in the single–product case after making use of a multidimensional version of 'integration by parts'. However, the resulting first–order conditions lose the simplicity of the single–product case since the optimal demand function depends upon the vector function $\boldsymbol{\mu}$, which acts in place of $-(1-F)$ as the integral of the density f. Since $\boldsymbol{\mu}$ is not explicitly determined we can no longer obtain the general explicit solution to the problem which was possible in the former case. The method was summarized in Result 4. Whenever this method worked we saw that, not only were the high demand consumers efficiently served (as was the case in the single–product case), but so were the low demand consumers (if they were served at all). This observation led to Result 5: a consumer either purchases all goods or nothing at all. Of course, we do not know how much weight to place on this result since we have little idea of how often the method of Result 4 will work. But it remains an attractive hypothesis that pure bundling will prove to be a characteristic feature of problems of this type.

We next found a class of cases which could be solved (Result 6). These involved an additional assumption that the cost function of the firm was homothetic and that types were distributed on the whole non–negative orthant. A change of variables was proposed which, in the two–good case with quadratic costs was equivalent to a change to polar coordinates. Provided that the density function of types was such that these two new variables were independently distributed, the problem could be solved essentially using the single–product method. In all of these cases the firm offers a tariff $T^*(\mathbf{q})$ which is a function only of its cost $C(\mathbf{q})$ of producing the bundle $\mathbf{q}$. This enabled us to find some examples, namely, when tastes were distributed Normally, as a t distribution and as a Weibull distribution, and these are perhaps the first of their kind in the literature. In the two–product case, the optimal demand functions proved to be rather simple; there was pure bundling and demand was increasing in all parameters. However, in the case of the Normal and the t distribution, when there were more goods the solution depended in a surprising manner upon whether the number of goods was even or odd, the case of an odd number proving particularly complicated. In terms of the original outputs $\mathbf{x}$, the assumption of quadratic costs which we made when calculating all of the examples is equivalent to consumers having a utility function in (1) of $\Sigma\, \alpha_i\sqrt{x_i} + y$. When the distribution of $\boldsymbol{\alpha}$ satisfied the conditions of Result 6, this choice of utility function implied that the profit–maximizing tariff was a function only of the sum of quantities demanded, $\Sigma\, x_i$. In such cases, then, allowing the firm to offer a full nonlinear tariff $T(\mathbf{q})$ need not result in very complex schedules.

Unfortunately, the conditions needed to invoke Result 6 are very restrictive, and in particular the result is only suited to the case where taste parameters had support consisting of the whole positive orthant. I think that extending this result to include a wider class of density function will prove to be difficult, and that computer simulation is likely to be the most efficient means to gain further insights. In this we will be following the honorable tradition of particle physics, where the many partial differential equations encountered are almost invariably solved by computer (the few cases which may be analytically solved again are one–dimensional). The problem considered in this chapter is easily solved on a (large) computer, and some simulations were described in Section 9. In the examples there was pure bundling, both high and low types were efficiently served, that demands increased with both taste parameters and the constraint that the surplus function needed to be convex was not binding. For the time being, then, we can take these to make up the typical characteristics of the solution.

Appendix.

Proof of Result 2. *By contradiction*: Let $s^*(\cdot)$ be the optimal surplus function with associated tariff schedule $T^*(\cdot)$ and suppose that all consumers participate in the market. Then the proof follows the method of the proof of Result 1 in Chapter 3, and in particular footnote 17 of that chapter is relevant to this case. Define

$$A(\epsilon) = \{ \alpha \mid \alpha \in A \text{ and } s^*(\alpha) \leq \epsilon \} .$$

Then $s^*(\cdot)$ convex and continuous implies that each $A(\epsilon)$ is compact and convex. If $T^*(\cdot)$ is optimal then $A(0)$ must be non–empty (otherwise the firm would benefit by increasing the tariff at every point by some fixed amount). The continuity of $s^*(\cdot)$ implies that $A(\cdot)$ is a continuous family of sets so that $A(\epsilon) \longrightarrow A(0)$ as $\epsilon \longrightarrow 0$.

We next show that $A(0) = \{\mathbf{a}\}$ (this is the only extra difficulty compared with the bundling case). Since A is symmetric, $\mathbf{a}$, the unique minimizer of $\Sigma\, \alpha_i$ on A, is a uniform vector: $\mathbf{a} = (a, \dots, a)$ for some scalar a. By symmetry s^* is a symmetric function and so $A(0)$ is a symmetric, convex, non–empty set. Therefore, $A(0)$ must contain a uniform vector $(t, \dots, t)$ in A, where $t \geq a$ from the definition of $\mathbf{a}$. But since s^* in an increasing function this implies $\mathbf{a} \in A(0)$. In addition, $\mathbf{a}$ is the only uniform vector in $A(0)$, because if $\mathbf{t} = (t, \dots, t) \in A(0)$ where $t > a$ then $A(0)$ also contains the line segment joining $\mathbf{a}$ to $\mathbf{t}$. This line segment lies in the interior of A (since A is of full dimension in $\mathbb{R}^n$) and is then also in the interior of $A(0)$ since s^* is increasing. But $A(0)$ is assumed to have measure zero and so cannot have an interior. Therefore, $\mathbf{a}$ is the only uniform vector in $A(0)$. Now suppose that $A(0)$ contains a second point $\hat{\alpha}$. By symmetry $A(0)$ contains all the 2^n points obtained by permuting the coordinates of $\hat{\alpha}$. All of these points lie on the same hyperplane $\Sigma\, \alpha_i = \Sigma\, \hat{\alpha}_i$ which, from the definition of $\mathbf{a}$, satisfies $\sum_i \hat{\alpha}_i > \sum_i a_i$. However, since $A(0)$ is convex it also contains the point $(\sum_k \hat{\alpha}_k)/2^n$, where $\{\hat{\alpha}_k\}$ is the collection of the 2^n vectors consisting of the permutations of $\hat{\alpha}$. But by construction

$$(\sum_k \hat{\alpha}_k)/2^n = (\sum_i \hat{\alpha}_i)(1/n, \dots, 1/n).$$

Since $A(0)$ contains no uniform vectors other than a we see that $A(0) = \{a\}$ as claimed.

Let $S(A(\epsilon))$ denote the surface area of the set $A(\epsilon)$. Let $g(t)$ and $G(t)$ be respectively the density and distribution of the scalar variable $s^*(\alpha)$ induced by the density function $f(\alpha)$ on A. Just as in the proof of Result 1 of Chapter 3, if B is the bound on f we must have

(A1) $$g(\epsilon) \leq B \times S(A(\epsilon))$$

and therefore

(A2) $$G(\epsilon) \leq \epsilon B \times S(A(\epsilon)) .$$

Let the firm modify the tariff to be $T^*(\cdot) + \epsilon$, where $\epsilon > 0$. Then number of consumers who leave the market as a result of this change is $G(\epsilon)$ and all others pay an extra ϵ. The change in profit, then, satisfies

$$\Delta\pi \quad \geq \epsilon(1 - G(\epsilon)) - M \times G(\epsilon)$$

where M is the maximum profit the firm made from any consumer in the set $A(\epsilon)$ with tariff T^* (so M is finite).

