

PRICING OF TELECOMMUNICATIONS
SERVICES UNDER THE PRESENCE OF
ASYMMETRIC INFORMATION

By
Sissel Jensen

A DISSERTATION SUBMITTED FOR THE DEGREE OF DR. OECON

AT

THE NORWEGIAN SCHOOL OF ECONOMICS
AND BUSINESS ADMINISTRATION,
BERGEN,
OCTOBER 2001.

Contents

Table of contents	iii
Acknowledgements	v
1 Introduction	1
1.1 A brief review of the literature	3
1.2 Outline of the thesis	6
2 Damaging Network Subscription	13
2.1 Introduction	13
2.2 A model with Damaging	17
2.2.1 Single quality	21
2.2.2 Damaging	23
2.3 Damaged goods in telecommunications	29
2.4 Concluding remarks	33
3 Two-Part Tariffs with Partial Unbundling	35
3.1 Introduction	35
3.2 Telecommunications services	38
3.3 A two-dimensional good	39
3.4 The model	42
3.5 Two-part tariffs	43
3.5.1 Uniform distribution	47
3.5.2 Beta distribution	48
3.6 Concluding remarks	52
Appendix	54
4 Three-Part Tariffs in a Duopoly	57
4.1 Introduction	57
4.1.1 Related literature	58
4.2 The model	60
4.2.1 Individual rationality	64
4.2.2 Incentive compatibility	64

4.2.3	Informational rents	65
4.3	Optimal allocations	66
4.3.1	The IR constraint is not binding	69
4.3.2	The IR constraint binds	70
4.4	Implementation	71
4.5	A numerical example with quadratic utility	74
4.6	Concluding remarks	77
5	Two-part Tariffs, Consumer Heterogeneity, and Cournot Competition	78
5.1	Introduction	78
5.2	The Model	80
5.3	Equilibrium outcomes	84
5.3.1	Market coverage	84
5.3.2	Market coverage versus exclusivity	87
5.4	Concluding remarks	93
	Appendix	95
	Bibliography	97

Acknowledgements

I am grateful to a number of people for their helpfulness in various respects during the course of completing my thesis. First of all I wish to express my profound thanks to my supervisor Lars Sjørgard for his very helpful comments and suggestions. I also wish to thank him for being so positive and optimistic on my behalf. I am also very grateful to the two other members of my supervising committee, Jon Vislie and Petter Osmundsen, for their useful comments at various stages of the research. Further, I am indebted to my fellow doctoral students at NHH for their comments and suggestions at our seminar series. During my work I have also benefited from the help of Fred Schroyen, who has always been available to help me on various kinds of problems encountered.

Anne Liv Scrase has helped me by proof-reading the English. This has undoubtedly improved the thesis substantially and I am very grateful. I would also like to express my gratitude to the staff and to the members of the Department for making me feel so welcome.

Special thanks go to Elisabeth and Claus for letting me live with them when I moved to Bergen, for marvellous cuisine, and for their friendship. Special thanks also to Karin and Ivar for their great hospitality, great friendship and for their support in every respect.

At an early stage of my writing Erik Øiolf Sørensen introduced me to the fun and challenges of working with \LaTeX . He did this without knowing the amount of questions and troubles it would bring him at the second stage. Nevertheless, he has patiently helped me with any problem I have run into during my writing.

Last, but not least, I would like to acknowledge my debt to Telenor Research and Development for financial support. The results and their interpretation as expressed in the thesis are solely my responsibility and do not represent the views of Telenor.

Bergen,
October 11, 2001

Sissel Jensen

Chapter 1

Introduction

by Sissel Jensen

Firms will always try to sell additional units of output, but are also reluctant to reduce the price of the units they are currently selling. There is always this trade-off between expanding the market and keeping the price-cost margin. For instance, a monopoly facing a downward sloping demand curve will expand the market by lowering the price as long as the marginal revenue exceeds the marginal cost. However, there are still some potential consumers who are willing to pay more than the cost of producing an extra unit.

Consumers typically differ in their willingness to pay for an increment of a firm's product. Two consumers will often have different willingness to pay for a given amount of the good, due to differences in income, taste, etc, and will also choose to buy different total quantities. Also, at a uniform price, because consumers' valuation of successive units is declining, consumers buying many units of the good enjoy consumer surpluses that are not captured by the firm. Both observations indicate that alternative pricing policies can raise profits. All methods of price discrimination attempt to expand output at a lower price without simultaneously offering all units at the same low price. If the firm can charge different consumers a different unit price, or if the firm can offer different units at a different price, it can increase profits. As to the second problem, the firm can increase profits if it can appropriate consumer surplus by other means than quoting a price per unit. For instance, the firm may announce a two-part tariff which encourages consumers to make larger purchases via a low marginal price, whereas the consumers' surpluses are captured via a fixed fee paid up-front.

The theoretical categorization of price discrimination is either as first, second, or third-degree price discrimination.¹ First-degree, or perfect, price discrimination occurs if the price charged for each unit is equal to the maximum valuation

¹See Philips (1983), Varian (1989), and Tirole (1988, chapter 3) for a general introduc-

for that unit. If the firm had exact knowledge about each consumer's valuation of the product, it could make a single take-it-or-leave-it offer to each consumer that appropriates the entire social surplus. Therefore, first-degree price discrimination entails no efficiency loss compared to perfect competition, since perfect competition maximizes social surplus (but it does affect the distribution of income compared to perfect competition).

Under second-degree price discrimination, or nonlinear pricing, the price per unit depends on the total number of units a consumer buys. A firm makes a universal announcement of its pricing policy, accessible for all consumers, but the tariff(s) may be designed so that a consumer's total expenditures do not increase proportionately with the amount purchased. If two consumers buy different amounts, because they value the product differently, they are also paying a different (average) price per unit.

Third-degree price discrimination refers to a situation where a firm charges different groups of consumers a different unit price, but charges a linear tariff for each group. Whether a consumer belongs to a certain group is determined on the basis of some exogenous information so that the firm can enforce the division easily, for instance age, occupation, sex, location, etc.

Perfect price discrimination is informationally extremely demanding and assumes that a supplier can offer different contracts conditionally on a consumer's "type", or willingness to pay. In practice, different contracts are offered conditionally on observable variable, where the observables are assumed to serve as imperfect estimates of a consumer's willingness to pay for the good. If the firm can only gather imperfect estimates about different consumers' valuation of the good, it cannot appropriate the entire surplus either. Second- and third-degree price discrimination will therefore yield second-best allocations.

Price discrimination is not a viable strategy unless the firm has some market power and hence, the ability to set price above marginal cost. In addition, it must also have the ability to sort consumers and to prevent resale. As to resale, consumers buying at a low price may resell the product to consumers facing a high price and thereby demolish the firm's attempt to charge different prices. However, resale may be difficult for consumers for several reasons, for instance; resale can be illegal, resale can lead to loss of warranty, or consumers must incur large transaction costs in order to resell the product (Carlton and Perloff, 1999). That there are impediments to resale is also supported by the fact that we observe price discrimination in so many markets.

Under third-degree price discrimination one assumes that a monopolist is able to prevent resale between groups, but not in any sense within a group (and must charge a uniform price). Under second-degree price discrimination one assumes

tion to price discrimination. For a comprehensive exposition of nonlinear pricing see Wilson (1993). Pigou (1920) provides the foundation for the classification of the different forms of price discrimination.

that consumers are anonymous in every respect but in the size of their purchases. There is only one single group in the above mentioned sense, but the monopolist can prevent all attempts to resell the product. However, the firm can not neglect the fact that consumers that have high willingness to pay have an incentive to pretend that their valuation is low. Generally, preventing resale (in one sense or another) is not regarded as a severe problem, whereas the problems associated with sorting consumers with different willingness to pay can be tremendous.

The simplest problem of screening can be formulated by letting a monopoly supply a single good to a group of consumers that are identical in all characteristics but their marginal willingness to pay for the monopolists product. This is the canonical model with single-dimensional private information and deterministic participation in Mussa and Rosen (1978) and Maskin and Riley (1984). The models also adopt the assumption that the firm observes a single-dimensional quantity (quality) variable, making quantity (quality) the only possible variable for the firm to trade against payment. The models serve as benchmarks in the literature on second-degree price discrimination.²

Sorting consumers becomes an increasingly difficult task to perform when the screening problem is multi-dimensional, or when the firm faces competition, and so far there are very few precise and robust predictions given in the literature. Nevertheless, these are important topics on the research agenda.

1.1 A brief review of the literature

Nonlinear pricing is a standard topic within the economic theory of the firm which is described in all textbooks in microeconomics. Although the way firms implement nonlinear pricing varies substantially, nonlinear pricing is practiced in many industries.³ Nonlinear pricing provides an efficient means of meeting a firm's revenue requirement, be it a regulated firm with the need to recover fixed costs or a profit maximizing monopoly, when compared to a uniform pricing regime. If the population of consumers is diverse, it is well established that the optimal tariff is generally nonlinear (whenever this is viable).

Let me just start with a situation where a monopoly offers a single two-part tariff to all consumers. The firm charges consumers a lump-sum fee for the right

²In Mussa and Rosen (1978) consumers have unit-demand and the product can be sold with different qualities, while in Maskin and Riley (1984) consumers purchase many units of a single-dimensional good. The two papers predict qualitatively identical results.

³There is a vast literature on nonlinear pricing and I do not intend to give anything close to a survey here. I will, however, try to sketch some of the main topics within the field together with the most important references. The bibliography section in Wilson (1993) provides a comprehensive list of references to the literature on nonlinear pricing together with a summary of the theoretical development. He also directs the reader to extensive bibliographies on the various subtopics of price discrimination. Rochet and Stole (2000) give a comprehensive survey of the literature on multidimensional screening.

to purchase goods and a uniform price per unit. Assuming that the demand side is described by a single representative consumer, a two-part pricing arrangement is equivalent to perfect price discrimination, which maximizes monopoly profits by appropriating all consumer surplus. This was pointed out in the classical articles by Oi (1971). In a market of many different consumers, the global profit maximum can be reached by offering equally many two-part tariffs with the price per unit equal to marginal cost and the lump-sum fees equal to each individual's surplus. If such discriminatory pricing is not within the firm's power, because it is not able to identify each individual's taste, or because making exclusive offers only to some consumers is unlawful, it will most likely offer a uniform two-part tariff with a unit price above marginal cost and a fixed fee that extracts all surplus from the consumer with the lowest willingness to pay. Later works on uniform two-part tariffs in a monopoly context include studies with nonzero income effects and with consumption externalities (see Schmalensee (1981), Ng and Weisser (1974), Wilson (1993, chapter 7) and Littlechild (1975) and the references therein).

Generally, let the demand side heterogeneity be captured in a single parameter which describes a consumer's intensity in the demand for the monopolist's product (it can reflect differences in taste, or in income, or simply represent an aggregate type parameter). If the demand of the different consumers can be ordered for each price and the ordering is preserved for any price, the firm profits when it designs equally many two-part tariffs as the number of consumers and lets all consumers choose from the "tariff menu", $((A_1, p_1), (A_2, p_2), \dots, (A_n, p_n))$. Since high demand consumers have the option to choose tariffs intended for low demand consumers, the profitability on high demand consumers is restricted by the relatively lower willingness to pay by low demand consumers. The firm designs its pricing structure to maximize profits subject to a self-selection constraint. (See for instance Goldman, Leland and Sibley (1984) and Wilson (1993, chapter 6). Sharkey and Sibley (1993) describe the properties of the optimal two-part tariffs chosen by a regulator when the welfare of different consumers has different weights.)

A menu of n different two-part tariffs mimics a single piecewise-linear n -part tariff, a block-declining price schedule with a fixed fee and $n - 1$ unit prices (Wilson, 1993, chapter 6.3 and 6.4).⁴ In the limit, with a continuum of types, a multipart tariff tends towards a fully nonlinear tariff with complete separation of types. Both approaches implement the allocation in Maskin and Riley (1984).

The results in the case of a single characteristic and a single instrument are qualified by taking the assumption that only the local downward incentive compatibility constraint is binding. Hence, it suffices to ensure that a certain con-

⁴This is only true, however, when consumers have perfect knowledge about their future demand, or when they can commit to future levels of consumption. Individual stochastic demands break the duality between consumption and choice of the corresponding self-selecting tariff, see Miravete (2000).

sumer will not mimic an adjacent type below his own type. When consumers' participation is deterministic, one can compute each consumer's expected surplus as a function of the allocation of goods, and maximize the firm's profit net of this expected surplus.⁵ Whenever the need to secure incentive compatibility does not conflict with the need to ensure voluntary participation, complete separation between different types can be reached, otherwise it may be optimal to exclude some consumers or to offer different types identical contracts (ironing, pooling, bunching). Rochet and Choné (1998) and Armstrong and Rochet (1999) explore whether these results can be extended to multidimensional contexts, with several characteristics and several instruments, and show what will be the likely properties of the optimum. Rochet and Stole (1999) and Armstrong and Vickers (1999) relax the assumption that the reservation utility is perfectly known by the firm and introduce stochastic participation. Multiple dimensions are difficult to handle, partly because the incentive compatibility conditions are frequently not only binding among local types. Discrete models with fewer incentive compatibility constraints can, on the other hand, be tractable. Rochet and Stole (2000) give a survey of the literature on multidimensional screening.

While keeping the assumptions that the private information concerns a single variable, that the single crossing condition is satisfied and that the reservation utility is independent of a consumer's type, other extensions of the benchmark models are to introduce more than one instrument or more than one observable variable.⁶ An example within the first class is Matthews and Moore (1987), examples within the second class are Sappington (1983) and Caillaud, Guesnerie, Rey and Tirole (1988).⁷ The latter framework generalizes the standard results with a single observable, whereas these are not necessarily found in the former. Especially, as in Matthews and Moore (1987), nonlocal incentive constraints may be binding.

In terms of economic applications, modelling imperfect competition between firms competing with nonlinear price schedules, is of great importance. It is, however, also very complicated. Among other factors, competitive models may naturally suggest at least two dimensions of heterogeneity, including uncertain participation (see section 7 in Rochet and Stole (2000), Rochet and Stole (1999), and Armstrong and Vickers (1999)). The hypothesis that firms practice nonlinear

⁵Participation is deterministic when the firm knows for certain that all consumers obtaining a nonnegative surplus will buy the good, i.e., consumers' reservation utility is perfect information to the firm. Especially, the reference models assume that the reservation utility is independent of a consumer's type. Under the single crossing condition global incentive compatibility reduces to the monotonicity condition.

⁶The approach of multidimensional types and a single instrument is done in Laffont, Maskin and Rochet (1987).

⁷Matthews and Moore (1987) extend the Mussa and Rosen (1978) model by allowing the monopolist to offer different levels of warranties as well as qualities. Sappington (1983) extends the Baron and Myerson (1982) model of incentive regulation by letting the regulated firm produce two instead of just one product.

pricing is supported by many real-life examples, see for instance Wilson (1993) and Michell and Vogelsang (1991). Also, the notion that oligopoly firms, one way or another, are implementing an outlay schedule with quantity discounts included, is also supported by Ivaldi and Martimort (1994).

Although we assume that all demand side heterogeneity is single dimensional, modelling strategic interaction into models with nonlinear pricing is not straightforward. In the simplest case with full information and N identical consumer, a monopoly would achieve perfect price discrimination by charging a two-part tariff (Oi, 1971). In a duopoly, however, the competing firm can always charge a uniform price that is preferred by all consumers and make nonnegative profit.⁸

Similarly to the strive at solving the uniform price Bertrand paradox, there are ways of getting round the similar problems of a nonuniform Bertrand equilibrium, (examples are given in Mandy (1992) and in Harrison and Kline (2001)). Even so, it may be necessary to add additional structure and restrictions to the models as compared to the uniform pricing case. For instance, adding capacity constraints will not alone solve the puzzle. Oren, Smith and Wilson (1983) model a case where firms compete in market shares and use nonlinear tariffs in an attempt to separate consumers with different willingness to pay. They find equilibria that resemble monopoly pricing as well as equilibria with aspects of the standard Cournot model. Harrison and Kline (2001) examine two-part tariffs in a Cournot oligopoly with homogeneous consumers (full information). In addition to committing to a capacity level at stage one, they assume that firms can commit to a fixed fee as well. According to their findings, the unit price is equal to marginal price, and the fixed fee may or may not extract all consumer surplus. Stole (1995) finds a separating screening equilibrium in an oligopoly with differentiated products and single-dimensional uncertainty. This is the model that is extended to a multidimensional context in Rochet and Stole (1999).

1.2 Outline of the thesis

The thesis consists of four essays on second-degree price discrimination. Firms operating in a market are supposed to announce a set of tariffs, it may be a single tariff or several tariffs. A tariff is to be understood as an announcement made by a firm describing their services and the payments to be charged for such services. In all four essays, the basic description of the demand side is that it consists of heterogeneous consumers deriving utility from consuming multiple units of a generic good q .⁹ A firm faces equally many downward-sloping demand curves and each individual obtains a surplus at a uniform price. Firms are assumed to

⁸See, for instance, Gasmı, Moreaux and Sharkey (2000).

⁹By multiple units we mean that consumers buy so many units that q is continuous rather than discrete. Hence, we do not treat the special case of unit demand, where one unit will be demanded by a consumer if the price is less than or equal to his reservation price.

have knowledge about the position of these demand curves, the intercept, the slope, and other relevant information. A firm does not, however, know which demand curve to associate with a given consumer. In all cases it is assumed that consumers have quasilinear utility. Thus, the Marshallian consumer surplus is an appropriate measure of individual welfare and this can also be measured in monetary terms. Obviously, it is this value firms want to capture via a fixed fee.

In all models that are presented, the decision makers are private and unregulated firms and the objective is always to maximize profits. Since we are using quasilinear utility, we do not regard how the firms' pricing affects different consumers' incomes and how this in turn might affect social welfare. Both first degree price discrimination and perfect competition produce efficient levels of output and the outcome cannot be ranked in a Pareto sense. From a welfare point of view then, the closer the pricing policy is to perfect price discrimination the more likely it is that the price discrimination leads to a more efficient outcome as compared to uniform pricing. However, a firm's pricing policy can lead to inefficiencies in different ways; when price exceeds marginal costs and output is restricted; when consumers have different marginal willingness to pay there are unexploited gains from further trade; and when firms or consumers spend resources that are of no benefit to any other party. Some of the models that are considered in this thesis produce effects related to all three sources and, hence, the net effect may be ambiguous.

The basic description of consumers' preferences is the same in all four essays. The two first essays analyze second-degree price discrimination in a simple two-type monopoly context (chapters 2 and 3). In the quantity framework with nonlinear pricing, the quantity variable is typically single dimensional. Hence, what is left to describe in the tariff is the payments to be charged for usage. The payment can, however, be a nonlinear function of usage, and the firm can offer several tariffs as well. If the quantity variable can be assigned a set of observable attributes, services assigned one set of attributes can be charged differently than services assigned another set of attributes. Hence, the part of the tariff describing the service becomes an important issue for the firm. In the first paper, in chapter 2, the monopoly is allowed to "damage" the portion of its product that is sold to low demand consumers. However, in this paper the firm must take the description of the damaged service as given. In the second paper, in chapter 3, we assume that the firm can observe each consumer's "mode of usage" as well as the individual purchase size. Hence, the firm is free to design the attribute dimension, as well as the payments to be charged.

The remaining two essays return to the case where the quantity variable is single dimensional and we consider second-degree price discrimination in a context with competition. Hence, the tariff is again a description of the payment to be charged for usage. Chapter 4 is concerned with how the optimal quantity-payment allocation can be truthfully implemented by optional piecewise linear tariffs in a differentiated goods duopoly. In the last essay, presented in chapter

5, we analyze firms' possibility to use a two-part tariff in a homogeneous goods oligopoly.

Chapter 2: Damaging Network Subscription

In the first paper we consider a monopoly model with heterogeneous consumer characteristics along both a quantity and a quality dimension. We examine the effects on the firm and consumers of introducing quality discrimination in a two-type model with quantity discounts. It is assumed that the monopoly can sell its product in two "versions", for instance a high-quality and a low quality version. By accepting certain restrictions in the use of the service, individuals are granted a reduction in the usage price. Even though producing low quality is at least as costly as producing high quality, and can be described as a practice of "damaging", it may be profitable. Either because it enables the firm to serve consumers it would otherwise exclude, or because it enables the firm to reduce the informational rent to consumers having the largest willingness to pay for the high-quality version.

We identify the private incentives to introduce a damaged version in the cases where the monopoly excludes and serves, respectively, the type with low willingness to pay. When the incentive behind damaging is to serve consumers that absent damaging would not be served, damaging enables the firm to extract rent from low demand types. Damaging will in this case lead to a Pareto-improvement. However, the firm might also want to introduce damaging because it enables the firm to reduce the information rent to high demand types, and thus, extract larger surplus from these consumers. With damaging, the firm will reduce the mark-up in the usage price towards low demand types, but the firm will also spend resources that are of no benefit to any of the consumer groups. Hence, the net welfare effects are ambiguous.

The model is closely related to Deneckere and McAfee (1996). They analyze damaging in a framework of uniform pricing, and show that the practice can lead to a Pareto-improvement. However, since nonlinear pricing has welfare effects that are different from the welfare effects under linear pricing, the consequences of damaging are also different in our model compared to Deneckere and McAfee (1996).

In the paper, we also give illustrations on the practice of damaging in telecommunications. The first illustration is optional calling tariffs that place an artificial restriction on consumers' call distribution. These are known as Calling circle tariffs or Friends and Family tariffs, in which a quality reduction is achieved by restricting the number of call termination points, i.e., phone numbers that can be reached by the tariff. The second illustration is specific calling plans in the cellular market that place a restriction on the consumer's mobility. By paying an additional monthly fee, mobile phone subscribers pay a lower usage price when the mobile phone is used from the subscriber's "home zone", which is the area

inside or near the home, than when it is used from outside the home zone. In this tariff, the quality reduction is achieved by placing a restriction on the possible points of call origination.

Chapter 3: Two-Part Tariffs with Partial Unbundling

The second paper also explores second degree price discrimination in a multi-dimensional good context. It is assumed that the firm is able to monitor consumers' use of the service, not just only the number of units but also consumers' mode of usage measured according to some observable service attributes. For instance, if we think of the generic good as telephony usage, i.e., the quantity variable being call minutes, the firm can assign each minute a unique list of observable attributes, such as time-of-day, day-of-week, call termination point (the phone number of the party being called), etc. Hence, we assume that it is possible for the firm to monitor each consumer's calling pattern.

Even though a consumer's willingness to pay is private information, his calling pattern can be observed by the firm. Hence, consumers having different calling patterns can be charged according to different schedules. In the present framework it is assumed that the demand side consists of two consumers, one is a low demand type, and one is a high demand type. The low demand type is assumed to have a concentrated calling pattern while the high demand type has a dispersed calling pattern. As in the traditional models of single dimensional screening, the contract meant for low types is distorted precisely in order to make it less attractive to high type consumers. The present context is similar to a multiproduct context since units assigned different sets of attributes can be treated as different products and the practice that is described refers to a situation with partial unbundling. If the heterogeneity in consumers' calling pattern is ignored, complete bundling occurs.

In models where the contract is two-dimensional, the single crossing condition is ensured when high demand types have larger marginal willingness to pay for all increments. If we impose any restriction on the mode of usage we must ensure that this property still holds. If the high demand type's contract is not distorted in any dimension, it is sufficient that the restriction has a nonpositive effect on the consumer's marginal willingness to pay.

The paper shows that the firm introduces distortions in the use of the service against a decrease in the quantity distortions in the low-type's contract. This may not come as a surprise; when the firm has two instruments at hand it will distort the contract in both relevant dimensions. If the heterogeneity on the demand side is large, and price-cost distortions are the sole instrument at hand, then it is also the case that a large fraction of the consumers pay a price well above marginal cost, and the welfare losses arising from this may be severe. Hence, any strategy that "increases the observability" of consumers' willingness to pay may potentially increase both profit and welfare. This feature of the contract has

analogous insight from theories on taxation. It is often optimal to use many, but smaller tax rates, because deadweight losses are convex functions of the tax rate.

Chapter 4: Three-Part Tariffs in a Duopoly

Nonlinear pricing plays an important role in competitive markets. Likewise, we can also observe that more sophisticated tariffs are replacing two-part tariffs. Especially, we often observe tariffs where the fixed monthly payment includes some “free” consumption allowance per month. In effect, firms are using three-part tariffs in addition to two-part tariffs to implement a nonlinear outlay schedule. One approach of the third paper is to study how nonlinear pricing can be implemented in a competitive market.

The model is based on Stole (1995), which shows how a duopoly can reach complete separation of types via direct mechanisms (take-it-or-leave-it contracts). Stole (1995) shows that the qualitative properties of the monopoly model are kept, that is, quantity purchases are downward distorted for all types but the one that values the service highest. Later works on competitive price discrimination in a similar setting, especially Armstrong and Vickers (1999) and Rochet and Stole (1999), consider a setting with multi-dimensional uncertainty. They find that many of the results achieved in the monopoly setting does not extend to a competitive framework. However, the model presented in chapter 4 maintains a setting with single dimensional uncertainty, as in Stole (1995). Especially, consumers choice with respect to participation can be treated as a deterministic decision. If the utility a consumer derives when he accepts one of the firms’ tariffs weakly exceeds the utility he derives when he rejects it, the firm knows with certainty that the consumer will participate.

As a consequence of the existence of a competing firm, consumers considering one of the firm’s tariffs have the option of accepting the competing firm’s tariff. The outside option will of course be of higher value for high demand types than for low demand types, and the reservation utility is therefore an increasing function of consumers’ type. Consequently, it is no longer sufficient to ensure that the individual rationality constraint binds for the worst type only, as in the monopoly version. The participation constraint might turn out to be binding for several types. If the two firms are not local monopolies, the participation constraint will bind in an interval in the lower end of the type-space.

Although product differentiation enables the firms to implement price discrimination, it is shown that competition has important effects on the tariff structure. In the monopoly case, the increasing hazard rate assumption is sufficient to ensure that the outlay schedule is implementable in two-part tariffs. The hazard rate assumption is important because it affects the quantity profile. In the present model, when the participation constraint binds it determines the quantity profile and, hence, the increasing hazard rate condition is no longer sufficient. Although it is difficult to prove, the model seems to suggest that a fully separating equi-

librium can only be reached if the firms are allowed to use three-part tariffs in addition to two-part tariffs.

Chapter 5: Two-Part Tariffs, Consumer Heterogeneity, and Cournot Competition

In the last paper we assume that firms competing in an oligopoly sell a homogeneous good in a market with heterogeneous demand. As mentioned, extending models of nonlinear pricing to a context with competition is not trivial. The paper by Harrison and Kline (2001) explores competition with two-part tariffs in a strategic oligopoly setting. They extend the basic problem of charging a group of N identical consumers according to a two-part tariff instead of a uniform unit price, i.e., they extend the first part of the model in Oi (1971) to oligopoly. An important property of their model is the assumption that firms commit to a fixed fee in addition to capacity. Without this assumption it is not possible to escape a situation where a competing firm charges a uniform price and captures all consumers.

A natural extension of Harrison and Kline (2001) is to introduce demand side heterogeneity. Equivalently, this is to extend the second part of Oi (1971) to a competitive context.¹⁰ In the model presented in chapter 5, we extend Harrison and Kline (2001) to a oligopoly context with two groups of consumers having different willingness to pay for the good firms are marketing. Apart from this, we keep all their assumptions apart from that. In the Harrison and Kline model, when the firms are able to commit to a fixed fee, the properties of the outcome are a modification of the monopoly model only with respect to the fixed fee. When we extend the model to include demand side heterogeneity, the properties of the monopoly model are modified in several other respects. The paper demonstrates that an extension from one consumer type to two types is quite different in a monopoly and in an oligopoly. A monopolist can discipline its conduct vis-à-vis the two consumer group, while a firm competing in an oligopoly is incapable of instructing other firms' pricing and the firms commit to a quantity, which they will sell in any case.

In the paper we show that the main results in Harrison and Kline (2001) are reversed when the model is extended from one to two types of consumers. In particular, we find that the unit price can exceed marginal costs, and that the fixed fee can be below fixed costs. As in the monopoly model, large demand side heterogeneity results in higher unit price, in order to extract surplus from high demand consumers. Then, the access fee is low even in a monopoly setting and competed away in a duopoly. We also show that two-part tariffs may collapse,

¹⁰An extension of this model again, would be models with competitive screening. However, as pointed out before, this involves large difficulties and is kept outside the scope of the present model.

because each firm would rather commit to a traditional Cournot price system with zero fixed fee.

Finally, numerical examples illustrate that both firms serving both types of consumers can be an equilibrium outcome in duopoly, in cases where a monopolist would serve only one type of consumers. The examples also demonstrate that there can be multiple equilibria.

Chapter 2

Damaging Network Subscription

by Øystein Foros, Sissel Jensen and Jan Yngve Sand*

2.1 Introduction

Nonlinear pricing, giving high demand consumers quantity discounts, is a well-known practice in the telecommunications market. From both a normative and a positive point of view it may be desirable to use the demand heterogeneity to achieve price discrimination. If consumers' preferences were perfectly known, the firm would simply offer a set of take-it-or-leave-it contracts depending directly on the taste of individual consumers. The different contracts would typically specify an amount to be paid by the consumers to the firm and a quantity level to be provided by the firm. The levels would be set so that the firm would extract the entire social surplus from producing the good. However, the presence of asymmetric information forces the firm to offer contracts depending purely on observable variables, which prevents it from extracting the entire surplus from every consumer. The seminal paper by Mussa and Rosen (1978) shows that a monopoly enlarges the quality spectrum to separate consumers that value quality differently.¹¹ When the firm has incomplete information about consumers' willingness to pay for quality, a quality reduction towards consumers with low

*We are indebted to Bjørn Hansen, Marit Hareland, Eirik Gaard Kristiansen, Petter Osmundsen, Patrick Rey, Lars Sjørgard, Jon Vislie, participants at the 10th Summer School of the EEA in Toulouse (September 1999), as well as to an anonymous referee for helpful comments. The paper is a revised version of Foros, Jensen and Sand (1999). An earlier draft was presented at the June 1998 conference of the ITS (International Telecommunications Society) in Stockholm and at the 25th annual E.A.R.I.E. conference (European Association for Research in Industrial Economics) in Copenhagen in August 1998.

¹¹This was captured early by Dupuit (1849). Ekelund (1970) gives an overview of Dupuit's contribution to the understanding of the practice of price discrimination.

willingness to pay is used to induce self-selection. Similarly, an enlargement of the quantity spectrum can be used to separate consumers that value equal quantities of the service differently.¹²

The question we ask in this paper is whether it may be profitable for a firm to use the heterogeneity on the consumer side with respect to calling pattern (or in more general terms the customers' use of the service) as an additional sorting device to quantity. We set up a model which allows us to examine the effects of introducing "quality discrimination" in a two type model with quantity discounts. The objective is to show that a monopoly that has verifiable information that is correlated with the consumer's type can gain a profit that is at least as large as what would have been achieved if he was able to design contracts that depend not only on quantity purchase, but also on a second observable variable. Hence, the variable being contracted upon may have two dimensions or attributes. We will focus on the telecommunications market when we think in terms of application of our model, and the observable variables will be thought of as a consumer's quantity purchase and his calling pattern. We will interpret the second variable as a quality variable and use the terms high and low quality, although it may be argued that this sometimes will represent a slight abuse of terminology. Throughout the paper, we will assume that the firm knows for certain what kind of calling pattern (quality) some given consumer (some given quantity type consumer) prefers and that each consumer has distinct different preferences over calling patterns. Examples of individual calling patterns are call dispersion i.e., how many different subscribers a consumer makes calls to, and time-of-use, i.e., daytime calling, evening-/nighttime calling.¹³

Telecom companies have traditionally provided consumers with perfect interconnectivity through a "fully featured" network subscription, i.e., with the possibility to communicate with any other member of the network. However, consumers' willingness to pay for different call termination points differs. Residential consumers most often pick up the phone to call a friend or a family member. On the other hand, a given business call could be terminating virtually anywhere, just think of a call from a phone marketing company. In addition, it is reasonable to believe that a business customer has a higher willingness to pay for a given quantity of the service compared to a residential consumer.

We start with the familiar model of second-degree price discrimination, pre-

¹²Maskin and Riley (1984) show that discrimination along a single dimension in quantity is very similar to discrimination along a single dimension in quality. They also show that the results in Mussa and Rosen (1978) apply to more general utility functions.

