COMPETITION, ACCESS PRICING AND REGULATION IN A SECOND DEGREE PRICE DISCRIMINATION SETTING

Maria Vagliasindi

Department of Economics

University of Warwick

This thesis is submitted at the Department of Economics University of Warwick for the degree of PhD in December 1995.

I certify that this thesis is my own original work, except where otherwise stated, and that is within the upper limit on length.

Acknowledgements

My principal debt is with Michael Waterson, my supervisor, who gave me invaluable help, encouragement and criticism throughout the period of work on this thesis. I have also received helpful comments from Jonathan Cave, Morten Hviid, Norman Ireland and John Vickers. Separate acknowledgements, relative to each chapter, follow.

Abstract

In broad terms, this work aims to gain a greater understanding of the particular features introduced in the regulatory set-up by competitive issues and vertically related markets. Specifically, we explore their impact on the profitability of the market and the possibility for the incumbent to maintain monopoly profits under different regulatory regimes. There was a time when utilities industries and in particular telecoms each seemed to be a natural monopoly. Most governments liked it that way because they owned the monopoly and siphoned off some of the profits. Nowadays, competition is spreading in most utilities market and it becomes imperative to assess its impact on the tariffs and in general on social welfare.

We deal with a second degree price discrimination model allowing the players -namely, an incumbent, who has a natural monopoly on the network, and a rival- to make use of non-linear pricing in intermediate and final goods. In this framework the entrant's choice of the customer types is endogenised in a sequential multistage game, where the incumbent, who is undoubtedly the most powerful player, acts as a first mover. We also show that cream skimming, contrary to the general wisdom, can be welfare enhancing. Particular attention is devoted to the access pricing problem which is becoming the key issue to the regulators, examining the relevance of simple pricing rules, such as the Baumol-Willig rule. Despite the presence of a growing literature in these areas, other models fail to incorporate the use of non-linear access pricing. Since price discrimination is common in practice this omission can lead to misleading results. Our analysis shows that the regulator should not allow competition for the low-demand consumers' types or by a less efficient entrant and should impose the adoption of socially optimal non-linear access tariffs. Therefore the general conclusion is that competition will not obviate the need of regulation.

CONTENTS

Chapter 0

INTRODUCTION

In countries, such as the UK, where previously held assumptions about the extent of network industries are being challenged, experiments have been made in introducing a degree of vertical separation. Particularly in telecommunication industries Oftel has been recently reviewing access pricing and, more generally, the whole terms of interconnection (see, for instance, Oftel 1993 and 1995). Also the European Commission has tried to open up the utilities market, through the intervention of industry related Directorates (e.g. the ones for Telecommunications, Energy and Transport) and the Competition Directorate. For example, think about the actual situation where Mercury (and others) compete with British Telecom to supply telephone calls, but exchange and local loop facilities are, in general, still owned by British Telecom. By introducing competition at one vertical level alone it is hoped to engineer a reduction in the need of regulation, allowing competition to do some of the work. However, can competition (at a vertical level alone) be really considered as a substitute to regulation? This is one of the central issues we address in this thesis. Although there is an extensive literature on topics such as access pricing regulation and vertical issues, many such interesting questions remain still to be analysed.

The mechanisms derived by these studies establish regulatory policies for a firm already awarded a franchise monopoly. This is the basic optimal regulation approach, as notably developed by Laffont and Tirole (1993), and it represents our starting point. To simplify matters, we ignore asymmetries of information between the regulator and the regulated firm, taking advantage of the similarities with the case of a public firm, which directly maximises social welfare. In fact, in the standard literature the amount of information regulators have is too much compared to the real situation, and the moral hazard issue is usually drastically simplified or entirely ignored. This perspective is not restrictive since regulation may be also considered as a two stage process, involving first the selection of a franchise monopolist from a set of potential suppliers and then the implementation of the regulatory contract, when the production stage takes place. Moreover, we introduce competition in the second stage of our game.

Specifically, we model the interactions between regulation and competition in a second degree price discrimination setting, where non-linear pricing is allowed for both in intermediate and final goods' markets. *Non-linear tariffs* are a means whose aim is to discriminate amongst consumers with different tastes when the regulator and the firm cannot directly observe consumers' tastes. This device ameliorates, at least to some extent, the problem caused by the asymmetry of information. In the canonical price discrimination model a monopolist (the principal) has incomplete information about the agent's types, namely the consumers' willingness to pay for his goods. He has to design a tariff schedule that determines the price to be paid as a function of the quantity purchased. We extend the canonical framework to the case of competition and to vertically related markets.

Our interest in these extensions derives from the consideration of some crucial policy issues. An issue that arose in November 1990, when the main duopoly review began in the UK, was how to open up the telecommunication industries to competition. In this case attention focuses on the terms on which rivals -Mercury and others- should gain *access* to British Telecom's local networks and the wider issue of British Telecom's *vertical structure*.

There are several relevant policy options that need to be analysed in a vertical setting, the most important of which are related to vertical *structure* and *conduct*. Regarding the first issue a network industry can be vertically integrated or separated (a policy of vertical separation involves divestiture if the incumbent starts off as a vertically integrated firm). Therefore, we also consider vertically separated structures in which the network is owned by an upstream monopolist, examining the circumstances under which the upstream monopolist (if allowed to price discriminate) is able to extract all the downstream industry's profits.

Our general aim is to gain a greater understanding of the particular features introduced in the regulatory (and procurement) set-up by vertical issues (more specifically different linkages between the upstream and downstream sector) and their impact on the profitability of the market and the possibility for the incumbent to maintain monopoly profits. In what follows the structure of the thesis is sketched.

Chapter 1 provides an overview of the state of the art of the economic literature on regulation and related issues in procurement. Specifically, we consider in detail the relevance of agency problems in the basic optimal regulatory setting. We then sketch the ways in which competition and regulation interact within different scenarios. Namely, regarding regulation we take into account regulatory constraints on final goods, such as price caps, and on intermediate goods. Finally, competition within the market is distinguished from competition for natural monopoly.

In chapter 2 we introduce our approach to model non-linear pricing and competition. Following a *positive* approach we first examine the *private incentives* of the economic players (basically an incumbent and a potential entrant) in the absence of regulatory constraints in horizontal and vertical settings (where the rival enters at the downstream level). Following the incumbent's point of view we analyse the conditions under which the incumbent can maintain monopoly profits while entry occurs at one vertical level. In doing this we also analyse whether it is in his interest to oblige the (eventual) competitor by the use of an appropriate access charge to be efficient. The case of cream skimming competition by a more efficient rival is analysed in detail, as well as its "desirability" from a productive and allocative point of view. Specifically, cream skimming competition is endogenised and it is shown to be welfare enhancing under a wide range of circumstances, in contrast with the normal wisdom. In fact, cream skimming is usually thought of as an inefficient form of competition, or it is assumed to be one of the major causes of the unsustainability of natural monopoly.

In Chapter 3 socially optimal regulation is introduced, so that the analysis is more normative. Information and incentive issues complicate the intervention of regulators. However, the amount of private information on the regulatory side built into the standard models is too limited compared with problems actual regulators face. Our approach is to leave aside problems of moral hazard and adverse selection on the regulatory side to make the problem more manageable and to reveal its fundamental structure. These assumptions, which lead us to analyse the full information benchmark, are quite usual in the case of a public firm (which directly maximises social welfare) as well as in the large part of the debate on access pricing. Despite the extensive literature, models usually fail to incorporate the use of non-linear pricing. Since price discrimination is common in practice, this omission can lead to misleading results. Welfare considerations are drawn for different specifications of the social welfare function, incorporating also distributional considerations and distortionary taxation. We also discuss the optimality of simple access pricing rules, such as the Baumol-Willig rule in the original setting and in our vertical game. Finally, we look further at the design of other welfare enhancing mechanism implicit in a vertical merger.

In chapter 4 we introduce vertically separated markets. For a long time regulatory reforms have gone in the direction of divestitures and deregulated structures. Yet the theoretical and applied literature on this intriguing area is very little. We first expand the model set out in Chapter 2 to see under which circumstances we can get the same outcomes as under vertical integration where the regulatory stage involves divestiture. We then model vertical separation in an alternative framework that portrays access price discrimination by an upstream monopoly towards downstream producers of different types (that is, level of efficiency). We complicate the model further by introducing second degree price discrimination also in the final demand market. As should be expected, there are no clear cut results in terms of the optimal pricing strategy to be applied in the intermediate and final demand markets. We discover a variety of different combinations of cases that can arise depending on the specification of the functional form of downstream producers' costs and final customers' demand. The general lesson to be drawn is that distortion can arise not only in the tariff schedule offered to final customers, but also in the customers' allocations between the downstream producers. Nevertheless, a modified monopoly result applies also in this framework, with quite relevant exceptions. Finally, we conclude by summarising the main results achieved in this thesis, pinpointing directions for future research.

Chapter 1

AN OVERVIEW ON THE NEW ECONOMICS OF REGULATION AND PROCUREMENT*

* I benefit from my previous work with John Vickers at the University of Oxford. In particular, sections 1.3 and 1.6 built upon my M.Phil thesis (Vagliasindi, 1994). Section 1.2 has been developed and partly revised in joint work with Michael Waterson (Vagliasindi and Waterson, 1995a).

Chapter 1

AN OVERVIEW OF THE NEW ECONOMICS OF REGULATION AND PROCUREMENT

1.1 Introduction

The aim of this chapter is to provide an overview of new developments in the field of regulation and related issues in procurement. The topics we are going to focus on are mainly concerned with competitive issues that have been subject of several recent controversies in the UK and in the USA related to the policies to apply in utility industries. Referring in particular to telecommunication industries, in the UK Oftel has been recently reviewing access pricing and, more generally, the whole terms of interconnection and experiments have been made in introducing a degree of vertical separation. Also the European Commission has being involved in dealing with such issues through the intervention of industry related Directorates and the Competition Directorate.

From a regulatory point of view we can rely on the results of an extensive literature on the standard regulatory set-up and more specifically on access pricing. An incredible number of variants of the standard analysis, early formulated in the 80s by Loeb and Magat (1979), Baron and Myerson (1982) and Laffont and Tirole (1986), has been proposed by the "new economics of regulation and procurement". In particular, several extensions of the basic approach introduced by Laffont and Tirole (1986) have been collected in a very impressive volume.[1] An alternative, more *positive*, approach to regulation has been developed in the UK by Littlechild (1983), Bradley and Price (1988), Vickers and Yarrow (1988) and Waterson (1988, 1992, 1994). New developments on access pricing -starting from the contributions of Willig (1979) and Baumol (1983)- introduced the so called Baumol-Willig rule,

---

[1] Laffont and Tirole's book (1993) develops a synthetic approach to regulation and procurement and provides a very detailed analysis of the most recent advances in this field. A particular emphasis, however, is put on regulation of natural monopolies. More attention to topics such as product market competition is devoted by Armstrong, Cowan and Vickers (1994).

recently reconsidered in potential applications (for instance in the case of New Zealand) by Baumol and Sidak (1994), Cave (1994), Armstrong and Doyle (1994), Armstrong and Vickers (1995) and Economides and White (1995).