(A3) $$\Delta\pi \quad \geq \epsilon(1 - G(\epsilon)) - \epsilon MB \times S(A(\epsilon)) \qquad \text{from (A2)} .$$

But since $S(\cdot)$ is a continuous function the fact that $A(\epsilon) \longrightarrow \{a\}$ as $\epsilon \longrightarrow 0$ implies that $S(A(\epsilon) \longrightarrow S(\{a\})$ as $\epsilon \longrightarrow 0$. Therefore, unless $n = 1$

(A4) $$S(A(\epsilon)) \longrightarrow 0 \text{ as } \epsilon \longrightarrow 0 .$$

Finally, provided that $n > 1$ the combination of (A3) and (A4) implies that $\Delta\pi > 0$ for all sufficiently small ϵ which contradicts the assumed optimality of T^*. □

Proof of Result 6. The proof mirrors the method used in constructing the Normal example of Section 6. Write surplus as $s^*(h(\alpha))$, with the corresponding demand function is $q^*(\alpha) = (s^*)'(h)\nabla h(\alpha)$. If h_{min} is the smallest h which participates, then for $h > h_{min}$ equation (26) implies

$$\mu(\alpha) + [\alpha - \nabla_q C(q^*(\alpha))]f_1(h(\alpha))f_2(\alpha/h(\alpha)) = 0.$$

But $\nabla_q C(q^*(\alpha)) = c'(H(q^*(\alpha)))\nabla H(q^*(\alpha))$

$$= c'(H((s^*)'(h)\nabla h(\alpha)))\nabla H((s^*)'(h)\nabla h(\alpha))$$

$$= c'((s^*)'(h)H(\nabla h(\alpha)))\nabla H(\nabla h(\alpha))$$ from the homogeneity of H

$$= c'((s^*)'(h))\frac{\alpha}{h(\alpha)},$$

the final equality following from the relation between h and H given in (22).[18] Therefore, μ must satisfy

$$\mu(\alpha) + \alpha[1 - \frac{c'((s^*)'(h))}{h(\alpha)}]f_1(h(\alpha))f_2(\alpha/h(\alpha)) = 0.$$

---

[18] Briefly, since $\nabla h(\alpha)$ is the optimal choice for q in (22) — which, in this footnote only, we denote by $q(\alpha)$ — non-satiation implies that $H(\nabla h(\alpha)) \equiv 1$. Similarly, $\nabla H(\nabla h(\alpha)) = \nabla H(q(\alpha)) = k\alpha$ for some k from the first-order conditions for $q(\alpha)$ in problem (22). But since homogeneity implies that $q'\nabla H(q) \equiv H(q)$, and $\alpha' q(\alpha) \equiv h(\alpha)$, we see that

$$1 = H(q(\alpha)) = q(\alpha)'\nabla H(q(\alpha)) = k\alpha' q(\alpha) = kh(\alpha)$$

and so $k = 1/h(\alpha)$. Therefore $\nabla H(\nabla h(\alpha)) \equiv \alpha/h(\alpha)$ as claimed.

Since (17) requires that $\operatorname{div}(\mu) = f(h)$ for $h > h_{min}$, the following differential equation in h should hold (after dividing through by f_1f_2):

(A5)
$$1 + n[1 - \frac{c'((s*)'(h))}{h}] + h[1 - \frac{c'((s*)'(h))}{h}]\frac{f_1'(h)}{f_1(h)} - c''((s*)'(h))(s*)''(h) + \frac{c'((s*)'(h))}{h} = 0$$

whenever $h > h_{min}$. (To obtain (A5) we used the fact that $\alpha'\nabla f_2(\alpha/h(\alpha)) \equiv 0$ since $f_2(\alpha/h(\alpha))$ is homogeneous of degree zero in α.) It may simply be calculated that $(s*)'(h)$ as given in (26) does indeed satisfy this equation (A5). h_{min} is given by the largest value of h in (26) which results in $(s*)'(h) = 0$. Since $h(\alpha)$ is convex, provided that (26) results in $(s*)'(h)$ being an increasing function of h — and condition (25) is exactly what is required to ensure this — the function $s*(h(\alpha))$ is convex in α. Optimal demands are given by

(A6)
$$q^*(\alpha) = \nabla h(\alpha)(s*)'(h(\alpha)).$$

The remainder of the proof follows the argument in Example 1A. The field μ is defined to be:

$$\mu(\alpha) = \begin{cases} -\alpha f_1(h(\alpha))f_2(\alpha/h(\alpha)) & \text{if } h(\alpha) \leq h_{min} \\ -\alpha[1 - \frac{c'((s*)'(h))}{h(\alpha)}]f_1(h(\alpha))f_2(\alpha/h(\alpha)) & \text{if } h(\alpha) \geq h_{min}. \end{cases}$$

With this definition μ is continuous and piecewise continuously differentiable on $\mathbb{R}^n_+$. By construction, $\operatorname{div}(\mu) = f$ when $h > h_{min}$, and the fact that $\operatorname{div}(\mu) \leq f$ for $h < h_{min}$ follows from assumption (24). Therefore, (17) is satisfied. Also, since $\frac{\partial h}{\partial \alpha_i}(\alpha) = 0$ whenever $\alpha_i = 0$ (something which follows from the assumption that $\frac{\partial H}{\partial q_i}(q) = 0$ whenever $q_i = 0$), μ satisfies $\mu'\hat{n} = 0$ on ∂A (where ∂A consists of the non–negative axes). Let A_i be defined by

$$A_i = \{ \alpha \mid h(\alpha) \leq i, \;\; \alpha \geq 0 \}.$$

The integral of $\mu'\hat{n}$ over any part of the axes is zero, and the integral of $s^*\mu'\hat{n}$ over ∂A_i is simply equal to its integral over the 'curved section', by which we mean

$$S_i = \{ \alpha \mid h(\alpha) = i, \;\; \alpha \geq 0 \} .$$

A point α on S_i has an outward–pointing unit normal vector equal to $\frac{\nabla h(\alpha)}{\|\nabla h(\alpha)\|}$. Therefore,

$$\int_{\partial A_i} s^*\mu'\hat{n}\, dS \;\; = \;\; \int_{S_i} -s^*(h)[h - c'((s^*)'(h))] \frac{f_1(h)}{\|\nabla h(\alpha)\|}\, dS$$

$$= k i^{n-1} s^*(i)[i - c'((s^*)'(i))] f_1(i)$$

(where we have used the fact that $h(\alpha)$ being homogeneous degree one implies that $\nabla h(\alpha)$ is homogeneous of degree zero, and so

$$\int_{S_i} \frac{1}{\|\nabla h(\alpha)\|}\, dS = i^{n-1} \times \int_{S_1} \frac{1}{\|\nabla h(\alpha)\|}\, dS .$$

Writing $k = \int_{S_1} \frac{1}{\|\nabla h(\alpha)\|}\, dS$ gives the above equality). Therefore, (26) implies that

$$\int_{\partial A_i} s^*\mu'\hat{n}\, dS = k s^*(i) \int_i^{\infty} \tilde{h}^{n-1} f_1(\tilde{h})\, d\tilde{h} .$$

First–best profits for the firm are obtained by charging a tariff equal to $c(H(q))$ and taking the resulting surplus function $s^{fb}(\alpha) \equiv \max: \{ \alpha' q - c(H(q)) \mid q \geq 0 \}$ back as a fixed charge. In fact, s^{fb} is simply a function of $h(\alpha)$ and given by

$$s^{fb}(h) = \max: \{ hq - c(q) \mid q \geq 0 \},$$

and so total first–best profit is $\int_0^\infty s^{fb}(\tilde{h})\tilde{h}^{n-1}f_1(\tilde{h})\, d\tilde{h}$ (up to a multiplicative constant). If, as we assume, the integral of first–best profits converges we have

$$s^{fb}(i) \int_i^\infty \tilde{h}^{n-1}f_1(\tilde{h})\, d\tilde{h} \;\leq\; \int_i^\infty s^{fb}(\tilde{h})\tilde{h}^{n-1}f_1(\tilde{h})\, d\tilde{h} \;\longrightarrow 0 \text{ as } i \longrightarrow \infty.$$