¹³Although calling pattern is *ex ante* private information, the firm is able to reveal any possible correlation between a consumer's quantity purchase and for example call dispersion by analyzing the customer's call records. We will assume that the two factors are positively correlated and that the problem can be reduced to a type-space that is single-dimensional. In particular, we shall assume that a consumer with low willingness to pay for quantity makes calls to a small number of subscribers, while a consumer with high willingness to pay for quantity makes calls to a large number of subscribers.

sented as a two-type case where the monopoly uses quantity as the single sorting device (following Tirole (1988) and Fudenberg and Tirole (1991), based on Maskin and Riley (1984)). Consumers have either high or low willingness to pay for some given quantity level of the service, and a consumer's willingness to pay for the monopoly's product is private information.¹⁴ In the two-type case, the firm might find it optimal to exclude low quantity types in order to reduce the information rent to the high-quantity types. Next, we expand the model and allow for consumer heterogeneity along a second dimension. This dimension takes into account that the service can be sold with different quality attributes, or on different terms.¹⁵ In our context, the interpretation of quality differences is a possible restriction on consumers' calling pattern, for example a restriction on the distribution of a consumer's outgoing calls.

It may be argued that the firm's way of introducing quality differences follows the notion of damaging. In fact, we will use damaging as a notation when the firm conditions a contract on the use of the service. Damaging occurs when a producer who initially offers a high quality product creates a low quality product by reducing the quality of his initial product. Deneckere and McAfee (1996) analyze damaging in a framework of uniform pricing and with the presence of two consumer types having high or low willingness to pay for some given quality.¹⁶ In such a framework low demand consumers might remain unserved when the firm offers only the high quality version. The assumption that the low demand segment is not served in the absence of the low quality product is crucial in the Deneckere and McAfee article. With the introduction of a low quality version, low demand types gain a positive utility since the demand function decreases in price. In addition, the firm may be forced to reduce the price of the high-quality good in order to deter consumers in this segment from buying the inferior good (the incentive constraint is binding).

In many markets where damaging occurs, it seems relevant to argue that linear pricing is an artificial restriction on the firm's pricing strategy. For that

¹⁴An alternative formulation is that the firm has complete information, but operates subject to a regulation saying that consumers can choose among all contracts offered by the firm. In other words, the firm cannot offer a take-it-or-leave-it contract exclusively to each consumer type.

¹⁵See Laffont and Tirole (1993, part 4.2.2) for a treatment of a regulation problem with multidimensional informational asymmetry, but where the unknown parameters enter the relevant functions through linear combinations. Introduction of a second type-dimension complicates the analysis considerably because the first-order conditions for consumers' maximization (incentive compatibility) can be difficult to incorporate as constraints in the firm's optimization problem, see Laffont et al. (1987), Rochet and Choné (1998), Armstrong and Rochet (1999), Wilson (1993).

¹⁶They also analyze a second case where the consumers' reservation price for high and low quality varies according to a continuous type parameter. They refer to the two different cases as "the dual use" case and "the single use" case. Basically, this gives the same conclusions about damaging leading to a Pareto improvement, but the conditions are more stringent and harder to arrive at with a continuous type parameter.

reason we let the monopoly firm charge nonlinear prices and search for the profit maximizing allocations in fully nonlinear take-it-or-leave-it contracts. This affects several aspects of the model in Deneckere and McAfee (1996). First, the likelihood of low demand consumers being served increases when the firm practices nonlinear pricing. Nevertheless, even if low demand consumers are served absent damaging the firm still has an incentive to reduce the quality towards low demand consumers. Hence, we can relax the crucial assumption about low demand consumers being unserved absent damaging. We show that the incentive to practice damaging is qualitatively different when low demand consumers are served absent damaging and when they are not. Secondly, since nonlinear pricing has welfare effects that are very different from the welfare effects under linear pricing, the consequences of damaging are also different in our model compared to Deneckere and McAfee.

Henriet, Henry, Rey and Rochet (1988) discuss an issue similar to the one we analyze. They examine the effects of discrimination with respect to price and the variety of the products offered in a model with unit demand. One of their main tasks is to examine the welfare effects of creating artificial differences in the characteristics of the product. They set up a model with variable quality and with production costs independent of quality, where consumers prefer a higher quality. They show that a welfare-maximizing monopoly would want to discriminate, whereas a profit-maximizing monopoly would not. Their results underline the fact that discriminating policies are not necessarily a result driven by profit concerns.

The model we present is an example of the fact that increased observability enhances the principals' ability to separate the different consumer types, as shown in Holmström (1979). A firm may find it profitable to make investments that enables it to reveal some information about a consumer's type, and subsequently use this information in the contract. Although in most situations a perfect estimation of a consumers' type is either impossible or prohibitively costly, imperfect estimation can improve profits – additional information is always of value to the firm because it allows a less costly separation of types. The profit a firm can gain by using two kinds of screening variables is at least as large as if it chooses not to use one of them. In our telecom example, the screening mechanisms are to monitor consumers' quantity purchase and calling pattern. Matthews and Moore (1987) address the screening problem in a context where consumers have one characteristic but where the monopoly can use several instruments. They consider a model where a good can be sold with different attributes (quality and warranty), where consumers vary in their evaluation of these attributes, and develop a technique for dealing with incentive compatibility between nonadjacent types. However, this matter is simplified in our model since we are dealing with two types only, so nonadjacent types do not exist.

In section 2 we present a simple model on damaging relevant for the telecommunications market. Suppose a monopoly selling the high quality product chooses to serve the high demand consumers only. When quality is introduced as an ad-

ditional sorting device, the firm may find it in its own interest to serve the low demand types. If so, the firm damages the version to extract rent from low demand consumers. Furthermore, nonlinear pricing creates an additional incentive to practice damaging. The firm can reduce the information rent achieved by high demand consumers if low demand consumers are offered a low quality product instead of a high quality product. The reason is that damaging reduces high demand consumers' incentive to mimic a low demand consumer.

Section 3 of this paper illustrates two examples of damaging in the telecommunications market. One example that fits the model is a set of specific optional calling plans, familiar to most people as *Friends and Family* tariffs. By accepting a restriction on the number of call termination points, a residential consumer could gain about the same discount as a large business customer. On the other hand, the restriction on call distribution implies that *Friends and Family* tariffs are less attractive to the business segment than they are to private subscribers.¹⁷ Another example is a practice known in the cellular market. Firms operating the DCS1800-technology have started to offer an optional calling plan denoted as a *home zone tariff*. By accepting a restriction on mobility, a consumer can call at a discounted rate. This enables the firm to compete with the fixed link technology. On the other hand, the restriction on mobility implies that the *home zone tariff* is less attractive to the segments that value mobility very highly. Finally, section 4 summarizes the main conclusions.

2.2 A model with Damaging

Implementation of price discrimination requires that the seller has some degree of monopoly power and that resale possibilities are limited or absent. For simplicity, we assume that the firm is a monopoly and that resale markets are absent. Hence, the monopoly simply offers take-it-or-leave-it contracts to the consumers without engaging in any negotiation with them. If the monopoly has complete information, it can discriminate perfectly and thus extract all surplus. However, we assume that the monopoly faces an information constraint and perfect price discrimination is then ruled out. The monopoly cannot tell consumers apart and this introduces the “self-selection” or “incentive-compatibility” constraint in the problem.

There are two types of consumers, type 1 in proportion λ_1 and type 2 in proportion $(1 - \lambda_1)$. Type 2 always has a higher marginal willingness to pay

¹⁷Friends and Family tariffs are usually seen as a mechanism for creating lock-in effects in the competition between network operators, see for instance Michell and Vogelsang (1991), 196. Once a subscriber's “community of interest” is subscribers of the same network, a calling circle tariff also creates switching costs. More generally, when a firm charges different prices for calls terminating on a subscriber's network than for those terminating on a rival's network it gives rise to network externalities. Such a pricing strategy is recently studied in Laffont, Rey and Tirole (1998).

for outgoing calls than type 1. Thus, outgoing calls are an increasing function of type. The monopoly can produce two vertically differentiated products, a product with high (H) and a product with low (L) quality. The number of possible call termination points represents the quality level of the products. With H one can reach the entire network and with L only a small fraction, the quality level of L is exogenous (i.e., the number of other subscribers one can reach with the L product). Furthermore, we assume that the consumer types differ in their use of the network subscription, i.e., calling pattern. The two types could be thought of as two distinct market segments. If we think of a type 1 consumer as a private customer he will have a relatively low willingness to pay for the possibility to make outgoing calls to the total network. If we think of a type 2 consumer as a business customer, like a phone marketing company, he will necessarily demand the possibility to call every subscriber in the network. We make the following assumption about the differences in calling patterns: type 1 consumers make calls to a small fraction of the network, whereas type 2 consumers make calls to the majority of the network. This implies that a type 2 consumer values high quality more than does a type 1 consumer and consumers' utility depends on the quantity purchased and the quality of the service. The assumptions on calling patterns are realistic if the business consumer is a phone marketing company. However, if the business consumer is a cab company or a pizza parlor the assumptions on calling patterns is violated.

To focus on price discrimination we have chosen to abstract from network externalities.¹⁸ The following assumptions eliminate both the access and the call externalities.

1. The network has N subscribers, N being fixed and exogenous. This assumption allows us to ignore the interdependencies between the pricing problem and the network size, i.e., the access externality. Even if type 1 consumers are excluded from making outgoing calls, or buying the low quality product, other consumers in the network can call them. This is equivalent to saying that all consumers have paid the one-time installation fee and the monthly fee for the basic service.
2. Both types have the same utility or disutility from receiving a call. Thus, the utility from receiving a call is independent of type (although the utility from outgoing calls is type-dependent). Initially, it seems realistic to assume that if the utility from an outgoing call is a function of type, the utility from an incoming call should be a function of type too. However, outgoing and incoming calls can be quite different services (see (3) below).

¹⁸Network externalities are an important issue in telecommunications and are often divided into access externalities and call externalities. The access externality encompasses the fact that a consumer's valuation of a network subscription increases as additional subscribers are connected. The call externality is the benefit of being called without paying for the call.

3. A consumer makes an active choice of consumption when he makes an outgoing call. With ordinary rationality assumptions he does not make outgoing calls that give him negative net utility. The situation is quite different for incoming calls. If for a moment we leave the digital age with number identification, answering a call may bring both positive and negative experiences for most people. Thus, our assumption is that the aggregate utility from receiving calls is zero for each consumer. In our context this implies especially that type 2 consumers are unaffected whether type 1 consumers are served or not (i.e., whether type 1 consumers can make outgoing calls or not).

Since the assumptions (1), (2) and (3) eliminate both the access and the call externalities, we can define \bar{U} as a type independent reservation utility and we assume that $\bar{U} \equiv 0$. \bar{U} represents the consumer surplus from being connected to the network (i.e., the utility from receiving calls).

Given that all consumers subscribe to the basic service, we want to explore the strategy of a firm which uses quality in addition to quantity as a means to separate the consumer groups. The high quality service gives a quantity discount to high quantity users, whereas the low quality service gives a “calling pattern discount” to low quantity users. Hence, the model restricts the consumer’s option by saying that he must buy either high or low quality, i.e., he cannot buy a combination of high and low.

The variable $q_i = \{q_i^L, q_i^H\}$ is a quantity vector for outgoing calls of the two qualities H and L ($i = 1, 2$) and, according to what we said previously, we restrict our attention to the case where $q_i = \{(0, q_i^H), (q_i^L, 0)\}$. If the firm were to sell only one quality, it would certainly choose to sell H . Also, if the firm finds it profitable to sell the low quality product, this will always be intended for type 1 consumers. The question is which of the bundles $q_1 = \{0, q_1^H\}$, $q_1 = \{0, 0\}$, or $q_1 = \{q_1^L, 0\}$, the monopoly will offer these consumers. In order to make the notation simple we define $q_1^L = \{q_1^L, 0\}$ and $q_1^H = \{0, q_1^H\}$. Total payment for the quantity q_i^j is T_i^j , ($i = 1, 2, j = H, L$). Since the monopoly will always sell high quality (H) to type 2 consumers, the contract designed for these always has $q_2 = q_2^H$ and payment T_2^H . With asymmetric information the firm has to design contracts for the two consumer groups in such a way that it is optimal for each consumer type to reveal private information through the choice of contract. When consumers can have either high or low willingness to pay for quantity, the consumer with low willingness to pay has no incentive to claim that he has high willingness to pay. The problem is to induce type 2 consumers to reveal their private information. The theoretical presentation takes advantage of the relevant theoretical results within the theory of mechanism design, see Fudenberg and Tirole (1991, chapter 7) and Tirole (1988, chapter 3).

The consumers have the following quasi-linear preferences for H and L re-

spectively¹⁹

$$U_1 = \begin{cases} \theta_1 V(q_1^H) - T_1^H & \text{if he is served with } H \\ \theta_1 V(q_1^L) - T_1^L & \text{if he is served with } L \\ \bar{U} = 0 & \text{otherwise} \end{cases}, \quad (2.1)$$

$$U_2 = \begin{cases} \theta_2 V(q_2^H) - T_2^H & \text{if he is served with } H \\ \alpha\theta_2 V(q_2^L) - T_2^L & \text{if he is served with } L \\ \bar{U} = 0 & \text{otherwise} \end{cases}. \quad (2.2)$$

$\theta_2 > \theta_1$, $\alpha \leq 1$. The $V(\cdot)$ function is common knowledge. The argument in the $V(\cdot)$ function is a quantity variable being of either high or low quality (one of the vectors defined above). The function captures consumers' type independent quality ordering. The parameter θ_i is private information and indicates consumer type i 's willingness to pay for quantity. The α -parameter captures that quality preferences are also type contingent – because of type contingent differences in calling patterns. With our previous assumptions on the calling patterns α has to be lower than 1. This takes into account the fact that type 2 consumers call a significantly larger fraction of the network than do type 1 consumers. Thus, type 2 consumers would be harmed more if they had to reduce their call distribution. Consequently, type 2 consumers value H relative to L higher than do type 1 consumers. Hence, if we suppose that type 1 weakly prefers a given bundle of H to a given bundle of L , then type 2 strictly prefers H to L .

Assumption 2.1 *We make the following assumptions related to consumers' preferences and the firm's costs (where subscript indicates partial derivative)*

$$\begin{array}{ll} (a) & V_q(q) > 0, V_{qq} < 0 \\ (b) & V_q(q^H) \geq V_q(q^L) \\ (c) & \alpha\theta_2 \geq \theta_1 \\ (d) & c^L \geq c^H \\ (e) & V_q(0) \geq c^H/\theta_1 \end{array}$$

Assumption 2.1(a) implies that the utility function is strictly increasing and strictly concave. Assumption 2.1(b) captures the consumers' preference ordering which implies that the marginal utility of one extra unit of H is at least as high as the marginal utility of one extra unit of L . In our context, this is quite intuitive, since the possibility set, i.e., the termination points available, is higher with H than with L . When a consumer buys H , he can always imitate the calling patterns he would choose with L . However, if he buys L , he cannot always imitate the calling patterns he may choose with H . Most of the time, we will simplify

¹⁹Income effects are excluded in the model. However, this does not rule out differences in income between the two consumer types. The differences in income may be embodied in the type-parameter. In effect, we assume that the amount of money spent on a network subscription and usage of quantity is small relative to the total income for each consumer. In addition, q^H and q^L are variants of the same product and are independent.

the results by assuming that it makes no difference which of the vectors defined above that enters as the argument in the $V(\cdot)$ -function, i.e., $V_q(q^H) = V_q(q^L)$ for $q^L = q^H$.

Assumption 2.1(c) ensures that the high demand consumers always get higher net surplus than the low demand consumers for equal quantity-outlay allocations. Assumptions 2.1(a), (b) and (c) ensure that the single crossing condition holds, i.e., that the indifference curves of the two types cross only once, even when L is offered.²⁰ Then, Assumptions 2.1(a), (b) and (c) imply that if the monopoly sells one quality only, it will sell H . Hence, the monopoly will always offer type 2 consumers high quality. Assumption 2.1(d) indicates that L is a damaged (altered) version of H . Thus, L is at least as costly to produce as H . For example, *Friends and Family* programs imply extra costs of monitoring, registration and billing, and the introduction of L is not motivated by cost reductions. Assumption 2.1(e) ensures that type 1 is served with full information.

2.2.1 Single quality

Traditionally, telecom companies have not discriminated along the quality dimension but offered the *high* quality service with perfect interconnectivity between all subscribers. We will take this as a starting point and assume that the monopoly only offers high quality (H). The monopoly uses quantity as a sorting device and offers two take-it-or-leave-it contracts, $\{q_1^H, T_1^H\}$ and $\{q_2^H, T_2^H\}$. In a single quality model, we apply the standard concept of second-degree price discrimination described in Tirole (1988).

The monopoly's profit is²¹

$$\pi_{12}^{HH} = \lambda_1 [T_1^H - c^H q_1^H] + (1 - \lambda_1) [T_2^H - c^H q_2^H]. \quad (2.3)$$

The monopoly maximizes profit subject to a restriction that the consumers participate voluntarily – the participation constraint (PC_i) – and that each consumer chooses the contract, $\{q_i^H, T_i^H\}$, intended for his type, the incentive constraint (IC_i). If the first restriction is satisfied for type 1, it is automatically satisfied for type 2 since type 2 can always choose the contract intended for type 1 and get a higher net surplus. Hence, PC_1 is the only binding participation constraint in the problem, when both types are served. Further, the incentive constraint is only downward-binding, i.e., type 2 consumers should not want to consume type 1 consumers' bundle, hence IC_2 is the only binding incentive

²⁰The single crossing condition holds if the marginal willingness to pay increases in a single dimensional parameter. With our definition of the support of θ_i and α , the type parameter collapses into a single-dimension. Here, single crossing follows directly from Assumption 2.1(c) (given (a) and (b)). See Fudenberg and Tirole (1991), chapter 7.

²¹Subscript 12 indicates the type, and superscript HH indicates that both types are offered high quality.

constraint.²² When these two constraints are satisfied, each consumer type will choose the contract $\{q_i^H, T_i^H\}$ intended for his type

$$\theta_1 V(q_1^H) - T_1^H = 0, \quad (2.4)$$

$$\theta_2 V(q_2^H) - T_2^H = \theta_2 V(q_1^H) - T_1^H. \quad (2.5)$$

The firm maximizes profit with respect to q_1^H and q_2^H , subject to 2.4, 2.5 and both q_1^H and q_2^H must be nonnegative. By substituting T_1^H and T_2^H in the profit we can write the maximization problem

$$\begin{aligned} \max_{q_1^H, q_2^H} \quad & \pi_{12}^{HH}(q_1^H, q_2^H) \\ \text{s.t.} \quad & q_1^H \geq 0, q_2^H \geq 0. \end{aligned} \quad (2.6)$$

The following Kuhn-Tucker conditions will describe a global maximum

$$\theta_2 V_q(q_2^H) = c^H, \quad (2.7)$$

$$\begin{aligned} [\lambda_1 \theta_1 - (1 - \lambda_1)(\theta_2 - \theta_1)] V_q(q_1^H) - \lambda_1 c^H & \leq 0, \\ & = 0 \text{ if } q_1^H > 0. \end{aligned} \quad (2.8)$$

If type 1 is served and $q_1^H > 0$, we have

$$\theta_1 V_q(q_1^H) = \frac{c^H}{1 - \frac{1 - \lambda_1}{\lambda_1} \frac{\theta_2 - \theta_1}{\theta_1}}, \quad (2.9)$$

This is the standard result of second-degree price discrimination known as “no distortion at the top”. Type 2 consumers are given socially optimal quantity, while type 1 consumers face an efficiency distortion. The only way to reduce the information rent to a type 2 consumer is to make the offer to type 1 less attractive for type 2. The monopoly does so by reducing q_1^H . We see a trade-off between allocative efficiency and rent extraction (reward to type 2) and the monopoly sacrifices efficiency in order to reduce the information rent (and thus increase profit). We suppose that $V_q(q_1^H) > c^H/\theta_1$ for $q_1^H = 0$. This ensures that the monopoly serves type 1 with H under complete information (i.e., the first best solution). Provided that the demand side heterogeneity is not too large, both

²²This has a simple economic intuition. Both consumer types have private information about their preferences, but only type 2 consumers have valuable private information. To prevent type 2 from “cheating” and choosing the contract intended for type 1 he must be given an information rent. The monopoly will minimize the level of the information rent and therefore IC_2 will bind in the profit maximization problem. A type 1 consumer has no incentive to misreport his type and his private information has no value, hence, PC_1 is binding. See Fudenberg and Tirole (1991).

types are served if H is the only quality offer. The monopoly chooses to serve both types if

$$\theta_1 > (1 - \lambda_1) \theta_2 + \lambda_1 \frac{c^H}{V_q(0)}. \quad (2.10)$$

If condition (2.10) is violated, type 1 is offered the bundle $\{q_1^H, T_1^H\} = \{0, 0\}$. Condition (2.10) is simple to obtain by taking the right-hand limit of (2.9), as q_1^H approaches zero. If the right hand side is larger than the left hand side it is profitable to increase q_1^H . If type 1 is excluded, PC_2 will be the only binding constraint and the information rent is zero (there will no longer be any binding incentive compatibility constraints in the problem).

2.2.2 Damaging

With damaging the monopoly may offer an inferior substitute with restrictions on call distribution. This service will only be intended for type 1 consumers. Damaging introduces quality as an additional sorting device. In the assumptions made above, the two consumer types are heterogeneous with respect to their use of network subscription, the quality level of L is exogenous and consumers have a choice between H and L . The monopoly offers a menu of take-it-or-leave-it contracts, $\{q_1^L, T_1^L\}$ intended for type 1 and $\{q_2^H, T_2^H\}$ intended for type 2.

The monopoly's profit is

$$\pi_{12}^{LH} = \lambda_1 (T_1^L - c^L q_1^L) + (1 - \lambda_1) (T_2^H - c^H q_2^H), \quad (2.11)$$

where T_1^L and T_2^H are determined by the relevant participation and incentive constraints

$$\theta_1 V(q_1^L) - T_1^L = 0, \quad (2.12)$$

$$\theta_2 V(q_2^H) - T_2^H = \alpha \theta_2 V(q_1^L) - T_1^L. \quad (2.13)$$

The maximization problem can be stated as

$$\begin{aligned} \max_{q_1^L, q_2^H} \quad & \pi_{12}^{LH}(q_1^L, q_2^H) \\ \text{s.t.} \quad & q_1^L \geq 0, q_2^H \geq 0. \end{aligned} \quad (2.14)$$

From the new incentive constraint 2.13 it appears that if the firm chooses to serve type 1 some given amount of L instead of H it can extract larger rents from type 2. Since the participation constraint for type 1 is unchanged, the profit contribution from type 1 is also unchanged. Hence, if the firm serves type 1 with L instead of H revenues are increased. Consequently, if damaging has no cost, the firm will always sell low quality to type 1, i.e., damaging is an effective screening instrument provided that it is not too costly to use.

The first order condition with respect to q_2^H is unchanged from (2.7) and type 2 consumers are still offered the socially optimal quantity. If type 1 is served, the Kuhn-Tucker condition translates into deciding on q_1^L such that

$$\theta_1 V_q(q_1^L) = \frac{c^L}{1 - \frac{1 - \lambda_1 \alpha \theta_2 - \theta_1}{\lambda_1 \theta_1}}. \quad (2.15)$$

The first order condition with respect to q_1^L differs from the first order condition with respect to q_1^H . The reason is the effect of vertical differentiation (assumptions 2.1(b) and (c)), and the cost effect (assumption 2.1(d)). We assume that $V_q(q_1^L) > c^L/\theta_1$ for $q_1^L = 0$. The monopoly serves both types if

$$\theta_1 > (1 - \lambda_1) \alpha \theta_2 + \lambda_1 \frac{c^L}{V_q(0)}. \quad (2.16)$$

Since $\alpha < 1$, condition (2.10) is a stronger condition than (2.16) if c^L is close to c^H . We can state that (2.10) is a stronger condition than (2.16) provided that

$$\frac{1 - \lambda_1}{\lambda_1} \theta_2 V_q(0) [1 - \alpha] \geq c^L - c^H. \quad (2.17)$$

Then, we might have a case where (2.10) is violated, whereas (2.16) is satisfied. That is, the monopoly excludes type 1 if it serves him with H , but offers him a positive amount if it serves him with L .²³

In assessing the implications of introducing an inferior substitute, it turns out to be crucial whether the monopoly serves both types or only type 2 consumers with H .

Case i: Type 1 is not served when only high quality is offered, i.e., type 1 is served with L *instead of being excluded*. This implies that (2.10) is violated, but (2.16) is satisfied. π_{12}^{0H} is the benchmark for the monopoly when it considers introducing a damaged good, and damaging is profitable if $\Delta\pi = \pi_{12}^{LH} - \pi_{12}^{0H} > 0$.²⁴

Case ii: Type 1 is served when high quality is the only offer, i.e., type 1 is served with L *instead of H* . This implies that both (2.10) and (2.16) will be satisfied. π_{12}^{HH} is the benchmark and damaging is profitable if $\Delta\pi = \pi_{12}^{LH} - \pi_{12}^{HH} > 0$.

²³Note that with the contract $\{q_1^H, T_1^H\} = \{0, 0\}$ type 1 has *de facto no* termination points available. Therefore, from type 1's point of view the contract $\{q_1^H, T_1^H\} = \{0, 0\}$ is even more damaged than the contract $\{q_1^L, T_1^L\}$ where $q_1^L > 0$.

²⁴ π_{12}^{LH} is the profit when type 1 is served with a positive quantity of L in (2.14), i.e., $\pi_{12}^{LH} = \lambda_1 (\theta_1 V(q_1^L) - c^L q_1^L) + (1 - \lambda_1) [\theta_2 V(q_2^H) - (\alpha \theta_2 - \theta_1) V(q_1^L) - c^H q_2^H]$. Whereas π_{12}^{0H} is the profit when type 1 is excluded and PC_2 is the only binding constraint, i.e., $\pi_{12}^{0H} = (1 - \lambda_1) (\theta_2 V(q_2^H) - c^H q_2^H)$.

The distinction between case *i* and *ii* represents two different incentives behind a practice of damaging. If it is to serve consumers that absent damaging would not be served, it extracts rent from type 1 consumers (case *i*) – although the information rent to type 2 increases. A different incentive to practice damaging exists because it enables the firm to reduce the information rent to type 2 – although the cost of serving type 1 increases (case *ii*).

Case i

Type 1 is served with L instead of being excluded, and damaging has two effects, (1) it enables the firm to extract surplus from type 1, but (2) increases the information rent to type 2.

Evaluated globally, damaging is profitable if $\pi_{12}^{LH} > \pi_{12}^{0H}$

$$\lambda_1 (\theta_1 V(q_1^L) - c^L q_1^L) > (1 - \lambda_1) (\alpha \theta_2 - \theta_1) V(q_1^L). \quad (2.18)$$

When quantity is the only sorting device q_1^H completely determines the information rent to type 2. The only way to reduce the information rent is to decrease the quantity to type 1. If the monopoly offers the damaged service L to type 1, this will influence the information rent to type 2 because the two goods are offered on different *terms*. Rather than using quantity distortions to reduce information rent, the monopoly may introduce quality distortions too. The likelihood of profitable damaging increases with the difference in preferences for L between type 1 and type 2, in other words, if α decreases. Since α is less than 1, the information rent to type 2 for a given quantity of L to type 1 is lower than if the same quantity of H were offered to type 1. The cost of damaging will of course influence the profitability of damaging. In addition to the reduction of the information rent, the monopoly may be able to increase profits from type 1 by offering a positive amount of the damaged service.

The firm's incentive to practice damaging is positive if it will serve type 1 with a marginal positive quantity Δq_1^L instead of $q_1^H = 0$, i.e., if²⁵

$$\theta_1 > (1 - \lambda_1) \alpha \theta_2 + \lambda_1 \frac{\frac{0 + \Delta q_1^L}{\Delta q_1^L} c^L}{V(0 + \Delta q_1^L) - V(0)}, \quad (2.19)$$

taking the limit as Δq_1^L goes to zero gives us a “serve type 1 with L instead of zero” condition, which is of course a re-statement of (2.16), the condition that type 1 is being served with L .

Hence, for θ_1 in some interval $[\theta_1', \theta_1'']$ the firm will serve type 1 with L instead of H , provided that this interval exists such that $\theta_1' < \theta_1''$. The boundaries are determined by $\theta_1' = (1 - \lambda_1) \alpha \theta_2 + \lambda_1 (c^L / V_q(0))$ and $\theta_1'' = (1 - \lambda_1) \theta_2 + \lambda_1 (c^H / (V_q(0)))$.

²⁵From now on we will use the assumption that $V(q^L) = V(q^H)$ for $q^L = q^H$.

The interval becomes smaller the higher is α and the higher is c^L , for α equal to 1 such an interval never exists, – obviously, since damaging has no effect. Formally, we can show that both $c^L \geq c^H$ and $\theta'_1 < \theta''_1$ cannot be true when $\alpha = 1$. If c^L is strictly above c^H , α can be strictly below 1 and damaging will still not be profitable.

If damaging occurs, we know that the monopoly makes more profit. Furthermore, type 2 is better off since he derives a positive surplus from consumption. Type 1 derives no net surplus in either case. Thus, there is a Pareto-improvement. The paradox here is that it is the practice of price discrimination and of costly damaging that leads to the Pareto-improvement. This case corresponds to the dual use case in Deneckere and McAfee (1996) where they show that a Pareto-improvement occurs when the type 2 segment is substantially more profitable than the type 1 segment. Hence, absent damaging, type 1 consumers would not have been served. However, a Pareto-improvement arises for different reasons. In the Deneckere and McAfee (1996) article, it arises from the fact that if low demand consumers are served, they have to extract a positive surplus when they are charged a linear price. In our model a Pareto-improvement arises not because type 1 consumers benefit, but entirely because type 2 consumers earn a positive information rent when type 1 consumers are served.

Case ii

What are the implications of introducing an inferior substitute L when both types would be offered a positive amount even when only H were offered? Type 1 is served with L instead of H , and damaging has two effects, (1) it increases the cost of serving type 1, but (2) decreases the information rent to type 2.

In case *ii*, evaluated globally damaging is profitable if $\pi_{12}^{LH} > \pi_{12}^{HH}$

$$(1 - \lambda_1) \theta_2 [V(q_1^H) - \alpha V(q_1^L)] - \theta_1 [V(q_1^H) - V(q_1^L)] > \lambda_1 (c^L q_1^L - c^H q_1^H). \quad (2.20)$$

The left-hand side of equation (2.20) represents the net gain from damaging. The cost effect appears on the right-hand side. The first part of the left-hand side is the change in the information rent and the second part is the change in type 1 consumers' willingness to pay. Type 1 consumers' willingness to pay is affected by a possible change in quantity ($q_1^H \neq q_1^L$) and by vertical differentiation, ($V(q_1^H) \geq V(q_1^L)$, for $q_1^H = q_1^L$). The right-hand side is a pure cost effect which is nonnegative if $q_1^L > q_1^H$ since $c^L \geq c^H$. If the effect of vertical differentiation is small and the cost difference between producing L instead of H is small, then the net effect of introducing damaging will more likely be positive, i.e., the reduction in information rent to type 2 will more likely outweigh the costs of damaging.