Previous approaches to regulation -which ignored agency problems due to the presence of moral hazard and adverse selection- were partially unsatisfactory, since they provided only a very stylised description of actual regulatory processes (considered as exogenous mechanisms) and they didn't pay enough attention to the opportunity for strategic behaviour on the part of both the regulator and the regulated firm. Specifically, the early regulatory literature focused on how traditional regulatory constraints induce profit maximising enterprises to misallocate resources, contrary to the central regulatory goal of replicating competitive conditions. The Averch and Johnson (1962) model, for instance, investigates the behaviour of a profit maximising monopolist operating under a rate of return on investment constraint (this system being adopted in the USA). The UK system of regulation, known as price cap, is based upon the notion that transfers are not feasible, so that the regulator must rely on *prices* to perform his objectives.

In reality both price caps and rate of return regulation may be seen as two polar cases of incentive contracts. Specifically, let us consider the basic trade off between providing incentives for efficiency and minimising the rent enjoyed by the regulated firm, because of asymmetric information. In this context price caps can be seen as high powered contracts (since they provide the best incentives for cost reductions), whilst rate of return constraints represent low powered schemes (but they are ideal for rent extraction). Having noted this, the following questions arise: a) do these regulatory mechanisms converge toward socially optimal pricing with a limited delay and social cost?; b) are the regulator's *positive* objectives and constraints different from the usual *normative* ones (e.g. optimal pricing) and has the regulator the power to implement them against the interest of the regulated firm?

Let us consider such questions in a static environment, where capacity is given and cost and demand conditions are stable. Basically, what the literature finds out is that rate of return regulation is likely to encourage inefficient productive choices and waste. Price regulation, instead, may lead to efficiency in the short run and eventually converge to socially optimal pricing, as shown by Vogelsang and Finsinger (1979). But the adjustment process may take many periods; in fact, the firm may have a weak incentive to reduce costs on the way to the optimum.[2] Another criticism moved by Sappington (1980) is the possibility for the firm to engage in strategic behaviour, manipulating the information on the cost and being intentionally inefficient.

Real word is much more complex and static models do not consider the evolution of the environment and the relation between the regulator and the firm. This leads us to modify question b) in order to deal with this type of problem. One specific problem arises with respect to the achievement of an optimal capacity path and socially efficient productive choices when the regulator can engage in strategic behaviour with respect to sunk investment. In a dynamic framework the previous conclusion may be reverted. While the cost plus characteristic of rate of return regulation makes its performance poor with respect to short run efficiency it makes it desirable in order to prevent opportunistic behaviour and to support an optimal investment path, as recently argued by Gilbert and Newbery (1994).

Notice how the mechanisms derived following a regulatory perspective establish regulatory policies for a firm already awarded a monopoly franchise. This is the kind of modelling approach we are going to develop in our thesis, using Laffont and Tirole's analysis as an interesting and relevant starting point.[3] In fact,

---

[2] Using more demanding mechanisms, such as Sappington and Sibley (1988), the convergence to Ramsey pricing occurs in a period.

[3] However, the regulatory process may also be considered as a two stage process, involving first the selection of a franchise monopolist from a set of potential suppliers and then the implementation of a regulatory policy, which occurs once the information on the firms' performance is known and the production stage takes place. This is the procurement approach. We believe that the auction theory developed for procurement can be extended to the regulation framework, so that the same modelling structure can be used both for procurement and regulation, following the lines of Laffont and Tirole (1993).

the normative framework developed by Laffont and Tirole (1986, 1990a, 1990b, 1993, 1994a, 1994b) has provided an important new paradigm for the analysis of optimal regulation under asymmetric information. Therefore, it becomes imperative to be clear about exactly what their framework does and does not do.

Focusing on asymmetric information issues the standard distinction is between the presence of *moral hazard* and *adverse selection* (some authors prefer to speak in terms of hidden action and hidden information).[4] In this chapter we first discuss the modelling implications of introducing adverse selection and moral hazard in the standard optimal regulation set-up, showing how, in practice, Laffont and Tirole's (1986) original approach does not involve any serious moral hazard issue. In reality, all Laffont and Tirole's regulatory analysis can be treated within a pure adverse selection model, with no loss of insight. This is true for their basic model, as for many of the extensions contained in Laffont and Tirole (1993) and (1994a). In that sense it is not a particular advance on Baron and Myerson (1982) that deals with the case of cost unobservability.[5] Specifically, the two formulations lead to similar qualitative results, despite the presence of several differences, even in the definition of the regulator's objective. For clarity of exposition, in standard first best analysis [Loeb and Magat (1979)] the regulator's objective, i.e. the social welfare function, is defined as the unweighted sum of consumers' and producer's net surplus, abstracting from *distributional considerations* and costs associated with monetary transfers. Baron and Myerson (1982) introduce the first extension, in a simple but "ad hoc" way, by putting more weight on consumers' surplus, whereas Laffont and Tirole (1986) follow the latter option, introducing the so called *cost of public funds*, a parameter that affects both terms of the welfare function and can be justified in the

---

[4] The firm is typically better informed then the regulator about the cost and demand conditions in the industry. This gives rise to problems of hidden information (or adverse selection). The problem of hidden action (or moral hazard) in the regulatory context depends on the unobservability of the firm's cost reducing effort and the consequent risk of managerial slackness.

[5] The techniques employed by these model are the ones developed by Mirrlees (1971).

ambit of a general equilibrium model. Furthermore, they also introduce in the managerial utility function (which has the same weight as profits and consumers' surplus) a moral hazard or effort parameter (distinguished from the adverse selection, which has a merely technical nature). However, the introduction of this new problem does not cause any trouble, as both the regulator and the firm's manager are risk neutral.

We then focus on *cream skimming* and optimal *bypass*, discussing the structure of the approach chosen by Laffont and Tirole (1990b) to model regulation and competition in a second degree price discrimination setting. Differently from Laffont and Tirole, who concentrate their analysis on socially optimal regulation under incomplete information, we present a simplified cream skimming game adopting their approach to competition, but leaving aside the *adverse selection* problem (between the regulator and the firm) and also regulation. In this setting we will derive all the main economic results, showing that in reality what matters is the consumer surplus offered to high demand customers (which corresponds to the relative *competitors' efficiency*) is exogenous. In practice, we have different regimes simply because competitors have a perfectly elastic offer curve, which modifies high demand consumers'partecipation constraint and not because of the presence of asymmetric information or the optimal regulatory structure imposed on the model.

A more detailed analysis shows that a crucial feature of Laffont and Tirole's model lies in the terms of the connection between competitors and high demand consumers. Specifically, as will be shown in chapter 2 (section 2.4.3), we can endogenise *cream skimming* competition by combining Laffont and Tirole (1990a, 1994a) and (1990b), with no recourse to any ad hoc assumptions. In fact, in a vertical framework with non-linear access pricing cream skimming becomes naturally the only type of competition allowed by the incumbent. By considering the situation where the incumbent is a monopolist and the entrant comes in at only one vertical level (a common event in network industries) we provide a rationale for entry concentrating on high-demand consumers.

Having clarified Laffont and Tirole's paradigm, we then mainly focus on the interactions between competition *and* regulation, abstracting from many interesting variants and extensions of the regulatory analysis. At first sight regulation may be regarded as the antithesis of competition; for instance one may argue that regulation simply diverts resources, which have more important alternative uses and produces undesirable side effects (such as the phenomenon of regulatory capture).[6] This may be true in a number of circumstances; however, there are several cases in which regulation has a positive role in preserving and extending the competitive process. In particular, regulation proved to be necessary in order to promote competition, as the market left to itself would not achieve this objective. The role of regulation becomes even more crucial in determining the terms of the interconnections between the regulated firm and competitors.

The modelling set-up in which we want to analyse the interactions between regulation and competition in the next chapters is a second degree price discrimination model -where non-linear pricing is allowed- and markets are vertically linked. *Non-linear pricing* allows to discriminate amongst consumers under asymmetric information on consumers' tastes. This self selecting device ameliorates to some extent the problem caused by the presence of asymmetric information. A similar strategy can be also used when there is asymmetric information between the regulator and the firm about industry conditions (i.e. the firm's cost or effort levels). A regulatory scheme should discriminate between firms with different level of efficiency. However, because firms with different costs must be offered the same contract, the firm's monopoly of information entails that the most efficient firms will enjoy rents. Therefore, if the profit of the regulated firm has a lower weight than the consumers' surplus (adopting the social welfare function of Baron and Myerson) or we introduce the cost of public funds there will be a trade-off between allocative efficiency and minimising informational rents, which reduces

[6] For a discussion on the regulatory capture phenomenon see Laffont and Tirole (1991).

overall welfare. As a result of this, it is optimal to set price above marginal cost.

A particular interest in *vertically related markets* derives from the consideration of some crucial policy issues. A relevant issue that arose at the beginning of the Duopoly Review in 1990 in the UK telecommunication industries was how to open up the industry to competition. The existence of economies of scale provided a possible case for limiting competition, whereas some competition (at least between networks) is likely to be beneficial in order to induce a greater efficiency level. In this case attention should focus on the terms on which rivals - Mercury (MCL) and others- should gain *access* to British Telecom's local networks and the wider issue of British Telecom's *vertical structure*. Another important question to analyse is whether this trade-off should be solved by the market forces or by regulation (for instance, by licensing just one or few additional competitors). Hence, when regulated firms operate in a competitive environment a fundamental issue is the determination of the access pricing, i.e. the pricing of an intermediate good that allows rivals to survive with the regulated firm in the final output market. This is not a merely theoretical question, as in many utility industries the feasibility of access constitutes a "conditio sine qua non" entry is blockaded.

From time to time the regulatory authorities have considered questions about *structure* (for instance, mergers and vertical integration, and the possibility of divestiture) and *conduct* (i.e. possible predatory behaviour). Therefore, the effectiveness and the efficiency of competition will be greatly influenced by regulation, and in particular by the terms of interconnection to BT's network. In particular, the Office of Telecommunication (OFTEL) commenced operations in mid 1984, a few months before the privatisation of BT. In the "Duopoly Review" it decided that a particular potential existed for competition at a local level, the benefits of competition being considered to outweigh the losses of economies of scale. At the start of competition BT was contributing to some of the costs of its local network through the prices for long distance calls. MCL was allowed to interconnect with BT's network without making an equivalent contribution to BT's local costs. This advantage helped MCL to offset its disadvantage from its lack of economies of scale in pricing its long distance calls. Furthermore, BT was practically constrained to charge uniform tariffs. This meant that MCL could concentrate its effort on the densest telephone traffic, with a limited ability of BT to respond.