Finally, since s^* as given in (26) satisfies $(s^*)'(h) \leq (s^{fb})'(h)$, and hence $s^*(i) \leq s^{fb}(i)$. Therefore

$$\int_{\partial A_i} s^*\mu'\hat{\mathbf{n}}\, dS \;\longrightarrow 0 \text{ as } i \longrightarrow \infty$$

so condition (18) is satisfied and hence Result 4 implies the result. □

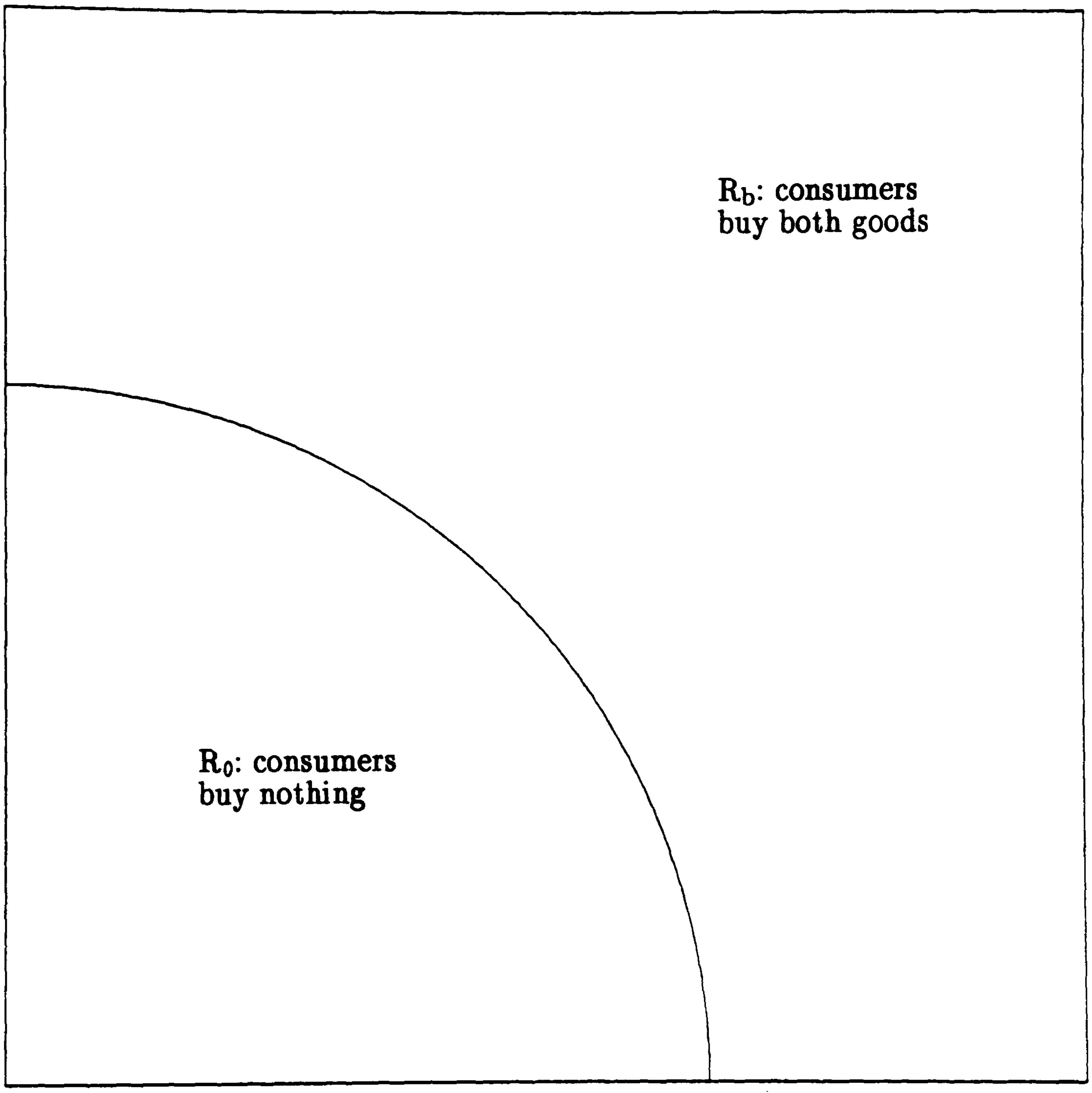

Figure 1: The typical pattern of demand

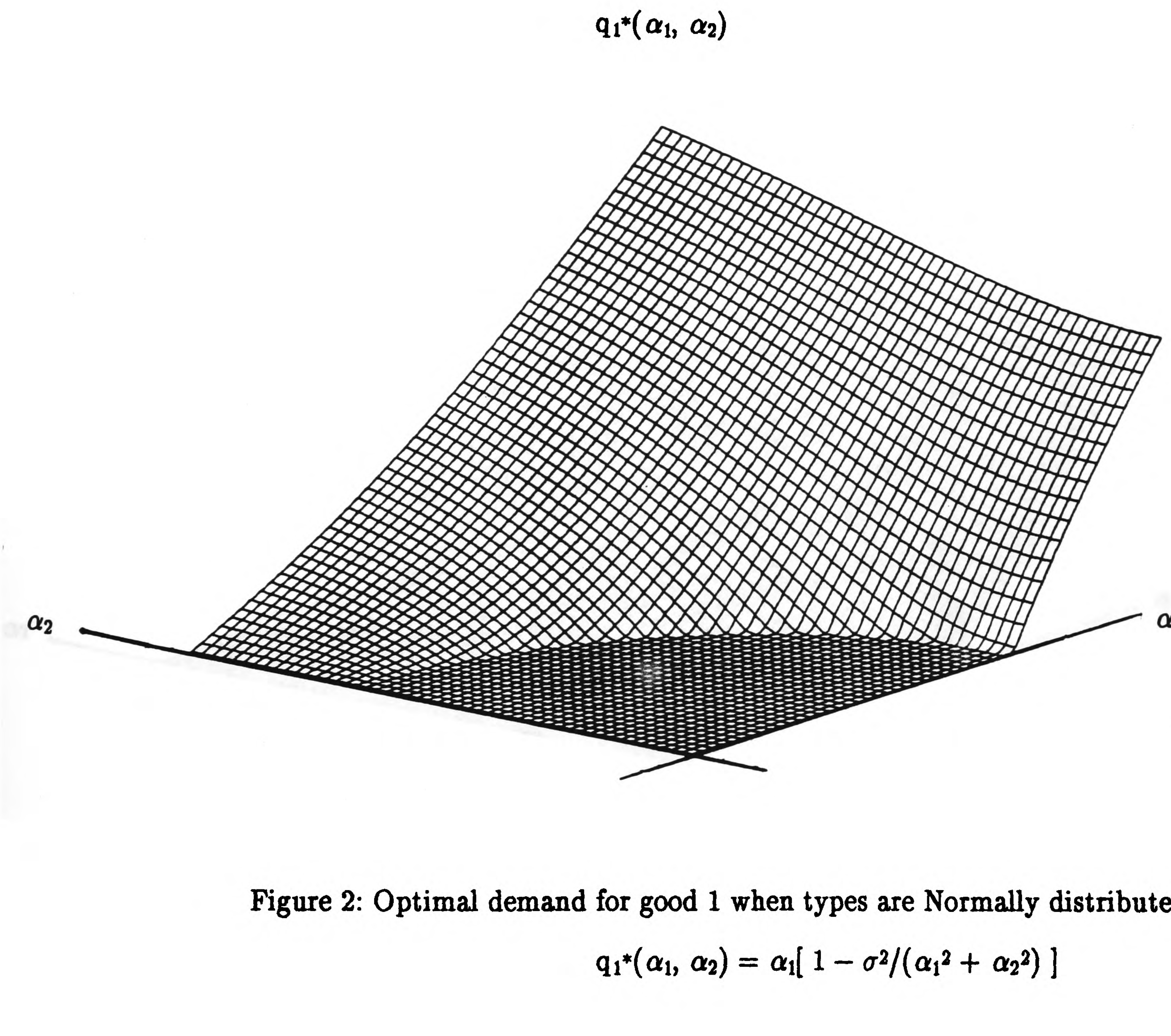