It turns out that the incentive to practice damaging is not as easy to identify as in the previous case. Whether the incentive to produce L is (marginally)

stronger than the incentive to produce H must be evaluated globally and will therefore give less intuition. Ideally, we would like to compare the marginal profits evaluated at $q = 0$ and use the condition $\partial\pi_{12}^{LH}/\partial q \geq \partial\pi_{12}^{HH}/\partial q$ to assess whether the monopoly will serve type 1 with L instead of H . (As long as we assume that q_2^H is chosen optimally, profit π_{12}^{jH} is a function of only q_1^j , $j = L, H$). However, it will be sufficient to do so only under the condition that π_{12}^{HH} is more concave as well. If π_{12}^{LH} is the most concave of the two profit functions, it might well reach its maximum at a lower profit level. The profit function π_{12}^{LH} is more concave than π_{12}^{HH} if

$$\frac{-\frac{d^2}{dq^2} [\pi_{12}^{LH}(q)]}{\frac{d}{dq} [\pi_{12}^{LH}(q)]} \geq \frac{-\frac{d^2}{dq^2} [\pi_{12}^{HH}(q)]}{\frac{d}{dq} [\pi_{12}^{HH}(q)]}. \quad (2.21)$$

This is a measure of concavity that is invariant to positive linear transformations. Necessary and sufficient conditions for L to be chosen by the firm can be stated as

$$\alpha \leq 1 - \frac{c^L - c^H}{\theta_2 V_q(0)} \frac{\lambda_1}{1 - \lambda_1}, \quad (2.22)$$

$$\alpha \geq \frac{c^L}{c^H} - \frac{c^L - c^H}{c^H} \frac{\theta_1}{\theta_2(1 - \lambda_1)}. \quad (2.23)$$

If the right-hand side in (2.22) is smaller than the right-hand side in (2.23), comparing marginal profits for $q = 0$ yields a sufficient condition. However, we can show that this is not sufficient and the following condition is never met

$$1 - \frac{c^L - c^H}{\theta_2 V_q(0)} \frac{\lambda_1}{1 - \lambda_1} \leq \frac{c^L}{c^H} - \frac{c^L - c^H}{c^H} \frac{\theta_1}{\theta_2(1 - \lambda_1)}. \quad (2.24)$$

Doing some manipulation on this inequality results in the following condition

$$\frac{c^L - c^H}{c^H} \geq \frac{c^L - c^H}{c^H} \left[\frac{1}{(1 - \lambda_1)} \right] \left[\frac{\theta_1}{\theta_2} - \lambda_1 \frac{c^H/\theta_2}{V_q(0)} \right], \quad (2.25)$$

which can never be true because the product of the two brackets on the right-hand side exceeds 1. To see this we can check whether the contrary can be true

$$\begin{aligned} \frac{1}{(1 - \lambda_1)} \left[\frac{\theta_1}{\theta_2} - \lambda_1 \frac{c^H/\theta_2}{V_q(0)} \right] &\leq 1, \\ \theta_1 &\leq (1 - \lambda_1) \theta_2 + \lambda_1 \frac{c^H}{\theta_2 V_q(0)}. \end{aligned} \quad (2.26)$$

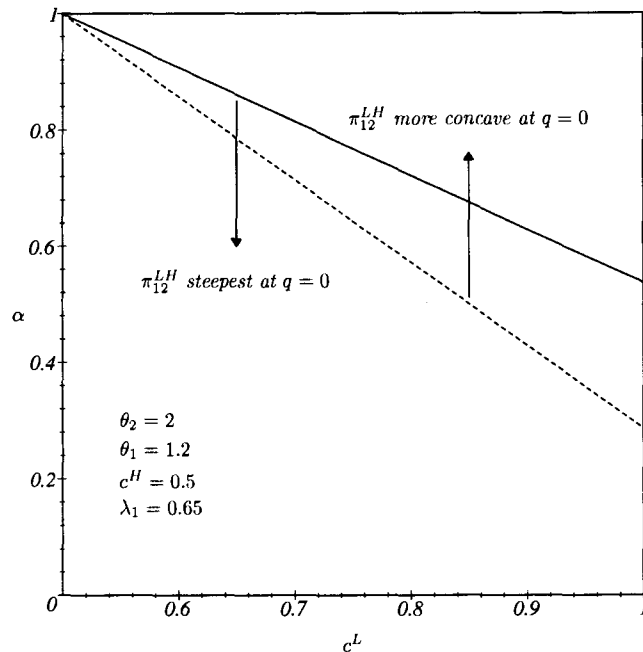


Figure 2.1: *Comparison of marginal profits and the concavity of the profit functions*

But since we are in a regime where type 1 would have been served with H , i.e., that $\theta_1 \geq (1 - \lambda_1)\theta_2 + \lambda_1 [c^H / V_q(0)]$, (2.26) yields a contradiction. The problem is illustrated in figure (2.1) in the case that $V(q) = q - \frac{1}{2}q^2$ and for a set of chosen parameter values.

Since the information rent is reduced in case *ii*, type 2 is worse off, and, contrary to case *i*, damaging cannot lead to a Pareto-improvement. Nevertheless, the monopoly may still find damaging profitable in this case. This incentive to practice damaging will only exist under nonlinear pricing and represents an extension of the analysis in Deneckere and McAfee (1996).

Let us briefly summarize the last section. If, absent damaging, the distortions towards type 1 consumers are so large that they are de facto excluded from consuming a positive amount, then the introduction of an inferior version increases the information rent to type 2. However, it might still be profitable for the firm because it can extract rent from type 1 consumers. Thus, such a strategy might lead to a Pareto-improvement. However, if type 1 consumers in fact consume a positive amount even absent damaging, the introduction of an inferior good is profitable because it leads to a decrease in the information rent to type 2. Because consumers value quality differently, a restriction on call distribution provides the firm with additional means of rent extraction. In the latter case, none of the consumer types gain, and type 2 consumers actually lose surplus from the pricing strategy under such circumstances. By adding a new type-dimension, the firm is

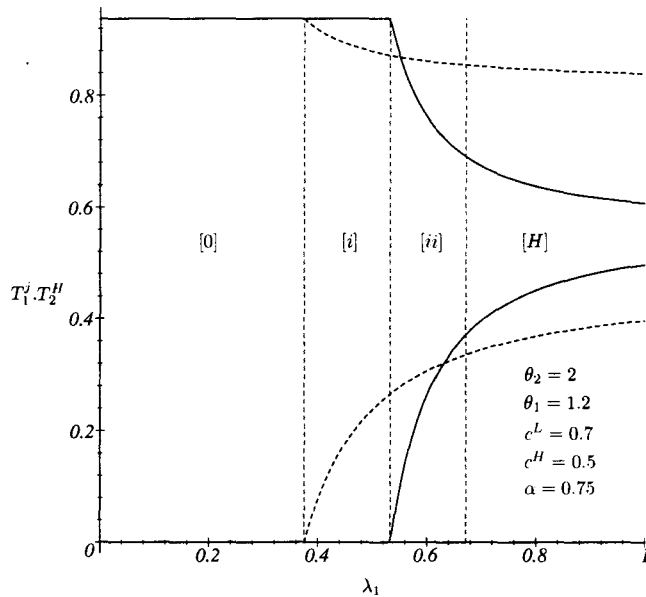


Figure 2.2: Payments from type 1 and 2, when $q_1 = q_1^H$ (solid line) and when $q_1 = q_1^L$ (dotted line). T_1^j is the lower schedules and T_2^H is the higher schedules.

able to reduce the price towards type 1 consumers without violating the incentive compatibility constraints.

The figures (2.2) and (2.3) illustrate our findings in the case that $V(q) = q - \frac{1}{2}q^2$. Figure 2.2 shows how the payments from the two types are affected by the introduction of L . In the interval marked $[0]$, where λ_1 is small, the firm excludes type 1 and extract all surplus from type 2. In the interval $[i]$ type 1 is served with H instead of being excluded. Type 2 pays less, but there is a net gain from introducing L that outweighs the cost. In $[ii]$ type 1 is served with L instead of H , and the firm extracts larger surplus from type 2. Finally, in $[H]$ the firm will not consider to serve type 1 with L .

2.3 Damaged goods in telecommunications

Telecom companies have traditionally offered a standard tariff which simply consists of a monthly fee (A) and a usage price (p). This is the standard tariff consumers need in order to be connected to the network and receive calls.²⁶ It is a high quality product in the sense that one can make calls to the entire network

²⁶If we abstract from variations in the usage price for long-distance/local calls and peak/off-peak hours, this is a two-part tariff as the *Standard* tariff in figure 2.4(a). The fee for being connected to the network and the monthly fee are set at a relatively low level. In most OECD-countries the telephony penetration is therefore very high, i.e. a large share of both the residential and business segments is connected to the network.

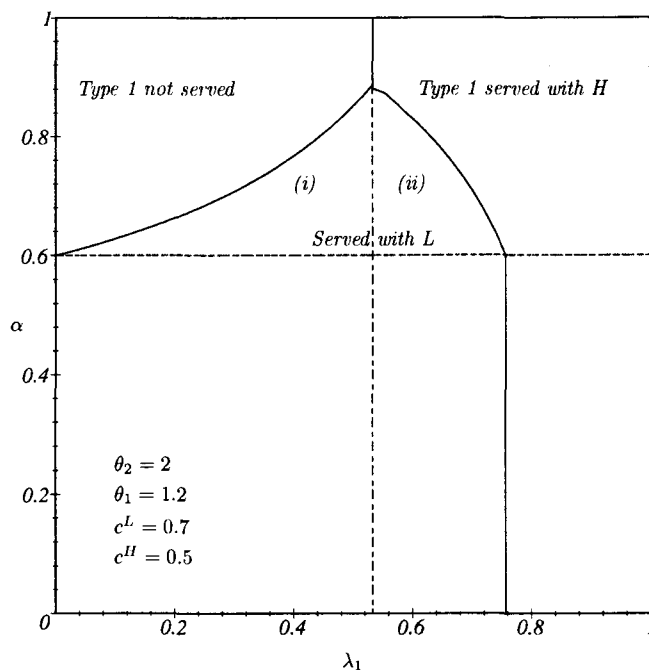


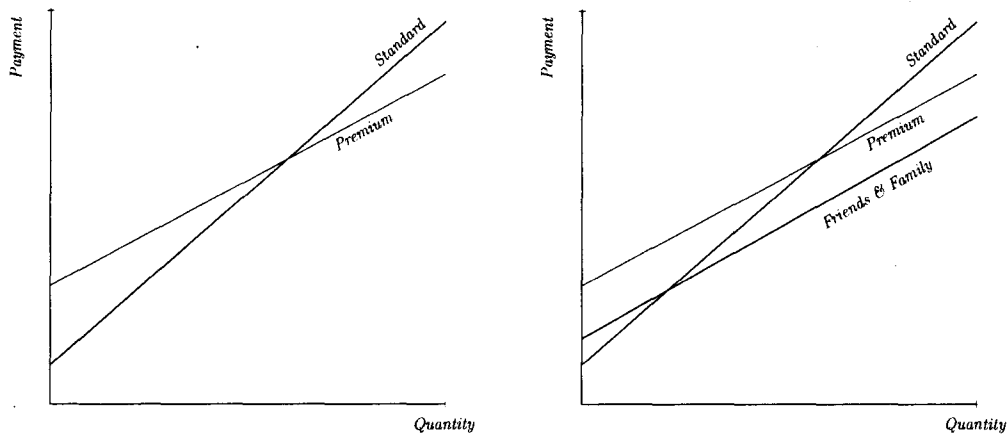
Figure 2.3: Configurations of λ_1 and α that lead the monopoly to exclude type 1, to serve him with L instead of exclusion (case (i)), with L instead of H (case (ii)), or to serve him with H. To keep the assumption that $\alpha\theta_2 \geq \theta_1$, α must be larger than 0.6.

and to interconnected networks. However, the usage price (p) has been relatively high, which to a large extent has excluded some network subscribers from making outgoing calls.

A series of additional tariffs have been introduced during the last decade, tariffs that are optional additions to the standard tariff. A large number of these tariffs provides pure quantity discounts to high quantity users. By paying an addition to the monthly fixed fee (A^P) the user obtains a reduction in the usage price (p^P), i.e. $p^P < p$. This is the *Premium* tariff in figure 2.4(a).

Another class of additional tariffs is what we denote *Calling circle* tariffs. Calling circle tariffs also provide quantity discounts at an additional monthly fee (A^{CC}). However, the quantity discount is only effective on a given number of pre-selected numbers, defined as the calling circle. Thus, the calling circle is a list of people that you call frequently, e.g. ten pre-selected numbers (members). Calls to these members are charged a discounted usage price (p^{CC}) and calls outside the calling circle are charged according to the *Standard* tariff (p).

In many real world examples of *Calling circle* tariffs both the monthly fee and the usage price are lower than in the *Premium* tariff, i.e. $A^P \geq A^{CC}$ and $p^P \geq p^{CC}$. Hence, the *Premium* tariff would not be implementable without the restrictions on call distribution in the *Calling circle* tariff. The restriction on call



(a) Single quality two-part tariffs

(b) Two-part tariffs with two qualities

Figure 2.4: *Nonlinear tariffs with damaging*

distribution is the reason why some customers prefer the *Premium* tariff rather than the *Calling circle* tariff.²⁷ For instance, if only the *Standard* tariff and the *Premium* tariff are offered, the private consumer would choose the *Standard* tariff, and his usage will be restricted because of the higher usage price. The phone marketing company is likely to buy the *Premium* tariff. The introduction of a *Calling circle* tariff enables the firm to reduce the usage price to the private consumer without tempting the phone marketing company to alter its tariff choice. Related to the model in chapter 2, a combination of the *Standard* tariff and the additional *Premium* tariff could be seen as the high quality product *H*. The combination of the *Standard* tariff and the *Calling circle* tariff will be analogous to the damaged product *L*.

The *Friends and Family* tariffs that are offered by most telecom firms are very similar to the original one offered by MCI. Although the specific features of the *Calling circle* tariffs in other countries may vary, the generic feature is very much the same. They all discount a certain calling pattern. The tariffs offered by Tele Danmark restrict the distribution of calls to six numbers, with discounts of 15 and 5 percent for off-peak and peak traffic respectively. The tariff discounts all off-peak traffic by 5 percent as well. The former Swedish monopoly, Telia, restricts call distribution to 3 numbers. Telia also offers tariffs that provide

²⁷The way of reducing the quality in the *Calling circle* tariff is in many ways analogous to that of rebated tickets in the airline industry where you are required to stay away from home Saturday night. This is more of a “punishment” to business travellers who want to get back home for the weekend, whereas it is not as big a problem for leisure travellers who often want to stay for the weekend. For leisure travellers, rebated tickets are almost perfect substitutes to the high quality product.

discounts on all calls to a specific geographic area.

The ever-growing market for information services may affect individuals' call distribution and calling pattern in general. Residential subscribers who are heavy users of the Internet demand at the extreme one point of call termination, the Internet Service Provider's point of presence. When local calls are billed by the minute users of the Internet become high quantity users of local calls. Our model suggests that it may be possible to design a flat rate plan (without a usage price) that fits the needs of the Internet users without creating significant revenue losses in the residential market as a whole (i.e., subscribers on the *Standard* tariff).

Empirical observations suggest that damaging is often introduced and continuously developed as a response to competition from a low quality product. The following section gives an example from the (competitive) cellular market in Denmark. The company Sonofon received its license to operate cellular services on a DCS 1800 network in Denmark in March 1997. In addition to offering mobile services, Sonofon wanted to offer a closer substitute to the fixed link telephony network subscription from Tele Denmark, the former PSTN monopoly. However, mobile services are priced to extract consumers' willingness to pay not only for quantity but also for mobility. In order to compete for Tele Denmark's fixed link subscribers (who presumably are subscribers who do not value mobility very highly) the mobile provider has to reduce the usage price in tariffs with low fixed fees. But – by doing so, this will cannibalize their revenues from high volume consumers of mobile telephony. Consequently, because Sonofon provides a service that has a high value to some consumers they became a “softer” competitor in the fixed link market.

To limit the cannibalization effect Sonofon has introduced a new tariff as a supplement to the standard mobile tariffs. The new tariff is designed such that a lower usage price is paid when the mobile phone is used from the subscriber's home zone or, more exactly, from the area inside or near the home, than from outside the home zone. The home zone is equal to the cell in which the subscriber's home is located. On calls from outside the home zone, the subscriber pays the standard price on the standard mobile tariff. The *Home zone* tariff is then an additional tariff where an additional monthly fee is paid in order to get a reduction in usage price for the calls made from home. In the DCS technology, the area that is covered by one cell is quite small.²⁸ This area becomes the home zone and is the area to which the *Home zone* tariff is related.

The novelty of the supplement is that it enables the firm to compete with the prices on fixed link telephony when consumers make calls from home, without making significant revenue losses from mobile calls originating from outside home.

²⁸Mobile or cellular networks consist of a grid of cells. The signals in DCS networks can not travel as far as the signals in GSM networks. Because the area covered by one cell is larger in GSM, this implies a greater cannibalization effect of mobile traffic (calls originated from outside home). Thus, the revenue loss from introducing a home zone tariff is probably greater in a GSM network than in a DCS network.

Because of the restrictions on the mobility in the tariff, the *Home zone* tariff can be viewed as a damaged product when mobility is regarded as a vertical quality dimension.

The *Home zone* tariff described above can be thought of as an “inverse” *Friends and Family* tariff. In a *Friends and Family* tariff a quality reduction is achieved by a restriction on call termination points, whereas a quality reduction in the *Home zone* tariff is achieved by a restriction on the point of call origination.

2.4 Concluding remarks

The paper focuses on the effects of a restriction on call distribution, used as a sorting device in a problem with nonlinear pricing. The firm can produce an inferior substitute by restricting the number of termination points (call distribution), i.e., a damaged service, where the damaging may be costly. If the damaged service is sold at a lower price than the high quality service, price discrimination occurs, and such a practice may under certain conditions give rise to a Pareto-improvement.

Whether damaging is profitable or not, depends upon the incentive behind it. If the incentive is to serve consumers that absent damaging would not be served, damaging enables the firm to extract rent from low demand types. This may happen in a case where the low demand consumers are excluded when only high quality is offered. Since the individual rationality constraint for these consumers binds in both cases, they are equally well off with or without damaging. High demand consumers have to be given a positive information rent if low demand consumers are served with a positive amount, and thus, they obtain increased utility. Hence, in this case the firm and high demand consumers are better off and damaging leads to a Pareto-improvement.

The other incentive to practice damaging exists because it enables the firm to reduce the information rent to high demand consumers. When low demand consumers are offered an inferior substitute instead of the high quality product, the incentive to mimic a low demand consumer is reduced. Hence, in this case the firm is at least able to extract larger surplus from high demand consumers. In addition, the efficiency distortions, via the pricing rule, towards low demand consumers are reduced. Depending on the costs of damaging and on low demand consumers' valuation of the damaged good relative to the high quality product, the firm is able to extract higher surplus from low demand consumers as well. So, in the end, damaging can not lead to a Pareto-improvement in this case, and both the profit and welfare effects are more ambiguous.

When firms interact strategically, damaging may be given other motivations than a way to separate consumers. As already mentioned, *Friends and Family* can be seen as creating lock-in effects or switching costs, and this may be part of the motivation behind *Friends and Family* in an oligopolistic market. In addition,

a firm that faces competition may want to introduce a new version (by damaging) to achieve a full line of products in order to deter entry (Deneckere and McAfee, 1996).

For simplicity, we have chosen to let the quality of L be exogenous, although this is not an accurate description in all cases. Sometimes the quality level is clearly endogenous. For example, the network operator sets the limitation on termination in the *Friends and Family* programs. On the other hand, it is natural to think of the quality level in the *home zone tariff* as exogenous.

From a more practical point of view, there are additional implications. A firm should design a high quality product such that it can be damaged, or versioned, later on. This point also stresses the importance of owning or controlling the network technology and ancillary systems, such as billing and registration, since this would enable firms to design products such as *Friends and Family* more easily (see also Varian (2000)).

Chapter 3

Two-Part Tariffs with Partial Unbundling

by Sissel Jensen*

3.1 Introduction

Consumers are often heterogeneous along numerous dimensions. In telecommunications, for example, consumers differ with respect to the quantity they purchase (minutes called) as well as in their calling pattern (whom they call, when they call, the duration of each call, etc.). With a few notable exceptions, the literature does not address the question on how a monopoly should price discriminate in a market with multidimensional heterogeneity.²⁹ The purpose of this paper is to

*I am grateful for comments from Petter Osmundsen, Fred Schroyen, Lars Sjørgard and Jon Vislie. I also thank seminar participants at the Norwegian School of Economics and Business Administration, and at the 3rd Nordic Workshop in Industrial Organization (NORIO III) in Helsinki for helpful comments.

²⁹Work on multi-dimensional screening includes different kinds of problems, see Rochet and Stole (2000). One polar case is when consumers are described by several characteristics but the firm has only one instrument at hand, references are Laffont et al. (1987), Lewis and Sappington (1989), Maggi and Rodriguez-Clare (1995), and Jullien (2000). The other polar case is when consumers are described by one characteristic but the firm can use several instruments, as in Matthews and Moore (1987) with risk-aversion, or as in Sappington (1983), and Caillaud et al. (1988) with several observables and instruments. Rochet and Choné (1998), and Armstrong and Rochet (1999) work within a model with several instruments and several characteristics, also providing an overview of the literature. Wilson (1993) provides definitions and examples of multidimensional goods and multidimensional pricing. Deneckere and McAfee (1996) and Foros et al. (1999) present models similar to the one presented in this paper; Deneckere and McAfee with uniform pricing and Foros et al. with nonlinear pricing, but in both articles the restriction on usage is exogenous.

explore how a monopoly might use two instruments to enhance the profitability from second degree price discrimination.

When consumers' willingness to pay is private information and the firm must condition the contract upon observable variables, it is most often assumed that the firm can observe only one variable. It is also common to assume that the observed variable is single dimensional, e.g., quality in Mussa and Rosen (1978) and quantity in Maskin and Riley (1984). Although being welfare improving compared to uniform pricing, the performance of second degree price discrimination relative to first best practice (marginal-cost pricing) depends upon the degree of the demand side heterogeneity.³⁰ In general therefore, any strategy that increases the observability of consumers' willingness to pay will increase profit and welfare.

In the present paper a monopoly firm sells a single generic good, for instance minutes of network usage. Each quantity increment can be assigned a unique list of observable attributes, such as time-of-day, distance, call termination point, etc, and this describes a consumer's calling pattern. By monitoring consumers' calling patterns, the firm is able to offer a tariff intended for low demand consumers on terms that differ from the terms on which high demand consumers make their purchases. Such practice can potentially improve "the observability" of consumers' characteristics in terms of self-selection, and thus implies less distortions towards low demand consumers.

It is possible to translate the multidimensionality implied by differences in calling patterns into a multiproduct setting by letting units assigned different sets of attributes be treated as different products. If the firm ignores the heterogeneity in consumers' calling patterns but charge all units the according to the same tariff, it aggregates all taste parameters and practice complete bundling. In the present context we let the firm bundle a subset of the products and charge units within this product bundle according to a different tariff. The firm do not debundle completely, and hence, we refer to this practice as partial bundling.³¹

We hold on to the assumption that consumers differ in their marginal willingness to pay for quantity and say that there are two types of consumers, high demand and low demand consumers. In addition, we assume that consumers with different willingness to pay also have distinctly different calling patterns. In particular, high demand consumers make calls to a large number of subscribers,

³⁰The price-cost margin depends on the range of the type-space and on the firm's prior beliefs about the distribution of types over this space. If the heterogeneity on the demand side is large then a large fraction of consumers pay a price well above marginal cost.

³¹In the multiproduct setting it would be the case that although the firm has imperfect knowledge about a given consumer's taste for one product, it knows that it is perfectly correlated with the taste for any other product. Miravete (2001) study multidimensional screening where different type components distinguish quality dimensions of products that can be aggregated. The ability to aggregate type components opens the possibility to reduce the dimensionality of the screening process. Sibley and Srinagesh (1997) explore the difference between screening the different dimensions of consumer types independently by means of two-part tariffs and the alternative of bundling all taste parameters to design a single two-part tariff.

whereas low demand consumers make calls to a small number of subscribers. High-dispersion subscribers can be thought of as business consumers while low dispersion subscribers can be thought of as residential consumers. The firm offers a menu of two-part tariffs, each specifying a fixed fee that must be paid up-front, a marginal usage price, and in addition use-of-service restrictions which consumers must obey. Customers choose their preferred tariff scheme and usage is subsequently billed according to this choice.

The predominant method of charging consumers for telecom usage has been to bill for the length of time a connection is used. All multi-dimensionality in the consumers' use of the service was translated into a single-dimensional quantity variable (pulses, and later minutes). The practice of sorting consumers with different willingness to pay for usage was handled by giving high demand consumers quantity discount, in consistence with single-dimensional screening models. Today the multi-dimensionality in usage patterns is to an increasing extent used to achieve separation. Tariff options known as Friends and Family and Best Friend are examples of discounts given on certain calling patterns. Other examples are telecom companies that offer discounts on dial-up internet access, in the form of discounts on standard calling rates or in the form of a monthly fixed fee for a fixed number of hours of usage, (flat rate dial-up internet access).³²

Firms' use of calling circle tariffs has received some attention in other areas in the economics literature. Wang and Wen (1998) consider a duopoly model with demand side heterogeneity, where such pricing behavior enables a new firm to enter the market despite the presence of consumer switching costs. This result is derived under specific assumptions about consumer calling patterns, specifically that low demand types make calls to other low demand types, whereas high demand types make calls to other high demand types. By relaxing this assumption one might conclude differently (see Klemperer (1995) for a survey on the switching cost literature). Laffont et al. (1998) examine the effects of discriminatory pricing on the negotiated interconnection agreements between rival network operators. When a network operator charges different prices for calls terminating on the subscriber's network and those terminating on a rivals network he can generate network externalities despite network interconnection.

Throughout the paper we shall hold on to a simple example applied to telecommunications and assume that consumers with different willingness to pay for the service have distinctly different calling patterns. Section 2 describes the generic features of telecommunications that are relevant for this paper. The aim in Section 3 is to give a definition of the quantity variable that the usage charge in a two-part tariff applies to. Section 4 presents the demand- and supply side conditions of the market as well as the informational constraints faced by the firm. Finally, in section 5 we draw conclusions from the analysis.

³²Dial-up internet access is in this way singled out as a separate product.

3.2 Telecommunications services

The telecommunications market has experienced rapid changes during the last decades. A large variety of services are nowadays provided on a common platform and the technology convergence gives rise to significant changes in the demand side of the market as well. New services at reasonable prices and more multifunctional customer premises equipment, for instance the world wide web, personal computers, and all applications on the web, have also led to large increases in the demand for transmission capacity (time length or more bandwidth). Built-in network intelligence and sophisticated monitoring of usage have enabled firms in the market to move from billing customers for single dimensional pulses to billing multidimensional minutes. The method of pricing a call used to be by a conversion from hour-of-day, day-of-week, distance, etc, to pulses by tables in the central office. The firm had no information about a consumer's demand other than the number of pulses consumed at the end of the billing period. Network technology and billing systems now price a call minute according to a detailed call record. Hence, the firm possesses very detailed information about a consumer's usage pattern.

The telecommunications network is a two-way network and a person or a machine that is present at one specific node asks for some type of communication with another specific node at some given hour, weekday, etc.³³ Even though a one minute call within a specific calling area is a perfectly standardized product, its point of destination is of vital importance to the consumer who makes the call. A consumer does not derive any benefit from a call which destines at a B-subscriber he did not intend to call.³⁴ The same feature also applies to information services generated at a specific network node. These are features of telecommunications that make it a multi-dimensional good. For instance, individual call records usually contain information about at what hour the call is made, who is the B-subscriber, and where the B-subscriber is located (local, long distance, international). Furthermore, subscribers are typically billed according to aggregate minutes (seconds) of peak-time long distance calling, off-peak long distance calling, peak-time international calling, etc. There are examples of service attributes that have an obvious ranking, e.g. if the attributes are different quality levels along a vertical dimension they are ranked the same way by all customers. However, this is not always the case with telecommunications. For instance, not all consumers prefer – at an equal price – to make calls at the same time of day. Service attributes such as the time-of-day or the node where a call

³³A phone call, an e-mail, a web-site etc with an objective to exchange, deliver or gather information). In this respect telecommunications is very different from other network industries, like electricity or water delivery. One kW/h of electricity injected at one point is a perfect substitute for one kW/h injected at any other point of generation.

³⁴The B-subscriber is the party being called whereas the A-subscriber is the party making the call.

terminates are attributes along a horizontal dimension and customers will rank them differently.

A widespread practice is to offer various kinds of calling circle tariffs. Under a calling circle tariff, a subscriber is billed according to aggregate minutes (seconds) of calling to specific B-subscribers (or specific network nodes), and the marginal price varies conditional on the node of termination. The following model aims at explaining and guiding the construction of such tariffs.

3.3 A two-dimensional good

Let q be a two-dimensional good $q = (n, x)$, where x is a quantity variable and n is a service attribute.³⁵ The vector q tells us how many units (x) with the given service attribute (n) a consumer did buy. When we sum over all possible service attributes (i.e., over every possible n) the sum is equal to a consumer's total demand, i.e., aggregate units of the generic good. This is to collect and sum up the x 's at every point in figure 3.1(a) (or 3.1(b)). Figure 3.1 gives two examples on representation of q . In the figure, x_a is the number of minutes a high demand consumer called network node n_a , and x'_a is the number of minutes a low demand consumer called network node n'_a . Note that the n -axis merely gives the identity of the party called (phone numbers) and is not ordered in any sense.

We represent a consumer's purchase set \mathbb{Q} by sorting along the attribute dimension, and describe this set with a "continuous boundary" $x(n)$. Using the telephone example again, and saying that the attribute assignment is network node (B-subscriber), sorting along the attribute dimension gives a presentation of the most called number, the second most called number, and so on.³⁶ We introduce heterogeneity on the demand side by assuming that a consumer's willingness to pay for the good is characterized by a privately known parameter θ , measuring the intensity of a consumer's valuation of quantity. A consumer θ includes in his purchase set \mathbb{Q} all increments for which his valuation $v(q, \theta)$ exceeds the marginal price p charged

$$\begin{aligned}\mathbb{Q}(\theta; p) &= \{q : v(q, \theta) \geq p\} \\ &= \{(n, x) : v((n, x), \theta) \geq p\}.\end{aligned}\tag{3.1}$$

³⁵A context with multidimensional products is similar to a multiproduct context since units assigned different sets of attributes can be treated as different products. The distinction between a multiproduct context and a multidimensional product context is that the consumer is allowed to custom design the service attributes by assigning each item his preferred attributes (termination node/B-subscriber, time of day, day of week etc) instead of choosing between a fixed and more constrained number of products. See Wilson (1993, part 3) for a description of multidimensional products and multidimensional pricing and for instance Armstrong (1996) on multiproduct pricing.

³⁶Such a presentation is only possible if the service attributes can be interpreted as cardinal levels and if n is continuous. The various attributes can not be along dimensions such as for instance color.

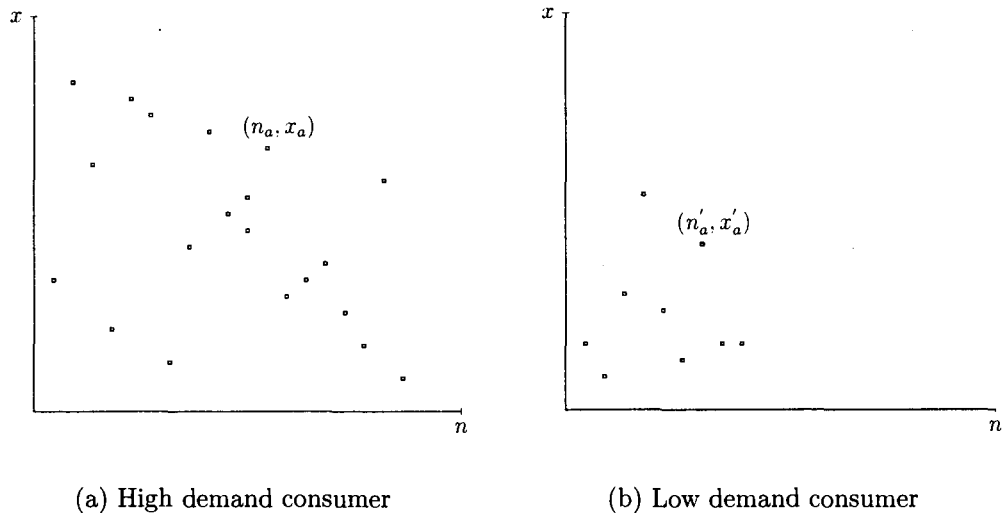


Figure 3.1: *Demand bundles. High demand consumers make calls of longer duration and to a larger number of network subscribers*

The boundary around the set given by (3.1) is those points where the marginal valuation equals the marginal price. Marginal valuation is given by

$$\begin{aligned}
 v((n, x), \theta) &= \frac{\partial^2 U(\cdot)}{\partial n \partial x}, \\
 v((n, x), \theta_1) &\leq v((n, x), \theta_2), \text{ when } \theta_1 < \theta_2.
 \end{aligned}
 \tag{3.2}$$

Type θ gains gross utility from consuming the purchase set $\mathbb{Q}(\theta; p)$ given by the double integral

$$U(\mathbb{Q}(\theta; p)) = \iint_{\mathbb{Q}} v((n, x), \theta) \, dn \, dx.
 \tag{3.3}$$

By saying that the boundary around the set can be represented by a monotonic function $x(n; \theta, p)$, which is continuous and everywhere differentiable, we can derive the demand from consumer type θ by solving a single integral over the attribute dimension. Aggregate demand for the generic good over all possible attribute levels is given by

$$Q(p, \theta) = \int_0^\infty x(s; \theta, p) \, ds,
 \tag{3.4}$$

and we define $Q_i(p) \equiv Q(p, \theta_i)$, $i = 1, 2$. Aggregate demand for the generic good with attribute level \bar{n} or lower, i.e., aggregate calls to the \bar{n} most called network nodes, is given by

$$\bar{Q}(p, \theta, \bar{n}) = \int_0^{\bar{n}} x(s; \theta, p) \, ds,
 \tag{3.5}$$

and similarly we define $\bar{Q}_i(p, \bar{n}) \equiv Q(p, \theta_i, \bar{n})$, $i = 1, 2$. If a consumer can customize demand freely, demand is given by (3.4). If he is to choose attribute levels within the interval $[0, \bar{n}]$, demand is given by (3.5), and $Q_i(p) \geq \bar{Q}_i(p, \bar{n})$, $i = 1, 2$. In the following we assume that $x(\bar{n}; \theta_1, p) < x(\bar{n}; \theta_2, p)$, $\forall p, \bar{n}$ for $\theta_1 < \theta_2$ and also that $x(\bar{n}; \theta, p)$ is monotonic. Hence $Q_2(p) > Q_1(p)$ and $\bar{Q}_2(p, \bar{n}) > \bar{Q}_1(p, \bar{n})$, $\forall p, \bar{n}$. Further, $\bar{Q}_i(p, \bar{n})$ is nonincreasing in p and nondecreasing in \bar{n} , and $Q_i(p)$ is also nonincreasing in p , $i = 1, 2$.