In Autumn 1993 Oftel published the determination of interconnection charges between BT and MCL and indicated how to deal with the Access Deficit Contribution (ADC) waivers. The ADC system aims to compensate BT for the deficit it incurs (compared with fully allocated costs) in the provision of access. However, access is meaningless by itself and its utility must be measured in relation to the expectation of its use, identifying the costs imposed by losing customers or sharing them with competitors. The main purpose of ADC is therefore to discourage inefficient entry. Major changes occurred in the market itself. Since mid 1993 competition developed in all market segments.[7] There has also been a series of price reduction by BT and the settlement of new prices for new or existing mobile services. The interesting feature of these development is that they entail competition between technologies with different cost structures. The intuition would suggest that customers will segment themselves according to the cost structures of the operators, which would also be reflected in their tariffs. For instance, a wire based technology characterised by substantial access costs and low usage costs seems to attract high users, whilst wireless technologies with low access costs but high usage charges would favour low users.[8]

Reasoning about *competition policies* requires better theoretical considerations on how different kinds of competition work in different circumstances. In particular, in order to enhance social welfare, we will pay attention to the conditions under which competition can be welfare reducing and therefore undesirable. Mankiw and

---

[7] Subscribers to telephony provided by cable television operators have developed and radio-based technologies penetrated into the local loops providing both fixed services such as Ionica and mobile services as Mercury One to One and Euro Digital.

[8] This intuition seems to be supported by the theoretical model that we introduce in the next chapter; in particular endogenising the choices of the customers' type for the competitor we get similar conclusions. See in particular section 2.4.2 and 2.4.3.

Whinston (1986) pointed out that part of the profits of entry comes not from the generation of additional consumers' surplus, but from the opportunity to steal profits from one's competitors. There is no welfare gain attached to the latter. Schwartz (1989) also pointed out a second source of inefficiency, when entry reallocates output from a low marginal cost incumbent to a high marginal cost entrant. Brennan (1991) showed how the similar reasoning applies to a special but common situation in which a regulated firm faces competition from firms offering differentiated products. To understand his reasoning consider the following example first in an unregulated setting. Suppose we have an incumbent who initially serves two types of customers, earning zero profits (as the revenues simply cover his costs). Assume now that competitive entry occurs in one of the markets driving the price down to the entrant's marginal cost. The incumbent no longer finds it profitable to stay in, causing a welfare loss in the other market. Under these circumstances, entry will reduce social welfare if the consumers' gain from the market in which entry occurred is less than the consumers' loss in the other market.

Let us now deal with multiproduct regulated markets. If the entrant's cost of serving the market exceeds the incremental cost to the regulated firm, entry is inefficient and is induced only by the price structure. This entry induced inefficiency is quite well known. However, as Brennan points out there may be welfare losses abstracting from welfare benefits associated with distribution, or favouring certain types of customers. Welfare losses in regulated markets can simply derive from the requirement that only the regulated firm's customers are expected to contribute to fixed costs. Hence, entry in regulated markets shrinks the set of markets from which cost recovery can be generated.

Competition may be used in order to reduce asymmetric information since it provides valuable information about the firm's performance. In fact, more precise information about the technology of the regulated firm can be collected through the comparison with other firms, by creating between them "yardstick competition". Even the mere threat of entry can be used as a sort of *endogenous regulatory* *mechanism*. In fact, the presence of an uncontrolled competitive fringe has a relevant influence both within the range of regulation, or outside it (leading to the shutdown of the regulated firm), as shown by Caillaud (1990).

Apart from product market competition, also competition through *auctioning* must be considered. In fact, from an allocative point of view procurement represents a major component of the public budget in many countries, so that its cost effectiveness is extremely relevant. The dominance of monopolistic practices and of arrangements with a single producer may lead to costly rents and overpricing. The role of regulation in this framework is to prevent the regulated firm from setting its price too far above cost and to encourage sufficient cost reducing activity. In a *static* context the introduction of competition can be used in order to improve such a situation and to realise savings through price reduction. In any case, as has been shown by Laffont and Tirole, among others, competition cannot perfectly substitute regulation; the best that can be done is just to auction a regulatory incentive contract able to achieve the second best level of effort.[9] Furthermore, the selection of the producer (through competition) is just the first step in the organisation of a natural monopoly. It should be followed by the aspects related to the production stage (such as the determination of the production level and of the firm's rent). In this enlarged regulatory context, apart from the further stages of the monitoring and enforcement of the contract, in which potential or yardstick competition may play a role (at least mimicking the audit process) reducing informational asymmetries, for long term activities (such as utility industries) a complete contract with commitment is quite inconceivable. This opens the problem of under-investment in specific assets, because the added value, generated by the investment, may be expropriated during the subsequent bargaining processes. Competition may be helpful also in a *dynamic* context, by providing incentives for investment or innovation. However, even if in this case important factors may limit the efficacy of competition, giving the

[9] The auction theory developed for procurement can be readily extended to the regulation framework, so that the same model can be used both for procurement and regulation.

incumbent a big advantage (as investments in R&D), any possibility to somehow weaken the monopolist's position should not be neglected.

The structure of this chapter is as follows. First of all, we discuss the modelling implications of introducing asymmetric information between the regulator and the firm in the standard optimal regulation set-up, showing how Laffont and Tirole's approach can be dealt within a pure adverse selection model with no loss of economic insight (section 1.2). In section 1.3 we present a simplified cream skimming game adopting Laffont and Tirole's (1990b) approach to competition but leaving aside the *adverse selection* problem (between the regulator and the firm) and regulation. In this setting we derive all the five different regimes, showing how in reality what matters in terms of outcomes is the consumers' surplus offered to high demand customers, which corresponds to the relative competitors' efficiency. Differently from Laffont and Tirole (1990b) we keep the marginal cost of the incumbent fixed, allowing the marginal cost of the entrant (and her efficiency) to vary. We believe this framework to be more realistic, since from a policy point of view the crucial question is whether or not to allow bypass, depending for instance on the relative efficiency of the competitor.

Section 1.4 examines the more positive approach to regulation proposed in the UK, following the price cap literature. Since the main question we address in this thesis has to do with the direct interactions between regulation and competition, *price cap regulation* will be introduced in the analysis, in order to explore its effects on competition. We then give some hints on important extensions relative to *vertically related markets* allowing competition to appear at one vertical level alone, rather than only "horizontally".

Our focus will then shift to the possible benefits deriving from the introduction of competition between firms, distinguish product market competition, that is competition *in* the market, analysed in section 1.5, from competition *for* natural monopoly, described in section 1.6. A final section (1.7) provide a sketch of the issues that we are going to tackle in the next chapters.

1.2 The relevance of agency problems in the regulatory framework

In this section we first provide a clear demonstration of the point that there is nothing of substance gained by including moral hazard in the form they introduce it within Laffont and Tirole's framework. The model may be collapsed to the pure adverse selection case. Our point in doing this is to clarify the literature on their model.[10] We do not claim that <u>any</u> introduction of moral hazard would leave results unchanged. In what follows we proceed to the proof of the absence of moral hazard in Laffont and Tirole's framework: first within the basic (1986) monopoly model with fixed output (Q=1) in the full information case, then under asymmetric information (in the two type case) and finally extending this result to a general framework with variable output.

1.2.1 The basic monopoly model in the full information case

Let us consider the basic (1986) model redefining the average costs only as a function of a pure adverse selection parameter (i.e. the technological parameter β):

$$v = c+\psi(e) = c+\psi(\beta\text{-}c)$$

where the original cost function is $C = c = \beta - e$, as in Laffont and Tirole, where c is the marginal (average cost) and e is the effort level (that is, the moral hazard parameter). Notice how v represents what Laffont and Tirole define as "the total cost of the project as perceived by the taxpayer" (1993, p. 56). Ignoring the moral hazard problem the disutility of effort $\psi(e)$ can be treated as an additional production cost. We do not even need to speak in terms of the manager's utility function, since we are only concerned with the definition of the incumbent's profit function. The latter can be expressed as the difference between transfers (tr) and costs (v):

$$\Pi = tr - v = tr - c - \psi(\beta\text{-}c)$$

---

10 See for instance, among others, Armstrong, Cowan and Vickers (1994): "this model [Laffont and Tirole (1986)] contains elements of both hidden information and hidden action and the regulator is not able to deduce whether observing a low realization of c is due to good luck ... or high effort", p. 33-4 and Baron (1989) "In the Laffont and Tirole model in which cost is observable, the regulator has the same incentive, but the regulator also faces a moral hazard problem", p. 1387.

It is then immediate to verify that the induced social welfare function, once we denote by S the consumers' gross surplus and by λ the shadow cost of public funds, is exactly the same as Laffont and Tirole's social welfare function:

$$W = S - (1+\lambda)tr + \Pi = S - (1+\lambda)v - \lambda\Pi = S - (1+\lambda)[c+\psi(\beta-c)] - \lambda\Pi$$

Therefore, in the absence of noise and with the specification of the cost function as $C = \beta\text{-}e$ Laffont and Tirole (1986)'s model is a pure adverse selection model, as the original one due to Baron and Myerson (1982). The only real difference has to do with the observability of c. Here, the newly defined cost v (inclusive of managerial disutility) is no longer observable.

1.2.2 The extension to the case of asymmetric information

Let us extend the reasoning to the presence of asymmetric information on two types of firms: an efficient ($\beta=\beta_L$, $c=c_L$) and an inefficient one ($\beta=\beta_H$, $c=c_H$). The efficient type's incentive compatibility condition can be written as:

$$\Pi_L \geq tr_H - c_H - \psi(\beta_L-c_H) = \Pi_H + \phi(\beta_H-c_H)$$

where the differential rent of the efficient type (relative to the inefficient type), i.e. $\phi(\beta_H-c_H) = \psi(\beta_H-c_H) - \psi(\beta_H-c_H-\Delta\beta)$, where $\Delta\beta = \beta_H-\beta_L$, is an increasing and convex function of the inefficient type effort (β_H-c_H). Since rents are costly, the previous constraint, as well as the inefficient type's individual rationality constraint, that is:

$$\Pi_H = tr_H - c_H - \psi(\beta_H-c_H) = 0$$

are binding, so that: $\Pi_L = \phi(\beta_H-c_H)$, $tr_H = c_H+\psi(\beta_H-c_H)$ and $tr_L = c_L+\psi(\beta_L-c_L)+\phi(\beta_H-c_H)$. Once we substitute these constraints into the expected social welfare function (where ν is the probability the firm being efficient) we get as in Laffont and Tirole's:

$$W = \nu\{S(\beta_L) - (1+\lambda)[c_L+\psi(\beta_L-c_L)] - \lambda\Pi_L\} + (1-\nu)\{S(\beta_H) - (1+\lambda)[c_H+\psi(\beta_H-c_H)] - \lambda\Pi_H\} =$$

$$\nu\{S(\beta_L) - (1+\lambda)[c_L+\psi(\beta_L-c_L)] - \lambda\phi(\beta_H-c_H)\} + (1-\nu)\{S(\beta_H) - (1+\lambda)[c_H+\psi(\beta_H-c_H)]\}$$

Therefore, the same conclusion as before can be derived: that is, in the absence of noise and with the specification of the cost function as $C = \beta\text{-}e$ Laffont and Tirole (1986)'s model is a pure adverse selection model.