Figure 2: Optimal demand for good 1 when types are Normally distributed

$$q_1^*(\alpha_1, \alpha_2) = \alpha_1[\, 1 - \sigma^2/(\alpha_1^2 + \alpha_2^2)\,]$$

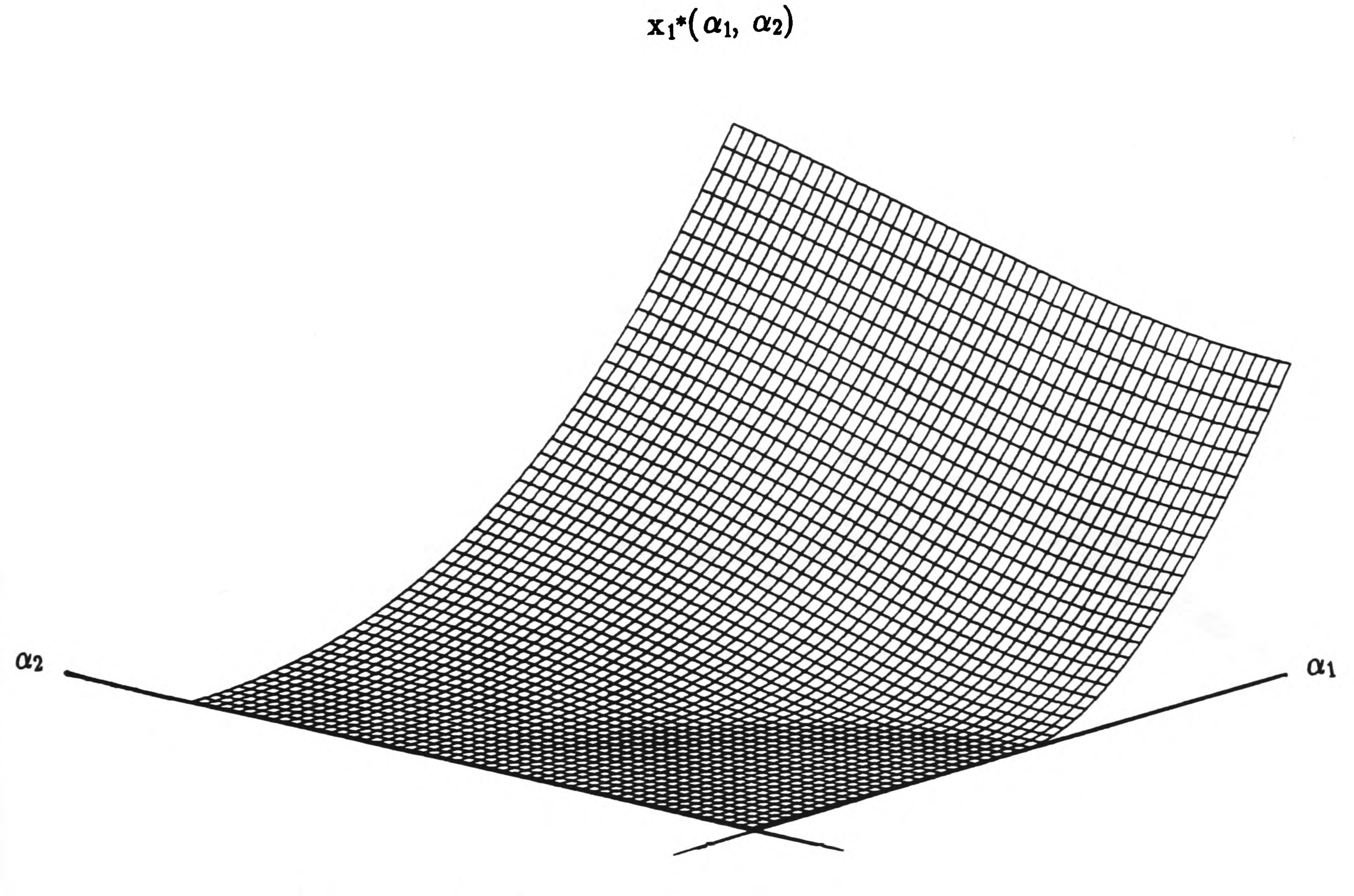

Figure 3: Optimal demand for good 1 when outputs are measured by $\mathbf{x}$ rather then $\mathbf{q}$

$$x_1^*(\alpha_1, \alpha_2) = \tfrac{1}{2}\alpha_1^2[\,1 - \sigma^2/(\alpha_1^2 + \alpha_2^2)\,]^2$$