Using telephony as an example, heterogeneity in consumer demand is given by differences in call duration and call dispersion. We define call dispersion according to a cumulative distribution $F_1(n) \geq F_2(n)$ with a probability density function $f_i(n)$, $i = 1, 2$.³⁷ Hence, we make the assumption that call dispersion is independent of the price per call minute and that type 2 has a more dispersed calling pattern compared to type 1. Since we are only interested in the calls made by these two consumers, we can without loss of generality normalize the “entire network” to 1, and say that type 2 always makes calls to the entire network whereas type 1 has a more concentrated calling pattern.

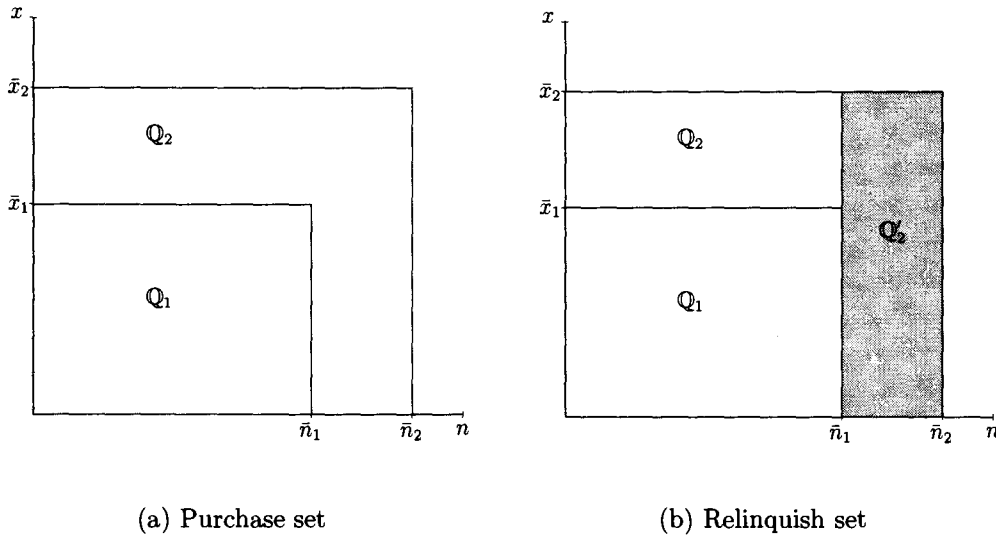


Figure 3.2: Rectangular purchase set (a) and the effect on type 2's purchase set of a restriction in call dispersion (b)

Figure 3.2 gives an illustration in the case of a rectangular purchase set. With a rectangular purchase set we have implicitly assumed that $f_i(n)$ is uniform on $[0, \bar{n}_i]$, $F_i(n) = n/\bar{n}_i$ ($\bar{n}_i = \{\bar{n}_1, 1\}$). The height of the rectangle measures the number of call minutes x to network node n . Type 1 makes calls to \bar{n}_1 different

³⁷The distribution of n conditional on θ_2 first-order stochastically dominates the distribution of n conditional on θ_1 , if $\theta_2 \geq \theta_1$. For notation we use $f_i(n) \equiv f(n; \theta_i)$, $F_i(n) \equiv F(n; \theta_i)$

network nodes, whereas type 2 makes calls to \bar{n}_2 different network nodes, i.e., \bar{n}_i is a measure of call dispersion. Figure 3.2 above reflects that there is heterogeneity in both call duration and call dispersion. If $\bar{x}_1 = \bar{x}_2$ all heterogeneity would be in call dispersion, whereas $\bar{n}_1 = \bar{n}_2$ would describe a situation with all heterogeneity in call duration. The shaded area \mathbb{Q}'_2 in figure 3.2 represents the part of type 2's ideal purchase set that he has to give up if he chooses a tariff with a restriction in call dispersion \bar{n}_1 .

3.4 The model

The market is served by a monopolist and resale opportunities are absent. The cost function is assumed to be linear, the fixed cost is excluded from the measure of profit and the marginal cost is normalized to zero. On the demand side there are only two consumers, type 1 with low willingness to pay and type 2 with high willingness to pay. A consumer's type is unobservable to the firm but each type's preferred calling pattern is known. We assume that type 2 has a more dispersed calling pattern than type 1. The types' call dispersion $f_i(n)$ is exogenous. The reservation utility is assumed to be equal for the two consumers and normalized to zero.

Because call dispersion is independent of the marginal price of a call minute, we can also write consumers' utility as a function independent of call dispersion. We use the following utility function that is quasilinear and quadratic in x ³⁸

$$U_i = \begin{cases} \theta_i x - \frac{1}{2}x^2 - T & \text{if they pay } T \text{ for } x \text{ minutes of calling,} \\ 0 & \text{if they do not buy.} \end{cases} \quad (3.6)$$

T is an increasing and continuous price schedule with a constant unit price $p = \{p_1, p_2\}$. Given information about each type's call dispersion, we derive expected call length to a network node n as $(\theta_i - p) f_i(n)$. Consumers' demand is thus given by

$$Q_i(p) = \int_0^1 (\theta_i - p) f_i(n) dn = (\theta_i - p), \quad (3.7)$$

$$\bar{Q}_i(p, n) = \int_0^n (\theta_i - p) f_i(n) dn = (\theta_i - p) F_i(n). \quad (3.8)$$

The density function f_i is positive and integrable on the support $n \in [0, 1]$ with a distribution function $F_i(n)$ with $F_1(n) > F_2(n)$, and $f_1 F_2 \leq f_2 F_1$. Aggregate

³⁸We abstract from the fact that some consumers may have positive utility even in the case when consumption is zero. A subscriber may want a network connection in order to receive calls only, or to be able to make emergency calls. Our assumption in this model is that if the expected net utility from making calls weakly exceeds a consumer's reservation utility he will find it beneficial to subscribe to the network. By assuming quasilinear utility we also ignore income effects.

demand for call minutes to the entire network is given by (3.7) and aggregate demand for call minutes to the n most frequently called nodes is given by (3.8). The latter case resembles the first, except that n affects the intercept and the slope of the individual demand curves. However, these are perfectly (negatively) correlated and the firm can infer about the slope when it knows the intercept (and vice-versa).³⁹

Consumer surplus under a two-part tariff $T_i = \{p_i, E_i\}$ for some given $n \leq 1$ is given by

$$CS_i(p_i, E_i, n) = \int_{p_i}^{\theta_i} (\theta_i - p) F_i(n) dp - E_i, \quad i = 1, 2, \quad (3.9)$$

$$CS_2(p, E, n) > CS_1(p, E, n). \quad (3.10)$$

When both types choose consumption subject to the same tariff, type 2 obtains a larger surplus given that $F_1(n)/F_2(n) \leq \theta_2/\theta_1$. Under this condition the demand curves of the two types never cross for any price. Since the demand curves are linear and $\theta_2 \geq \theta_1$, it is sufficient to evaluate the condition $(\theta_2 - p)F_2(n) \geq (\theta_1 - p)F_1(n)$ as p approaches zero.

When we solve the model we proceed in two steps. First, we solve for the optimal two-part tariffs, T_1 intended for type 1 and T_2 intended for type 2, treating n as exogenous. Next, having obtained a reduced form profit as a function of n we solve for the optimal restriction in call dispersion in the two-part tariffs T_1 and T_2 .

3.5 Two-part tariffs

Given the slopes of the demand curves and asymmetric information over θ the practice that maximizes profit is to offer different two-part tariffs intended for the two consumer types. We know equilibrium in this model as a solution where $p_1 > 0$ and $p_2 = c$. The fixed fee in type 1's tariff is chosen in such a way that he receives his reservation utility and the fixed fee in type 2's tariff is chosen such that type 2 does not choose the tariff intended for type 1. More formally, consider the model as follows. A two-part tariff is characterized by a triple $\{p_i, E_i, n_i\}$, p_i is the marginal price, E_i is a fixed fee and $n_i \leq 1$ is the fraction of the network that can be reached with the tariff. When the reservation utility is normalized to zero, it is individually rational to accept any tariff $\{p, E, n_i\}$ that yields nonnegative consumer surplus. The two individual rationality constraints are

$$CS_i(p_i, E_i, n_i) \geq 0, \quad i = 1, 2. \quad (IR_i)$$

³⁹Laffont et al. (1987) solve for the optimal nonlinear price schedule when a monopolist is uncertain about both the slopes and the intercepts of the individual demand curves it faces, assuming a continuum of types and that the distributions of slopes and intercepts are independent.

Since $CS_2(\cdot) > CS_1(\cdot)$, IR_2 can not bind whenever IR_1 is weakly met. Hence if type 1 is served, IR_1 is the only binding individual rationality constraint. The other relevant constraints are the incentive compatibility constraints

$$CS_i(p_i, E_i, n_i) \geq CS_i(p_j, E_j, n_j), \quad i, j = 1, 2, \quad i \neq j. \quad (IC_i)$$

The incentive constraint requires that a consumer buys the bundle intended for his type. IC_1 can never bind if IC_2 is weakly met. Hence, the incentive constraint is downward binding only.⁴⁰

It is never profitable to restrict type 2's demand and any restriction in call dispersion will only occur in the tariff intended for type 1. Henceforth we use the notations $n_2 = 1$ and $n_1 = n$. The firm is searching for two-part tariffs $\{p_1, E_1, n\}$ and $\{p_2, E_2, 1\} \equiv \{p_2, E_2\}$ in order to maximize profit. If the restriction on n is fixed we have the following maximization problem

$$\Pi = \max_{p_1, p_2, E_1, E_2} \left\{ E_1 + p_1(\theta_1 - p_1)F_1(n) + E_2 + p_2(\theta_2 - p_2) \right\} \quad (3.11)$$

subject to $p_i \geq 0, E_i \geq 0$ ($i = 1, 2$), IR_1 , and IC_2

$$E_1 = \int_{p_1}^{\theta_1} (\theta_1 - p) F_1(n) dp, \quad (3.12)$$

$$E_2 = E_1 + \int_{p_2}^{\theta_2} (\theta_2 - p) dp - \int_{p_1}^{\theta_2} (\theta_2 - p) F_2(n) dp. \quad (3.13)$$

The outcome is unique with $p_1 \geq p_2 = 0$, and $E_2 > E_1$, whenever $\theta_2 > \theta_1$ and both types are served. The last term in (3.13) illustrates the two instruments that can be used to reduce the information rent. The firm can increase p_1 or decrease n . If the firm chooses not to serve type 1, the unique outcome is a cost-plus-fixed fee tariff, $p_2 = 0$, and the entire consumer surplus is extracted via the fixed fee.

We can now turn to the question of how severe the restriction in call dispersion in type 1's tariff should be. As a benchmark however, we first repeat the profit maximizing two-part tariffs in the single-dimensional case with $n = 1$.

If the firm has no ability to monitor call dispersion, or to condition a tariff on a restriction in call dispersion, Q is treated as a single dimensional good, $n_1 = n_2 = 1$. This is the canonical model with two-types and single-dimensional screening which is examined in, for instance, Sharkey and Sibley (1993).

Lemma 3.1 (Single-dimensional screening) *A monopoly that is unable to observe anything but individual quantity purchases will increase the unit price in type 1's tariff above marginal cost in order to reduce the information rent to type 2. If consumer heterogeneity is too large, the monopoly will exclude type 1 from buying.*

⁴⁰See for instance Tirole (1988) pp 153-154, and Fudenberg and Tirole (1991), pp 247-248.

(i) For $\frac{\theta_2}{\theta_1} \in [1, \frac{3}{2}]$ the monopoly will serve both types and offer two different two-part tariffs $\{p_1, E_1\}$ and $\{0, E_2\}$ given by

$$p_1 = \theta_2 - \theta_1, \quad E_1 = \frac{1}{2} (2\theta_1 - \theta_2)^2, \quad E_2 = \frac{1}{2} (2\theta_1 - \theta_2)^2 + \frac{1}{2} (\theta_2^2 - \theta_1^2).$$

(ii) For $\frac{\theta_2}{\theta_1} > \frac{3}{2}$ the monopoly will exclude type 1 and offer a cost-plus-fixed-fee tariff $\{0, E_2\}$ and extract all surplus from type 2. The tariff is given by

$$E_2 = \frac{1}{2} \theta_2^2.$$

Lemma 3.1 is simple to verify by substituting for $F_1(n) = F_2(n) = 1$ in the above maximization problem (3.11)–(3.13). The information rent to type 2 is exactly balanced against the gain from serving type 1 when $\theta_2/\theta_1 = 3/2$, i.e., type 1 is served only if $\theta_2/\theta_1 \leq 3/2$ (cut-off rate).

Now we turn to the case of a wider strategy set, i.e., where the tariff intended for type 1 may have a restriction in call dispersion. Type 1 can only reach a limited number of call termination points (a fraction n of the full network). According to (3.7) and (3.8) a restriction in call dispersion causes a negative horizontal shift in the demand curves. Type 2's gross surplus from consuming the good is evaluated according to type 2's true willingness to pay, $Q_2(p)$, while he is given an information rent as if the heterogeneity was described according to the demand curves $\bar{Q}_1(p, n)$ and $\bar{Q}_2(p, n)$. A distortion in type 1's tariff makes it less tempting for the high demand type to mimic the low demand type. Type 2 is less tempted by type 1's tariff if he cannot reach the entire network and he is less tempted when the unit price in type 1's tariff is high. Although type 1 also suffers under such distortions, he is not as seriously affected as type 2. In both cases the means is to restrict type 2's consumption if he selects type 1's tariff, by way of a high unit price or access to a smaller network (reduced opportunity set).

Lemma 3.2 (Two-dimensional screening) *If consumers' calling patterns are type dependent, and can be monitored by the monopoly, a restriction on type 1's call dispersion serves as an alternative to a distortion in the unit price to type 1. For a given restriction n , type 2 is offered a cost-plus-fixed-fee tariff $\{0, E_2^n\}$ and type 1 is offered a two-part tariff $\{p_1^n, E_1^n, n\}$, $n \leq 1$*

$$p_1^n = \theta_2 - \theta_1 \frac{F_1(n)}{F_2(n)},$$

and where the fixed fees E_1^n and E_2^n are determined by (3.12) and (3.13).

Under our assumptions on F_1 and F_2 , p_1^n is nondecreasing in n , continuous, and differentiable whenever $p_1^n > 0$. Because type 2 consumers suffer more both from a restriction in call dispersion and from an increase in the unit price, they

serve as alternative instruments to relax the incentive constraint. This is reflected in the result that p_1^n is decreased (increased) when n is decreased (increased).

On the other hand, both instruments are costly to use in the sense that type 1's consumption is de facto restricted (whereas type 2's consumption is restricted only if he opts for type 1's tariff). In either case the consequence is that type 1 will make fewer calls. The firm loses income from these calls and since type 1 loses surplus on these calls he is not willing to participate unless the fixed fee is reduced. On the other hand, type 1's tariff is no longer as tempting for type 2 and the fixed fee from type 2 can be increased. The optimal trade-off in the firm's use of the two instruments depends on the relative effect they have on the two types' demand. From the pricing rule in Lemma 3.2 we see that larger heterogeneity in call duration (θ_2 is large relative to θ_1) results in a larger unit price.

Assuming that both types are served we use part (i) of Lemma 3.2 and write the expected profit as a function of n as

$$\Pi(n) = \begin{cases} \frac{1}{2}\theta_2^2 + \frac{1}{2}\theta_1^2 \frac{F_1(n)^2}{F_2(n)} - F_1(n)\theta_1(\theta_2 - \theta_1) & \text{if } p_1^n > 0, \\ \theta_1^2 F_1(n) + \frac{1}{2}\theta_2^2(1 - F_2(n)) & \text{if } p_1^n = 0. \end{cases} \quad (3.14)$$

The firm maximizes profit with respect to n and the tariffs are determined by Lemma 3.2. If the heterogeneity in quantity type is large relative to the heterogeneity in call dispersion, the firm will offer type 1 consumers a two-part tariff with a restriction in call dispersion together with a distorted unit price. In the opposite case the firm will offer type 1 consumers flat-rate pricing with restriction in call dispersion. Whenever there is heterogeneity in the types' calling pattern the firm will restrict type 1's calling.

Lemma 3.3 (Restriction in call dispersion) *The firm separates between high and low demand consumers by distorting type 1's tariff with respect to call dispersion, alone or together with a distortion in the unit price.*

(i) *Type 1 is offered a cost-plus-fixed-fee tariff with a restriction in call dispersion $\tilde{n} \in (0, 1]$ if \tilde{n} exists such that*

$$\frac{F_1(\tilde{n})}{F_2(\tilde{n})} \geq \frac{\theta_2}{\theta_1} \geq \sqrt{\frac{2f_1(\tilde{n})}{f_2(\tilde{n})}}$$

(ii) *Type 1 is offered a two-part tariff with a unit price distortion and a restriction in call dispersion $\hat{n} \in (0, 1]$ if \hat{n} exists such that*

$$\frac{\theta_2}{\theta_1} \geq 1 + \frac{F_1(\hat{n})}{F_2(\hat{n})} \left(1 - \frac{1}{2} \frac{f_2(\hat{n})}{f_1(\hat{n})} \frac{F_1(\hat{n})}{F_2(\hat{n})}\right) \geq \frac{F_1(\hat{n})}{F_2(\hat{n})}$$

The tariffs are subsequently determined according to Lemma 3.2.

The firm chooses to place a restriction in call dispersion in order to satisfy the condition $\partial\Pi/\partial n \leq 0$. The first inequality in part (i) of Lemma 3.3 states the condition for $p_1^n > 0$, whereas the last inequality in part (ii) of Lemma 3.3 states the condition for $p_1^n = 0$. In the latter case, the firm only has to trade-off how an increase in n affects the fixed fees. Hence, if the heterogeneity in call duration is low relative to the heterogeneity in call dispersion, type 1 is more likely to be served with a cost-plus-fixed-fee tariff, i.e., when θ_2/θ_1 is small and/or F_1/F_2 is large. Since the tariff intended for type 2 has no restriction in call dispersion, the demand curves $\bar{Q}_1(p, n)$ and $Q_2(p)$ never cross if $\theta_2/\theta_1 \geq F_1(n)$, which is always met. It does not matter whether the demand curve $\bar{Q}_2(p, n)$ crosses $\bar{Q}_1(p, n)$ since type 2 is not expected to make his purchases along $\bar{Q}_2(p, n)$.

When call dispersion conditional on consumer type θ is known, we can characterize the firm's pricing policy. We do this in the following two sections. For simplicity we assume that type 2 makes calls of equal length to all nodes, i.e., $f_2(n)$ is uniformly distributed on the interval $[0, 1]$. Regarding type 1's call dispersion we assume two different cases, call dispersion is described either by the uniform distribution or by a Beta distribution.

3.5.1 Uniform distribution

Call dispersion for type 1 is uniformly distributed on the interval $[0, \bar{n}_1]$, and call dispersion for type 2 is uniformly distributed on the interval $[0, 1]$, $0 < \bar{n}_1 < 1$.

The marginal unit price is $p_1^n = \theta_2 - \theta_1 \frac{1}{\bar{n}_1}$. For $p_1^n > 0$, the derivative of the firm's profit with respect to n can be written as

$$\frac{d\Pi}{dn} = \begin{cases} -\frac{\theta_1^2}{2n^2} & \text{if } \bar{n}_1 \leq n \leq 1, \\ -\frac{\theta_1(\theta_2 - \theta_1)}{\bar{n}_1} + \frac{\theta_1^2}{2\bar{n}_1^2} & \text{if } 0 < n < \bar{n}_1. \end{cases} \quad (3.15)$$

And if $p_1^n = 0$ we have $\Pi = E_1^n + E_2^n$, and the derivative with respect to n is

$$\frac{d\Pi}{dn} = \begin{cases} -\frac{1}{2}\theta_2^2 & \text{if } \bar{n}_1 \leq n \leq 1, \\ \frac{1}{\bar{n}_1}\theta_1^2 - \frac{1}{2}\theta_2^2 & \text{if } 0 < n < \bar{n}_1. \end{cases} \quad (3.16)$$

The profit function is linear for $n \in [0, \bar{n}_1)$ but the sign of the derivative is ambiguous, for $n \in (\bar{n}_1, 1]$ profit decreases in n . The optimal restriction in call dispersion will be one of the extremes $n^* = 0$ or $n^* = \bar{n}_1$. In the first case type 1 is *de facto* excluded. Henceforth, we define a variable $t \equiv \frac{\theta_2}{\theta_1}$. The propositions that follow describe the monopoly's pricing strategy.

Proposition 3.1 *If heterogeneity in call dispersion is sufficiently large relative to heterogeneity in call duration, $n^* = \bar{n}_1$ and type 1 is served with a cost-plus-fixed-fee tariff $\{0, E_1^n, \bar{n}_1\}$. For $t \in [1, 2]$ this occurs for $\bar{n}_1 \leq \frac{1}{t}$, for $t > 2$ it occurs for $\bar{n}_1 \leq \frac{2}{t^2}$.*

Proposition 3.1 shows that a restriction in call dispersion in type 1's tariff may be sufficient to separate the types. Consumers with different willingness to pay are charged identical unit price, but type 2 pays a larger fixed fee. In terms of pricing, this looks like first degree price discrimination. For $t \leq \frac{3}{2}$, type 1 is served with a restriction in call dispersion instead of with a distortion in the unit price, for $t > \frac{3}{2}$, type 1 is served with a restriction in call dispersion instead of being excluded.

Proposition 3.2 *If demand side heterogeneity is more moderate and balanced, type 1 is served with a two-part tariff $\{p_1^n, E_1^n, \bar{n}_1\}$, and $p_1^n > 0$. This occurs for $\bar{n}_1 > 0.5$ and $t < 2$, and t such that $\frac{1}{\bar{n}_1} \leq t \leq 1 + \frac{1}{2\bar{n}_1}$.*

A restriction in call dispersion will always be used, either alone ($p_1^n = 0$) or in combination with a restriction on usage via distortionary pricing.

Proposition 3.3 *If heterogeneity in call duration is sufficiently large relative to heterogeneity in call dispersion, type 1 is excluded from making purchases. This occurs for $t > \sqrt{2/\bar{n}_1}$ for $\bar{n}_1 \in [0, 0.5)$, or for $t > 1 + \frac{1}{2\bar{n}_1}$ for $\bar{n}_1 \in [0.5, 1]$. Type 2 is served with a cost-plus-fixed-fee tariff, $\{0, E_2, 1\}$, which extracts the entire social surplus. Type 1 is served in more cases relative to the single-dimensional case.*

Although increased heterogeneity in call dispersion reduces the incentive to exclude type 1, proposition 3.3 states that this incentive still exists.⁴¹ The proofs of the propositions are given by simple calculations that are shown in the appendix. Figure 3.3 illustrates the results.

The effect of a reduction in call dispersion is that the firm can give informational rent to type 2 as if the types were described according to the demand curves $\bar{Q}_i(p, \bar{n}_1)$, but extract gross surplus from type 2 according to the demand curve $Q_2(p)$. Typically, the possibility of type 1 being served increases as \bar{n}_1 decreases because this increases the 'observability' of the two types. The generalization of this is the fact that the firm is always better, or at least equally well, off with an additional observable and instrument at hand.⁴²

3.5.2 Beta distribution

The Beta-distribution allows for the possibility that the call length may vary over n , i.e., over points of call termination. That is, call termination points are

⁴¹Instead of saying that $n^* = 0$ we could say that $n^* = 1$ but let p_1^n be sufficiently high to ensure that $Q_1(p_1) = 0$.

⁴²Sappington (1983) shows this in a regulation model. A regulator that is uncertain about a multiproduct firm's production technology achieves additional information by observing the production level of each product. Caillaud et al. (1988) generalize the case with several observable variables.

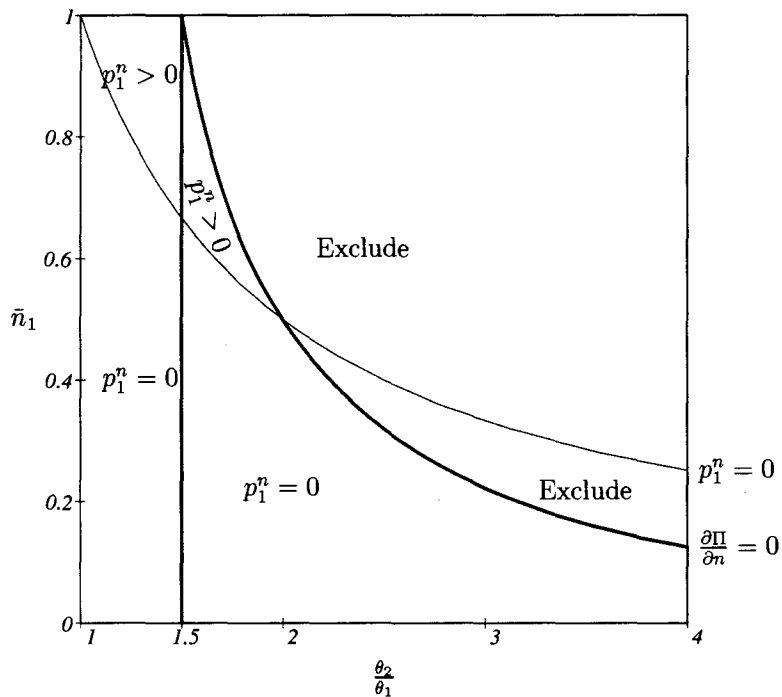


Figure 3.3: Pricing policy towards type 1 depending on the heterogeneity along the two dimensions. The larger the heterogeneity in call dispersion (low \bar{n}_1) the larger is the possibility that type 1 is served and that he is served with an efficient tariff, i.e., a cost-plus-fixed-fee tariff.

ordered according to the most called number, the second most called number etc. We keep the simplification that type 2's calling is uniformly distributed on $[0, 1]$ but say that type 1 has a more concentrated calling pattern by using the Beta distribution and placing more probability weight to the left tail of the distribution. Figure 3.4 illustrates this difference between the types.

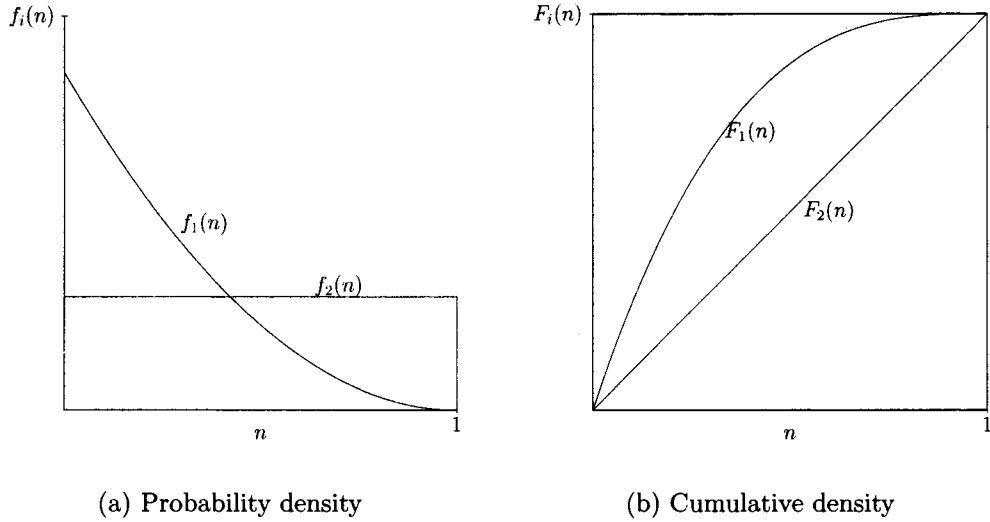


Figure 3.4: *Probability distribution over n , the uniform distribution and the Beta distribution with $v = 1$*

The probability density function for the beta distribution is

$$f(n, v, w) = \begin{cases} \frac{n^{v-1} (1-n)^{w-1}}{B(v, w)} & \text{if } 0 \leq n \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (3.17)$$

where the shape parameters v and w are positive numbers. The denominator $B(v, w)$ is the Beta function. With $v = 1$, the shape of the distribution is determined by w , the higher is w the larger is the mass for low n . We can redefine the distributions for type 1 by fixing v to be 1 and letting w vary ($w = 1$ is the uniform distribution on $[0, 1]$). The p.d.f and the c.d.f. are defined by

$$f_1(n, w) = \begin{cases} w(1-n)^{w-1} & \text{if } 0 \leq n \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (3.18)$$

$$F_1(n, w) = \begin{cases} 1 - (1-n)^w & \text{if } 0 \leq n \leq 1. \\ 0 & \text{otherwise.} \end{cases} \quad (3.19)$$

The probability density and cumulative density functions $f_2(n)$ and $F_2(n)$ are the same as before. The firm seeks to maximize profit with respect to n according to the optimality condition in Lemma 3.3. The monopoly's pricing strategy is given in the following propositions.

Proposition 3.4 *If heterogeneity in call dispersion is sufficiently large relative to the heterogeneity in call duration, both types are served with a cost-plus-fixed-fee tariff $\{0, E_1^n, n^*\}$, $n^* \in [n', n'']$. This occurs for $t \leq t' \leq t''$. n' and n'' decrease whereas t' and t'' increase as the heterogeneity in call dispersion increases (w increases).*

Proposition 3.4 is a replication of proposition 3.1, the larger the heterogeneity in call dispersion, the more powerful is a restriction in call dispersion as an instrument to separate the types. This can be utilized by the firm in two different ways. The firm can achieve less costly separation by decreasing n (reflecting that n'' decreases as w increases), or serve more types (reflecting that t'' increases as w increases).

Proposition 3.5 *When the heterogeneity is more moderate and balanced, type 1 consumers are offered a two-part tariff $\{p_1^n, E_1^n, n^{**}\}$, $p_1^n > 0$, $n^{**} \in [0, n']$. This occurs for $t' \leq t \leq t''$, and $w \leq 2$.*

Proposition 3.5 is a replication of 3.2. When the heterogeneity in call duration increases, it is necessary to increase the restriction in call dispersion in order to restore incentive compatibility.

Proposition 3.6 *If heterogeneity in call duration is sufficiently large relative to heterogeneity in call dispersion, type 1 is excluded from making purchases. This occurs for $t > \sqrt{2w}$ if $w < 2$ or for $t > 1 + \frac{1}{2}w$ if $w > 2$. Type 1 is served in more cases relative to the single-dimensional case. If $w = 2$, then $t' = t'' = 2$ and all types that are served are served with a cost-plus-fixed-fee tariff.*

Finally, proposition 3.6 is a replication of proposition 3.3. The incentive to exclude low demand consumers still exists when the heterogeneity in call duration is sufficiently large. The propositions 3.4, 3.5, and 3.6 are proved in the Appendix. Figure 3.5 illustrates the results.

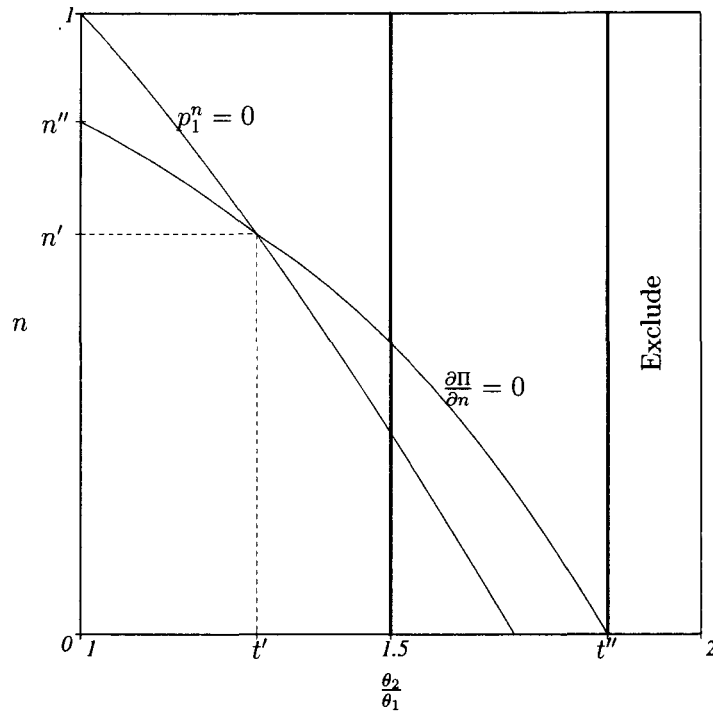


Figure 3.5: Pricing policy towards type 1, $w = 1.7$. The larger the heterogeneity in call dispersion (high w) the larger is the possibility that type 1 is served and that he is served with an efficient tariff, i.e., a cost-plus-fixed-fee tariff.