1.2.3 The extension to variable output

Let us now consider the extension of the basic procurement model in a setting with variable quantity. In practice, ignoring fixed costs and remembering that in Laffont and Tirole $C = (\beta\text{-}e)Q=cQ$ we may rewrite the total average cost function as $v = c+\psi(e)/Q = c+\psi(\beta\text{-}c)/Q$ and hence the monopolist's cost function as:

$$C = vQ = cQ+\psi(\beta - c)$$

The incumbent's profits Π are simply given by the difference between transfers (tr) and costs:

$$\Pi= tr - vQ$$

In the monopoly case the welfare function is just the original one present in Laffont and Tirole who are dealing with a private regulated firm:

$$W = S(Q) - (1+\lambda)tr + \Pi$$

In this new setting with variable quantity following the previous procedure we can rewrite the binding efficient type's incentive compatibility condition as:

$$\Pi_L = tr_H - C_H - \psi(\beta_L\text{-}c_H) = \Pi_H + \phi(\beta_H\text{-}c_H)$$

But since the inefficient type's individual rationality constraint is binding $\Pi_H = tr_H - C_H - \psi(\beta_H\text{-}c_H) = 0$ the transfers from the public budget can be written as: $tr_H = C_H + \psi(\beta_H\text{-}c_H)$ and $tr_L = C_H + \psi(\beta_L\text{-}c_L) + \phi(\beta_H\text{-}c_H)$. Hence, only the efficient firm enjoys a rent given by $\Pi_L = \phi(\beta_H\text{-}c_H)$.

Substituting these constraints in the objective function the latter becomes exactly the same as Laffont and Tirole's social welfare function:

$$W = v\{S(q_L)\text{-} (1+\lambda)[C_L+\psi(\beta_L\text{-}c_L)]\text{-}\lambda\phi(\beta_H\text{-}c_H)\} + (1\text{-}v)\{S(q_H)\text{-}(1+\lambda)[C_H + \psi(\beta_H\text{-}c_H)]\}$$

Hence, to sum up introducing the moral hazard issue (through the effort variable e) in the presence of a risk neutral firm keeps us in the ambit of a pure adverse selection model whenever we make use of Laffont and Tirole's deterministic or stochastic cost function. The crucial point is that the managerial disutility ψ becomes ex post observable, once the adverse selection parameter β is observed. It is also worth noticing how linear schemes are robust in that they are optimal regardless of the distribution of accounting and forecast errors. Hence, even Laffont and Tirole's specification of the cost function as $C = \beta - e + \varepsilon$ gives rise a pure adverse selection model. In any case, the presence of moral hazard with cost unobservability would not had changed matters, as long as the manager is risk neutral, as shown by Baron (1989).[11]

1.2.4 A simplified model

The problem can be further simplified, as we will show for simplicity's sake in the monopoly case, if we assume that a public firm directly maximises welfare. In fact, in this case we can deal within a full information setting where the level of β, the adverse selection parameter, is known by the regulator who now coincides with a benevolent public manager. In this simplified setting we may more easily deal with the most interesting problems of the entry of a competitor in horizontal and vertical settings and of non-linear access pricing regulation. The basic structure of the model presented below will be kept in the following chapters where we examine these problems, allowing for competition.

If one analyses in depth the accounting convention present in Laffont and Tirole (in practice, the fact that transfer and costs are paid by the public sector, which directly receives all revenues from sales, a convention that we will continue to adopt) it is natural to think of a public firms whose deficit is automatically part of the public budget. Hence, one may wonder what happens if, differently from Laffont and Tirole, we are dealing with a public firm. It seems natural, in this case, following part of the literature, to suppose that the manager directly maximises social welfare

---

[11] Baron (1989) shows that the introduction of moral hazard into a generalised version of Baron and Myerson's model would not change matters. "The choice of effort in a managerial model in which there is no ex post observables is thus efficient ... and the price is distorted from marginal cost only as a result of the marginal information ... That is, there is no moral hazard problem when there is no ex post observable and the manager is risk neutral" Baron (1989), p. 1379. For an analysis of the implications of the introduction of different types of errors see Baron and Besanko (1987).

with respect to the tariff system $\{T_H, q_H, T_L, q_L\}$.[12] In fact, the question regarding who controls the public manager is similar to the one who regulates the regulator. The authority may impose a different welfare function than the public manager's one getting a better outcome (especially in the context of simultaneous games).

Following this line of reasoning we show below that in this case, in which there are simply payment-transfers to the public firm's manager, who only cares about social welfare, the model just collapses into what Laffont and Tirole often call the full information benchmark. In chapter 3 where we follow a normative (rather than positive) approach, analysing socially optimal regulation, we will then only consider the case of a public firm even if the results still hold for a private regulated firm under full information.

In fact, assuming that the regulator and manager's utility functions coincide we do not have any problem of asymmetric information and the entire model simplifies, if for simplicity's sake, one also keeps on assuming that competitor's cost and demand functions are known, given that the manager belongs to the same industry. However, we still have a principal-agent problem between the firm and the customers.

In practice, the manager's objective function is no longer given by his private satisfaction {the difference between transfers and costs: $\Pi = tr - vQ = tr - C - \psi(\beta - c)$}, but coincides with W the social welfare function (used by Laffont and Tirole who are dealing with a private regulated firm) in which the manager's private satisfaction Π has the same weight as the consumers' surplus. Denoting by S(Q) the consumers' gross surplus, the social welfare function is exactly the same as Laffont and Tirole's function: $W = S(Q) - (1+\lambda)[c+\psi(\beta-c)] - \lambda\Pi$.

It is simple but instructive to show how the very same results as Laffont and Tirole may be obtained under the standard and optimal regulatory setting in this model, limiting ourselves, for simplicity's sake to the monopoly case, under the

---

12 See De Fraja and Del Bono (1990), Vogelsang (1990), Pint (1991) and De Fraja (1993) for useful references.

following simplified hypotheses.

On the **consumers' side** we have two types of customers present in the same number N: high and low-demand customers (t = H, L), as in Laffont and Tirole (1990b). For the same quantity, the surplus function of the high-demand consumer can be expressed as a multiple of the low-demand one. In particular, $u_H = \theta u(q_H)$ and $u_L = u(q_L)$, where the taste parameter $\theta > 1$ is the factor of proportionality.

On the **industry's side** the incumbent makes use of *fully non-linear tariff*:

$$T_t = T(q_t).$$

where T_t is the amount paid to the firm by the customer of type t.

In the absence of entry the incumbent acts as a monopolist and is characterised by the following *cost* and *revenue functions*:

$$C(Q) = v\,Q$$

$$R(Q) = [T_L + T_H]\,N$$

where v is the new defined marginal cost function, which includes the social costs due to the disutility of managerial effort, and $Q = N(q_L + q_H)$ denotes the total output.

As before, due to the costs of public funds the consumers' surplus may be written as $S(Q)+\lambda R(Q)$, but differently from Laffont and Tirole we have a public firm that maximises social welfare with respect to the tariff system $\{T_H, q_H, T_L, q_L\}$:

$$\max W^* \equiv \quad S+\lambda R(Q)-(1+\lambda)vQ =$$

$$N[\theta u(q_H)+u(q_L)]+\lambda N(T_H+T_L)-(1+\lambda)[cN(q_H+q_L)+\psi(\beta-c)]$$

subject to:

$$[IR_L] \qquad T_L = u(q_L)$$

$$[IC_H] \qquad T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$$

With full information (that is, when v, i.e. the level of β, the adverse selection parameter, is known by the regulator) W can be maximised with respect to q_H and q_L:

$$\max W^* = N[\theta u(q_H)+u(q_L)]+\lambda N[(2-\theta)\, u(q_L)+\theta u(q_H)]-(1+\lambda)[cN(q_H+q_L)+\psi(\beta-c)]$$

Making use of the optimality conditions for the consumers $[p_t = \theta_t u'(q_t)]$ from the previous first order condition with respect to c we can write the following condition of optimal effort:

$[c^\lambda]$ $\psi'(\beta\text{-}c) = Q$ Optimal effort condition

This last condition [that can be also rewritten as $(\beta\text{-}c) = e^*$] can be clearly interpreted as the Laffont and Tirole's "optimal effort condition", which equates the marginal disutility of the public managers' effort and the marginal cost saving [that is, the generalisation of the condition $\psi'(\beta\text{-}c) = 1$ to a model in which the output is variable]. Thus, the introduction of moral hazard (represented by the variable effort) with full information doesn't change the result obtained in the absence of vertical issues: in fact we have some distortion at the bottom, since the value of the high type's marginal price is: $p_H = \theta\, u'(q_H) = c/[1+\lambda(2\text{-}\theta)/(1+\theta)]$, being costly to get public revenue from taxes. Furthermore, the previous condition $[c^\lambda]$ holds for a regulated private firm (not just for a public one maximising social welfare) in the absence of adverse selection. In the following section we also get rid of adverse selection showing that the same outcomes of the cream skimming model can be derived in a simplified setting.

1.3 Modelling cream skimming competition a la Laffont-Tirole: a graphical approach

"Cream skimming" labels the special case in which competition focuses on the most profitable part of the demand market served by a regulated firm that is engaged in price discrimination. In the case in which competitors are successful, we are in the presence of the so called bypass regimes: high-demand consumers purchase from the entrant bypassing the regulated monopoly, which is left with the less lucrative part of the market ("skimmed milk").

The existence of a bypass regime with no capacity constraint (for a high level of the incumbent's intrinsic marginal cost) introduces discontinuities in the control of the regulated firm, given the reduction in production, because the firm serves only low-demand customers.[13] However, even the mere threat of this potential competition interferes with the pricing policy of the monopolist (in the context of second or third degree price discrimination). In particular, with second degree price discrimination the possibility of bypass complicates the incentive compatibility problems. The monopoly may consequently call for a restrictive policy toward this aggressive competition that serves the low-cost and high-return part of the market.

Laffont and Tirole (1990b) examine this case under the following hypotheses: **(i)** individual consumption of two customer types (q_L, q_H), respectively low and high-demand consumers (indexed by t= L and H and whose proportion are α_L and α_H) can be monitored so that the use of fully *non-linear* tariff $T_t = T(q_t)$ is feasible for the established firm;

(ii) the surplus function of high-demand consumers is a fixed multiple of the surplus function of low-demand consumers (here we have put θ_L=1 just to simplify notation);

$$u_H = \theta\ u(q_H)$$
$$u_L = u(q_L)$$

with $u'(q_t)>0$ and $u''(q_t)<0$.

(iii) the regulated firm's cost function is linear in total output $Q = c(\alpha_L\ q_L + \alpha_H\ q_H)$:

$$C(Q) = (\beta\text{-}e)Q = c(\alpha_L\ q_L + \alpha_H\ q_H)$$

where β represents the adverse selection parameter (i.e. the intrinsic marginal cost) and e the moral hazard parameter. The managerial utility function is given by the difference between transfers tr and the disutility of effort $\psi(e)$; that is, U_m=tr-$\psi(e)$;

(iv) competitors' bypass technology allows a positive net surplus only to high-demand customers and it is always optimal for the incumbent to serve low-demand consumers (no shutdown hypothesis);

---

[13] Furthermore, as Laffont and Tirole argue, in general, the presence of asymmetric information between the regulator and the firm, rising the actual cost, increases the probability of bypass in comparison with the optimal complete information level.