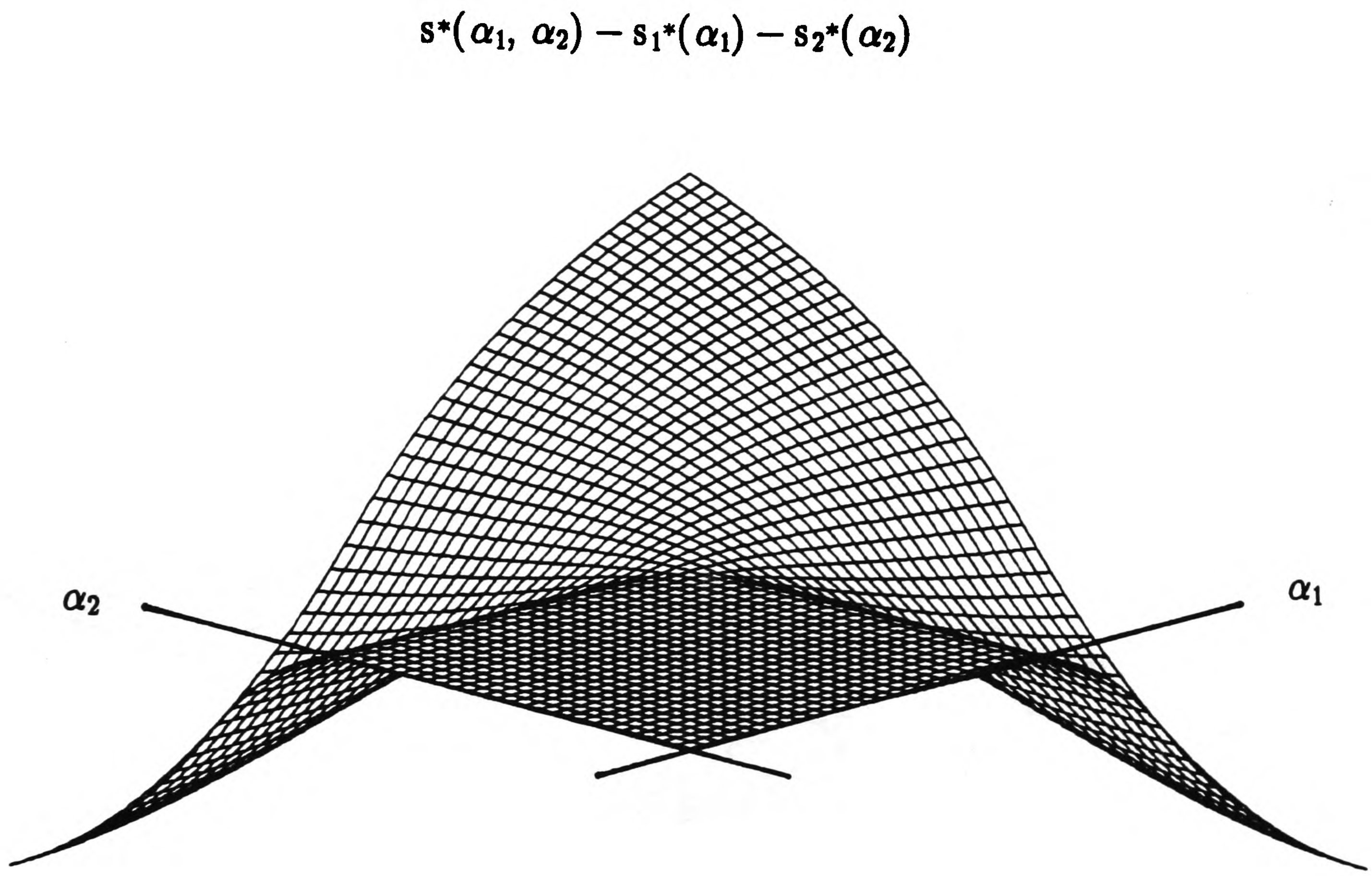

Figure 4: Who wins and who loses in the move from a separable tariff to a fully optimal tariff