3.6 Concluding remarks

In the model presented in this paper, we have assumed that a monopoly firm can use two instruments to achieve second-degree price discrimination. The firm can introduce quantity distortions towards low demand types, according to the well-known model with nonlinear pricing. Another instrument is to introduce a restriction on the use of the service in such a way that high demand consumers are punished more than low demand consumers. The firm typically finds it optimal to combine distortions along the two dimensions. Then, type 1 consumers face a two-part tariff with a marginal price above marginal cost, together with a restriction on usage. However, the restriction on usage allows the firm to reduce the distortion in the pricing rule in the low-demand type's tariff. Whenever the monopoly firm finds it profitable to serve type 1 and there is observable heterogeneity in the use of the service, it will always impose a restriction on usage in type 1's contract. Sometimes, imposing a restriction on usage is sufficient to achieve separation. We also show that the results are qualitatively the same in the case when calls are distributed according to the uniform distribution and the Beta-distribution.

The theoretical model contributes to explain the practice of optional tariffs such as calling circle tariffs, in which the restriction is really severe. However, it should be remarked that promotion of calling circle tariffs might also serve as a strategy to create lock-in effects in duopolistic competition.

Further, the model suggests that it might be possible to practice a pricing strategy closer to flat rate pricing by separating consumers by other means than price-cost distortions. Hence, the outcome would be closer to first degree price discrimination. Although this paper applies the model to a very simple example within telecoms, the pricing principles derived are of general validity.

Appendix

A.1 Proof of Propositions 3.1–3.3

For $p_1^n(\bar{n}_1) \geq 0$ the profit function in (3.15) is increasing in n if

$$\frac{\theta_2}{\theta_1} \leq 1 + \frac{1}{2\bar{n}_1}. \quad (3.20)$$

For $p_1^n(\bar{n}_1) = 0$ the profit function in (3.16) is increasing in n if

$$\frac{\theta_2}{\theta_1} \leq \sqrt{\frac{2}{\bar{n}_1}}. \quad (3.21)$$

The unit price is positive if

$$\frac{\theta_2}{\theta_1} \geq \frac{1}{\bar{n}_1}. \quad (3.22)$$

Both conditions (3.20) and (3.21) define a curve that is steeper in the $(\theta_2/\theta_1, \bar{n}_1)$ space than does the condition $p_1^n = 0$. Also, $p_1^n = 0$ and $d\Pi/dn = 0$ are binding jointly for $(\theta_2/\theta_1, \bar{n}_1) = (2, \frac{1}{2})$. The slopes are given by the following

$$\left. \frac{d\bar{n}_1}{d(\theta_2/\theta_1)} \right|_{p_1^n=0} = -\bar{n}_1^2, \quad (3.23)$$

$$\left. \frac{d\bar{n}_1}{d(\theta_2/\theta_1)} \right|_{\Pi_n=0} = \begin{cases} -2\bar{n}_1^2 & \text{for } \theta_2/\theta_1 < 2, \\ -\left(\sqrt{\frac{2}{\bar{n}_1}}\right)\bar{n}_1^2 & \text{for } \theta_2/\theta_1 > 2. \end{cases} \quad (3.24)$$

Proposition 3.1 is derived by solving for \bar{n}_1 in (3.22) (or (3.21)) respectively for $t < (>)2$. Proposition 3.2 is simply given by (3.20) and (3.22). Proposition 3.3 is derived by turning the inequality in (3.20) for $\bar{n}_1 \in [0, 0.5)$, and by turning the inequality in (3.21) for $\bar{n}_1 \in [0.5, 1)$. Since $\lim_{\bar{n}_1 \rightarrow 1} (1 + \frac{1}{2\bar{n}_1}) = \frac{3}{2}$, type 1 is served in more cases relative to the single-dimensional case.

A.2 Proof of Propositions 3.4–3.6

From Lemma 3.2 and Lemma 3.3 we derive the conditions $p_1^n = 0$ and $\Pi'_n = 0$, which are the two curves in figure 3.5. The slopes of these are given by

$$\left. \frac{dn}{dt} \right|_{p_1^n=0} = \frac{n^2}{nf_1 - F_1} \leq 0, \quad (3.25)$$

$$\left. \frac{dn}{dt} \right|_{\Pi'_n=0} = \begin{cases} \frac{\sqrt{2f_1}}{f_{1n}} \leq 0 & \text{if } p_1^n = 0, \\ \frac{2n^3 f_1^2}{2f_1(nf_1 - F_1)^2 + nF_1^2 f_{1n}} \leq 0 & \text{if } p_1^n > 0, \end{cases} \quad (3.26)$$

with notation $f_{1n} \equiv df_1(n, w)/dn$, $f_{1w} \equiv df_1(n, w)/dw$ and so on.

When w increases there will be a positive shift in the curve defining $p_1^n = 0$,

$$\left. \frac{dt}{dw} \right|_{p_1^n=0} = \frac{F_{1w}}{n} \geq 0. \quad (3.27)$$

The shift in the curve defining $\Pi'_n = 0$ is negative for larger values of n and positive for smaller values of n ,

$$\left. \frac{dt}{dw} \right|_{\Pi'_n=0} = \begin{cases} -\frac{f_{1w}}{f_{1n}} & \text{if } p_1^n = 0. \\ \frac{F_{1w}}{n} - \frac{1}{2n^2} \frac{F_1(2F_{1w}f_1 - f_{1w})}{f_1^2} & \text{if } p_1^n > 0. \end{cases} \quad (3.28)$$

When w increases it places more probability weight to the lower end. Hence, f_{1w} is positive for smaller values of n and negative for higher values of n , while f_{1n} is negative for all $n \in [0, 1]$.

Next, we evaluate the shift along the t -axis

$$\lim_{n \rightarrow 0^+} \left[\left. \frac{dt}{dw} \right|_{\Pi'_n=0} \right] = \begin{cases} \frac{1}{\sqrt{2w}} & \text{if } p_1^n = 0, \\ \frac{1}{2} & \text{if } p_1^n > 0. \end{cases} \quad (3.29)$$

Hence, since the shift is positive along the t -axis, the shift along the n -axis must be negative, implying that t'' is increasing and n'' is decreasing in w .

We can show that n' decreases when the heterogeneity in call dispersion increases by differentiating the condition

$$\frac{F_1(n', w)}{n'} = \sqrt{2f_1(n', w)}, \quad (3.30)$$

which gives us

$$\frac{dn'}{dw} = -\frac{n^2 f_{1w} - n F_{1w} \sqrt{2f_1}}{n^2 f_{1n} - \sqrt{2f_1}(n f_1 - F_1)} \leq 0. \quad (3.31)$$

Since we have $t' = \frac{F_1(n', w)}{n'}$, which is monotonic with $dt'/dn' < 0$ (by 3.25), t' is increasing in w . By inspection we can conclude that the firm offers a cost-plus-fixed-fee tariff for $t < t'$ and $n > n'$. This completes the proof of Propositions 3.4 and 3.5

When $w = 2$ the curves are tangent at the point $(t, n) = (2, 0)$ and $t' = t''$

$$\lim_{n \rightarrow 0^+} \left[\frac{F_1}{n} \right] = w, \quad (3.32)$$

$$\lim_{n \rightarrow 0^+} \left[\sqrt{2f_1} \right] = \sqrt{2w}, \quad (3.33)$$

$$\lim_{n \rightarrow 0^+} \left[1 + \frac{F_1}{n} \left(1 - \frac{1}{2} \frac{F_1}{f_1} \right) \right] = 1 + \frac{1}{2w}. \quad (3.34)$$

The shift in the curve defining $p_1^n = 0$ along the t -axis is given by

$$\lim_{n \rightarrow 0^+} \left[\frac{dt}{dw} \Big|_{p_1^n=0} \right] = 1. \quad (3.35)$$

The shift in (3.35) is larger than (3.29). Since $w > 1$ type 1 is for certain served when $t < 3/2$. Together with the preceding statements this completes the proof of Proposition 3.6.

Chapter 4

Three-Part Tariffs in a Duopoly

by Sissel Jensen*

4.1 Introduction

We observe nonlinear pricing in many markets, that is, pricing arrangements where payment is not strictly proportional to the quantity of purchases. In the literature, implementation of nonlinear pricing is typically studied as a single two-part tariff or as a menu of two-part tariffs. Further, with a few notable exceptions, the existing literature applies a setting with a monopoly firm where nonlinear pricing is implemented by two-part tariffs. However, it is easy to verify that this does not sufficiently describe the practice of nonlinear pricing. Firstly, nonlinear pricing is a common practice in duopoly and oligopoly markets as well as in monopolies. Secondly, we frequently observe that other tariff arrangements rather than just two-part tariffs are used. The purpose of this paper is to make a contribution in the second part of the gap between theory and practice within the field of nonlinear pricing. We examine whether the fact that there is competition between two firms instead of a monopoly significantly changes the tariff structure. We find that implementation by two-part tariffs may not be a feasible strategy in a duopoly, but if a firm can use a combination of two-part and three-part tariffs, a fully nonlinear pricing schedule can be implemented. Three-part tariffs are used for small quantity purchases while two-part tariffs are used for large quantity purchases. Furthermore, quantity discounts are given for larger

*I am indebted to Øystein Fos, Kåre Petter Hagen, Bjørn Hansen, Petter Osmundsen, Atle Seierstad, Knut Sydsæter, Lars Sjørgard, Steinar Vagstad, Jon Vislie, and two anonymous referee for their helpful comments. I have also benefited from comments of seminar participants at the University of Bergen, and at the Norwegian School of Economics and Business Administration. I also thank Marit Hareland and Hans Olav Husum for contributions in the discussion of this problem at an early stage.

purchases only. Finally we show that this is in fact what firms actually do in the telecommunications market, where we observe competition rather than monopoly.

The market perception of what are reasonable tariff structures would vary according to what kind of market one is studying. However, menus of two-part and three-part tariffs are frequently used and it seems natural to restrict the analysis to menus of piecewise linear tariffs. A firm confronts consumers with a menu of tariffs and consumers make their optimal quantity choice subject to the tariff chosen and are also billed according to this tariff. Under two-part tariffs consumers receive larger quantity discounts if they are willing to pay a larger fixed fee in advance. Three-part tariffs can be implemented in two different ways; Consumers may commit to a specific minimum usage level and pay a flat fee until this level is reached. The higher the minimum usage consumers commit to the higher discount they get. Another way to implement a three-part tariff is to apply larger discounts when realized usage exceeds some specific threshold level during a billing period.

4.1.1 Related literature

In a monopoly context models on optimal nonlinear pricing often assume that it is sufficient to ensure that the individual rationality constraint is satisfied for the worst type only. If the worst type finds it weakly rational to participate, then all types will indeed participate. Under the monotone hazard rate condition, a menu of two-part tariffs is sufficient to implement a fully nonlinear outlay schedule with complete separation of types. The underlying assumptions behind this result are that the agent's participation decision is deterministic; the reservation utility is independent of consumer type and the private information is single-dimensional. There is an increasing amount of literature that explores how the weakening of the modelling assumptions affects the results. Within the part of incentive theory where an agent contracts with only one principal, i.e., models with only a single principal or models with delegated common agency, richer models incorporate either multi-dimensional types or type-dependent participation constraints. Rochet and Stole (2000) give a review of the literature on multidimensional screening.

Several papers have incorporated nonlinear pricing into models with imperfect competition, but few study tariff design and tariff implementation under asymmetric information about individual quantity-type. The papers by Stole (1995), Armstrong and Vickers (1999) and Rochet and Stole (1999) model nonlinear pricing in a differentiated oligopoly. In Stole's paper the qualitative property of the monopoly model with downward distortion for all types but the highest is kept, while Rochet and Stole (1999) and Armstrong and Vickers (1999) find conditions that imply that efficient two-part tariffs emerge as an equilibrium. The divergence between these two results is partly relying on how transportation costs enter the model. In Stole's model transportation costs depend on the quantities consumed (and on taste) whereas the transportation costs are assumed to be

lump-sum costs in the two others. However, Stole (1995) leaves the question of implementation aside.⁴³ Other papers that study two-part tariffs under competition often do this in a Cournot or Bertrand game, but with focus on two-part tariffs versus linear tariffs rather than on how the informational problem affects the tariff design.⁴⁴

There is literature that deals with multi-dimensional screening where the informational asymmetry relates directly to the variable being contracted upon (e.g., consumers' willingness to pay for different quality attributes, or an agent's efficiency type when performing different tasks for a principal). The work by Armstrong and Rochet (1999), Rochet and Choné (1998) provides an overview of the literature and represents the status on how far the techniques are developed.⁴⁵ Another view on multi-dimensionality in mechanism design is taken in Rochet and Stole (1999), who work on a general model of nonlinear pricing where the informational asymmetry is present in the consumers' reservation utility as well as in their preferences, i.e., with a more general modelling of the participation decision. The methodology developed in Rochet and Stole (1999) paper with randomness in the agents' outside option fits a situation where consumers' location is *not* perfectly known. They demonstrate the difficulties of working on multidimensional problems.

The literature on type-dependent participation constraint includes the work by Lewis and Sappington (1989), Biglaiser and Mezzetti (1993), Ivaldi and Martimort (1994), Maggi and Rodriguez-Clare (1995), Stole (1995) and Jullien (2000). Type-dependent participation constraint may arise in a situation with multiple principals (but where an agent contracts exclusively with one of them e.g., Biglaiser and Mezzetti (1993), Stole (1995)) or it can arise because of other reasons, i.e., it is for some reasons natural to model a type's outside option as a function of the privately known type parameter (e.g., Lewis and Sappington (1989)).⁴⁶ Ivaldi and Martimort (1994) provide empirical research that support that nonlinear pricing prevail under oligopolistic competition (energy distribution). Equilibrium pricing schemes are concave and depend on unknown private valuation and on the rivals contract parameters. They restrict the regression of payments to second-order polynomials on quantities. Hence, we cannot rule out

⁴³Valletti (1999) derives similar results in a model with discrete types.

⁴⁴Examples of such work are Calem and Spulber (1984), Gasmi et al. (2000), Hayes (1987), Oren et al. (1983). Wilson (1993) provides a comprehensive survey of the literature and the practice of nonlinear pricing, Michell and Vogelsang (1991) provide a survey of the pricing of telecommunications in the U.S. during the 70s and 80s. Stole (1995) also provide a brief overview of the literature.

⁴⁵Literature includes Laffont et al. (1987), Matthews and Moore (1987), Wilson (1993), Armstrong (1996).

⁴⁶Models on common agency can be found in Stole (1992), Martimort (1992), Martimort (1996), Mezzetti (1997) and Olsen and Osmundsen (1998). These are cases describing a situation where each principal requires that a task be performed by a common agent. The agent's ability or effort in performing the two tasks is unobservable but is privately known by the agent.

a hypothesis that the true outlay schedule has convex parts, although the overall shape is concave.

Insights from these papers show that many of the results achieved earlier in nonlinear pricing are not robust. In models with multi-dimensional screening it is shown that the “no distortion at the top” result may appear together with distortion, no distortion or bunching at the bottom, as opposed to the Mussa and Rosen (1978) result with downward distortions for all types except the highest. The literature on type-dependent participation constraints demonstrates the possibility of a non-monotonic informational rent, i.e., countervailing incentives may arise. The incentive constraint can be downward binding for some types and upward binding for other types.

The model presented in this paper falls into a situation with asymmetric information along a single (vertical) dimension and with a type-dependent participation constraint. The basic model is identical to the model in Stole (1995). But, while he solves for an equilibrium in fully nonlinear tariffs, the model we present here searches for an implementable *tariff structure*. Further, given the difficulties of involving multidimensional screening, we keep the assumption that the agent’s participation decision is deterministic. There are no gains from joint consumption and this eliminates the “competitive externality” in the incentive constraint and one source of countervailing incentives.⁴⁷ The informational rent on the other hand has to be evaluated net of an outside option, which is the maximal utility a consumer gains if he rejects the firm’s contract. Generally, it is not sufficient to ensure that the individual rationality constraint is satisfied for the worst type only. A priori, the sign of the marginal information rent can be positive, zero, negative or even change sign over the type space, creating a second source of countervailing incentives.⁴⁸ Countervailing incentives do not occur in this model, the participation constraint is binding only in the lower part of the distribution of types (or maybe only for the very lowest type) and the information rent is strictly increasing elsewhere.

4.2 The model

The model is closely related to Stole (1995). However, the focus is distinctly different from his. The main issue in this paper is implementation of nonlinear prices, an issue not raised in Stole (1995). While Stole in his paper lets consumers buy a single unit of a good but with variable quality, the present paper sets up the alternative quantity framework. However, impose the restriction that a consumer

⁴⁷A competitive externality exists when the utility from buying q units from firm 0 is evaluated net of the (foregone) utility from *not* buying the same amount of q from another firm. This will be similar to the models in Lewis and Sappington (1989) and Maggi and Rodriguez-Clare (1995).

⁴⁸This will be similar to the models in Biglaiser and Mezzetti (1993) and Jullien (2000).

must choose a single tariff. Hence, we exclude the possibility that consumers buy from both firms, and we also exclude the possibility that consumers choose more than one tariff as well.⁴⁹

The model describes a case where two firms, denoted by firm 0 and firm 1, offer one product each and the products are spatially differentiated. The firms are located at the two extremes on a line of length 1, firm 0 at extreme 0 and firm 1 at the other extreme, 1. Each individual's preferences over the two firms are identified according to each individual's location $\gamma \in [0, 1]$ on the interval, referred to as brand preference. Total length of the distance between a consumer and firms 0 and 1 is $|0 - \gamma|$ and $|1 - \gamma|$ respectively. Transportation costs are normalized to unit, hence, the total loss from not being able to buy the ideally preferred product is γ and $(1 - \gamma)$. Brand preferences are common knowledge and firms practice first-degree price discrimination over the horizontal dimension. Both firms face constant and identical marginal costs, $c_0(q) = c_1(q) \equiv c$.

Consumers' taste varies over a vertical dimension, which we interpret as a quantity-preference parameter, referred to as quantity-type (θ) subject to private knowledge.⁵⁰ The firms have common prior beliefs about the distribution of types $\theta \in [\underline{\theta}, \bar{\theta}]$ described by a cumulative distribution function $F(\theta)$. The corresponding density function $f(\theta)$ is strictly positive on the support. Thus, $F(\theta)$ is the objective distribution over a population of buyers having identical brand preferences γ .⁵¹ We will assume that the distribution satisfies the monotone hazard rate condition.

The first assumption, i.e. about product differentiation, can be justified by considering that identical services – with respect to the communication capabilities they provide – are sold or bundled with different ancillary services or quality levels that consumers value differently. This could for example be differences in billing features (more detailed billing) and in support services, but it could also be features *perceived* as differences in the quality of the service provided.⁵²

⁴⁹This is a simplification to keep the similarity to Stole's model. As pointed out by Stole (1995) it is plausible to restrict a consumer to purchasing from a single firm under the quality framework with unit demand. In the alternative quantity framework it requires additional technical restrictions to ensure that a consumer is not better off by buying two times $q/2$ than one time q . The restriction we impose on consumers' behavior is for instance plausible when we think of telephony, or the mobile phone, industry, where consumers subscribe to a particular tariff option. If they subscribe to more than one option they must also have more than one phone number, which is by most people regarded as undesirable. If this restriction is binding, it indicates that the quantity-quality framework are not as intimately related in the duopoly as in the monopoly framework and that one should be more careful in the modelling and interpretations.

⁵⁰Since both θ and γ are taken to be continuous, we drop all subscripts for location and consumers' quantity type throughout the paper. However, we use superscript 0 and 1 to denote the location of the two firms.

⁵¹The distribution over quantity-types θ is independent of γ , i.e., for each γ -value the corresponding density function $f(\theta | \gamma) \equiv f(\theta)$ for all possible $\gamma \in [0, 1]$.

⁵²Examples on differences in quality may be found in AT&T marketing of "AT&T True

The second assumption can be rationalized by taking into account the fact that consumers have different needs for communication, e.g. residential and business customers.

Consumers' preferences are represented by a utility function $u(q, \theta, \gamma)$ and $u(q, \theta, 1 - \gamma)$ when he buys from firm 0 and 1 respectively. If a consumer buys a quantity q and pays an amount T , his net utility is $U = u(q, \theta, \gamma) - T$.

Assumption 4.1 *The utility function is at least three times continuously differentiable and strictly concave in q . We make the following assumptions about the derivatives of the utility functions $u(q, \theta, \gamma)$ and $u(q, \theta, 1 - \gamma)$*

- | | |
|---|---|
| (a) $u(0, \theta, \cdot) = 0$ | (e) $u_{q\theta}(\cdot) > 0$ |
| (b) $\lim_{q \rightarrow 0} u_q(q, \theta, \cdot) = \infty$ | (f) $u_{\theta\theta}(\cdot) \leq 0$ |
| (c) $\lim_{q \rightarrow \infty} u_q(q, \theta, \cdot) = 0$ | (g) $u_\gamma(q, \theta, \gamma) < 0$ |
| (d) $u_\theta(\cdot) > 0$ | (h) $u_\gamma(q, \theta, 1 - \gamma) > 0$ |

To satisfy sufficient conditions, we will also make assumptions about the third order derivatives, and say that $u_{\theta qq} \leq 0$ and that $u_{\theta\theta q} \leq 0$. Further, we will make use of the following definition on consumers' indirect utility

Definition 4.1 *Let $U^k(\theta, \cdot)$ be the net utility (surplus) for a consumer located at γ , with quantity type parameter θ when he is faced with a general price schedule $T^k(q_k)$ and buys firm k 's product. The surplus he obtains is*

- | |
|---|
| (a) $U^0(\theta, \gamma) \equiv \max_q \{u(q, \theta, \gamma) - T^0(q)\}$ |
| (b) $U^1(\theta, 1 - \gamma) \equiv \max_q \{u(q, \theta, 1 - \gamma) - T^1(q)\}$ |

where $T^k(q)$ is a general price schedule ($k = 0, 1$).

Assumptions 4.1(a)-(c) secure the existence of a unique solution in consumers' choice of consumption q_k as long as there exists a continuous and appropriate outlay schedule $T(q)$.

The necessary single crossing condition together with assumption 4.1(d), implies that the indifference curves of consumers with different quantity preferences cross at most once, i.e., assumption 4.1(e). High-quantity type consumers value a marginal quantity increase higher than low-quantity types, regardless of brand preferences. Assumptions 4.1(g)-(h) follow from the fact that the products are horizontally differentiated.

In a first-best situation consumers would be confronted with prices equal to marginal cost, and under our assumptions this yields unique quantity allocations and consumer surplus.

Voice". Examples on differences in billing features can be many. Telecom companies undertake large investments to be able to support detailed billing towards business consumers. This can be to break down the cost of telecommunications to different business departments, and/or to different services (fixed link communications, mobile communications, 800-services (Premium Rate Services), etc.

Definition 4.2 *The first-best quantity level \bar{q}_k ($k = 0, 1$) is the optimal quantity purchase when consumers buy at marginal cost and the corresponding utility, denoted as first best utility, is given by*

- (a) $\bar{q}_k(\theta, \cdot) \equiv \arg \max_{q_k} \{u(q_k, \theta, \cdot) - cq_k\}$, $k = 0, 1$
- (b) $\underline{U}^0(\theta, \gamma) \equiv u(\bar{q}_0(\theta, \gamma), \theta, \gamma) - c\bar{q}_0(\theta, \gamma) > 0$
- (c) $\underline{U}^1(\theta, 1 - \gamma) \equiv u(\bar{q}_1(\theta, 1 - \gamma), \theta, 1 - \gamma) - c\bar{q}_1(\theta, 1 - \gamma) > 0$.

It follows from assumptions 4.1 that the first-best quantity and utility, $\bar{q}(\theta, \cdot)$ and $\underline{U}^k(\theta, \cdot)$ are both increasing in θ .

The two firms' products are perfect substitutes, except that they are of different brands. There are no gains from joint consumption (i.e., utility is not subadditive), and, for $0 < \gamma < 1/2$, the gains from purchasing good q_1 in addition to q_0 will never exceed the surplus from purchasing good q_0 . The implication of this is that the quantity purchases of q_0 are always largest when q_0 are bought alone. The opposite apply for $1/2 < \gamma < 1$

According to assumption 4.1(g), if a consumer chooses to purchase the good from firm 0, utility is decreasing in location, $u_\gamma(q_0, \theta, \gamma) < 0$. Hence, buying from the closest firm will always give largest *first best* utility. For all parameter values $\theta, \gamma \in [\underline{\theta}, \bar{\theta}] \times [0, 1/2)$ we have that $\underline{U}^0(\theta, \gamma)$ is strictly larger than $\underline{U}^1(\theta, 1 - \gamma)$

We will also assume that the first-best utility is convex

$$\frac{\partial^2 \underline{U}^k(\theta, \cdot)}{\partial \theta^2} = \frac{[u_{\theta q}(\bar{q}_k, \theta, \cdot)]^2}{-u_{qq}(\bar{q}_k, \theta, \cdot)} + u_{\theta\theta}(\bar{q}_k, \theta, \cdot) > 0, \quad k = 0, 1. \quad (4.1)$$

With such characterizations of consumers' preferences, the firm located at 0 has a competitive advantage in serving consumers located in the interval $[0, 1/2]$, whereas the firm located at 1 has a competitive advantage in the interval $[1/2, 1]$. Also, with symmetric marginal costs, price competition between the two firms will force the fixed fee down to zero and the marginal price down to marginal cost toward consumers being indifferent between buying from firms 0 and 1. Also, it is an equilibrium strategy for firm 1 to offer marginal cost pricing towards *every* consumer located in the interval $[0, 1/2]$. The problem is solved within a framework where an agent contracts with a single principal, the other firm's presence does only affect the individual rationality constraint.

At stage one of the game, each firm offers a fully nonlinear tariff with an ordered pair of take-it-or-leave-it contracts. At stage two, consumers make a choice of whether to buy from firm 0 or 1 (or from none) and also a choice of q_k ($k = 0, 1$). This is equivalent to assuming that the firm announces a menu of distinct tariffs at stage one, and letting consumers choose a tariff from this menu at stage two. Then, formally there is a stage three where consumers decide on individual quantity purchase and are billed according to the tariff choice at stage two. As long as the tariffs considered in the second type of game truthfully

implement the fully nonlinear tariff in the first game, the two formulations yield identical equilibria. Formally, the solution to the first game is analyzed in section 3, whereas section 4 characterizes the set of tariffs that truthfully implement this solution.

In the game, the firms implement their contracts subject to the incentive compatibility and individual rationality constraints. The consumers' choice of firm and quantity is *de facto* equivalent to announcing a type, which is in line with traditional mechanism-design. Further, since marginal-cost pricing is the single offer from firm 1 inside firm 0's turf (for $\gamma \in [0, 1/2]$), it is only necessary to secure truth-telling mechanisms in a single-dimensional space. That is, we can ignore the complications of a common agency case, in which an agent might misreport his type differently to the two principals. Therefore, we can solve the delegated problem as if it is a single-principal case. Under the single crossing condition, monotonicity is sufficient for local- and global second-order conditions to be satisfied under quasi-linear preferences (Fudenberg and Tirole (1991), theorem 7.1 and 7.2).

4.2.1 Individual rationality

As a consequence of the existence of a competing firm, consumers in firm 0's turf $[0, 1/2]$ have an outside option. The reservation utility is defined as the maximum utility obtained by not purchasing, which is normalized to zero, and the utility from buying the less preferred good. The latter was in the previous section termed $\underline{U}^1(\theta, 1 - \gamma)$.

Lemma 4.1 *The individual rationality constraint is given by*

$$U^0(\theta, \gamma) \geq \max \{ \underline{U}^1(\theta, 1 - \gamma), 0 \}. \quad (4.2)$$

The proof of Lemma 4.1 is standard, see for example Fudenberg and Tirole (1991, chapter 7).

Thus, given that the other firm practices marginal cost pricing within firm 0's turf, the individual rationality constraint is a function of consumer type.

Furthermore, since an outside option is of higher valuation for more distant consumers (closer to $1/2$), the individual rationality constraint will differ according to consumers' preferences over the two firms' goods. Generally, the value of the outside option is increasing and convex in θ , since the first-best utility is increasing and is assumed to be convex in θ . From (4.2) we also observe that if $\gamma = 1/2$, the only way to fulfill the *IR* constraint is to offer marginal cost pricing. Otherwise the firms have some market power in their respective market turfs.

4.2.2 Incentive compatibility

Consumers choose contracts that maximize their net utility. Under a direct-revelation mechanism approach, a consumer of type θ maximizes utility with

respect to a type announcement θ' . By definition

$$U^0(\theta, \theta', \gamma) = u(q(\theta', \gamma), \theta, \gamma) - t(\theta', \gamma), \quad (4.3)$$

$$U^0(\theta, \theta, \gamma) \equiv U^0(\theta, \gamma). \quad (4.4)$$

Global incentive compatibility requires

$$U^0(\theta, \theta, \gamma) \geq U^0(\theta, \theta', \gamma), \quad \forall \theta', \theta \in [\underline{\theta}, \bar{\theta}]. \quad (4.5)$$

Hence

$$\begin{aligned} U^0(\theta, \gamma) &= u(q(\theta, \gamma), \theta, \gamma) - t(\theta, \gamma), \\ &= \max_{\theta'} \{u(q(\theta', \gamma), \theta, \gamma) - t(\theta', \gamma)\}. \end{aligned} \quad (4.6)$$

Lemma 4.2 *Under the condition of Single Crossing, $u_{q\theta}(\cdot) > 0$, necessary and sufficient conditions for global incentive compatibility are given by*

$$\frac{\partial U^0(\theta, \gamma)}{\partial \theta} = u_{\theta}(q_0, \theta, \gamma), \quad (4.7)$$

$$q_0(\theta, \gamma) \text{ nondecreasing}. \quad (4.8)$$

The proof of Lemma 4.2 is also standard and is omitted.⁵³

Hence, (4.2), (4.7) and (4.8) are necessary and sufficient conditions for implementation. As is usual in the literature, we will ignore (4.8) at the first stage but subsequently check that it is met.

4.2.3 Informational rents

Before we proceed it might be convenient to determine the sign on the marginal informational rent to a type θ consumer that truthfully reveal his type.

Lemma 4.3 *A consumer of type θ that buys exclusively from firm 0, receives an informational rent*

$$R(\theta, \gamma) = U^0(\theta, \gamma) - \underline{U}^1(\theta, 1 - \gamma) \geq 0, \quad (4.9)$$

$$\frac{\partial R(\theta, \gamma)}{\partial \theta} = u_{\theta}(q, \theta, \gamma) - u_{\theta}(\bar{q}_1(\theta, 1 - \gamma), \theta, 1 - \gamma) \geq 0. \quad (4.10)$$

When the informational rent is unambiguously increasing in type, we can rule out the presence of countervailing incentives. To see that this is the case consider the following reasoning. When the *IR* constraint is binding in a neighborhood of $+\theta$, we have $R(\theta, \gamma) = 0$ and $R'_{\theta} = 0$. Choosing among the possible solutions

⁵³When the Single Crossing condition is satisfied, local (adjacent) incentive compatibility is also sufficient for global incentive compatibility. See for instance Fudenberg and Tirole (1991).

in q that meets (4.10) (if more than one exist) we select the schedule that also satisfy (4.6). Hence, (4.10) determine a quantity schedule $q(\theta, \gamma) = \tilde{q}(\theta, \gamma)$.

Hence, if the *IR* constraint is *not* binding, we must follow a quantity schedule satisfying the condition $q(\theta, \gamma) > \tilde{q}(\theta, \gamma)$. Consequently, since $u_{\theta q}(\cdot) > 0$ the information rent is nondecreasing in θ , and $R'_\theta \geq 0$. When the derivatives with respect to θ and the quantity schedule in the equilibrium are continuous, the *IR* constraint can only be binding in the left part of the distribution over θ , (or for $\underline{\theta}$ only), i.e., $U^0(\underline{\theta}, \gamma) = \underline{U}^1(\underline{\theta}, 1 - \gamma)$ and $U^0(\bar{\theta}, \gamma)$ is free. Note as well that it is sufficient to check whether $\tilde{q}(\theta, \gamma)$ is nondecreasing.