(v) the policy choices of the regulator maximise the expected social welfare, defined as in their basic (1986) model; that is, introducing the social cost of public funds, a parameter that affects both the consumers' surplus, profits and the managerial utility: $W= S+\lambda R-(1+\lambda)[c+\psi(e)]- \lambda U_m$. S denotes the consumers' surplus and R the revenue function of the incumbent. We will get the standard Loeb Magat (1979) welfare function for λ equal to zero. For λ close to infinity the case of an unregulated monopoly can be derived.

The Laffont and Tirole model can be stated as:

$$\max W \equiv \quad \alpha_L u(q_L)+\alpha_H\theta u(q_H)+\lambda(\alpha_L T_L+\alpha_H T_H) - (1+\lambda)[c(\alpha_L q_L+\alpha_H q_H)+\psi(e)] - \lambda U_m$$

subject to:

$[IR_L]$ $\quad u(q_L) \geq T_L$

$[IR'_H]$ $\quad \theta\, u(q_H) - T_H \geq S_H$

$[IC_L]$ $\quad u(q_L) - T_L \geq u(q_H) - T_H$

$[IC_H]$ $\quad \theta\, u(q_H) - T_H \geq \theta u(q_L) - T_L$

$[U_m]$ $\quad U_m = tr - \psi(e) \geq 0$

No particular comments are needed for the individual rationality $[IR_L]$ and incentive compatibility $[IC_L]$ $[IC_H]$constraints since they are the standard ones ($[U_m]$ is just the managerial individual rationality constraint). The modified participation constraint $[IR'_H]$ instead tells that a H consumer will not buy from the incumbent, unless he is allowed to enjoy a rent greater or equal to the exogenous surplus offered by the entrant S_H. In the absence of bypass, we may have a sequence of five regimes that maximises the expected social welfare, as the intrinsic cost of the regulated firm -that is identified with a technological parameter β (which represents also the adverse selection parameter)- increases; that is, his level of efficiency becomes lower and lower.

Regime 1 is the usual one with **no distortion at the top** (i.e. the marginal price for high-demand customers is set equal to the marginal cost) where only the low-demand customers' individual rationality constraint and the high-demand customers' incentive compatibility constraint are binding. Only a very efficient firm will offer to the H type a very high net surplus (greater than the one offered by the bypass technology). Low-demand consumers will be constrained to obtain the optimal separating equilibrium. With a less efficient firm (in regime 2) high-demand customers must be given incentives not to bypass; so their individual rationality constraint becomes binding. Therefore, the distortion at the bottom will be gradually reduced as consumption of the H type remains constant. Subsequently, in regime 3, both marginal prices are equal to marginal costs because the incentive compatibility constraint of the H type is no longer binding. Later, with a further increase in marginal costs, in regime 4, the high-demand customers need more incentives and the L type should be prevented from buying his bundle. In this way, the incentive compatibility constraint of the L type becomes binding and the distortion at the top should be gradually increased. Thus, the H type consumption remains constant, till it is possible to extract all surplus from low-demand consumers. With very high intrinsic costs, in regime 5, the bundle offered to high-demand customers becomes so attractive that low-demand customers must be prevented from consuming it. In this last regime, clearly, low-demand customers enjoy a positive net surplus, since their individual rationality constraint is no longer binding.

It can be demonstrated that in regimes 4 and 5 it becomes optimal to charge marginal prices below marginal cost to high-demand consumers. This form of *predatory behaviour* is put in practice in order to avoid bypass and to prevent low-demand customers from buying the high-demand bundle.[14] Naturally, not all these regimes need exist.[15]

---

[14] Because the managerial disutility and rent are just a function of the realised marginal cost (through effort and government transfers), it is possible to maximise welfare separately with respect to the non-linear schedule of the two types of customers. Hence, the regimes that we are going to examine exist even in absence of the adverse selection and moral hazard problem at the firm level. We will discuss this issue in the following section.

[15] For instance, when the firm is constrained to charge a *two-part* tariff, the number of regimes is reduced, since there are no incentive compatibility constraints to be satisfied. However, even in this case it is still possible for price to fall below marginal cost. In fact, if only the individual rationality constraint for high-demand

Finally, the passage to the bypass regime introduces a discontinuity in the incumbent's high type consumers' demand that disappears. The regulated firm serves only low-demand customers, taking all their surplus. This introduces discontinuities in effort and in the marginal cost of the regulated firm. In fact, with a low supply there is less interest in efficiency issues, and more incentive to reduce costs and rents of low cost type firms. In what follows, we will derive all these regimes graphically in a stylised version of the model, with no asymmetric information both in the unregulated case and under different types of "optimal" regulation. Notice how the entrant will simply serve the high type whenever she can bypass the incumbent.

1.3.1 A graphical representation of the cream skimming paradigm

In fig. 1.1 the indifference curves (or individual rationality constraint, IR) for each type (passing through the origin) are represented in the (q, T) space, where q and T, as usual, denotes respectively the quantities and the tariffs of the regulated monopolist addressed to each type. The curve IR'_H does not pass through the origin simply because he enjoys a positive surplus given by the competitors.

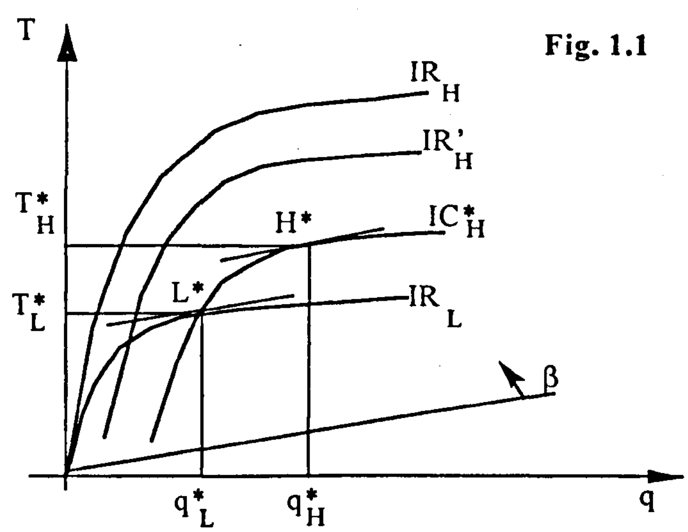


Fig. 1.1

The optimal non-linear tariffs and bundles (q_L^*, T_L^*) (q_H^*, T_H^*) (under asymmetric information) are derived in fig. 1.1 above assuming: a) the social welfare function is the unweighted sum of consumers' surplus and profit -that is, in the absence of

consumers is binding, the firm is obliged to reduce the fixed fee in order to keep this type of customers.

public fund costs (i.e. when λ is equal to zero); b) the relevant binding incentive compatibility constraint of type H passes through (q_L^*, T_L^*). Notice how in fig. 1.1 below an increase in the value of β implies the anticlockwise rotation around the origin of the variable cost of the incumbent (namely, a decrease in his level of efficiency). Moreover, also the lines tangent to IR_L and IC_H (parallel to the variable cost curve) will be subject to the same rotation.

As the intrinsic marginal cost β of the regulated firm increases we can have different non-bypass regimes as shown in the following figures, which are drawn for simplicity's sake, following Maskin and Riley (1984), in the (q, π) space for different welfare specifications, where π is the per customer profit. Not all Laffont and Tirole's five regimes need to exist; moreover, as we will see, the existence of some regimes depends crucially on the social welfare function adopted by the regulator.

In fig. 1.2 below, following the Maskin and Riley convention, the indifference curves (the locus of points for which the utility functions $u(q_L, \theta_L)$ and $u(q_H, \theta_H)$ are constant) represent also the profits that the regulated firm can extract from each customer.

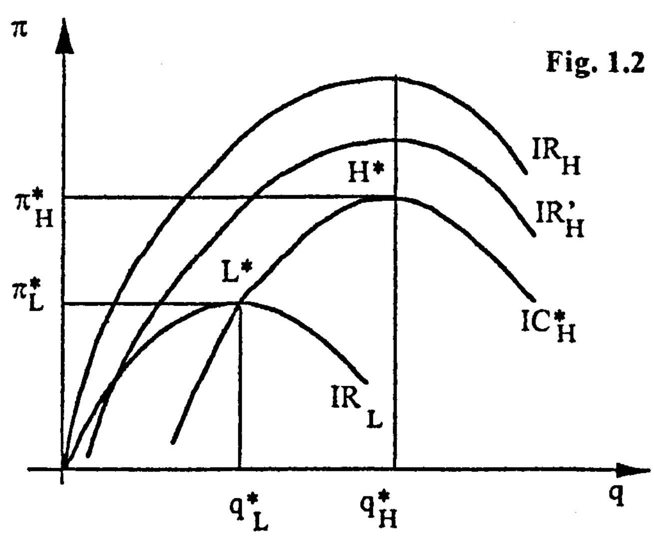
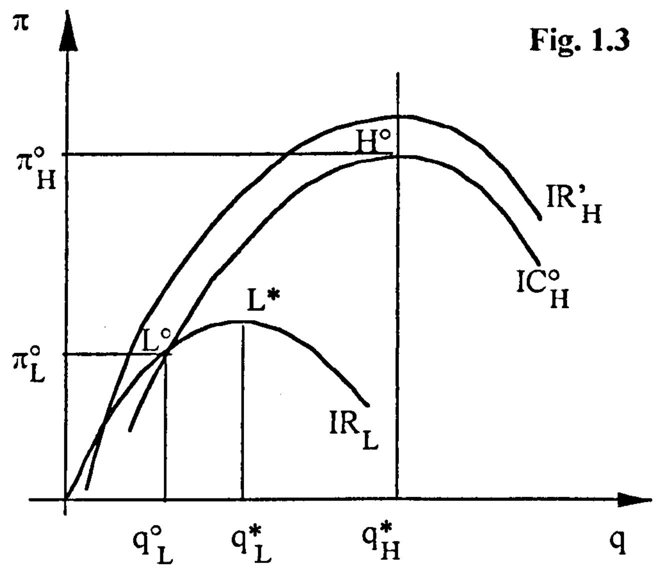


Fig. 1.2

In the absence of additional public fund costs ($\lambda = 0$), the socially optimal non-linear tariffs -under asymmetric information- are the bundles (q_L^*, π_L^*) (q_H^*, π_H^*) of the separating equilibrium for which the unweighted social welfare is maximised under the incentive compatibility constraint of type H, related to bundle (q_L^*, π_L^*), and the participation constraint of the low type.

On the other hand, with $\lambda > 0$, welfare increases if the L type consumes less than the efficient bundle; that is, there is no distortion at the top, but some distortion at the bottom, as represented in the fig 1.3 below. A similar reasoning holds for the unregulated monopolist (a case that occurs as λ tends to infinity), which reduces to a greater extent the low customers' bundle. Thus the relevant incentive compatibility constraints of type H pass through (q_L°, π_L°). In both cases ($\lambda = 0$, $\lambda > 0$) there is **no distortion at the top** (i.e. marginal price is set equal to marginal cost for high-demand customers, so that they consume the efficient bundle).