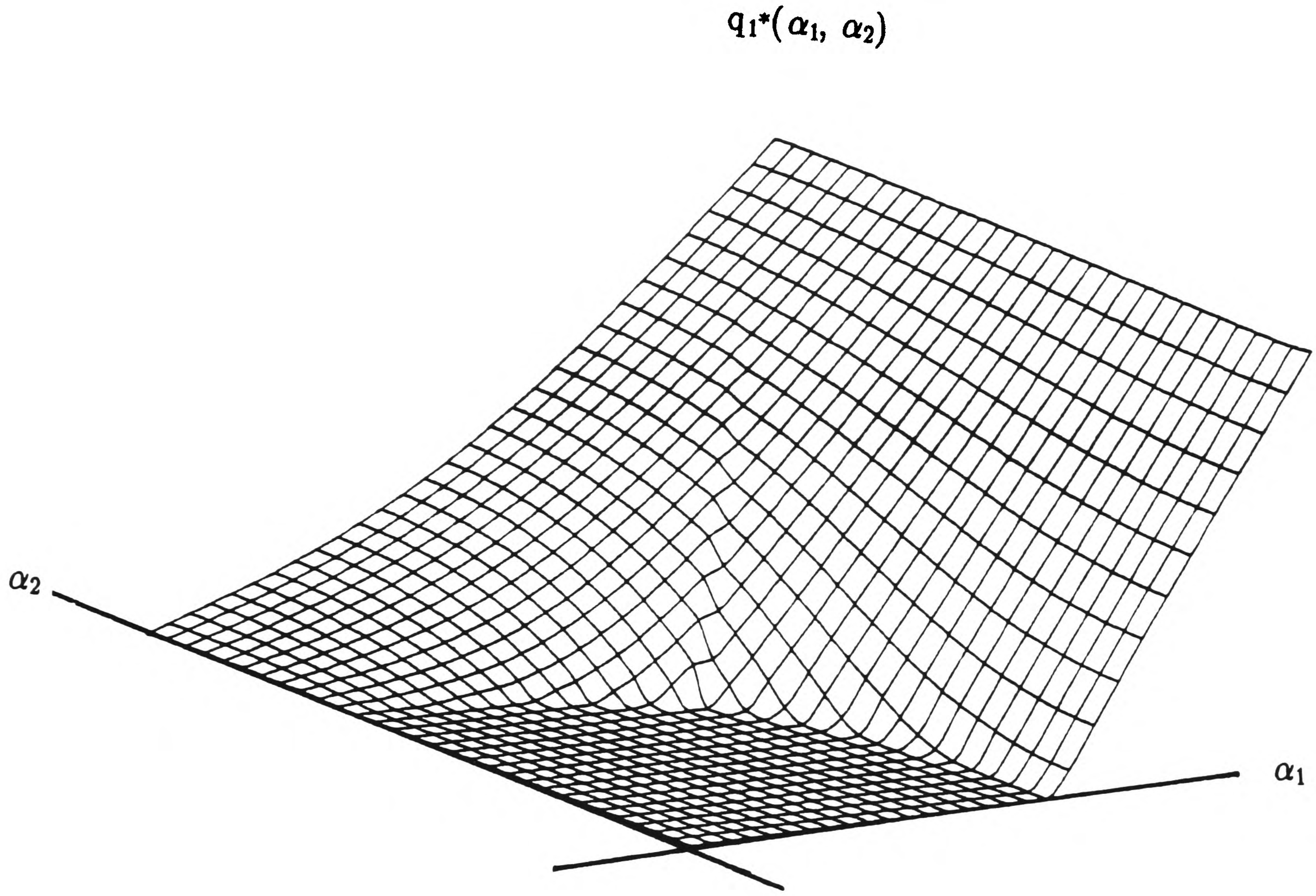

Figure 5A: Demand for good 1 when types are uniformly distributed

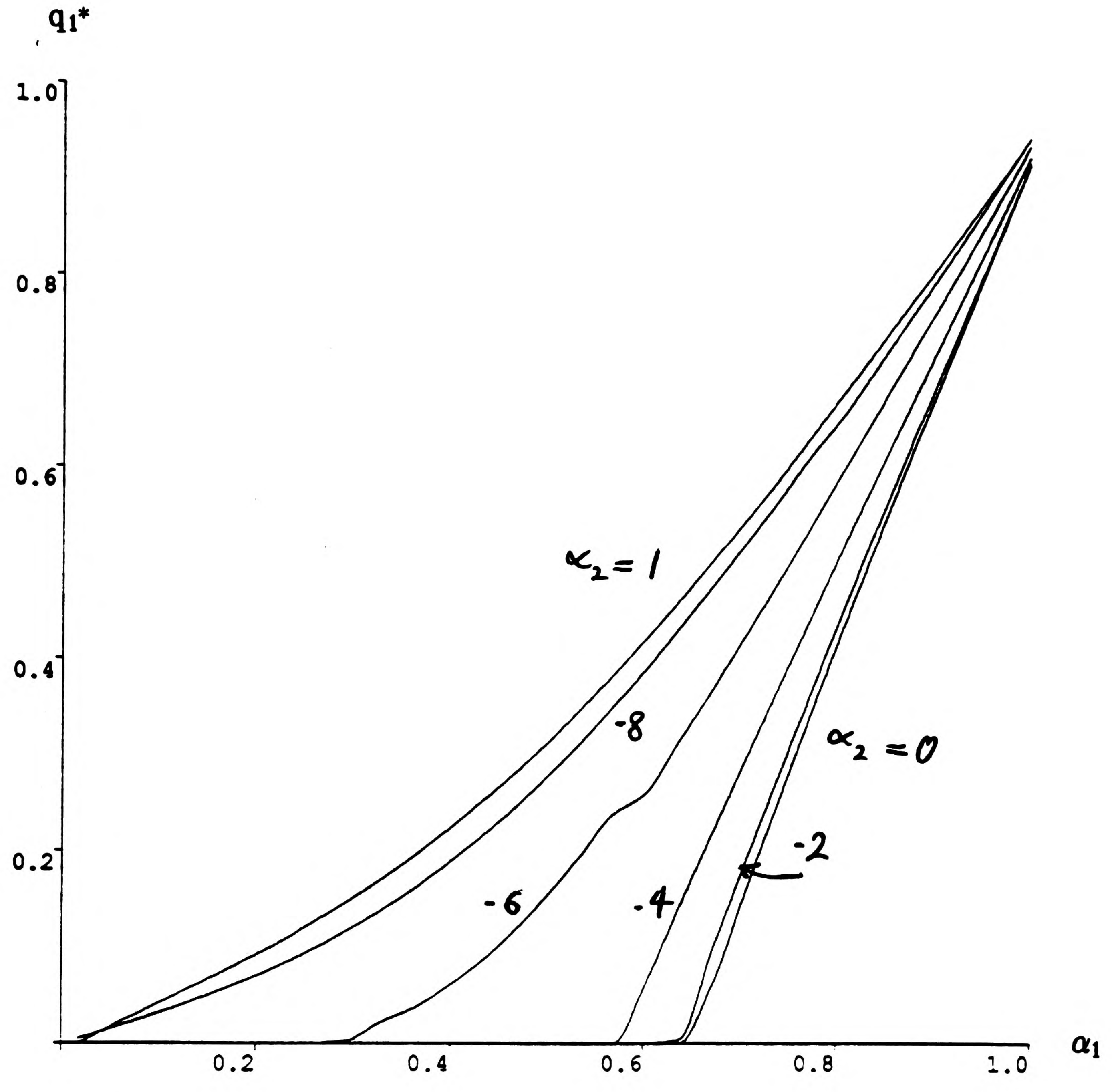

Figure 5B: Demand for good 1 when types are uniformly distributed as a function of α_1 for various fixed values of α_2