Without loss of generality we normalize the value of an outside option to zero for the *lowest* type, i.e., $\underline{U}^1(\underline{\theta}, 1 - \gamma) = 0$ (in practical terms we subtract this constant from $\underline{U}^1(\theta, 1 - \gamma)$, which is assumed to be positive). We make the following *redefinition* of the outside option

$$\underline{U}^1(\theta, 1 - \gamma) \equiv u(\bar{q}_1, \theta, 1 - \gamma) - c\bar{q}_1 - \underline{U}^1(\underline{\theta}, 1 - \gamma) \geq 0. \quad (4.11)$$

The justification behind doing so is that the individual rationality constraint is *binding* for the lowest type. Secondly, in this setting we can also compare the strategies of implementing in the duopoly solution and the monopoly solution respectively. In the latter, the value of an outside option is normalized to zero for the lowest type, and for every other type as well.⁵⁴ If the reservation utility profile is implementable, i.e., if q is nondecreasing when consumers receive their reservation utility, it might be the case that the individual rationality constraint binds for several types at the low end of the type space.

4.3 Optimal allocations

Firm 0's objective is to maximize profit subject to the individual rationality constraint and the (downward binding) incentive constraint. Profit maximization is a separate problem for each $\gamma \in [0, 1/2]$. The objective is

$$\text{Max} \int_{\underline{\theta}}^{\bar{\theta}} [t(\theta, \gamma) - cq(\theta, \gamma)] f(\theta) d\theta \quad (4.12)$$

s.t. *IR* and *IC*.

We use optimal control to solve the problem, imposing only the first order condition for incentive compatibility at the first stage (4.8). When we know the sign of the information rent, we are able to state the initial and terminal values of the state variable U^0 . From now and onwards, we drop the subscript on q , since the

⁵⁴See also Jullien (2000). If all types are served, the global level of the reservation utility does not really matters, what matter is the slope of reservation utility. If $U^{1*}(\theta, \gamma)$ is the solution to the problem when the reservation utility is U^0 , then $U^{1*}(\theta, \gamma) + c$ is the solution to the problem when the reservation utility is $U^0 + c$ for any constant c .

only q we are talking about is q_0 except when we denote the quantity level in the outside option $\bar{q}_1 = \bar{q}$. The objective is

$$\max_{q \geq 0} \int_{\underline{\theta}}^{\bar{\theta}} [u(q, \theta, \gamma) - U^0 - cq] f(\theta) d\theta \quad (4.13)$$

subject to

$$\begin{aligned} \partial U^0 / \partial \theta &= u_{\theta}(q, \theta, \gamma) \quad (a.e.), \\ U^0(\underline{\theta}, \gamma) &= 0, \quad U^0(\bar{\theta}, \gamma) \text{ free}, \\ U^0(\theta, \gamma) &\geq \underline{U}^1(\theta, 1 - \gamma), \\ \forall \theta &\in [\underline{\theta}, \bar{\theta}]. \end{aligned}$$

q is the control variable and U^0 is the state variable. This is a control problem with a pure state constraint.⁵⁵

The *Lagrangian* or *generalized Hamiltonian* L is

$$\begin{aligned} L &= [u(q, \theta, \gamma) - U^0 - cq] f(\theta) \\ &+ \lambda(\theta) u_{\theta}(q, \theta, \gamma) + \mu(\theta) [U^0 - \underline{U}^1], \end{aligned} \quad (4.14)$$

where $L = L(\theta, q, U^0, \lambda, \mu) = H(\theta, q, \gamma, U^0, \lambda) + \mu[U^0 - \underline{U}^1]$. The costate variable is $\lambda(\theta)$ and $\mu(\theta)$ is the multiplier of the state constraint. The Hamiltonian $H(\theta, q, U^{0*}(\theta, \gamma), \lambda(\theta))$ is strictly concave in q and the maximized Hamiltonian, $\hat{H}(\theta, U^0, \lambda(\theta)) = \max_{q \geq 0} H(\theta, q, U^0, \lambda(\theta))$ is concave in $U^0(\theta, \gamma)$. In addition the state constraint is quasiconcave in U^0 .⁵⁶

Let $(q^*(\theta, \gamma), U^{0*}(\theta, \gamma))$ be an admissible pair in the problem (4.13). Further, we assume that there exists a *continuous* function $\lambda(\theta) (\leq 0)$, with a piecewise continuous derivative $\lambda'(\theta)$, and a piecewise continuous function $\mu(\theta) \geq 0$ in the interval $[\underline{\theta}, \bar{\theta}]$. Then, we can use the Arrow sufficiency theorem to state the following additional conditions for a solution to the problem⁵⁷

$$(u_q - c) f(\theta) + \lambda(\theta) u_{\theta q}(q, \theta, \gamma) = 0, \quad (4.15)$$

$$\partial \lambda(\theta) / \partial \theta = -\frac{\partial L}{\partial U^0} = f(\theta) - \mu(\theta), \quad (4.16)$$

$$\lambda(\bar{\theta}) = 0 \quad (4.17)$$

⁵⁵See Seierstad and Sydsæter (1977) and Seierstad and Sydsæter (1987) for a treatment on optimal control theory with mixed and pure state constraint.

⁵⁶Although γ is certainly an argument in the H and L functions, the parameter is omitted in the writing of these functions as well as the λ and μ functions in order to make the notation easier. As long as $0 \leq \gamma \leq 1/2$, the value of γ has only the effect of shifting the level of the outcome whereas the characterization of the outcome remains the same regardless of γ .

⁵⁷See Seierstad and Sydsæter (1977, theorem 7 p. 377) and Seierstad and Sydsæter (1987, chapter 5)

$$\partial U^0(\theta, \gamma)/\partial \theta = u_\theta(q, \theta, \gamma), \quad (4.18)$$

$$\mu(\theta) [U^0 - \underline{U}^1] = 0, \quad \mu(\theta) \geq 0, \quad [U^0 - \underline{U}^1] \geq 0, \quad (4.19)$$

A configuration $(U^0(\theta, \gamma), q(\theta, \gamma), \lambda(\theta), \mu(\theta))$ that satisfies (4.15) – (4.19), $U^0(\theta, \gamma)$, $q(\theta, \gamma)$ and $\lambda(\theta)$ being continuous and piecewise differentiable, $\mu(\theta)$ piecewise continuous, is also an optimum. In addition we have to allow for optimal configurations in which $\lambda(\theta)$ is only piecewise continuous and has a finite number of jumps in the domain over θ . Under such circumstances we must apply the additional condition

$$\lambda(\theta_i^-) - \lambda(\theta_i^+) = \beta \left(\frac{\partial}{\partial U^0} (U^0 - \underline{U}^1) \right) = \beta, \quad (4.20)$$

$$\beta \geq 0 \quad (= 0 \text{ if } U^0 > \underline{U}^1), \quad (4.21)$$

where $\underline{\theta} < \theta_1 < \dots < \theta_k \leq \bar{\theta}$ are the discontinuity points of $\lambda(\theta)$, and β is a positive number. Since the jump must be from above ($\lambda(\theta^-) - \lambda(\theta^+) \geq 0$) we can rule out the case that there is a jump at $\theta = \bar{\theta}$, measured by $\lambda(\theta^-) - \lambda(\bar{\theta}) = \lambda(\theta^-) \geq 0$. If we allow $\lambda(\theta^-)$ to be positive it implies that firm 0 sells its' product at a price below marginal cost, since $\lambda(\theta) = -[(u_q - c)f(\theta)]/u_{q\theta}$. But under the assumption that the firms are symmetric with respect to marginal cost the individual rationality constraint can never impose such a strategy. If the *IR* constraint stops binding for some $\theta < \bar{\theta}$, conditions (4.20)-(4.21) apply (Seierstad and Sydsæter (1987, theorem 8, p. 380)). Because $R'_\theta \geq 0$ this leaves only one possible discontinuity point, the point where the state constraint stops being binding. If we find a solution with a continuous $\lambda(\theta)$ we focus on this and do not elaborate further on solutions where $\lambda(\theta)$ is not continuous.

First, from the optimality condition (4.15) the distortion is proportional to $\lambda(\theta)$, which is necessarily negative since setting a price below marginal cost can never be a part of the equilibrium strategy.

By differentiating the optimality condition with respect to θ we obtain the following condition for the monotonicity constraint to be met

$$\frac{dq}{d\theta} = - \frac{u_{q\theta}(f(\theta) + \lambda') + (u_q - c)f'(\theta) + \lambda u_{q\theta\theta}}{u_{qq}f(\theta) + \lambda u_{qq\theta}} \geq 0. \quad (4.22)$$

The denominator is negative under the assumption that the Hamiltonian $H(\theta, q, U^{0*}(\theta, \gamma), \lambda(\theta))$ is strictly concave in q . The likelihood of $dq/d\theta$ being positive increases as the slope of $\lambda(\theta)$ increases. When $\lambda'(\theta)$ is negative, there is a chance that the numerator becomes negative. Note that if we assume that third derivatives are indeed small, the slope of $\lambda(\theta)$ rather than $\lambda(\theta)$ itself will be important in the monotonicity constraint. Generally, we need

$$f(\theta) + \lambda'(\theta) \geq - \left[\frac{(u_q - c)}{u_{q\theta}} f'(\theta) + \lambda \frac{u_{q\theta\theta}}{u_{q\theta}} \right]. \quad (4.23)$$

When third derivatives are zero and θ is uniformly distributed so $f'(\theta) = 0$, the condition can be reduced to

$$\lambda'(\theta) \geq -f(\theta). \quad (4.24)$$

If the *IR* constraint does not bind, the costate equation states that $\lambda'(\theta)$ is equal to $f(\theta)$ and the monotonicity condition is met when $\mu(\theta) = 0$. On the other hand, if the *IR* constraint is binding we have $\lambda'(\theta) = f(\theta) - \mu(\theta)$, $\mu(\theta) \geq 0$, and therefore $f(\theta) \geq \lambda'(\theta)$. Hence a necessary condition for monotonicity is

$$f(\theta) \geq \lambda'(\theta) \geq - \left[f(\theta) + \left\{ \frac{(u_q - c)}{u_{q\theta}} f'(\theta) + \lambda(\theta) \frac{u_{q\theta\theta}}{u_{q\theta}} \right\} \right]. \quad (4.25)$$

Although we will check whether the candidate for a quantity schedule meets the monotonicity constraint, we can tell by now that there is a fairly good chance that it does. The expression in the bracket parenthesis is zero or positive so the condition expresses that the marginal distortions when the *IR* constraint bind can be more than opposite the marginal distortions when the constraint is not binding.

4.3.1 The *IR* constraint is not binding

Since λ is continuous at $\bar{\theta}$ we can integrate up the costate equation (4.16), which gives us $\hat{\lambda}(\theta) = -(1 - F(\theta)) = \lambda$ as a candidate for $\lambda(\theta)$.

A candidate solution for $\hat{q} = q(\theta, \gamma)$ determined by (4.15) is given by

$$u_q(\hat{q}, \theta, \gamma) = c + \frac{1 - F(\theta)}{f(\theta)} u_{\theta q}(\hat{q}, \theta, \gamma). \quad (4.26)$$

This is the schedule we know from a monopoly nonlinear pricing problem.

Onwards the notation is simplified by writing the accent (e.g. bar, hat, or tilde) on the symbol for the function to denote that the function is to be evaluated at a point where $q(\theta, \gamma)$ has the relevant accent. Henceforth, $\hat{u} = u(\hat{q}, \theta, \gamma)$, $\bar{u} = u(\bar{q}, \theta, \gamma)$, and $\tilde{u} = u(\tilde{q}, \theta, \gamma)$. We can then write the slope of the quantity schedule as

$$\begin{aligned} \frac{\partial \hat{q}}{\partial \theta} &= \frac{(\hat{u}_q - c) \mathcal{H}' - \hat{u}_{\theta\theta q} + \mathcal{H} \hat{u}_{q\theta}}{-[\mathcal{H} \hat{u}_{qq} - \hat{u}_{\theta qq}]} \geq 0, \\ \mathcal{H} &= \frac{f(\theta)}{1 - F(\theta)}. \end{aligned} \quad (4.27)$$

Together with assumptions 4.1, when the hazard rate \mathcal{H} is increasing in θ , the Hamiltonian $H(\theta, q, U^{0*}(\theta, \gamma), \lambda(\theta))$ is strictly concave in q , $\partial \hat{q} / \partial \theta$ must be positive since our assumptions guarantee that both the numerator and the denominator is positive.

4.3.2 The IR constraint binds

Since the nonnegativity constraint is binding, we have $\partial U^0/\partial\theta = \partial \underline{U}^1/\partial\theta$, which implies that a candidate for $q(\theta, \gamma)$ is given by

$$u_\theta(\tilde{q}, \theta, \gamma) = u_\theta(\bar{q}_1, \theta, 1 - \gamma). \quad (4.28)$$

Let (4.28) determine $\tilde{q}(\theta, \gamma)$, and let (4.15) define a solution to $\tilde{\lambda}(\theta)$. The solution in $\mu(\theta)$ is determined by the costate equation.

Differentiating (4.28) yields a solution to $\partial\tilde{q}/\partial\theta$

$$u_{\theta q}(\bar{q}_1, \theta, 1 - \gamma) \frac{\partial\bar{q}_1}{\partial\theta} + u_{\theta\theta}(\bar{q}_1, \theta, 1 - \gamma) = u_{\theta q}(\tilde{q}, \theta, \gamma) \frac{\partial\tilde{q}}{\partial\theta} + u_{\theta\theta}(\tilde{q}, \theta, \gamma),$$

and by Definition 4.2(a)

$$\frac{\partial\bar{q}_1}{\partial\theta} = -\frac{\bar{u}_{q\theta}}{\bar{u}_{qq}} \geq 0, \quad (4.29)$$

so

$$\frac{\partial\tilde{q}}{\partial\theta} = \frac{\frac{(\bar{u}_{q\theta})^2}{-\bar{u}_{qq}} - [\tilde{u}_{\theta\theta} - \bar{u}_{\theta\theta}]}{\tilde{u}_{\theta q}}. \quad (4.30)$$

For $\tilde{q}(\theta)$ to be an increasing function it is necessary that

$$\left[\frac{(\bar{u}_{q\theta})^2}{-\bar{u}_{qq}} + \bar{u}_{\theta\theta} \right] - \tilde{u}_{\theta\theta} \geq 0. \quad (4.31)$$

Because the expression in the bracket is in fact $\partial^2 \underline{U}^1/\partial\theta^2$ and \underline{U}^1 is convex, the condition is certainly met when $\tilde{u}_{\theta\theta} \leq 0$.

Last, the IR constraint binds in the interval $[\underline{\theta}, \theta_1]$ where θ_1 is the solution in θ to the equation $\hat{q}(\theta, \gamma) = \tilde{q}(\theta, \gamma)$ (or equivalently $\hat{\lambda}(\theta) = \tilde{\lambda}(\theta)$), or $\theta_1 = \underline{\theta}$ if a solution in θ to $\hat{q}(\theta, \gamma) = \tilde{q}(\theta, \gamma)$ fails to exist (we can determine θ_1 this way only because we have assumed that $\lambda(\theta)$ is continuous).⁵⁸

The optimal allocation can now be characterized. Quantity-outlay allocations are described by the following characteristics (Stole, 1995)

$$q^*(\theta, \gamma) = \begin{cases} \tilde{q}(\theta, \gamma) & \text{if } \theta \in [\underline{\theta}, \theta_1] \\ \hat{q}(\theta, \gamma) & \text{if } \theta \in [\theta_1, \bar{\theta}] \end{cases}, \quad (4.32)$$

$$\theta_1 = \{\theta : \tilde{q}(\theta, \gamma) = \hat{q}(\theta, \gamma)\}, \quad (4.33)$$

⁵⁸Using the fact that $\lambda'_\theta = f(\theta) - \mu(\theta)$, $\mu \geq 0$ could lead us to the same conclusion. For $\theta = \bar{\theta}$, $\tilde{\lambda}(\bar{\theta}) - \hat{\lambda}(\bar{\theta}) < 0$ since we cannot have a jump at the right end of the distribution. Since μ is positive if IR binds, we must have $f(\theta) \geq \lambda'_\theta$. Thus, if the IR constraint is binding in any subinterval, this is always in the lower part, for some interval $[\underline{\theta}, \theta_1]$ – either $\tilde{\lambda}(\theta)$ crosses $\hat{\lambda}(\theta)$ once or not at all.

and finally

$$t^*(\theta, \gamma) = u(q(\theta, \gamma), \theta, \gamma) - U^{0*}(\theta, \gamma), \quad (4.34)$$

$$\begin{aligned} U^{0*}(\theta, \gamma) &= \underline{U}^1(\underline{\theta}, \gamma) + \int_{\underline{\theta}}^{\theta} u_{\theta}(q(s, \gamma), s, \gamma) ds, \\ &= \int_{\underline{\theta}}^{\theta} u_{\theta}(q(s, \gamma), s, \gamma) ds. \end{aligned} \quad (4.35)$$

This is proved in Stole (1995).

4.4 Implementation

The outlay function is the upper envelope of a family of indifference curves $u(q, \theta, \gamma) - t = U^0(\theta, \gamma)$. Since $q^*(\theta, \gamma)$ is strictly increasing in θ , there exists an inverse function $\theta^*(q, \gamma)$.⁵⁹

Using (4.34) we can define the outlay schedule $T(q^*, \gamma)$ by

$$T(q^*(\theta, \gamma), \theta, \gamma) \equiv t^*(\theta^*, \gamma) = u(q, \theta^*, \gamma) - U^0(\theta^*, \gamma), \quad (4.36)$$

and the slope of the outlay schedule $T(q^*, \gamma)$ is given by

$$\frac{dT}{dq} = u_q + (u_{\theta} - U_{\theta}^0) \frac{\partial \theta^*}{\partial q} = u_q(q, \theta^*, \gamma) \geq 0, \quad (4.37)$$

which is positive ($(u_{\theta} = U_{\theta}^0)$ by the envelope theorem).

The curvature of $T(q^*, \gamma)$ is given by

$$\frac{d^2T}{dq^2} = u_{qq}(q, \theta^*(q, \gamma), \gamma) + u_{q\theta}(q, \theta^*(q, \gamma), \gamma) \frac{1}{\partial q^*/\partial \theta}. \quad (4.38)$$

$u_{qq}(\cdot)$ is negative and the last term is positive, and $T(q^*, \gamma)$ is concave if

$$\frac{\partial q^*}{\partial \theta} \geq \frac{u_{q\theta}(q, \theta^*(q, \gamma), \gamma)}{-u_{qq}(q, \theta^*(q, \gamma), \gamma)} \geq 0. \quad (4.39)$$

Hence, concavity of the outlay schedule imply a stronger restriction than monotonicity with respect to $q^*(\theta, \gamma)$.

When the participation constraint is not binding, substituting $\partial \hat{q}/\partial \theta$ into (4.39), reorganizing and evaluating the condition for $q^*(\theta, \gamma) = \hat{q}(\theta, \gamma)$ yields

$$\frac{(-\hat{u}_{qq})(\hat{u}_q - c)H' + \hat{u}_{\theta qq}\hat{u}_{qq} - \hat{u}_{q\theta}\hat{u}_{\theta qq}}{(\hat{u}_{qq} - H\hat{u}_{\theta qq})\hat{u}_{qq}} \geq 0, \quad (4.40)$$

⁵⁹An early paper on implementation is Laffont and Tirole (1986)

and the outlay schedule is certainly concave for $q \in [q^*(\theta_1, \gamma), q^*(\bar{\theta}, \gamma)]$.

When the participation constraint binds we have to evaluate the condition for $\partial \tilde{q}/\partial \theta$. Rewriting condition (4.39) for $q^*(\theta, \gamma) = \tilde{q}(\theta, \gamma)$ yields

$$\frac{[u_{q\theta}(\bar{q}_1, \theta, 1 - \gamma)]^2}{-u_{qq}(\bar{q}_1, \theta, 1 - \gamma)} + u_{\theta\theta}(\bar{q}_1, \theta, 1 - \gamma) \geq \frac{[u_{q\theta}(\tilde{q}, \theta, \gamma)]^2}{-u_{qq}(\tilde{q}, \theta, \gamma)} + u_{\theta\theta}(\tilde{q}, \theta, \gamma). \quad (4.41)$$

The left-hand side in (4.41) is the second order derivative of the outside option (the first best utility) with respect to θ . This is assumed to be positive. The right-hand side is to be evaluated under a quantity distortion, i.e., $\tilde{q}(\theta, \gamma) \leq \bar{q}(\theta, \gamma)$, but at a more favorable location, i.e., $\gamma \leq (1 - \gamma)$, (for $\gamma \leq 1/2$). Hence, it is ambiguous whether the condition is met or not. At $\gamma = 1/2$, the left-hand side equals the right-hand side. If we are able to show that the right-hand side increases when γ decreases, we can conclude that (4.41) imply a contradiction. Thus, we differentiate the right-hand side at $\gamma = 1/2$ and evaluate the sign of this (the negative of the sign since $d\gamma < 0$)

$$-\frac{\partial}{\partial \gamma} \left\{ \frac{[u_{q\theta}(\tilde{q}(\theta, \gamma), \theta, \gamma)]^2}{-u_{qq}(\tilde{q}(\theta, \gamma), \theta, \gamma)} + u_{\theta\theta}(\tilde{q}(\theta, \gamma), \theta, \gamma) \right\}. \quad (4.42)$$

Hence, if (4.42) is positive the outlay schedule $T(q^*, \gamma)$ is convex, i.e., if

$$\left\{ \frac{2u_{q\theta} \left(u_{qq\theta} \left(-\frac{d\tilde{q}}{d\gamma} \right) + u_{q\theta\gamma} \right)}{u_{qq}} - \left(u_{qqq} \left(-\frac{d\tilde{q}}{d\gamma} \right) + u_{qq\gamma} \right) \left(\frac{u_{q\theta}}{u_{qq}} \right)^2 \right\} - \left\{ u_{\theta\theta q} \left(-\frac{d\tilde{q}}{d\gamma} \right) + u_{\theta\theta\gamma} \right\} \geq 0. \quad (4.43)$$

We can now formulate the following

Proposition 4.1 *If (4.43) is met, the outlay schedule $T(q^*, \gamma)$ defined by (4.36) is strictly convex for any θ in the interval $[\underline{\theta}, \theta_1)$ and consequently for any q in the interval $[q^*(\underline{\theta}, \gamma), q^*(\theta_1, \gamma))$ and strictly concave elsewhere, for $q(\theta, \gamma) > q(\theta_1, \gamma)$. Otherwise, $T(q^*, \gamma)$ is concave everywhere.*

Proposition 4.1 is proved by the preceding discussion. The sign of the expression in (4.43) is hard to evaluate using a general utility function. In the case with quadratic utility, $u = \theta(1 - \gamma)q - \frac{1}{2}q^2$, (4.43) reduces to $2(1 - \gamma) > 0$. With a logarithmic utility function, $u = \theta(1 - \gamma) \ln q$, (4.43) reduces to $1/\theta > 0$. Hence, for this two important cases, the outlay schedule is convex in the lower part.

Next, we turn to the problem of how to implement the outlay schedule. Instead of announcing the complete set of take-it-or-leave-it contracts, or announcing the fully nonlinear tariff $T(q^*, \gamma)$, the firm try to implement it via a menu of optional tariffs. These are described by the following Lemma

Lemma 4.4 *If the outlay schedule $T(q^*, \gamma)$ is to be implemented by a menu of tariffs defined by $T_\Lambda(q, \theta, \gamma)$, these tariffs must meet the following conditions*

$$\begin{aligned}
(i) \quad & T_\Lambda(q(\theta, \gamma), \theta, \gamma) = T(q^*, \gamma) = t^*(\theta, \gamma), \\
(ii) \quad & T_\Lambda(q, \theta, \gamma) \geq T(q^*, \gamma), \\
(iii) \quad & T_\Lambda(q, \theta, \gamma) \geq 0, \quad \forall q \geq 0.
\end{aligned}
\tag{4.44}$$

The conditions in Lemma 4.4 follow from the individual rationality constraint and the incentive compatibility constraints. With these characteristics, the outlay function is the lower envelope of the family of tariffs $T_\Lambda(q, \theta, \gamma)$. Implementation requires that type θ (with brand preference γ) finds it optimal to consume an amount $q^*(\theta, \gamma)$, and that he pays an amount $t^*(\theta, \gamma)$ for this consumption. When a consumer of type θ announces a type parameter θ' , it is equivalent to selecting a tariff $T_\Lambda(q, \theta', \gamma)$ and purchasing a quantity $q(\theta', \gamma)$. Expected utility is $u(q(\theta', \gamma), \theta, \gamma) - t(\theta', \gamma)$, and by construction of $t(\theta, \gamma)$ this is maximized when $\theta' = \theta$.

If $T(q^*, \gamma)$ is everywhere concave, we know that it can be represented by the lower envelope of its tangents. Hence, a menu of two-part tariffs will meet the incentive compatibility constraint and, of course, by construction, the individual rationality constraint. The following definition characterizes a menu of two-part tariffs.

Definition 4.3 *A menu of two-part tariffs (subscript 2P) is described by*

$$\begin{aligned}
T_{2P}(q, \theta, \gamma) &= u(q, \theta^*, \gamma) - U^0(\theta^*, \gamma) + u_q(q, \theta^*, \gamma)(q - q^*(\theta, \gamma)), \\
&= t^*(\theta, \gamma) + u_q(q^*, \theta, \gamma)[q - q^*(\theta, \gamma)].
\end{aligned}
\tag{4.45}$$

If $T(q^*, \gamma)$ is concave the menu of two-part given by definition 4.3 meet the requirements in Lemma 4.4. However, if $T(q^*, \gamma)$ is convex, or has convex parts, a two-part tariff that is the tangent to $T(q^*, \gamma)$ at a point $(q^*(\theta, \gamma), t(\theta, \gamma))$ would intersect $T(q^*, \gamma)$ at one or more points and, hence, it would violate part (ii) of Lemma 4.4. Alternatives to pooling tariffs, i.e., tariffs such that different quantity types are confronted with the same tariff, have to involve a more complicated scheme. The following definition characterizes a menu of three-part tariffs.⁶⁰

⁶⁰A three-part tariff can be considered as a moderated version of a “knife-edge” mechanism. In the absence of any uncertainty in demand, the allocation can always be implemented by a “knife-edge” mechanism, where a consumer pays $t(\theta, \gamma)$ if he announces θ and consumes $q(\theta, \gamma)$, otherwise he has to pay ∞ . But, with even very small demand disturbances present such a mechanism is not implementable. Picard (1987) shows that a menu of quadratic tariffs might implement the optimal solution in a situation where a menu of linear tariffs cannot. See also Laffont and Tirole (1993) pp. 107-109 for a reference to Picard in the case of quadratic transfer schemes in a regulation model. However, quadratic tariffs seem difficult to commercialize, and will therefore be of little interest in this context. Three-part tariffs on the other hand are a fairly good approximation to quadratic tariffs and are sufficiently simple to be understood by the market as well.

Definition 4.4 A menu of three-part tariffs (subscript 3P) is described by

$$T_{3P}(q, \theta, \gamma) = \begin{cases} t^*(\theta, \gamma) & \text{if } q \leq q^*(\theta, \gamma), \\ t^*(\theta, \gamma) \\ + u_q(\hat{q}(\theta, \gamma), \theta, \gamma) [q - q^*] & \text{otherwise.} \end{cases} \quad (4.46)$$

Although $T_{3P}(q, \theta, \gamma)$ is not differentiable at $q = q^*$, it is continuous and both the right side and left side limits are unique and equal to $t^*(\theta, \gamma)$. The menu described by 4.4 meets the requirements in Lemma 4.4 given that $u_q(\hat{q}, \theta, \gamma)$ is sufficiently large to satisfy part (ii) in Lemma 4.4. If not, we can substitute any schedule of marginal prices in the menu three-part tariffs that is decreasing in type.

A three-part includes in the fixed payment $t^*(\theta, \gamma)$ some “free” consumption allowance $q^*(\theta, \gamma)$, subsequent purchases are charged according to a unit price $u_q(\hat{q}, \theta, \gamma)$.

Finally, the following Proposition characterizes the solution in a (possibly) mixed tariff regime.

Proposition 4.2 (i) If the outlay schedule has a convex part in the lower quantity end, it can be implemented by a mixed tariff regime with a menu of three-parts and two-part tariffs. A mixed tariff regime is characterized by the following solution

$$T^*(q, \theta, \gamma) = \begin{cases} T_{3P}(q, \theta, \gamma) & \text{if } \theta \in [\underline{\theta}, \theta_2], \\ T_{2P}(q, \theta, \gamma) & \text{if } \theta \in (\theta_2, \bar{\theta}]. \end{cases} \quad (4.47)$$

θ_2 is the minimal solution to $\{\theta : T_{2P}(q, \theta, \gamma) = t^*(\theta, \gamma)\}$, which is given by $\{\theta : T_{2P}(q^*(\underline{\theta}, \gamma), \theta, \gamma) = t^*(\underline{\theta}, \gamma)\}$. (ii) Otherwise, the outlay schedule is concave everywhere and can be implemented by a menu of two-part tariffs, $T^*(q, \theta, \gamma) = T_{2P}(q, \theta, \gamma), \forall \theta$.

Proposition 4.2 is proved by the preceding discussion and by applying Lemma 4.4.

4.5 A numerical example with quadratic utility

Let us consider a numerical example with linear transportation costs and quadratic demand. For reasons of comparison, the assumptions are identical to those used in Stole (1995). The quantity parameter θ is distributed uniformly on the interval $[1, 2]$. Each firm’s marginal cost is equal to zero. Utility is specified by the function $u(q, \theta, \gamma) = \theta(1 - \gamma)q - \frac{1}{2}q^2$. The value of an outside option is $\underline{U}^1(\theta, 1 - \gamma) = \frac{1}{2}\theta^2\gamma^2$, when we normalize this to be zero for the very lowest type we get $\underline{U}^1(\theta, 1 - \gamma) = \frac{1}{2}\gamma^2(\theta^2 - 1)$. The utility function is linear in quantity type and localization and the reservation utility is convex in θ and γ .

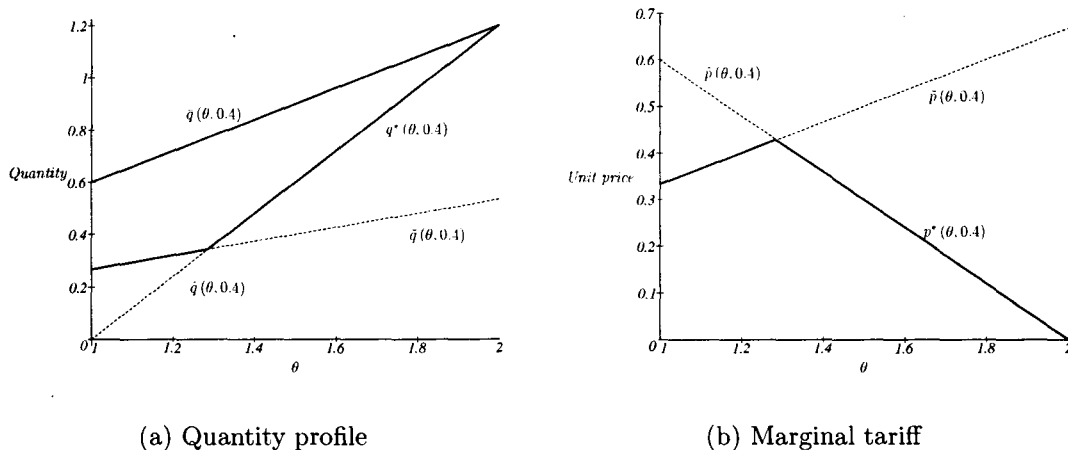


Figure 4.1: *The quantity schedule is implementable but not with two-part tariffs*

Figure 4.1 represents the solution with respect to q , p , λ and μ . The quantity schedule (the bold upper envelope) in figure 4.1 is the solution to the problem with direct revelation mechanisms. This is continuous and nondecreasing, and therefore truthfully implementable. In a monopoly context, the dual problem is to implement the allocation by offering the consumers to choose a tariff from a menu of two-part tariffs. But as we see in figure 4.1, this is not implementable since the outlay schedule is not everywhere a concave function.

The graphics in figure 4.2 illustrate the allocation in the (t, q) space and what implementation of $T(q^*, \gamma)$ looks like in the numerical example described above, for $\gamma = .4$. The fully nonlinear bold line represents $T(q^*, .4)$, the solid lines are some selected three-part and two-part tariffs, the dotted lines are consumers' indifference curves in the (t, q) space. The three-part tariff $T_{3P}(q, \theta, \gamma)$ will truthfully implement $q(\theta, \gamma)$, hence it is tangent to the indifference curve $U(\theta, \gamma) = t - u(q, \theta, \gamma)$ at the point $q = q(\theta, \gamma)$. Similarly, a two-part tariff $T_{2P}(q, \theta_2, \gamma)$ will truthfully implement $q(\theta_2, \gamma)$ because it is tangent to the indifference curve $U(\theta_2, \gamma) = t - u(q, \theta_2, \gamma)$ at $q = q(\theta_2, \gamma)$.