Fig. 1.3



Low demand customers are constrained to consume less (with contract L°, with $q_L^\circ < q_H^*$) to increase the public profits (from the high type with contract H°, where $\pi_H^\circ > \pi_H^*$) which have a greater weight $1+\lambda$. No problem of competition arises when the regulated firm is very efficient and the H type's participation constraint IR'_H is between IR_H and IC_H° (or IC_H^* when $\lambda = 0$).

In regime 2 (which exists only for $\lambda > 0$ or for an unregulated monopolist) the individual rationality constraint IR'_H is below IC_H° and becomes binding, even if it is higher than IC_H^*.

As represented in fig. 1.4 below, the relevant incentive compatibility constraint IC'_H of type H (which is no longer IC_H°) coincides with IR'_H and the distortion at the bottom is gradually reduced as consumption q'_L increases till it reaches q_L^*.

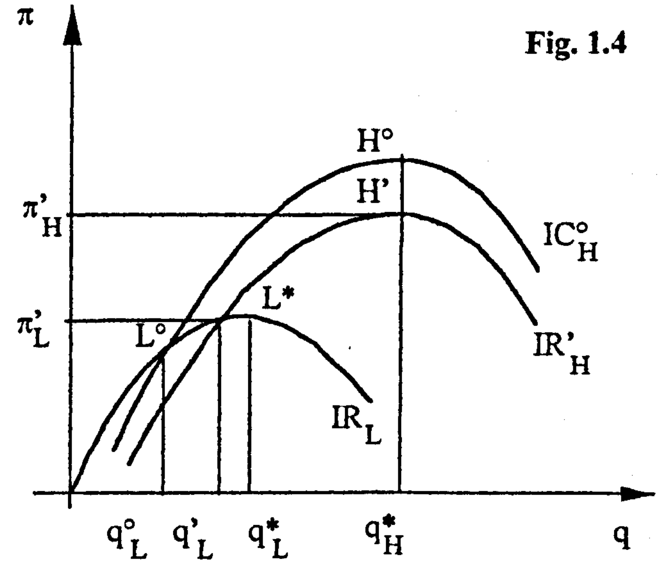


Fig. 1.4

Then, in regime 3, no incentive compatibility constraint is binding and the firm can directly extract the entire rent (apart from the net surplus guaranteed by the bypass technology from the H type) with non-linear pricing, exactly as in a first degree price discrimination setting, as fig. 1.5 below shows.

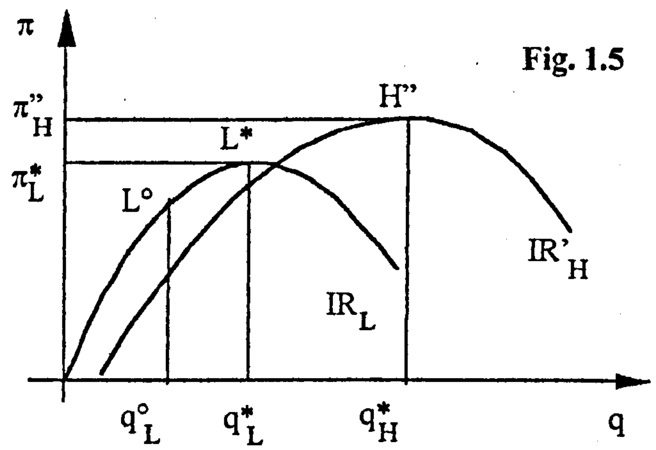


Fig. 1.5

Subsequently, in regime 4 the indifference curve IR_L cross IR'_H on the right of its maximum profit point, as shown in fig. 1.6 below.

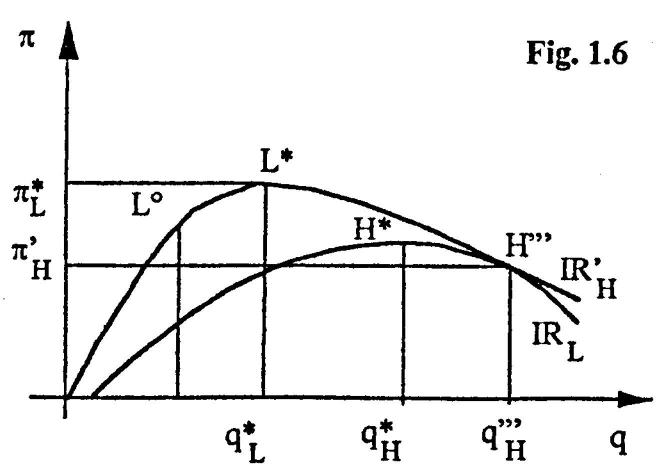


Fig. 1.6

The L type incentive compatibility constraint IC_L is binding; L is prevented from buying the H's bundle but enjoys no net surplus while there is some distortion at the top. Notice how this regime exists only for $\lambda > 0$ or for the unregulated case in the absence of entry.

Finally, with high marginal costs, in regime 5, represented in fig. 1.7 below, while the incentive compatibility constraint IC_L is binding the L type participation constraint IR_L is no longer binding. Thus, low-demand customers enjoy a positive net surplus. As in regimes 4 marginal prices are below marginal costs for high-demand consumers, to avoid bypass and dissuade L customers from buying the H bundle.

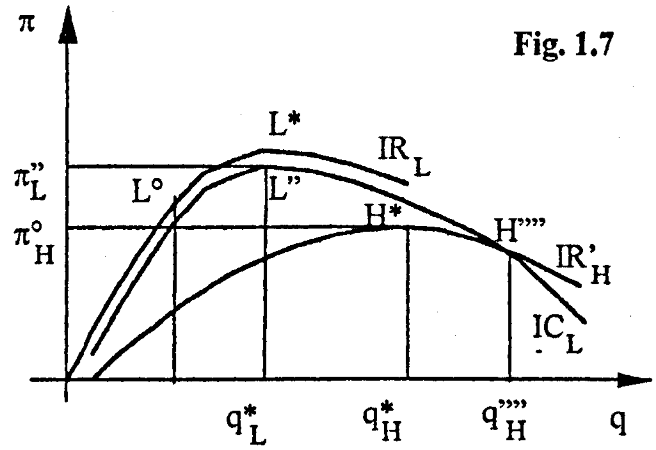


Fig. 1.7

It should be noticed how for λ equal to zero -namely, regulating a la Loeb and Magat- the points H* and H'''' do coincide and we have no distortion in all the possible regimes, apart from the bypass one (that is, regime 1, 3 and 5). Our conclusion is that there are just three non distortionary regimes when social welfare is the unweighted sum of consumers' surplus and profits a la Loeb and Magat (1979). We need to introduce the distortion due to the shadow cost of public funds in order to derive the five regimes as described in Laffont and Tirole (1990b) with optimal regulation.

Only in this respect the cream skimming paradigm proved not to be so robust. No major changes will instead take place without regulation. Thus, paradoxically, this paradigm proves to be so robust that the very same distortion arises as in an unregulated monopoly, the only difference being that the distortion is not always a negative factor. In regime 1 and 2 monopoly tariffs are far from optimal, the marginal price of the L type being higher than the optimal one.

Let us now examine in more depth the *consequences* of cream skimming competition *upon consumers*. First of all, notice how the same sequence of regimes can be derived as competitors become more efficient (and the net surplus of the H type increases), taking as given the monopolist's marginal cost, as we will do in the LT game. Laffont and Tirole claim that (if bypass cannot be controlled by the regulator) the presence of efficient potential competitors may be indirectly advantageous to low-demand customers. Their argument is that, when very efficient competitors are prevented to operate, there is less (or no) distortion at the bottom and in some case even a positive net surplus can be allowed to the L type. However, we must take into account the fact that Laffont and Tirole's model is restricted to the analysis of two types of consumers. It is therefore not surprising that the low-demand type is *not hurt* by the presence of bypass.[16]

In any case, in Laffont and Tirole's model it can be shown that:

(a) supply to both types increases as the competitor becomes more efficient, and

(b) the difference between the two type surpluses is greater when the incentive compatibility is upward binding. Hence, even if there is less or no distortion at the bottom, the differential surplus of high demand customer increases as the surplus left to high demand customers does, and, if public revenues are not to be reduced, low-demand customers will *generally be hurt*.[17]

Let us now briefly look at whether the *rent* of the regulated firm is always reduced by competition. In this type of model, the production function resulting from

---

[16] If we just allow for more than two types it becomes evident that some intermediate ones can be hurt by the desertion of some others (through a surplus reduction). Another case in which low-demand consumers are worse off arises when the firm is subject to a budget constraint, since they must bear the losses in revenues associated with the bypass.

[17] Denoting by S the surplus functions (which differ from a factor of proportionality θ) and by T the tariff charged for the two types, from the incentive compatibility constraint for the high-demand customers (binding in regime i) $\theta S(q^i_H)-T^i_H - \theta[S(q^i_L)-T^i_L]=0$ we obtain the difference between the two surpluses $\theta S(q^i_H)-T^i_H - [S(q^i_L)-T^i_L]=(\theta-1)S(q^i_L)$, whereas from the upward binding incentive compatibility $S(q^i_H)-T^i_H - [S(q^i_L)-T^i_L]=0$ we get $\theta S(q^i_H)-T^i_H - [S(q^i_L)-T^i_L]=(\theta-1)S(q^i_H)$.

optimal regulation shows increasing returns to scale, even if the intrinsic cost function is linear in total output, because the regulator is more interested in reducing total costs as output grows. Laffont and Tirole calculate the new values of the rents left to the incumbent in the case in which the regulator can prohibit bypass. They show how rents could be: **(1)** lower for mediocre regulated firms (the explanation is that, as the bypass region is reduced, the interest in efficiency is increased, as are the rents of lower cost firms), **(2)** higher for the most efficient ones (since it is no longer required to increase output to contrast bypass in regimes 2-5 and the interest in efficiency is reduced, as are the rents of lower cost firms).

We may repeat their reasoning, taking as given the monopolist's marginal cost, and introducing two types of competitors: one efficient (that leads to a bypass regime) and the other inefficient. With a ban on bypass, it is likely that in the first case rents are reduced in the bypass regime and hence in regime 1, whereas in the second case nothing happens (as regime 1 is the only regime and bypass is not allowed). When bypass cannot be banned, clearly in the presence of efficient competitors the rent-reduction effect certainly predominates for mediocre regulated firms. Instead, with inefficient competitors the rent surely increases for all values of marginal cost if there is more than one region. Therefore, we may conclude that in any case it is always desirable to have efficient competitors because the more perverse effects may happen when bypass is not optimal; that is, in the presence of inefficient competitors.[18]

1.3.2 The LT game

In what follows we present the LT game in which we adopt the same approach to competition as Laffont and Tirole (1990b), leaving aside asymmetric information and "optimal" regulation. In other words, in this section we will also get rid of

---

[18] In the simplest case of third degree price discrimination Laffont and Tirole show how cream skimming may also interfere with *redistributional concerns*. However, even if this type of competition interferes with optimal regulation, it should be borne in mind that, by assumption, Laffont and Tirole exclude the presence of any positive effects of competition (such as the reduction of information asymmetries).

adverse selection. In this framework the incumbent faces an entrant who is perfectly competitive and is unable to enter the low demand customers' market. If one is not yet convinced that the richness of possible regimes does not depend on the presence of asymmetric information or, more generally, that it does not crucially hinge on the optimal regulatory structure imposed in the model our present analysis shows that this is the case. In fact, in the LT game we have different regimes only because competitors have a perfectly elastic offer curve and modify high demand consumers' incentive compatibility constraint (by offering them a given net surplus S_H).