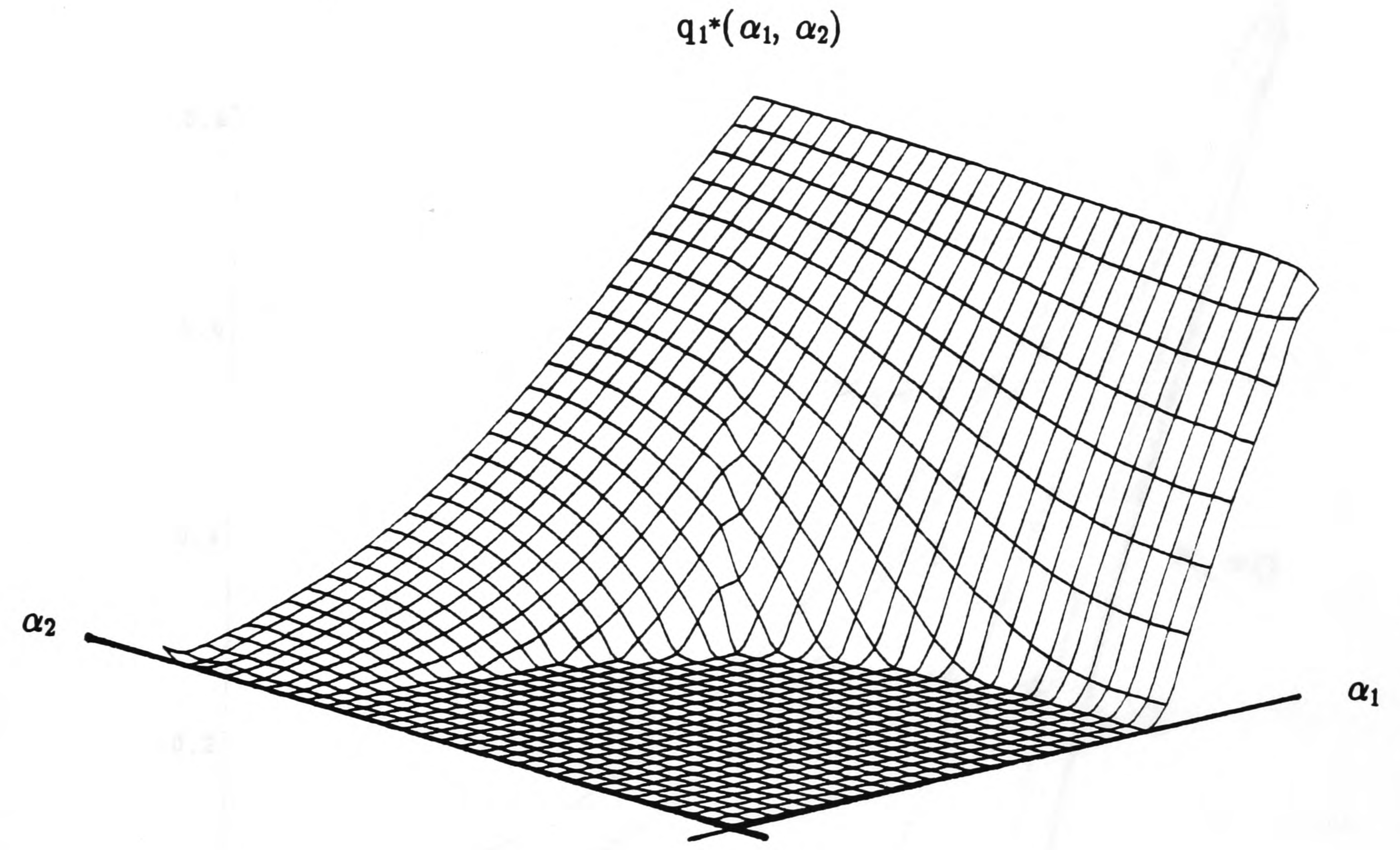

Figure 6A: Demand for good 1 when $f(\alpha_1, \alpha_2) = \alpha_1\alpha_2$

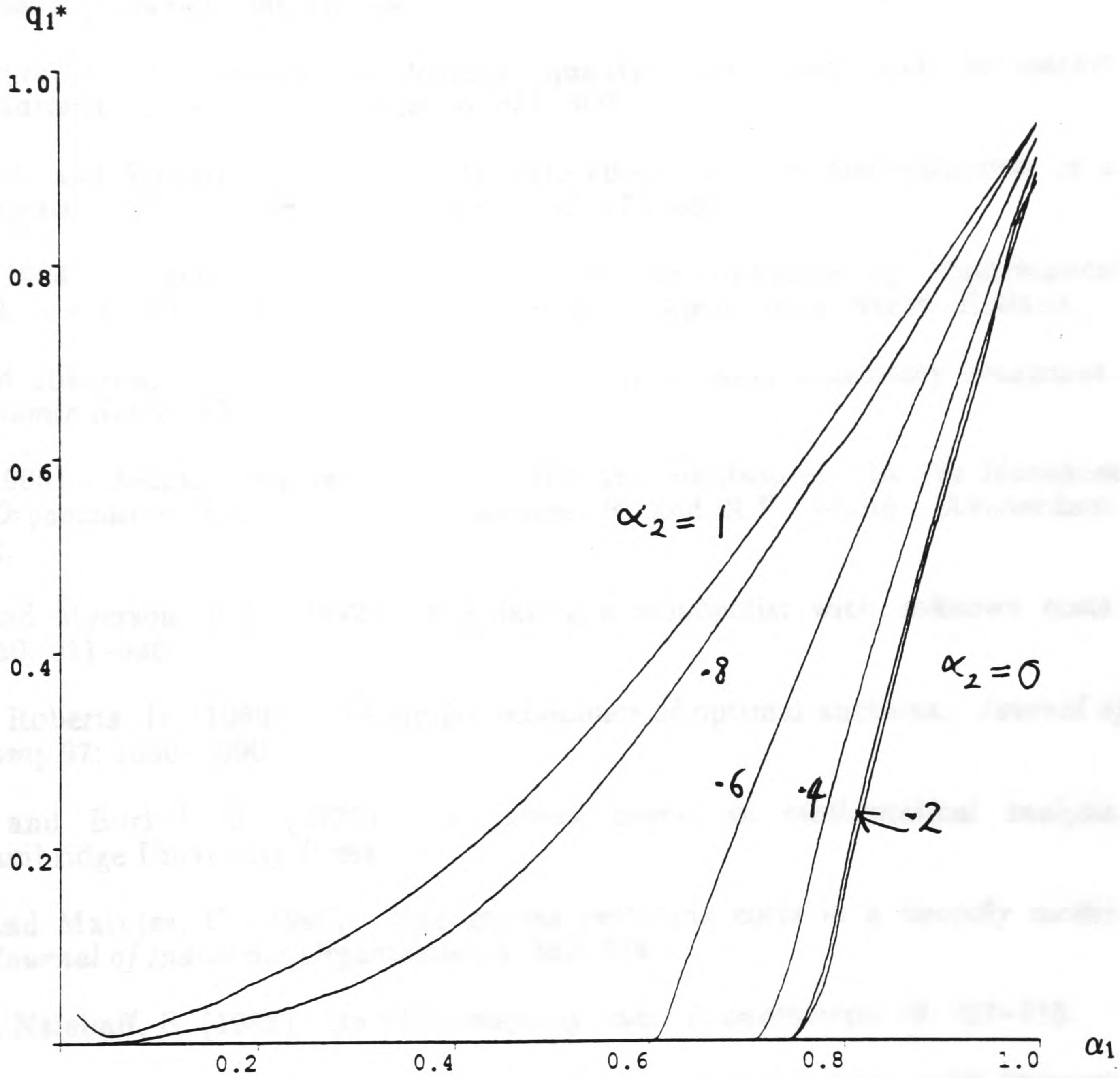

Figure 6B: Demand for good 1 when $f(\alpha_1, \alpha_2) = \alpha_1\alpha_2$ as a function of α_1 for various fixed values of α_2

References

Adams, R.B. (1978). *King C. Gillette: the man and his wonderful shaving device.* Boston: Little, Brown.

Adams, W.J. and Yellen, J.L. (1976). Commodity bundling and the burden of monopoly. *Quarterly Journal of Economics* 90: 475–98.

Akerlof, G.A. (1970). The market for 'lemons': quality, uncertainty, and the market mechanism. *Quarterly Journal of Economics* 84: 488–500.

Armstrong, C.M. and Vickers, J.S. (1991). Welfare effects of price discrimination by a regulated monopolist. *RAND Journal of Economics* 22: 571–580.

Arrow, K.J. (1986). Agency and the market. In *The Handbook of Mathematical Economics*, Vol. 1, eds. Arrow, K.J. and M.D. Intriligator. Amsterdam: North–Holland.

Averch, H. and Johnson, L. (1962). Behaviour of the firm under regulatory constraint. *American Economic Review* 52: 1052–1069.

Baron, D.P. (1988). Design of regulatory mechanisms and institutions. In *The Handbook of Industrial Organization*, Vol. 2, eds. Schmalensee, R. and R.D. Willig. Amsterdam: North–Holland.

Baron, D.P. and Myerson, R.B. (1982). Regulating a monopolist with unknown costs. *Econometrica* 50: 911–930.

Bulow, J. and Roberts, D. (1989). The simple economics of optimal auctions. *Journal of Political Economy* 97: 1060–1090.

Burkill, J.C. and Burkill, H. (1970). *A second course in mathematical analysis.* Cambridge: Cambridge University Press.

Caminal, R. and Matutes, C. (1990). Endogenous switching costs in a duopoly model. *International Journal of Industrial Organization* 8: 353–374.

Caplin, A. and Nalebuff, B. (1988). On 64%–majority rule. *Econometrica* 56: 787–815.

Caplin, A. and Nalebuff, B. (1991). Aggregation and social choice: a mean voter theorem. *Econometrica* 59: 1–23.

Crémer, J. (1984). On the economics of repeat buying. *RAND Journal of Economics* 15: 396–403.

Crémer, J. and McLean, R. (1985). Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica* 53: 345–361.

Crémer, J. and McLean, R. (1988). Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica* 56: 1247–1258.

Dana, J.D. (1992). The organization and scope of agents. *Journal of Economic Theory*, forthcoming.

Demski, J. and Sappington, D.E.M. (1984). Optimal incentive contracts with multiple agents. *Journal of Economic Theory* 33: 152–171.

Eggleston, H.G. (1958). *Convexity.* Cambridge: Cambridge University Press.

Engers, M. (1987). Signalling with many signals. *Econometrica* 55: 663–674.

Epstein, L.G. and Tanny, S.M. (1980). Increasing generalized correlation: a definition and some economic consequences. *Canadian Journal of Economics* 14: 16–34.

Farrell, J. (1986). Moral hazard as an entry barrier. *RAND Journal of Economics* 17: 440–449.

Fudenberg, D. and Tirole, J. (1991). *Game Theory.* Boston: MIT Press.

Gilbert, R.J. (1992). Symposium on compatibility: incentives and market structure. *Journal of Industrial Economics* 40: 1–8.

Goldman, M.B., Leland, H.E. and Sibley, D.S. (1984). Optimal nonlinear pricing. *Review of Economic Studies* 51: 305–320.

Green, J. and Stokey, N. (1983). A comparison of tournaments and contracts. *Journal of Political Economy* 91: 349–364.

Grossman, S.J. and Hart, O.D. (1983). An analysis of the principal–agent problem. *Econometrica* 51: 7–45.

Harris, C. (1986). *On strategic taxation.* Nuffield Discussion Papers in Economics No. 7, Oxford.

Harris, M. and Raviv, A. (1989). The design of securities. *Journal of Financial Economics* 24: 255–287.

Hart, O. (1975). On the optimality of equilibrium when markets are incomplete. *Journal of Economic Theory* 11: 418–443.

Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics* 10: 74–91.

Jewitt, I. (1988). Justifying the first–order approach to principal–agent problems. *Econometrica* 56: 1177–1190.