Empirical observations do support the theoretical results. The examples drawn in figure 4.3 seem to represent a trend for tariff arrangements in competitive markets.⁶¹ The first graphic shows examples on tariffs in the US long distance fixed telephony market, represented by some of AT&T's tariff offerings in the residential market, AT&T Basic and AT&T Savings.⁶² The latter graphic shows examples on tariffs effective in the Norwegian cellular market, represented

⁶¹The figures are based on assumptions about daytime-, evening-time and weekend-time usage, as well as usage patterns with respect to distance bands, and are illustrations rather than precise tariff computations.

⁶²See also Michell and Vogelsang (1991) and Wilson (1993) for a survey of the practice on telecommunications pricing during the seventies and eighties.

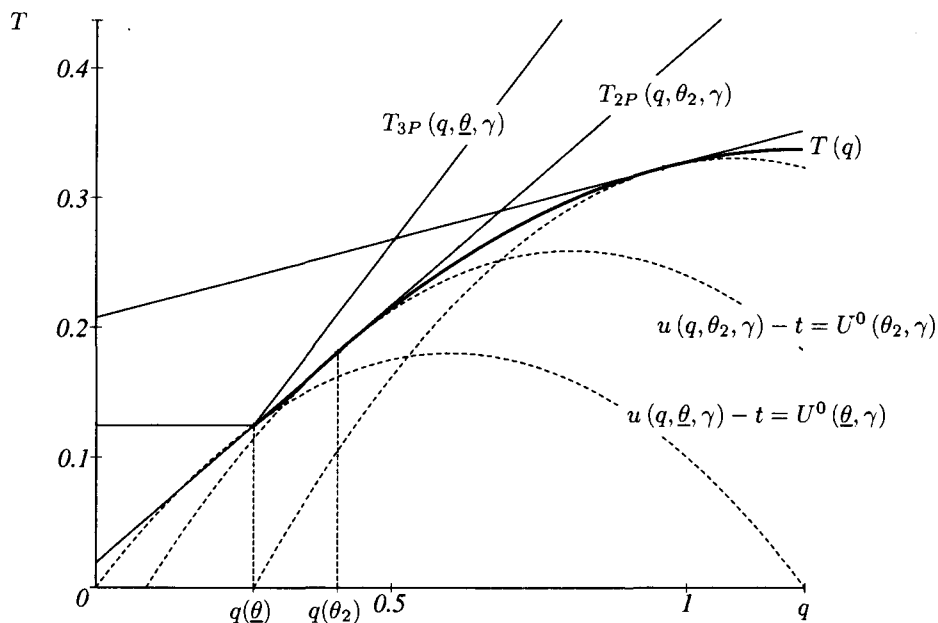


Figure 4.2: *Utility, outlay, three-part and two-part tariffs, $\gamma = .4$*

by the tariffs of Telenor Mobil.

The idea of using a combination of three-part and two-part tariffs seems more appealing when the outlay schedule is convex for low quantities. Three-part tariffs are communicated to the market as discounts conditional on a minimum usage level and such an idea would be hard to introduce towards high quantity users.

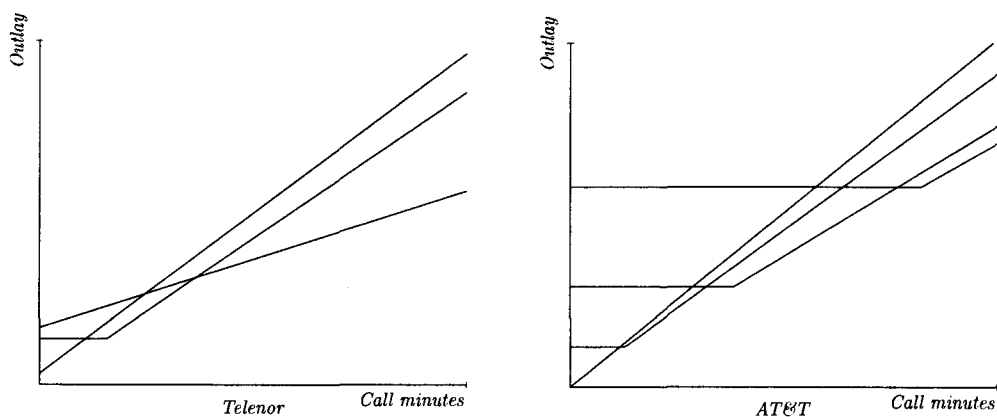


Figure 4.3: *Pricing of telecom services, AT&T and Telenor Mobil*

4.6 Concluding remarks

Linear contracts, as two-part tariffs, are attractive because under many conditions they implement the optimal contracts in an easy way. However, the paper shows that the problem is not as straightforward in a duopoly as it is in a monopoly setting. In a monopoly model, the monotone hazard rate condition is sufficient for the payment function to be concave, and hence for a menu of two-part tariffs to implement the outlay function. Although the monotone hazard rate condition is still a necessary condition in our duopoly model, it is shown that under reasonable assumptions two-part tariffs are outruled for low quantity purchases. In a monopoly the firm will balance the magnitude of downward quantity distortions below the first best level in order to reduce the information rent to better types (and all consumer surplus net of the transfer to the firm is informational rent). In the duopoly, however, the existence of an outside option places a restriction on consumers' net surplus. This will in turn change the magnitude of downward quantity distortions. This produces a convexity in the outlay schedule when the individual rationality constraint is binding and prevent the firm from using two-part tariffs for small purchases.

By analyzing the pricing strategies of the firms, one could draw conclusions about the competitiveness in the market. If the firms to a large extent are using three-part tariffs, this indicates that the market is more competitive. Although one should be careful in making comparisons of different markets, the US long distance market seems to be more competitive than the Norwegian cellular market.

Chapter 5

Two-part Tariffs, Consumer Heterogeneity, and Cournot Competition

by Sissel Jensen and Lars Sørsgard*

5.1 Introduction

Nonlinear prices are common in many industries, and have been studied extensively in the economic literature. However, most theoretical studies use a monopoly setting. In contrast, we observe nonlinear prices not only in monopoly markets, but also in other market settings such as oligopoly. The purpose of this article is to help bridging this gap. We analyze two-part tariffs in a Cournot-like setting by extending the seminal model of Harrison and Kline (2001).

Nonlinear pricing may not be sustainable in oligopoly. For example, Mandy (1992) finds that in a traditional Bertrand oligopoly with homogeneous products – where we allow the firms to set nonlinear prices – all prices may collapse to a uniform price. The finding illustrates that, except for some special cases which he explores, some of the assumptions in the traditional Bertrand model have to be relaxed in order to make nonlinear prices sustainable in oligopoly. This has been done in the emerging literature on nonlinear prices. One extension of the traditional Bertrand model is to introduce product differentiation, see Callem and Spulber (1984), Castelli and Leporelli (1993), Economides and Wildman (1995),

*We are grateful for comments from Petter Osmundsen. We are also indebted to participants at the Nordic Workshop on ICT-related research at Norwegian School of Economics and Business Administration in June 2001, and to participants at the doctoral seminar series at the Norwegian School of Economics and Business Administration for helpful comments.

Shmanske (1991), and Young (1991). Another extension is to introduce capacity constraints, as is done in Harrison and Kline (2001), Oren et al. (1983), Scotchmer (1985a, 1985b) and Wilson (1993). Scotchmer (1985a, 1985b) only considers existence when the number of firms becomes large, while both Oren et al. (1983) and Wilson (1993) assume that the firms predict the market shares of their rivals. In contrast, Harrison and Kline (2001) model quantity as a strategic variable and consider the strategic interaction between a small (or a large) number of firms.

In Harrison and Kline (2001) each firm commits to a certain quantity, as is the case in a traditional Cournot model. In addition, each firm sets its fixed fee while the unit prices are determined endogenously by market forces. The latter is analogous to what is the case in a traditional Cournot model. In their paper, Harrison and Kline also provide some examples where we do observe that fixed fees are less flexible than prices per unit.⁶³ It is found that in equilibrium price is set equal to marginal costs, and the fixed fee is positive for a given number of firms. Furthermore, it is found that fixed fees extract the entire consumer surplus if the number of firms is sufficiently small. Finally, they found that when the number of firms approaches infinity the fixed fee tends toward zero.

We extend the model introduced in Harrison and Kline (2001) by assuming two instead of one type of consumers. It turns out that none of the conclusions referred to above is robust to such an extension of the model. If both types of consumers are served, we find that price per unit is above marginal costs. Furthermore, fixed fees can be zero or even negative for a finite number of firms. In fact, firms can be better off committing to traditional Cournot competition where the firms can only charge a unit price. If the firms can choose whether to serve one or both types of consumers, they may choose to serve only the large consumers. Then the equilibrium outcome replicates the one shown in Harrison and Kline (2001), except there are now some consumers that are not served. However, by using a numerical example we show that there can be multiple equilibria. Moreover, it is shown that both firms serving both types of consumers can be an equilibrium duopoly outcome in cases where the monopolist would have preferred to serve only one type of consumers. The driving force is that the rival, non-deviating firm supplies a given quantity which it is committed to sell, acting as a constraint on the deviating firm's price setting.

The article is organized as follows. In Section 2 we formulate our model, and report optimal pricing strategies given that all firms serve either both types of consumers or only one type. In section 3 we explore the equilibrium outcomes of the model. First, we consider the case with full market coverage, that is, both firms are restricted to sell to both types of consumers. Second, we consider the case where each firm chooses either to sell to one or both types of consumers.

⁶³One example is a consumer club like Costco. The membership fee corresponds to a fixed fee and the prices of the products a member buys when he visits the store may vary considerably. For more examples, see Harrison and Kline (2001).

Finally, in Section 4 we offer some concluding remarks.

5.2 The Model

We consider a setup with k identical firms, $k \geq 2$, supplying a homogeneous product. The cost function is characterized by constant returns to scale, $C(Q) = cQ$ where $c > 0$ is the marginal cost and Q is output. For simplicity we omit fixed costs. The number of firms is exogenous and the question of entry is left outside the scope of this paper.⁶⁴

There are two groups of consumers with a total of N . Consumers with taste parameter θ_1 are in proportion λ and consumers with taste parameter θ_2 are in proportion $(1 - \lambda)$.⁶⁵ Preferences are defined by a quasi-linear utility function

$$\begin{aligned} V &= \begin{cases} u(q, \theta_\ell) - T & \text{if they pay } T \text{ and consume } q \text{ units} \\ 0 & \text{if they do not buy} \end{cases}, \\ \theta_\ell &= \{\theta_1, \theta_2\}, \\ u(q, \theta_2) &\geq u(q, \theta_1), \quad \forall q. \end{aligned} \tag{5.1}$$

The utility function is assumed to be increasing and strictly concave in q , $u(0, \theta) = 0$, $\lim_{q \rightarrow 0} u_q(q, \theta) \geq c$, $\lim_{q \rightarrow \infty} u(q, \theta) \leq 0$. For any tariff $T = A + pq$, where A is a fixed fee that is paid up-front and p is a unit price, utility maximization yields a downward sloping demand curve for each individual which is independent of income and therefore also of the fixed fee. Indirect utility gross of the fixed fee is

$$\begin{aligned} q_\ell(p) &\equiv q(p, \theta_\ell) = \max_q u(q, \theta_\ell) - pq - A, \\ V(p, \theta_\ell) &= u(q_\ell(p), \theta_\ell) - pq_\ell(p), \\ V'_p &= -q_\ell(p), \\ \ell &= 1, 2. \end{aligned} \tag{5.2}$$

With quasilinear utility we can measure the indirect utility in monetary terms. Consumers choose to buy if they obtain a nonnegative net surplus at some firm i , that is, iff $V(p_i, \theta_j) - A_\ell \geq 0$, $i \in \{1, 2, \dots, k\}$ and $\ell = 1, 2$. They buy from the firm providing them with the highest surplus, $V(p_i, \theta_\ell) - A_i \geq V(p_j, \theta_\ell) - A_j$, $(i, j \in \{1, 2, \dots, k\}, i \neq j, \ell = 1, 2)$. When the two consumer types are charged the same tariff, a type 2 consumer obtains a surplus that is at least as large as the surplus a type 1 consumer obtains. Thus, if type 1 is able to obtain a nonnegative surplus, type 2 obtains a strictly positive surplus.

Firms act to maximize profit by choosing a strategy $s_i = (Q_i, A_i)$, with $Q_i > 0$ for all $i = 1, 2, \dots, k$, and we assume that firms are able to commit to this strategy. The firm cannot exclude any consumer from buying. In our model, we use

⁶⁴See Harrison and Kline (2001) on entry in this model.

⁶⁵We refer to the first group as type 1 consumers or low demand consumers and to the other group as type 2 consumers or high demand consumers.

the assumption that for a given strategy combination there exists a consumer equilibrium defining a consumer-price profile $((n_1, \dots, n_k), (p_1, \dots, p_k))$. This is formally defined in Harrison and Kline (2001). Although we define a firm's strategy in capacity and the fixed fee, from a consumer's point of view he chooses the quantity that maximizes his utility for a given A_i and p_i . The notion behind this reasoning is that it is a competitive equilibrium where a large number of consumers without market power trade, given the fixed fees and quantities from each firm. If all firms leave each consumer with equal and nonnegative surplus, we assume that all firms serve an equal share of each consumer type, $n_i = \lambda N/k + (1 - \lambda)N/k$.

If there are at least two active firms, the relevant participation constraints in firm i 's optimization problem are given by

$$V(p_i, \theta_\ell) - A_i \geq V(p_j, \theta_\ell) - A_j, \quad \ell = 1, 2, \quad j \in \{1, 2, \dots, i-1, i+1, \dots, k\}. \quad (5.3)$$

Profit for firm i is given by

$$\Pi_i = n_i A_i + (p_i - c) Q_i. \quad (5.4)$$

Since the fixed fee is a lump sum transfer from consumers to the firm, the unit price in firm i 's tariff is adjusted in such a way that aggregate demand for firm i 's product is equal to firm i 's supply. Hence, the unit price is independent of the fixed fee. Whenever the fixed fee is positive, consumers will make all or nothing purchases at firm i . When firm i serves a total of n_i consumers, the unit price is adjusted to satisfy the following market clearing condition

$$Q_i = n_i [\lambda q_1(p_i) + (1 - \lambda) q_2(p_i)]. \quad (5.5)$$

In line with Harrison and Kline (2001), let us assume that all firms charge the same fixed fee and the same unit price. Firm i maximizes profit subject to the condition that the unit prices charged by rival firms are adjusted to satisfy the market clearing condition and subject to voluntary participation. When every other firm but i serves both consumer types the unit price p charged by every other firm must satisfy the condition

$$\begin{aligned} Q_{-i} &= (N - n_i) [\lambda q_1(p) + (1 - \lambda) q_2(p)], \\ Q_{-i} &= \sum_{j \neq i} Q_j. \end{aligned} \quad (5.6)$$

When rival firms charge their consumers according to the tariff $T = A + pq$ consumer θ_ℓ is indifferent between buying from firm i and one of the other firms when the participation constraint is binding. If the firm leaves the consumer with additional surplus, it sacrifices profit. We therefore expect

$$V(p_i, \theta_\ell) - A_i = V(p, \theta_\ell) - A, \quad \ell = 1, 2. \quad (5.7)$$

Henceforth, superscript 12 denotes that both consumer types are served and superscript 2 denotes that type 1 (the “small” type) is excluded. Let us first suppose that both consumer types are served. Then there is at least one additional active firm where both consumers buy a strictly positive quantity. When the best alternative option for a type 1 consumer is represented by a tariff T^{12} , the relevant participation constraint is given by

$$V(p_i^{12}, \theta_1) - A_i^{12} = V(p^{12}, \theta_1) - A^{12}. \quad (5.8)$$

Taking rival firms’ tariffs as given and maximizing profit with respect to p_i^{12} give the following optimality condition for the unit price in a two-part tariff

$$p_i^{12} = c + \frac{(1 - \lambda)[q_2 - q_1]}{-[\lambda q_1' + (1 - \lambda)q_2']}, \quad (5.9)$$

$$q_\ell' \equiv \frac{dq_\ell}{dp}.$$

Next, firm i must choose the strategy (Q_i^{12}, A_i^{12}) in such a way that p_i^{12} satisfies the market clearing condition. To attract additional consumers from rival firms, firm i has to adjust the fixed fee. Hence, a marginal increase in market share affects firm i ’s profit via the fixed fee. Finding the profit maximizing strategy reduces to finding the optimal number of consumers to serve.

The effect on the firm’s profit of a marginal increase in market share is

$$\frac{\partial \Pi_i^{12}}{\partial n_i} = A^{12} - \frac{p^{12} q_1}{\varepsilon(Q)} \left(\frac{1}{k-1} - (1 - \lambda) \frac{q_2 - q_1}{q_1} \right). \quad (5.10)$$

If all firms exclude type 1 and serve type 2 alone, the participation constraint when the best alternative option for type 2 consumers is represented by a tariff T^2 becomes

$$V(p_i^2, \theta_2) - A_i^2 = V(p^2, \theta_2) - A^2. \quad (5.11)$$

The optimal tariff is a cost-plus-fixed-fee tariff and firm i chooses a strategy (Q_i^2, A_i^2) in such a way that the market clearing condition is satisfied when $p^2 = c$.

Again, applying symmetry, the effect on the firm’s profit of a marginal increase in market share is

$$\frac{\partial \Pi_i^2}{\partial n_i} = (1 - \lambda) \left(A^{12} - \frac{1}{k-1} \left[\frac{c q_2(c)}{|\varepsilon(q_2(c))|} \right] \right). \quad (5.12)$$

Notice that if firm i takes the number of consumers it serves as given, for any tariff charged by rival firms the reservation utility is defined as a constant and will not affect the optimization with respect to unit price. The problem then resembles the monopoly problem, and the marginal price in our model is identical to that in a monopoly.

The following two Lemmas state the pricing strategies in a k -firm oligopoly, given that they either serve both types or exclude type 1.

Lemma 5.1 (Two consumer types) (i) *Let us assume that both consumer types are served by all firms. Then the pricing strategy in two-part tariffs in a k -firm oligopoly is given by*

$$\begin{aligned}
A_i^{12} &\equiv A_{TT}^{12} = \min \left\{ V(p^{12}, \theta_1), \frac{p^{12} q_1}{\varepsilon(Q^{12})} \left(\frac{1}{k-1} - (1-\lambda) \frac{q_2 - q_1}{q_1} \right) \right\} \\
p_i^{12} &\equiv p_{TT}^{12} = c + \frac{(1-\lambda)[q_2 - q_1]}{-(\lambda q'_1 + (1-\lambda)q'_2)} \\
Q_i^{12} &\equiv Q_{TT}^{12} = \frac{N}{k} (\lambda q_1 + (1-\lambda)q_2) \\
(q_\ell &= q_\ell(p_{TT}^{12}), q'_\ell = q'_\ell(p_{TT}^{12}), \ell = 1, 2)
\end{aligned} \tag{5.13}$$

(ii) *If both consumer types are served, the pricing strategy in a traditional Cournot game is given by*

$$\begin{aligned}
A_i^{12} &\equiv A_{UP}^{12} = 0 \\
p_i^{12} &\equiv p_{UP}^{12} \geq c \\
Q_i^{12} &\equiv Q_{UP}^{12} = \frac{N}{k} (\lambda q_1 + (1-\lambda)q_2) \\
(q_\ell &= q_\ell(p_{UP}^{12}), q'_\ell = q'_\ell(p_{UP}^{12}), \ell = 1, 2)
\end{aligned} \tag{5.14}$$

where p_{UP}^{12} is the (standard) price when both types are served in a Cournot game with k identical firms charging a uniform price.

Lemma 5.2 (Harrison and Kline) *If one of the consumer types is excluded from purchasing, the pricing strategy in two-part tariffs in a k -firm oligopoly is given by*

$$\begin{aligned}
A_i^2 &\equiv A_{TT}^2 = A = \min \left\{ V(c, \theta_\ell), \frac{c q_\ell}{(k-1) |\varepsilon(q_\ell)|} \right\} \\
p_i^2 &\equiv p_{TT}^2 = c \\
Q_i^2 &\equiv Q^2 = \frac{N}{k} \lambda_\ell q_\ell \\
\lambda_\ell &= \lambda \text{ if } \ell = 1, \lambda_\ell = (1-\lambda) \text{ if } \ell = 2, \\
(q_\ell &= q_\ell(c), \ell = 1 \text{ or } 2)
\end{aligned} \tag{5.15}$$

with $\theta_2 \geq \theta_1$ type 2 will always be served.

Lemma 5.2 is the result in Harrison and Kline (2001) when the tariffs are symmetric. Lemma 5.1 is the extension of this to the two-type case, and the proof is given by the previous calculations. According to Lemma 5.2 the fixed fee in the single-type case converges toward zero as the number of firms approaches infinity. Moreover, note that the price per unit is set equal to marginal costs in the case with one type. As Lemma 1 indicates, these results are reversed when we extend the model from one to two types.

Harrison and Kline (2001) give a thorough treatment of Cournot competition with two-part tariffs and a single consumer type, and they also guide the reader

through all proofs in that case. They show that the pricing described in Lemma 5.2 is a unique Nash equilibrium in pricing strategies for the game. All k firms produce. In addition to the equilibrium with symmetric market, shares there also exist equilibria that are asymmetric in market shares.

In what follows, we consider first the firms' pricing in a symmetric equilibrium when all consumers are served. Next, since low demand types may be excluded we consider the prospects for a unique equilibrium with symmetric pricing in a duopoly with respect to market coverage.

5.3 Equilibrium outcomes

To illustrate the equilibrium outcomes, we have chosen to focus on a case where consumer preferences are represented by a quadratic utility function. We let the reservation utility be zero for both consumers. $V = \theta_\ell q - \frac{1}{2}q^2 - T$, $\ell = 1, 2$, if they pay T and consume q units, otherwise they obtain zero utility. Each consumer has a linear demand function $q_\ell = \theta_\ell - p$, $\ell = 1, 2$. Letting $\theta \equiv \lambda\theta_1 + (1 - \lambda)\theta_2 \geq \theta_1$, expected demand is $\lambda q_1 + (1 - \lambda)q_2 = \theta - p$. The indirect utility exclusive of the fixed fee for a consumer paying a unit price of p is $V(p, \theta_\ell) = \frac{1}{2}(\theta_\ell - p)^2$, $\ell = 1, 2$. Because we are interested in how equilibrium strategies are affected by heterogeneity in demand, the example is somewhat simplified by letting $\theta_1 = 1$ and $c = \frac{1}{2}$. Increased demand side heterogeneity is captured by variations in λ and θ_2 . Large heterogeneity can then come about either by an increase in the number of type 2 consumers (λ decreases), or because a type 2 consumer has larger willingness to pay relative to a type 1 consumer (θ_2 increases). Hence, increased demand side heterogeneity is captured by an increase in θ . We use Lemmas 5.1 and 5.2 to characterize the equilibrium in terms of pricing and expected profit per consumer. All these computations are given in the appendix.

5.3.1 Market coverage

Let us first consider the case where both types are served by all firms.⁶⁶ This could be due to some institutional restrictions, forcing them to provide a universal service. Given such a restriction, which combination of fixed fee and price per unit would each firm choose?

Proposition 5.1 *Let us assume that both types of consumers are served and each firm sets a two-part tariff. If (i) $0 \leq \lambda \leq \lambda^* \equiv \frac{4\theta_2 - 5}{4\theta_2 - 4}$, or (ii) $k > k^* \equiv \frac{1}{2(\theta_2 - 1)(1 - \lambda)}$, then $A_{TT}^{12} < 0$ and $p^{12} = \lambda + (1 - \lambda)\theta_2 - \frac{1}{2} \equiv p_{TT}^{12} > c$. Otherwise, $A_{TT}^{12} > 0$ and $p^{12} = p_{TT}^{12} > c$.*

⁶⁶In the next section we show that this can be the equilibrium outcome for a large number of parameter values.

The critical values λ^* and k^* are derived in the appendix. First, we see that each firm would set a price per unit that exceeds marginal costs. In contrast, Harrison and Kline (2001) found that each firm would set a price per unit equal to marginal costs. Obviously, the extension of the model – from one to two types of consumers – explains the change in the result. It is well known from a monopoly model that a firm that serves two types of consumers with one two-part tariff should let the unit price exceed marginal costs, see Oi (1971). By doing so it is able to extract more profits from the high demand consumer, and this outweighs the loss in profit extraction from the low demand consumer as long as the price-cost margin is not above a certain threshold level. The price-cost margin is higher the larger the difference between the consumer types (θ_1 versus θ_2), and the larger the proportion of the high demand consumers (λ approaches zero). This is natural, since a large difference between those two groups of consumers would lead to a relatively high price-cost margin to extract profits from the larger group.

Second, note that the price-cost margin is *not* influenced by the number of firms. At first glance, this may come as a surprise. Why do they not compete on prices? The reason is that they compete on access prices, not prices per unit. The prices per unit are set to balance the revenues from the two consumer groups, *after* they have competed on fixed fees to attract consumers. Note that our result is in line with the result in Harrison and Kline (2001), where the price per unit is always equal to marginal costs since the unit price in both cases just replicates the monopoly price.

Third, we see that each firm's fixed fee can be set below the fixed cost of serving a consumer (which is normalized to zero in our setting). In contrast, Harrison and Kline (2001) found that the fixed fee is always above costs, but approaches costs when the number of firms approaches infinity. In their setting, as well as in ours, profits approach zero when the number of firms approaches infinity. But the fact that we have a positive price-cost margin, implies that the fixed fee is competed away even for a finite number of firms. In fact, if the demand side heterogeneity is sufficiently large, the fixed fee is competed away even in a duopoly.

Obviously, the existence of many firms would lead to fierce competition on fixed fees. But even with two firms, fixed fees can be negative if the fraction of the high type consumer is large or when the difference in consumers' type is large. In such a case the price per unit is high, to extract profits from the "large" consumer. Then the fixed fee is low even in a monopoly setting, and competed away in a duopoly setting.

An interpretation of a negative fixed fee in our model is that the fixed fee is positive, but below costs. This is what we observe in some cases. In Norway for example, mobile phones have been sold at a price of NOK 1 each, while some retailers have received a payment of approximately NOK 2000 from the producer. The producer then incurs a loss of approximately NOK 2000 for each consumer

it captures, and earns revenues on the same consumer from what he pays for the use of the mobile phone.⁶⁷ What is labelled *loss leaders* in the grocery sector can be interpreted in a similar way. Grocery stores advertise low prices on certain products in order to attract consumers to the store, and the consumers end up buying both the advertised product as well as other products. It has been shown that the grocery store should then set a price below costs on the advertised products, and a high price-cost margin on other products (see Lal and Matutes (1994)).

In some instances, however, access can be cost free (or close to cost free), for instance joining some kind of club as the examples referred to in Harrison and Kline (2001). Hence, an obvious question is whether the firm would have been better off constraining its tariff policy to uniform pricing. What, then, if the firm sets a fixed fee equal to zero rather than a negative fixed fee? It can then be shown that the following would emerge as equilibrium outcomes

Proposition 5.2 *Let us assume that both types of consumers are served and each firm can choose either to set a two-part tariff or a uniform price (fixed fee equal to zero). Then each firm chooses a uniform price if the fixed fee in a two-part tariff would be negative (see the previous Proposition), where $p_{UP}^{12} < p_{TT}^{12}$. Otherwise, it chooses a two-part tariff with $A_{TT}^{12} > 0$ and $p_{UP}^{12} > p_{TT}^{12}$.*

First, we see that as long as the fixed fee is above costs in a setting with a two-part tariff, the firm would set a two-part tariff rather than restrict its pricing policy to a uniform price. A uniform price, which equals the traditional Cournot price, would in that case be higher than the unit price in a two-part tariff. This suggests that a firm would find it profitable to deviate from an outcome where both firms set a uniform price. It could deviate by setting a lower price per unit, and extract the gross consumer surplus it generates through a positive fixed fee. Therefore, we would expect that the firms would end up with a two-part tariff with a positive fixed fee.

Second, we see that each firm would choose a uniform price if the alternative is that both firms set a two-part tariff with a negative fixed fee. To understand this, note that in such a case the price per unit in a two-part tariff is higher than the traditional Cournot price (a uniform price). In our model, the firms compete in utility levels. Then if other firms hold a high unit price and generate consumer surplus via a negative fixed fee, it will be profitable to match other firms' offer by restricting the fixed fee to zero and lowering the price per unit, thereby increasing consumer surplus.

Note that competition between the firms leads to a low price per unit: The equilibrium outcome is a uniform (Cournot) price if that price per unit is lower

⁶⁷Strictly speaking, the tariff structure is more complicated than the one with a fixed fee and a price per unit. The user pays a fixed fee in addition to a monthly fixed fee and a price per unit. Then the fixed fee is followed by a two-part tariff, not a uniform price as in our model.

than the price per unit in a two-part tariff, and vice versa. This is illustrated in Figure 1, where the solid lines show the price per unit in equilibrium.

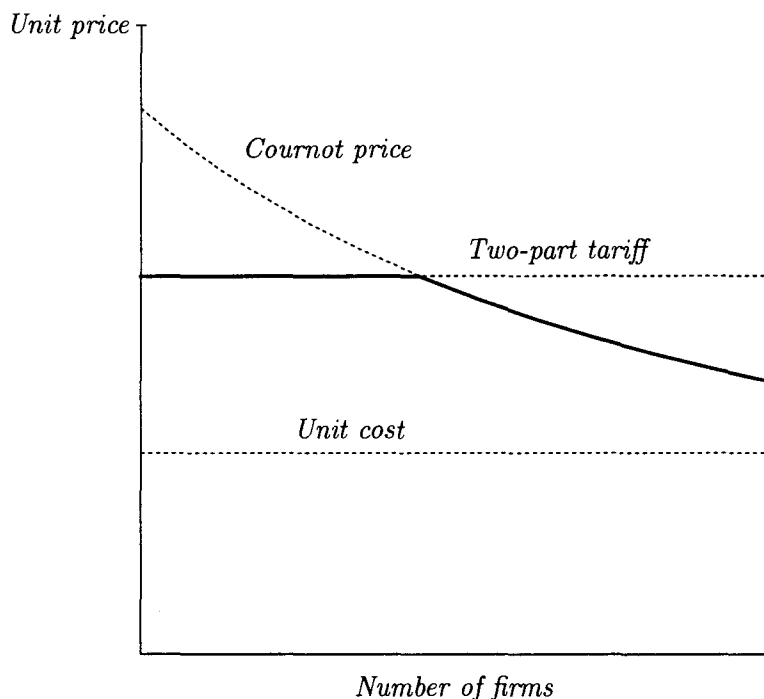


Figure 5.1: *Price per unit in equilibrium.*

As explained above, in some instances the institutional setting is such that the firms are forced to set a fixed fee. In other instances, though, firms are more flexible. If the choice is either to set a negative fixed fee and a relative high price per unit or a low uniform price, each firm may end up choosing the latter price system because that would generate a larger sale and thereby a larger profit. This suggests that there is no conflict between public policy and private incentives concerning the choice of tariff structure. Each firm has incentive to choose the tariff structure with the lowest price per unit, which is beneficial for consumers and leads to only a limited dead weight loss.

5.3.2 Market coverage versus exclusivity

In the previous section, we assumed that each firm served both types of consumers. This may not be the equilibrium outcome. As is well known from monopoly, in some cases it is beneficial for a firm to exclude the type with low willingness to pay and in other cases it is preferable to serve both types of consumers. Would the same be true in oligopoly? It turns out to be hard to obtain closed form solutions when we assess the firm's incentive to deviate from an equilibrium with symmetric tariffs and market shares. That is, to decide whether the

case where type 1 is served or excluded, respectively, is a stable equilibrium or not. We have therefore chosen to present some numerical examples to illustrate possible equilibrium outcomes.