The game's structure is as follows. In the first stage of the LT game the entrant chooses the tariff for which she is committed to serve any customer. In the second stage the incumbent chooses his strategy, but for him it is always optimal to serve low-demand consumers (due to the "no shut-down" hypothesis). The analytical derivation of all the regimes is given in Appendix 1. Here after sketching the analytical problem we just provide a simplified graphical exposition.

The LT problem

max	$\alpha_L T_L + \alpha_H T_H - c(\alpha_L q_L + \alpha_H q_H)$	subject to:
$[IR_L]$	$u(q_L) \geq T_L$	
$[IR'_H]$	$\theta u(q_H) - T_H \geq S_H$	
$[IC_L]$	$u(q_L) - T_L \geq u(q_H) - T_H$	
$[IC_H]$	$\theta u(q_H) - T_H \geq \theta u(q_L) - T_L$	

All the five regimes already described in our analysis of Laffont and Tirole's cream skimming model and the correspondingly pricing rules (p'_H, p'_L, p''_H and p''_L) are represented graphically in the case of linear marginal utility $u'(q) = 1-q$.

Regime 1 is the usual one with **no distortion at the top** (i.e. marginal price for high-demand customers is set equal to marginal cost). It is obtained with a very efficient firm able to offer to the H type a very high net surplus (greater than the one offered by the bypass technology). Low demand customers will be constrained to obtain the optimal *separating* equilibrium.

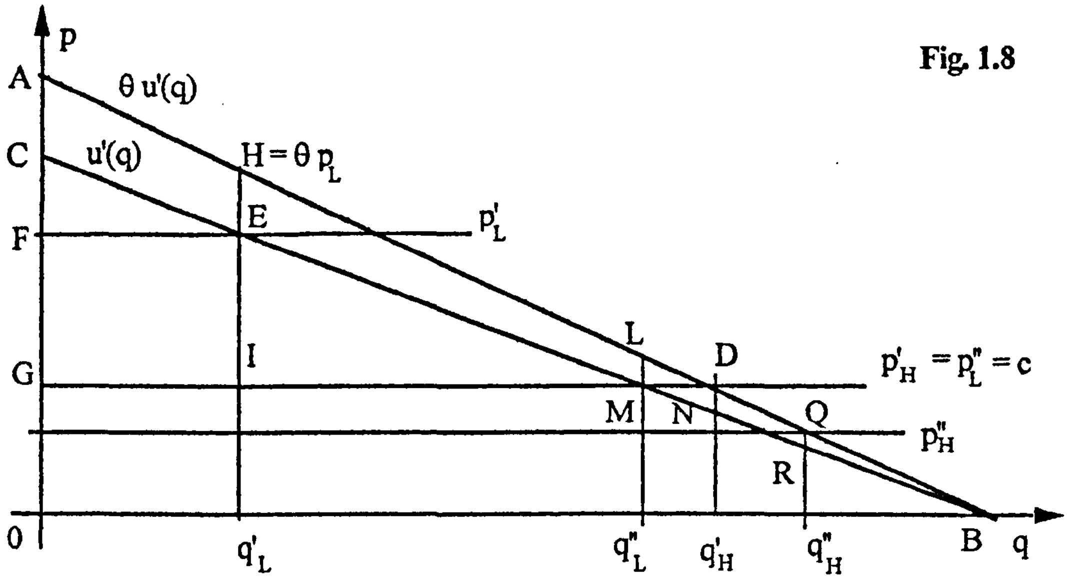


Fig. 1.8

In fig. 1.8 above the linear marginal utility functions u'(q) and θu'(q) are shown as well as the optimal discriminatory marginal prices p_H=OG=c and p_L=OF>c. The quantities allocated per unit of customer are obtained in correspondence of the intersection between the curves representing the marginal utility and marginal price for each type. The tariffs T_L and T_H give a measure of the gross revenue per unit of customer. T_L is simply given by the area COq'_LE, while to obtain T_H the area $Hq'_Lq'_HD$ should be added. Net revenues per customer are obtained subtracting the marginal cost c: graphically CGIE and HID measure the respective net revenue obtained from a customer of type L and H. The H type enjoys a net surplus $(\theta-1)T_L$ (the area ACEH) in order to satisfy his incentive compatibility constraint.

With a less efficient firm (in regime 2) high-demand customers must be given incentives not to bypass; so, their individual rationality constraint becomes binding. This implies that the distortion at the bottom can be gradually reduced as consumption of the H type remains constant. Graphically, while the H type marginal price stays the same (p'_H =OG=c) the L type optimal discriminatory marginal price varies between p'_L and c. The L bundle the related tariff T_L are increased. Instead, in correspondence of q'_H the tariff T_H is reduced as the related net surplus $(\theta-1)T_L$ (necessary to satisfy the H type's modified participation constraint).

Subsequently, in regime 3, both marginal prices are equal to marginal costs (p'_H = p''_L) because the incentive compatibility constraint of the H type is no longer binding. Now, in correspondence of q'_H the H customer pays a lower tariff and his net surplus varies between the area ACML and ACND. The H type consumption remains constant, till the surplus given by competitors reaches ACND.

Later, with a further increase in marginal costs, in regime 4, the high-demand customers need more incentives and the L type should be prevented to buy his bundle. In this way, the incentive compatibility constraint of the L type becomes binding and the distortion at the top should be gradually increased.

Finally, while the L type marginal price p''_L =c and the related tariff T_L stays the same the H type optimal discriminatory marginal price varies between p'_H and p''_H below c. The quantity q_H increases as the surplus T_H till the level ACRQ is reached in correspondence of q''_H. Then in regime 5 the H and L types' consumption remains constant at the levels q''_H and q''_L and it is no longer possible to extract all the surplus form low-demand customers who will subsequently enjoy a net surplus. It can be demonstrated that in regimes 4 and 5 it becomes optimal to charge marginal prices below marginal cost to high-demand consumers. This form of *predatory behaviour* is put in practice in order to avoid bypass and to prevent low-demand customers from buying the high-demand bundle.

With very high marginal costs, in regime 5, the bundle offered to high-demand customers becomes so attractive that low-demand customers must be prevented from consuming it. In this last regime, clearly, low-demand customers enjoy a positive net surplus, since their individual rationality constraint is no longer binding. Finally, in the bypass regime high-demand consumers' demand disappears and the firm serves only low-demand customers, taking all their surplus, without any distortion.

To demonstrate the existence of the five regimes we need also to show that regime 5 is more profitable than the bypass regime. This requires just that the average price for the H type is greater than the average cost (at the end of regime 4), or to verify the inequality $u(q''_L)/q''_L > c$. In practice, we require:

$$u(q''_L)/q''_L = (1+c/[1+(\alpha_H/\alpha_L)(\theta-1)/\theta])/2 > c$$

which is clearly verified for θ sufficiently close to 1.

In reality, for a competitor there are no obvious good reasons to increase the surplus of high-demand customers (apart from the theoretical case of perfect competition). Furthermore, in a more general model, the type of customer should be selected endogenously, on the basis of the profit maximisation and not assuming a priori that one type will never be served for technical reasons. The strength of this model seems to lie in its weakness: that is, assuming competitors with unlimited capacity and restricting competition for the H type all the problems connected to the entry and pricing decision on the competitor's side do not arise. Furthermore, many complications in the welfare maximisation problem are eliminated as only the H type can be served by perfectly competitive entrants who do not enjoy any profit.

These ad hoc hypotheses lead to the most striking feature of this type of game: that is, the regulated firm (which is implicitly assumed to be a dominant firm, in any case big enough to find it convenient to serve all customers of type L) seems really to play the role of the "entrant" for the H type. In fact, apart from the first regime, in which he responds to the tariff fixed by the competitors, in all the other regimes he behaves exactly as a "surplus taker competitor".

In chapter 2 we follow a complete different approach to competition, showing how, when the incumbent remains the monopolist of an intermediate good, which is consumed both internally and by any potential competitors, cream skimming turns out to be the only strategy of competition allowed by the incumbent (cf. section 2.5). Therefore, in this context cream skimming does not derive from ad hoc hypotheses, but is simply due to the vertical structure of the market.

To develop this regulatory research topic, in chapter 2 we will start from the original model proposed by Laffont and Tirole (1990a) and (1994) tackling the problem of entry as a crucial one. In fact, regarding competitive issues their model turns out to be quite restrictive, as it simply bypasses the most interesting problems related to strategic competition. As in the basic model (abstracting from vertical issues) competitors are assumed to have an unlimited capacity, and they don't really take any type of economic decisions. Furthermore, the perfectly competitive entrants are only allowed to charge linear pricing schedules and also the regulated firm and the authority make use of a linear access pricing policy. Hence, it may be interesting to introduce the possibility of price discrimination. Since this pricing strategy is a more general one, always superior to the linear pricing strategy, it is not clear why a firm (and in particular the incumbent) should reject it a priori.

Our approach will focus on competition and price discrimination. In the following section we will summarise the main insights provided by the literature on price cap regulation, following a positive approach to regulation.

1.4 Interactions between price regulation and competition

1.4.1 Alternative approaches

The optimal regulatory models mentioned in the previous sections should be regarded just as sources of insights rather than solutions to any actual regulatory problem. For instance, the amount of private information built into the models seems too limited compared with the problems actual regulators face. There is no reason why the regulator should be merely uncertain about the level of the cost function and not about its shape, or the demand condition of the industry.

Lewis and Sappington (1989) introduced asymmetric information on the demand side, assuming that the firm has better knowledge of demand than the regulator. They derive in this setting the optimal regulatory policy in two scenarios, namely when marginal production costs increase or decrease with output. In the first case price decisions can be delegated to the firm, which can be induced costlessly to set efficient prices (enjoying no rent from its superior knowledge). This result contrasts sharply with the one established for asymmetric information on the cost side, among the others by Laffont and Tirole (1986), where the regulator induces prices above marginal cost in order to limit the firm's informational rent. In the case in which marginal costs decrease with output, instead, no delegation takes place and the optimal policy involves the setting of a single price invariant to demand. This optimal price turns out to be greater than marginal costs for small demand realisation and below marginal costs otherwise. Informational rents are enjoyed by the firm only for intermediate demand levels (rather than in extreme cases as it occurs in the case of cost uncertainty).

Biglaiser and Ma (1995) introduce competition (in the form of a Stackelberg follower) in a model very close to Lewis and Sappington (1988) in the informational structure, characterising the optimal policy in the case of constant marginal production costs. They isolate a trade off between the efficient distribution of consumers across firms and the excess profit, a trade off present even under complete information. The conclusion is that the presence of competitors implies some relevant constraints on regulation, so that the pricing decision can no longer be delegated to the firm. Optimal policies under asymmetric information can be pooling, constant or separating, depending on the distribution of the private information.