Katz, M.L. (1983). Non–uniform pricing, output and welfare under monopoly. *Review of Economic Studies* 50: 37–56.

Katz, M.L. and Shapiro, C. (1992). Product introduction with network externalities. *Journal of Industrial Economics* 40: 55–84.

Klemperer, P. (1990). *Competition when consumers have switching costs: an overview.* 1990 Review of Economic Studies Lecture and St. Catherine's College, Oxford, Working Paper.

Laffont, J.–J. (1989). *The economics of uncertainty and information.* Cambridge: MIT Press.

Laffont, J.–J. and Tirole, J. (1988). The dynamics of incentive contracts. *Econometrica* 56: 1153–1175.

Laffont, J.–J., Maskin, E. and Rochet, J.–C. (1985). Optimal nonlinear pricing with two dimensional characteristics, in *Information, Incentives and Economic Mechanisms,* ed. T.Groves, R.Radner, and S.Reiter. Oxford: Basil Blackwell.

Loeb, M. and Magat, W. (1979). A decentralized method for utility regulation. *Journal of Law and Economics* 22: 399–404.

Long, J.B. (1984). Comments on 'Gaussian Demand and Commodity Bundling'. *Journal of Business* 57: S235–246.

Loomis, L.H. and Sternberg, S. (1990). *Advanced Calculus*. Boston: Jones and Bartlett.

McAfee, R.P. and McMillan, J. (1987). Auctions and bidding. *Journal of Economic Literature* 25: 699–738.

McAfee, R.P. and McMillan, J. (1988). Multidimensional incentive compatibility and mechanism design. *Journal of Economic Theory* 46: 335–354.

McAfee, R.P., McMillan, J. and Whinston, M. (1989). Multiproduct monopoly, commodity bundling, and the correlation of values. *Quarterly Journal of Economics*: 371–83.

McAfee, R.P. and Reny, P.J. (1992). Correlated information and mechanism design. *Econometrica* 60: 395–422.

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

Milgrom, P. and Roberts, J. (1986). Prices and advertising signals of product quality. *Journal of Political Economy* 94: 796–821.

Mirman, L.J. and Sibley, D.S. (1980). Optimal nonlinear prices for multiproduct monopolies. *Bell Journal of Economics* 11: 659–670.

Mirrlees, J.A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38: 175–208.

Mirrlees, J.A. (1974). Notes on welfare economics, information and uncertainty. In *Essays in economic behaviour under uncertainty*, ed. M. Balch, D. McFadden and S. Wu. Amsterdam: North–Holland.

Mirrlees, J.A. (1975). The theory of moral hazard and unobservable behaviour — Part I. Mimeo, Nuffield College, Oxford.

Mirrlees, J.A. (1976). Optimal tax theory. *Journal of Public Economics* 6: 327–358.

Mirrlees, J.A. (1986). The theory of optimal taxation. In *The Handbook of Mathematical Economics*, ed. Arrow, K.J. and M.D. Intriligator. Amsterdam: North–Holland.

Monopolies and Mergers Commission (1976). Indirect electrostatic reprographic equipment. HC 47. London: HMSO.

Monopolies and Mergers Commission (1981). Full–line forcing and tie–in sales. HC 212. London: HMSO.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory* 18: 301–317.

Myerson, R.B. (1981). Optimal auction design. *Mathematics of Operations Research* 6: 58–73.

Oftel (1992). *Future controls on British Telecom's prices*. Oftel: London.

Ordover, J. and Panzar, J. (1980). On the non–existence of Pareto superior outlay schedules. *Bell Journal of Economics* 11: 351–354.

Palfrey, T.R. (1983). Bundling decisions by a multiproduct monopolist with incomplete information. *Econometrica* 51: 463–484.

Polinsky, A.M. and Rubinfeld, D.L. (1991). A model of optimal fines for repeat offenders. *Journal of Public Economics* 46: 291–306.

Quinzii, M. and Rochet, J.–C. (1985). Multidimensional signaling. *Journal of Mathematical Economics* 14: 261–284.

Rees, R. (1985). The theory of principal and agent — Part I. *Bulletin of Economic Research* 37: 3–26.

Roberts, K. (1979). Welfare considerations of nonlinear pricing. *Economic Journal* 89: 66–83.

Roberts, K. (1984). The theoretical limits to redistribution. *Review of Economic Studies* 51: 177–195.

Rochet, J.–C. (1987). A necessary and sufficient condition for rationalizability in a quasi–linear context. *Journal of Mathematical Economics* 16: 191–200.

Rogerson, W. (1985). The first order approach to principal agent problems. *Econometrica* 53: 1357–1367.

Rothschild, M. and Stiglitz, J. (1970). Increasing risk: I. A definition. *Journal of Economic Theory* 2: 225–243.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly Journal of Economics* 90: 629–650.

Rubinstein, A. and Yaari, M. (1983). Insurance and Moral Hazard. *Journal of Economic Theory* 14: 441–452.

Sappington, D.E.M. and Sibley, D.S. (1992). Strategic nonlinear pricing under price–cap regulation. *RAND Journal of Economics* 23: 1–19.

Schmalensee, R. (1981). Monopolistic two–part pricing arrangements. *Bell Journal of Economics* 12: 445–466.

Schmalensee, R. (1982). Commodity bundling by a single–product monopolist. *Journal of Law and Economics* 25: 67–71.

Schmalensee, R. (1984). Gaussian demand and commodity bundling. *Journal of Business* 57: S211–230.

Shavell S. (1979). Risk–sharing and incentives in the principal–agent relationship. *Bell Journal of Economics* 10: 55–73.

Shleifer, A. (1985). A theory of yardstick competition. *RAND Journal of Economics* 16: 319–327.

Spence, A.M. (1973). *Market Signalling.* Cambridge: Harvard University Press.

Spence, A.M. (1977). Nonlinear prices and welfare. *Journal of Public Economics* 8: 1–18.

Spence, A.M. (1980). Multi–product quantity–dependent prices and profitability constraints. *Review of Economic Studies* 47: 821–42.

Srinagesh, P. (1986). Nonlinear prices and the regulated firm. *Quarterly Journal of Economics* 101: 51–68.

Stigler, G. (1963). United states v. Loew's Inc.: A note on block booking. *Supreme Court Review* 152: 152–157.

Tirole, J. (1988). *The theory of industrial organization.* Cambridge: MIT Press.

Varian, H. (1985). Price discrimination and social welfare. *American Economic Review* 75: 870–875.

Whinston, M. (1990). Tying, foreclosure and exclusion. *American Economic Review* 80: 837–859.

Willig, R.D. (1978). Pareto–superior non–linear outlay schedules. *Bell Journal of Economics* 9: 56–69.

Wilson, R.B. (1985a). Incentive efficiency of double auctions. *Econometrica* 53: 1101–1117.

Wilson, R.B. (1985b). Multidimensional signalling. *Economics Letters* 19: 17–21.

Wilson, R.B. (1991). Multiproduct tariffs. *Journal of Regulatory Economics* 3: 5–26 [and correction, page 211].

Wilson, R.B. (1992a). Design of efficient trading procedures. In *The Double Auction Market*, eds. Friedman, D., J. Geanakoplos, D. Lane and J. Rust. SFI Studies in the Sciences of Complexity, Proc. Vol. XV. Addison–Wesley.

Wilson, R.B. (1992b). *Nonlinear Pricing.* Oxford: Oxford University Press. Forthcoming.