To simplify, let us consider duopoly. Consider the two equilibrium candidates in pure strategies where the firms announce identical tariffs and serve the same customer base. In the first equilibrium candidate, both consumer types are served with a tariff $(A_{TT}^{12}, p_{TT}^{12})$ and each firm earns a profit per consumer π_{TT}^{12} . In the second equilibrium candidate, low demand consumers are excluded from making purchases and type 2 is served with a tariff (A_{TT}^2, c) . Each firm earns a profit per consumer π_{TT}^2 . For now we assume that the firms have equal market shares, *i.e.*, $n_a = n_b = \frac{1}{2}N$. Expected profit in each of the two possible equilibrium outcomes is

$$\Pi_{TT}^{12} = \frac{N}{2} \pi_{TT}^{12}, \quad (5.16)$$

and

$$\Pi_{TT}^2 = \frac{N}{2} (1 - \lambda) \pi_{TT}^2. \quad (5.17)$$

If demand side heterogeneity is not too large, a duopoly is able to extract all surplus from type 1 when both consumer types are served. They would generate the same profit in each of the symmetric cases when $\Pi_{TT}^{12} = \Pi_{TT}^2$, *i.e.*, if

$$\lambda \equiv \lambda^{**} = \frac{1}{2} + \frac{3 - 4\theta_2 + \sqrt{(4\theta_2^2 - 3)(4\theta_2^2 - 8\theta_2 + 5)}}{8(\theta_2 - 1)^2}. \quad (5.18)$$

Since the duopoly extracts all surplus from type 1 provided that $\lambda \geq (2\theta_2 - \sqrt{2} - 1)/(2\theta_2 - 1)$ (which is smaller than λ^{**}), λ^{**} is also the monopolist cutoff value: If $\lambda < \lambda^{**}$ it serves only type 2 consumers, while if $\lambda > \lambda^{**}$ it serves both types of consumers.

Let us use λ^{**} as a reference point for our numerical examples. If $\lambda < \lambda^{**}$, demand side heterogeneity is large and we conjecture that the firms would tend to exclude type 1. Conversely, we conjecture that each firm would tend to serve both types of consumers if $\lambda > \lambda^{**}$. Note, however, that it is not at all obvious that the cutoff point is the same in duopoly as in monopoly. A monopoly can exclude type 1 consumers by designing a tariff they would never accept, while this is not possible in a duopoly. To find the Nash equilibrium, we check for unilateral deviations from each of those two possible equilibrium candidates, for different values of λ . Then we can compare the equilibrium outcome in duopoly with the equilibrium outcome in monopoly.

First, let us consider the equilibrium candidate where both firms serve only type 2 and the firms' tariffs are given by (A_{TT}^2, c) . Type 1 is excluded and the firms extract the entire surplus from type 2 via the fixed fee, and $A_{TT}^2 = V(c, \theta_2)$. The two firms split the base of type 2 consumers equally, $n_a = n_b = (1 - \lambda)N/2$.

Would a unilateral deviation from an outcome where both firms serve only type 2 be profitable? One firm, say firm a , could deviate by setting a tariff that type 1 is just willing to accept and capture all type 1 consumers, λN . However,

since low demand consumers derive nonnegative surplus, high demand consumers will derive strictly positive surplus by switching to low demand types' tariff. The deviating firm will then serve a mix of type 1 and type 2 consumers, it will serve all type 1 consumers and more than half of all type 2. Since firm a captures some of the high demand types as well, this tends to make such a deviation profitable. Let the deviating firm choose a strategy $(\tilde{Q}_{TT}^{12}, \tilde{A}_{TT}^{12})$, or equivalently charge a tariff $(\tilde{A}_{TT}^{12}, \tilde{p}_{TT}^{12})$ in order to maximize profit subject to individual rationality and firm b 's strategy (Q_{TT}^2, A_{TT}^2) . The problem is to maximize

$$\begin{aligned} \tilde{\Pi}_{TT}^{12} | \Pi_{TT}^2 = [N - \bar{n}_b] \tilde{A}_{TT}^{12} + \\ (\tilde{p}_{TT}^{12} - c) [N\lambda\tilde{q}_1 + (N(1-\lambda) - n_b)\tilde{q}_2] \end{aligned} \quad (5.19)$$

subject to

$$V(\tilde{p}_{TT}^{12}, 1) \geq \tilde{A}_{TT}^{12}, \quad (5.20)$$

$$V(\tilde{p}_{TT}^{12}, \theta_2) - V(\tilde{p}_{TT}^{12}, 1) = V(\bar{p}_{TT}^2, \theta_2) - V(c, \theta_2), \quad (5.21)$$

$$\frac{N}{2}(1-\lambda)q_2 \geq \bar{n}_b\bar{q}_2, \quad (5.22)$$

where $\tilde{q}_i = q_i(\tilde{p}_{TT}^{12})$, $\ell = 1, 2$, $\bar{q}_2 = q_2(\bar{p}_{TT}^2)$, and $q_2 = q_2(c)$. Firm b , the non-deviating firm, will then lose type 2 consumers. This leads to a price reduction at firm b in order to restore individual rationality, the unit price falls to $\bar{p}_{TT}^2 < c$. Since a unit price reduction in turn leads to an increase in a type 2 consumer's demand, $q_2(\bar{p}_{TT}^2) > q_2(c)$, the capacity supplied by firm b becomes insufficient to serve all type 2 consumers, and $\bar{n}_b < (1-\lambda)N/2$ is adjusted to restore market clearing at firm b . Formally, the individual rationality constraint (5.21) and the market clearing condition (5.22) jointly determine firm b 's share of type 2 consumers as a function of firm a 's strategy, $\bar{n}_b = \bar{n}_b(\bar{p}_2(\tilde{p}_{TT}^{12}))$.

Although firm a obtains lower profit per consumer when it deviates, it expands its market. When λ is low or θ_2 is high, the market expansion effect is less likely to cover the per-consumer-loss in profit. In that case there are few type 1 consumers to serve and expected profit per consumer is significantly lower when firm a deviates. Conversely, we expect that a deviation is profitable when demand side heterogeneity is low. For λ close to λ^{**} the expected revenue per consumer is identical and we therefore conjecture that it is profitable to deviate.

In Table 1 we have reported some numerical examples for $N = 100$ and $c = \frac{1}{2}$. Hence, $p_{TT}^2 = \frac{1}{2}$ and $A_{TT}^2 = V(c, \theta_2)$. The results in Table 1 confirm our conjecture. Note that when $\lambda < \lambda^{**}$, the monopolist would serve only type 2 consumers. This particular case therefore suggests that a Nash equilibrium in a duopoly where both firms serve only one type of consumers to a large extent coincides with the case where a monopolist prefers to serve only one type of consumers.

Second, let us consider the equilibrium candidate where both firms serve both types of consumers, where the firms' tariffs are given by $(A_{TT}^{12}, p_{TT}^{12}) (> (0, c))$.

Table 5.1: *Deviation from a symmetric equilibrium where type 1 is excluded.*

θ_2	λ	λ^{**}	\tilde{p}_{TT}^{12}	\tilde{A}_{TT}^{12}	\tilde{p}_{TT}^2	\bar{n}_b	$\bar{n}_b / [N(1 - \lambda)]$	Π_{TT}^2	$\tilde{\Pi}_{TT}^{12}$
1.2	.2	.467	.618	.073	.374	33.9	.42	9.8	8.9
1.2	.4	.467	.580	.088	.365	25.1	.42	7.4	9.7
1.2	.47	.467	.568	.093	.362	22.1	.42	6.5	10.0
1.2	.8	.467	.522	.114	.351	8.2	.41	2.5	11.5
1.5	.4	.732	.714	.041	.261	24.2	.40	15.0	11.6
1.5	.7	.732	.597	.081	.214	11.7	.39	7.5	11.5
1.5	.74	.732	.583	.087	.209	10.1	.39	6.5	11.6
1.5	.9	.732	.531	.110	.189	3.8	.38	2.5	12.1
2	.4	.883	.927	.003	.156	24.4	.41	33.8	17.8
2	.7	.883	.699	.045	.037	11.5	.38	16.9	13.0
2	.8	.883	.630	.068	.003	7.5	.38	11.3	12.4
2	.9	.883	.564	.095	0	3.7	.37	5.6	12.3
3	.8	.959	.762	.028	0	7.5	.37	31.3	15.0
3	.92	.959	.603	.079	0	2.9	.36	12.5	12.7
3	.97	.959	.593	.107	0	1.1	.36	4.7	12.4

Then type 2 enjoys positive surplus, and type 1 receives his reservation utility. Again, assume that the firms have equal market shares so that they each serve $N/2$. Consider, again, a unilateral deviation by firm a , and keep the strategy for firm b fixed (Q_{TT}^{12}, A_{TT}^{12}).

In this case firm a can deviate by using one of two strategies. Firm a can aim for all type two consumers $N(1 - \lambda)$, but leave them a positive surplus, hence setting $\tilde{A}_{TT}^2 < V(c, \theta_2)$. Or, knowing that firm b has a limited capacity, firm a could act as a monopoly on any residual demand. He will then serve less than the pool of type 2 consumers $N(1 - \lambda)$ but extract all surplus $\tilde{A}_{TT}^2 = V(c, \theta_2)$.

Consider the first strategy. Firm a announces a tariff (\tilde{A}_{TT}^2, c) that is strictly preferred by type 2 consumers. It will extract as much as possible from type 2 consumers via the fixed fee and will maximize

$$\tilde{\Pi}_{TT}^2 | \Pi_{TT}^{12} = N(1 - \lambda) \tilde{A}_{TT}^2 \quad (5.23)$$

subject to

$$V(c, \theta_2) - \tilde{A}_{TT}^2 \geq V(\tilde{p}_{TT}^{12}, \theta_2) - A_{TT}^{12}, \quad (5.24)$$

$$\frac{N}{2} (\lambda q_1 + (1 - \lambda) q_2) \geq N \lambda \bar{q}_1, \quad (5.25)$$

where $q_i = q_i(p_{TT}^{12})$, or $q_i = q_i(p_{UP}^{12})$ if $A_{TT}^{12} = 0$, ($\ell = 1, 2$), and $\bar{q}_1 = q_1(\tilde{p}_{TT}^{12})$. The unit price \tilde{p}_{TT}^{12} is adjusted to account for the fact that firm b is now left with only type 1 consumers instead of a mix of type 1 and type 2. Given that type 1 consumers receive exactly their reservation utility, the unit price that clears the

market at firm b cannot exceed p_{TT}^{12} , (instead, type 1 consumers are rationed at firm b). Hence, $0 < \bar{p}_{TT}^{12} < \min\{p_{TT}^{12}, p_{UP}^{12}\}$. This restricts the fixed fee in (5.24), which in turn will restrict the profitability earned on type 2 consumers.

From (5.23) it would seem that a deviation is profitable when λ is small. However, when λ is small, \bar{p}_{TT}^{12} is low as well in order to restore market clearing at firm b . Hence, \tilde{A}_{TT}^2 is also low in this case. The more intensely firms compete, either via a low fixed fee or a low unit price, the more binding is the restriction on \tilde{A}_{TT}^2 . This suggests that in duopoly an outcome where both firms serve both types of consumers can be an equilibrium outcome in situations where a monopolist would have preferred to serve only one type of consumers. In our numerical example, the second effect always dominates the first and a deviation is never profitable. In Table 2 we have reported some numerical examples, again using $N = 100$, and $c = \frac{1}{2}$, hence $p_{TT}^{12} > c$, $A_{TT}^{12} = V(p_{TT}^{12}, 1)$.

Table 5.2: Deviation from a symmetric equilibrium $(Q_{TT}^{12}, A_{TT}^{12})$, the fixed fee in type 2's tariff is restricted.

θ_2	λ	λ^{**}	p_{TT}^{12}	A_{TT}^{12}	\tilde{p}_{TT}^2	\tilde{A}_{TT}^2	$V(c, \theta_2)$	\bar{p}_{TT}^{12}	Π_{TT}^{12}	$\tilde{\Pi}_{TT}^2$
1.2	.2	.467	.66	.058	.5	-.417	.245	0	6.9	-33.4
1.2	.4	.467	.62	.072	.5	-.023	.245	.38	6.6	-1.4
1.2	.47	.467	.61	.078	.5	.055	.245	.47	6.5	2.9
1.2	.8	.467	.54	.106	.5	.133	.245	.54	6.3	2.7
1.5	.4	.732	.8	.020	.5	-.183	.5	.38	7.1	-11.0
1.5	.7	.732	.65	.061	.5	.194	.5	.64	6.8	5.8
1.5	.74	.732	.63	.069	.5	.190	.5	.63	6.7	4.9
1.5	.9	.732	.55	.101	.5	.150	.5	.55	6.3	1.5
2	.7	.883	.8	.020	.5	.154	1.125	.64	7.1	4.6
2	.8	.883	.7	.045	.5	.309	1.125	.69	7.3	6.2
2	.9	.883	.60	.080	.5	.225	1.125	.60	6.5	2.3
3	.8	.959	.9	.005	.5	.301	3.125	.69	9.0	6.0
3	.92	.959	.66	.058	.5	.445	3.125	.66	6.9	3.6
3	.97	.959	.56	.097	.5	.245	3.125	.56	6.3	0.7

The other possible deviation strategy in this situation was for firm a to act as a monopoly on any residual demand from type 2. This time, consider a deviation where firm a announces a tariff that extracts all surplus from type 2, $(V(c, \theta_2), c)$. Type 2 enjoys positive surplus by switching to firm b 's tariff. Hence, type 2 consumers will crowd out type 1 consumers at firm b since capacity at firm 1 is insufficient to meet all demand. Firm a earns monopoly profit on each type 2 consumer it serves and aggregate profit is given by

$$\tilde{\Pi}_{TT}^2 | \Pi_{TT}^{12} = [N(1 - \lambda) - \bar{n}_b] V(c, \theta_2), \quad (5.26)$$

where \bar{n}_b is the number of type 2 consumers that can be served by firm b . Type 2

is indifferent between the two firms' tariffs when he receives zero surplus. Hence, the unit price in firm b 's tariff must be adjusted in order to restore individual rationality for type 2, \bar{p}_{TT}^2

$$V(\bar{p}_{TT}^2, \theta_2) - A_{TT}^{12} \geq 0, \quad (5.27)$$

$$\frac{N}{2} (\lambda q_1 + (1 - \lambda) q_2) \geq \bar{n}_b \bar{q}_2, \quad (5.28)$$

$$N(1 - \lambda) \geq \bar{n}_b. \quad (5.29)$$

This time, firm b is left with type 2 consumers only, instead of with a mix of type 1 and type 2. Again, we would have thought it is profitable to deviate when λ is small. But now, when λ is small, the fixed fee A_{TT}^{12} is low. And therefore, type 2 consumers will gain considerably if they switch to firm b . Hence, the unit price \bar{p}_{TT}^2 is high and demand from type 2 is restricted. This means that \bar{q}_{TT}^2 is low and that \bar{n}_b is large in order to restore market clearing.

In Table 3 we report some numerical examples, still using $N = 100$ and $c = \frac{1}{2}$. As shown, we find no examples where such a deviation is profitable. Again, the fact that the non-deviating firm has committed itself to sell a certain quantity acts as a constraint on the deviating firm's behavior. If there are few type 2 consumers, the non-deviating firm would serve them all and the deviating firm would have no residual demand. If there are many type 2 consumers, the price per unit would be close to marginal costs. If so, there is a limited scope for the deviating firm to generate additional consumer surplus from type 2 by setting price per unit equal to marginal costs.

Table 5.3: *Deviation from a symmetric equilibrium $(Q_{TT}^{12}, A_{TT}^{12})$, acting as a monopoly on the residual demand from type 2:*

θ_2	λ	λ^{**}	p_{TT}^{12}	A_{TT}^{12}	\bar{p}_{TT}^2	\hat{A}_{TT}^2	\bar{n}_b	$N(1-\lambda)$	Π_{TT}^{12}	$\hat{\Pi}_{TT}^2$
1.005	.02	.196	.505	.123	.510	.128	50	98	6.25	6.1
1.05	.17	.168	.542	.105	.592	.151	55	83	6.29	4.3
1.1	.3	.292	.570	.093	.670	.180	58	70	6.37	2.1
1.2	.2	.467	.660	.058	.860	.245	74	80	6.89	1.6
1.2	.4	.467	.620	.072	.820	.245	60	60	6.61	0
1.2	.47	.467	.606	.078	.806	.245	53	53	6.53	0
1.2	.8	.467	.540	.106	.740	.245	20	20	6.29	0
1.5	.4	.732	.800	.020	1.3	.500	60	60	8.5	0
1.5	.7	.732	.650	.061	1.15	.500	30	30	6.81	0
1.5	.74	.732	.630	.069	1.13	.500	26	26	6.67	0
1.5	.9	.732	.550	.101	1.05	.500	10	10	6.31	0
2	.7	.883	.800	.020	1.80	1.125	30	30	8.5	0
2	.8	.883	.700	.045	1.70	1.125	20	20	7.25	0
2	.9	.883	.600	.080	1.60	1.125	10	10	6.5	0
3	.8	.959	.900	.005	2.90	3.125	20	20	10.25	0
3	.92	.959	.660	.058	2.66	3.125	8	8	6.89	0
3	.97	.959	.560	.097	2.56	3.125	3	3	6.34	0

5.4 Concluding remarks

Harrison and Kline (2001) have shown how we can extend the traditional Cournot model to a setting with not only a unit price, but also a fixed fee. They found that each firm sets a price per unit equal to marginal costs, and a positive fixed fee that approaches zero when the number of firms becomes large. Thus, we extend their model from one to two types of consumers. It turns out that the conclusions in Harrison and Kline (2001) are not robust to such an extension. Let us assume that both types are served. We then find that price per unit exceeds marginal costs and the fixed fee can be negative. If the firms can choose between a traditional Cournot pricing (a uniform price) and a two-part tariff, they may choose a uniform price.

We have also explored the case where the firms can choose whether to serve both types of consumers or only one type. It turns out that this case is difficult to solve analytically. We have therefore chosen to illustrate the possible equilibrium outcomes with numerical examples. The examples suggest that there might be multiple Nash equilibria. First, both firms serving only one type of consumers can be an equilibrium outcome. The numerical examples suggest that this equilibrium outcome to a large extent coincides with the cases where the monopolist chooses to serve only one type of consumers. Second, we find that both firms serving

both types of consumers can be an equilibrium outcome for a large number of parameter values. In fact, we find no examples where the firms would deviate from such an outcome. The intuition is that the rival, non-deviating firm's given quantity acts as a constraint on the deviating firm's behavior. Although this is just a numerical example, it illustrates that there are instances where a duopoly serves both types of consumers while the monopoly would prefer to serve only one type.

Appendix

Calculation of pricing and profit

In the following we derive the firms' pricing in the case when they announce identical tariffs, as given in Lemma 5.1 and Lemma 5.2. Superscript 12 is used when both types are served (superscript 2 when type 1 is excluded) and k is an argument used to describe the number of active firms.

A.1 Both consumers are served

Pricing is given by Lemma 5.1. With two active firms we have

$$A^{12}(2) = \begin{cases} \frac{1}{8}(3-2\theta)^2 & 1 \leq \theta < \frac{1}{2}(\sqrt{2}+1) \\ \frac{5}{4} - \theta & \frac{1}{2}(\sqrt{2}+1) \leq \theta < \frac{5}{4} \\ 0 & \theta \geq \frac{5}{4} \end{cases} \quad (5.30)$$

$$p^{12}(2) = \begin{cases} \theta - \frac{1}{2} & 1 \leq \theta < \frac{5}{4} \\ \frac{1}{3}(\theta+1) & \theta \geq \frac{5}{4} \end{cases} \quad (5.31)$$

$$\pi^{12}(2) = \begin{cases} \frac{1}{8} + \frac{1}{2}(\theta-1)^2 & 1 \leq \theta < \frac{1}{2}(\sqrt{2}+1) \\ \frac{1}{2}\left(\frac{3}{2} - \theta\right) & \frac{1}{2}(\sqrt{2}+1) \leq \theta < \frac{5}{4} \\ \frac{1}{18}(2\theta-1)^2 & \theta \geq \frac{5}{4} \end{cases} \quad (5.32)$$

With three active firms we have

$$A^{12}(3) = \begin{cases} \frac{1}{4}\left(\frac{7}{2} - 3\theta\right) & 1 \leq \theta < \frac{7}{6} \\ 0 & \theta \geq \frac{7}{6} \end{cases} \quad (5.33)$$

$$p^{12}(3) = \begin{cases} \theta - \frac{1}{2} & 1 \leq \theta < \frac{7}{6} \\ \frac{1}{4}\theta + \frac{3}{8} & \theta \geq \frac{7}{6} \end{cases} \quad (5.34)$$

$$\pi^{12}(3) = \begin{cases} \frac{3}{8} - \frac{1}{4}\theta & 1 \leq \theta < \frac{7}{6} \\ \frac{3}{64}(2\theta-1)^2 & \theta \geq \frac{7}{6} \end{cases} \quad (5.35)$$

With more than three firms we have

$$A^{12}(k) = \begin{cases} \frac{1-2k(\theta-1)}{4(k-1)} & 3 < k < \frac{1}{2(\theta-1)} \\ 0 & k \geq \frac{1}{2(\theta-1)} \end{cases} \quad (5.36)$$

$$p^{12}(k) = \begin{cases} \theta - \frac{1}{2} & 3 < k < \frac{1}{2(\theta-1)} \\ \frac{2\theta+k}{2(k+1)} & k \geq \frac{1}{2(\theta-1)} \end{cases} \quad (5.37)$$

$$\pi^{12}(k) = \begin{cases} \frac{3-2\theta}{4(k-1)} & 3 < k < \frac{1}{2(\theta-1)} \\ k \frac{(2\theta-1)^2}{4(k+1)^2} & k \geq \frac{1}{2(\theta-1)} \end{cases} \quad (5.38)$$

In Proposition 5.1 the critical value λ^* solves the inequality $\frac{5}{4} - \lambda - (1-\lambda)\theta_2 \leq 0$ from (5.30). k^* solves the inequality $\frac{1-2k(\lambda+(1-\lambda)\theta_2-1)}{4(k-1)} \leq 0$ from (5.36).

A.2 Only type 2 is served

Pricing is given by Lemma 5.2. The unit price is always equal to marginal price, $p^2(2) = p^2(3) = c$, and the firms' profit per consumer is whatever they manage to capture via the fixed fee $A^2(k)$. With less than 3 active firms we have

$$A^2(2) = \pi^2(2) = A^2(3) = \pi^2(3) = \frac{1}{8}(2\theta_2 - 1)^2 \quad (5.39)$$

With more than 3 firms we have

$$A^2(k) = \pi^2(k) = \frac{(2\theta_2 - 1)^2}{4(k-1)} \quad (5.40)$$

A.3 Uniform Cournot price

When both types are served in a k -firm oligopoly and all firms charge a uniform price, we have

$$p_{UP}^{12}(k) = \frac{2\theta + k}{2(k+1)} \quad (5.41)$$

and

$$\pi_{UP}^{12}(k) = k \frac{(2\theta - 1)^2}{4(k+1)^2} \quad (5.42)$$

Proposition 5.2 can be verified by comparing the firms' profit in the two relevant cases. When A_{TT}^{12} is negative $\pi^{12}(k)$ (from (5.37)) is equal to or greater than $\pi_{UP}^{12}(k)$ (from (5.42)).

The monopolist's cut-off rate λ^{**} solves the equality $\pi^{12}(2) = \pi^2(2)$ in (5.32) and (5.39) respectively, given that the duopoly extracts all surplus from type 1 when both types are served.

Bibliography

- Armstrong, M. (1996). "Multiproduct nonlinear pricing." *Econometrica*, **64**(1), 51–75.
- Armstrong, M. and J.-C. Rochet (1999). "Multi-dimensional screening: A user's guide." *European Economic Review*, **43**(4-6), 959–979.
- Armstrong, M. and J. Vickers (1999). "Competitive price discrimination." Mimeo, Oxford.
- Baron, D. P. and R. B. Myerson (1982). "Regulating a monopolist with unknown costs." *Econometrica*, **50**(4), 911–930.
- Biglaiser, G. and C. Mezzetti (1993). "Principals competing for an agent in the presence of adverse selection and moral hazard." *Journal of Economic Theory*, **61**, 302–330.
- Caillaud, B., R. Guesnerie, P. Rey and J. Tirole (1988). "Government intervention in production and incentives theory: A review of recent contributions." *RAND Journal of Economics*, **19**(1), 1–26.
- Calem, P. S. and D. F. Spulber (1984). "Multiproduct two part tariffs." *International Journal of Industrial Organization*, **2**, 105–115.
- Carlton, D. W. and J. M. Perloff (1999). *Modern Industrial Organization*. Addison-Wesley Longman, Reading, Mass.
- Castelli, F. and C. Leporelli (1993). "Critical mass of users versus mass of services in a multiproduct information service system." *Information Economics and Policy*, **5**, 331–355.
- Deneckere, R. J. and R. P. McAfee (1996). "Damaged goods." *Journal of Economics and Management Strategy*, **5**(2), 149–174.
- Dupuit, J. (1849). "On the tolls and transport charges." Translated in *International Economic Paper* (London, Macmillan 1952). Original version in *Annales des Ponts et Chaussées* 17.

- Economides, N. and S. Wildman (1995). "Monopolistic competition with two-part tariffs." Working Paper EC-95-10, New York University, Stern School of Business.
- Ekelund, R. B. (1970). "Price discrimination and product differentiation in economic theory: An early analysis." *Quarterly Journal of Economics*, **84**, 268–278.
- Foros, Ø., S. Jensen and J. Y. Sand (1999). "Damaging network subscription." *Communications and Strategies*, **33**, 37–58.
- Fudenberg, D. and J. Tirole (1991). *Game Theory*. MIT Press, Cambridge.
- Gasmi, F., M. Moreaux and W. W. Sharkey (2000). "Strategic nonlinear pricing." *Journal of Economics*, **71**(2), 109–131.
- Goldman, M. B., H. E. Leland and D. S. Sibley (1984). "Optimal nonuniform pricing." *Review of Economic Studies*, **51**, 304–319.
- Harrison, M. and J. J. Kline (2001). "Quantity competition with access fees." *International Journal of Industrial Organization*, **19**, 345–373.
- Hayes, B. (1987). "Competition and two-part tariffs." *Journal of Business*, **60**(1), 41–54.
- Henriet, D., C. Henry, P. Rey and J. Rochet (1988). "Intérêt public, intérêt privé et discrimination." In G. Dionne, ed., *Incertain et Information, Collection, Société Canadienne de Science Économique*. Economica.
- Holmström, B. (1979). "Moral hazard and observability." *Bell Journal of Economics*, **19**, 74–91.
- Ivaldi, M. and D. Martimort (1994). "Competition under nonlinear pricing." *Annales d'Économie et de Statistique*, **34**, 71–114.
- Jullien, B. (2000). "Participation constraints in adverse selection models." *Journal of Economic Theory*, **93**, 1–47.
- Klemperer, P. (1995). "Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade." *Review of Economic Studies*, **62**(4), 515–539.
- Laffont, J.-J., E. Maskin and J. Rochet (1987). "Optimal nonlinear pricing: The case of buyers with several characteristics." In T. Groves, T. Radner and S. Reiter, eds., *Information, Incentives and Economic Mechanisms: In Honour L. Hurwicz*. University of Minnesota Press.

- Laffont, J.-J., P. Rey and J. Tirole (1998). "Network competition: II. Price discrimination." *RAND Journal of Economics*, **29**(1), 38–56.
- Laffont, J.-J. and J. Tirole (1986). "Using cost observations to regulate firms." *Journal of Political Economy*, **94**(3), 614–641.
- Laffont, J.-J. and J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation*. MIT Press, Cambridge.
- Lal, R. and C. Matutes (1994). "Retail pricing and advertising strategies." *Journal of Business*, **67**, 345–370.
- Lewis, T. and D. Sappington (1989). "Countervailing incentives in agency problems." *Journal of Economic Theory*, **49**, 294–313.
- Littlechild, S. C. (1975). "Two-part tariffs and consumption externalities." *The Bell Journal of Economics*, **6**, 661–670.
- Maggi, G. and A. Rodriguez-Clare (1995). "On countervailing incentives." *Journal of Economic Theory*, **66**, 238–263.
- Mandy, D. M. (1992). "Nonuniform Bertrand competition." *Econometrica*, **60**(6), 1293–1330.
- Martimort, D. (1992). "Multi-principaux avec selection adverse." *Annales d'Economie et de Statistique*, **28**, 1–38.
- Martimort, D. (1996). "Exclusive dealing, common agency, and multi-principals incentive theory." *RAND Journal of Economics*, **27**(1), 1–31.
- Maskin, E. and J. Riley (1984). "Monopoly with incomplete information." *RAND Journal of Economics*, **15**(2), 171–196.
- Matthews, S. and J. Moore (1987). "Monopoly provision of quality and warranties: An exploration in the theory of multidimensional screening." *Econometrica*, **55**(2), 441–467.
- Mezzetti, C. (1997). "Common agency with horizontally differentiated principals." *RAND Journal of Economics*, **28**(2), 323–345.
- Michell, B. and I. Vogelsang (1991). *Telecommunications Pricing. Theory and Practice*. Cambridge University Press, Cambridge.
- Miravete, E. J. (2000). "Quantity discounts for time-varying consumers." CEPR Discussion Paper No. 2699, London.
- Miravete, E. J. (2001). "Screening through bundling." CARESS Discussion Paper No. 01-01.

- Mussa, M. and S. Rosen (1978). "Monopoly and product quality." *Journal of Economic Theory*, **18**, 301–317.
- Ng, Y. and W. Weisser (1974). "Optimal pricing with a budget constraint – The case of the two part tariff." *Review of Economic Studies*, **41**, 337–345.
- Oi, W. Y. (1971). "A Disneyland dilemma: Two-part tariffs for a Mickey Mouse monopoly." *Quarterly Journal of Economics*, **85**, 77–96.
- Olsen, T. E. and P. Osmundsen (1998). "Common agency with outside options: The case of international taxation of an mne." Mimeo, University of Bergen.
- Oren, S., S. Smith and R. Wilson (1983). "Competitive nonlinear tariffs." *Journal of Economic Theory*, **29**, 49–71.
- Phlips, L. (1983). *The Economics of Price Discrimination*. Cambridge University Press, Cambridge.
- Picard, P. (1987). "On the design of incentive schemes under moral hazard and adverse selection." *Journal of Public Economics*, **33**, 305–331.
- Pigou, A. (1920). *The Economics of Welfare*. Macmillan, London.
- Rochet, J.-C. and P. Choné (1998). "Ironing, sweeping, and multidimensional screening." *Econometrica*, **66**(4), 783–826.
- Rochet, J.-C. and L. A. Stole (1999). "Nonlinear pricing with random participation constraints." Mimeo, Chicago.
- Rochet, J.-C. and L. A. Stole (2000). "The economics of multidimensional screening." Presented at the World Congress of the Econometric Society, August 15, 2000, Seattle, Washington.
- Sappington, D. (1983). "Optimal regulation of a multiproduct monopoly with unknown technological capabilities." *Bell Journal of Economics*, **14**, 453–463.
- Schmalensee, R. (1981). "Monopolistic two-part pricing arrangements." *Bell journal of Economics*, **12**(2), 445–466.
- Scotchmer, S. (1985a). "Profit-maximizing clubs." *Journal of Public Economics*, **27**, 25–45.
- Scotchmer, S. (1985b). "Two tier pricing of shared facilities in a free entry equilibrium." *RAND Journal of Economics*, **16**(4), 456–486.
- Seierstad, A. and K. Sydsæter (1977). "Sufficient conditions in optimal control theory." *International Economic Review*, **18**(2), 367–391.

- Seierstad, A. and K. Sydsæter (1987). *Optimal Control Theory with Economic Applications*. North Holland, New York.
- Sharkey, W. W. and D. S. Sibley (1993). "Optimal nonlinear pricing with regulatory preference over customer type." *Journal of Public Economics*, **50**, 197–229.
- Shmanske, S. (1991). "Two-part pricing in differentiated duopoly." *Journal of Institutional and Theoretical Economics*, **147**, 441–458.
- Sibley, D. S. and P. Srinagesh (1997). "Multiproduct nonlinear pricing with multiple taste characteristics." *RAND Journal of Economics*, **28**(4), 684–707.
- Stole, L. A. (1992). "Mechanism design under common agency." Working Paper, University of Chicago.
- Stole, L. A. (1995). "Nonlinear pricing and oligopoly." *Journal of Economics and Management Strategy*, **4**(4), 529–562.
- Tirole, J. (1988). *The Theory of Industrial Organization*. MIT Press, Cambridge.
- Valletti, T. M. (1999). "Location choice and price discrimination in a duopoly." Discussion Paper No. 2322. CEPR.
- Varian, H. R. (1989). "Price discrimination." In R. Schmalensee and R. Willig, eds., *Handbook of Industrial Organization, volume I*. North-Holland, Amsterdam.
- Varian, H. R. (2000). "Versioning information goods." In B. Kahin and H. R. Varian, eds., *Internet publishing and beyond. The economics of digital information and intellectual property*. MIT Press, Cambridge.
- Wang, R. and Q. Wen (1998). "Strategic invasion in markets with switching costs." *Journal of Economics and Management Strategy*, **7**(4), 521–549.
- Wilson, R. (1993). *Nonlinear Pricing*. Oxford University Press, Oxford.
- Young, A. R. (1991). "Transaction costs, two-part tariffs, and collusion." *Economic Inquiry*, **29**, 581–590.