Another shortcoming of the *normative* approach is that it focuses on the theoretical case of "optimal" (i.e. social welfare maximising) regulation and examines only very special situations: we are referring in particular, as should be clear from the previous section, to the modelling of competition. In practice, Ramsey pricing is seldom implementable, because of the huge amount of information required both on the demand and the cost side (for a derivation of the Ramsey formula see section 1.5.3). The alternative approach that we are going to examine in this section is a more pragmatic one, whose purpose is to analyse the properties of regulatory schemes used *in practice*. Since this approach leaves the question of why transfers are ruled out unexplained, it can be criticised as "ad hoc". However, it is a useful complement to approaches that assume that regulators have the power to make such transfers, contrary to what is observed in reality. If transfers are not feasible the regulator must rely on *prices* to perform his objectives. Laffont and Tirole (1994b) recognise the very demanding requirements of the Ramsey approach and seem to reach a compromise with price regulation, proposing the "global price cap", which we are going to analyse in the final subsection (1.4.3) on access pricing.

The *positive* regulatory literature has first focused on how traditional regulatory constraints induce profit maximising enterprises to misallocate resources, contrary to the central regulatory goal of replicating competitive conditions. The Averch and Johnson model, for instance, investigates the behaviour of a profit maximising monopolist operating under a rate of return on investment constraint. A key assumption is that the allowed rate of return exceeds the actual cost of capital. The so called *A-J effect* is basically an over-investment result, as it implies that the ratio of capital to labour is too high to minimise costs in producing any observed level of output.

A shift from profit to price level regulation affects a shift of risks and benefits between the firm and its customers. In pure theory *profit level* regulation assigns to consumers the risks of cost increases and the benefits of cost reductions, while *price level* regulation reassigns both to the firms. By breaking the equation between the firm's actual internal costs and allowed revenues, price regulation regenerates the range of managerial incentives for profit maximisation (subject to the constraint of maximum price levels). However, price levels erroneously set too low will encourage output contraction and eventual withdrawal from the market, whereas price levels just enough to maintain the firm's viability will discourage new investments. The conclusion is that price caps methods are more likely to encourage innovation under certain circumstances, because the firm can keep the benefits (at least for a limited period of time); but this may also occur under rate of return regulation (henceforth ROR) whenever a regulatory lag is present. A comparative assessment of these two methods of regulation and an interesting discussion is contained in Waterson (1992).

Essentially the same conditions for efficient regulation are required for cost of service regulation and price caps. Without truthful revelation and commitment by the regulator neither will be efficient. Claims for the superior efficiency of price cap regulation over ROR stem from the notion that ROR to the extent that it is a cost driven system, does not provide an incentive to minimise costs. However, the decoupling of price and cost under price caps is by no means guaranteed. As proposed in telecommunications there is a problem about setting the initial level of the price caps and the adjustment formula over time.

Waterson (1992) examines in detail the theoretical and practical implications of the use of these two methods in the US and UK. Following the evaluation of Littlechild (1983) for the telecommunication industry in the UK one might conclude that if protection against monopoly is not important there would be no need for regulation. In sum, regulation would apply only if the industry is particularly relevant, or there are substantial barriers to entry. The more recent assessment of Beesley and Littlechild (1989) provides several argument in favour of price cap regulation over ROR. The advantages of price caps are due to its greater flexibility and simplicity but also to the fact of being less vulnerable to the A-J effect.

Let us now consider very briefly the originality of the framework adopted by Waterson (1992). First of all, he analyses some of the distortions introduced in the application of price cap. An *output distortion* effect is due to the fact that generally the price cap applies as a constraint on average price over a range of products. A sort of A-J effect can arise with price caps, if they include certain cost passthrough elements. Moreover, in practice as the price cap formula can be revised over time, there are usually regulatory lags in ROR. Therefore, also non pricing issues need to be reexamined. Finally, a desirable criterion emphasised in this paper is the openness of decision taking which not only is useful in order to ameliorate the problem of asymmetric information but also provides the opportunity for concerned people affected by the actions of the regulated firm to put their cases.

A quite general framework to deal with these questions is provided by Waterson (1994), who examines alternative regulatory tools to be used in practice, comparing price cap regulation with ROR and the approach proposed by Braeutigam (1993). Price cap regulation has been initially proposed in order to ensure consumers to be better off, providing at the same time to the regulated firm the incentive to engage in cost reductions and quality and service enhancement. In reality, many variants have been proposed and not all of them ensure these desirable properties. Moreover, the very real problem has to do with the promotion of competition in all aspects of the network utilities. For instance, over time there is allocative inefficiency where costs fall more rapidly than the price formula reduces prices. ROR could solve this problem; sometimes regulators use rate of return earned in order to assess what revisions to the price cap formula are required.

Another alternative framework would be relying upon the existing institutions such as the Office of Fair Trading and Monopolies and Mergers Commissions, which cover competition issues in a broader context; i.e. in the overall economy. A serious problem that we would have to face is that the anticompetitive practices are too slow. A system that seems more appropriate to solve medium term problems has been proposed by Braeutigam. According to Braeutigam's procedure there is a constraint between the maximum profit made by the firm overall and the price allowed in the regulated sector. In particular, only if prices fall increases in profits are permitted. This approach allows to overcome some of the difficulties connected with ROR and price cap regulation. In particular, it would no longer be possible for a firm to subsidise unregulated competitive activities from regulated, or to avoid the consequences of the revisions of a stringent price cap. The only real problem with this procedure is that it implicitly assumes that the regulator can observe demand shifts and that it is not clear at all how it can deal with innovation and quality issues.

Waterson argues strongly for the need of regulation in order to promote competition, even after the privatisation and/or liberalisation of public utilities. In fact, competition is not guaranteed in a framework in which there is still a monopoly at a stage of supply. A regulator is needed to oversee access, charging, maintenance of competition, promotion of consumer interests, network benefits etc. Clearly there would be always present a tradeoff between the losses of economies of scale (associated with the network nature of public utilities industry) and the promotion of competition.

1.4.2 Price discrimination and competition

Let us now examine how the pricing structure (especially when is subject to regulation) influences social welfare, in a direct way, or indirectly, by modifying the nature of the competition faced by a multiproduct incumbent. In the pure monopoly case (without regulation) it is well known that price discrimination has an ambiguous welfare effect, which can be positive only when there is an increase in *total output* (the negative consequence of price discrimination is that it induces marginal utilities to differ; we will refer to this phenomenon as the *inequality* effect).[19]

In what follows we will show how the introduction of price regulation (in particular an average revenue constraint) can induce the incumbent to set price below marginal cost, with unintended negative effects on potential competition. Furthermore, the incumbent's response to entry is likely to be even more aggressive when price discrimination is allowed. We can then conclude that the presence of this kind of predatory behaviour provides a strong case *against* freedom to price discriminate.

The analysis of a regulated discriminating monopolist of Armstrong and Vickers (1991) and Armstrong, Cowan and Vickers (1995) represents the first step to follow in order to examine the direct interactions between competition and price regulation. Thus, it pays to give a sketch of the main results, some of which can be directly applied in the presence of a given scale of entry and no redistributional concerns, as shown by Armstrong and Vickers (1993). In the case of *third degree price discrimination* Armstrong and Vickers (1991) show how freedom to price discriminate, subject to average revenue constraint, reduces consumers' surplus, but has an ambiguous net welfare effect (as in the unregulated case). The net welfare effect is negative if total output does not increase enough or the price cap is close to marginal cost. Furthermore, in this latter case the firm is induced to charge price below marginal cost whenever sufficiently different elasticities of demand are

---

[19] Excellent surveys on the literature on price discrimination are Phlips (1983), Tirole (1988) and Varian (1989).

present. Similar results are found with non-linear pricing, a form of *second degree price discrimination.*[20] In this framework Armstrong, Cowan and Vickers (1995) show how the results achieved by the literature on pure monopoly (such as the optimality of offering quantity discounts) still hold.

Average price regulation, being equivalent to minimum output regulation, induces the discriminating monopolist to produce a higher output and in most circumstances improves welfare in comparison with the unregulated case, as shown by Katz (1983). But, because of the increase in the level of output, the highest types face a marginal price lower than marginal cost (a result due to a very high elasticity of the demand function). Thus, the negative *inequality* effect is likely to be stronger (than in the unregulated case) and only under particular circumstances it can be offset by a rise in the total output. In any case, the possibility to engage in non-linear pricing unambiguously reduces consumers' surplus. In fact if the average revenue constraint is binding the "average" consumer has no benefit, as he can purchase at the uniform price the same quantity (thus, only higher types may be better off).

Finally, the use of optional tariffs (that leave open to the consumer the possibility to purchase at the uniform price) is Pareto improving, as demonstrated by Willig (1978). However, this result holds only if the firm faces no competition and we are not dealing with an intermediate good. The firm is worse off in this regime compared with average price regulation; however, it is still better off, and so are the consumers, in comparison with the uniform tariff's case.[21]

Let us now consider the effects of price discrimination upon competition. For simplicity's sake, suppose that the regulated incumbent operates in two markets, one of which is potentially open to competition. If the prices are required to be equal and are restricted from above by the upper bound given by the regulated price cap, the

---

[20] For a technical exposition of non-linear pricing see Maskin and Riley (1984) and Wilson (1993).

[21] In the case in which the firm does not offer optional tariff it is obliged to offer all consumers a lower tariff, being not allowed to offset a higher tariff for low users with a lower tariff for high consumers.

firm has no incentive to engage in pricing predatory behaviour, since whenever it reduces the price in the competitive market it automatically does so in the other market, losing profits. The incumbent's response to entry is likely to be more aggressive if price discrimination is allowed. As potential competitors take this circumstance into account, price discrimination has a relevant effect on the entry's occurrence and on its level. Once we consider the residual demand (instead of the total demand) in the competitive market we can directly apply the previous results introducing an exogenous scale of entry and putting the same weight on consumers' surplus and profit. Hence, the welfare consequences of price discrimination are, once again, ambiguous: while the incumbent benefits from the possibility to reduce price in the competitive market and to increase it in the captive one, both the competitors and the consumers are worse off.

This situation is analysed by Armstrong and Vickers (1993) allowing also for endogenous degree of competition. The structure of the game is as follows: in the first stage the government decides whether or not to allow price discrimination; the second stage involves the rivals in the decision of whether, and on what scale, to enter the market and the final stage consists in the determination of the pricing rule followed by the incumbent (with the same price in both markets when price discrimination is banned). The optimal pricing rule chosen by the incumbent turns out to be strictly decreasing with respect to the scale of entry (a result due to the enhanced elasticity of the residual demand curve). It follows that when fixing the scale of entry, the price in the competitive market is lower allowing for price discrimination (while the opposite is true in the captive market).

The authors' analysis starts from a situation in which regulation is absent. Prohibiting price discrimination encourages entry, since it restrains the incentive of the incumbent to reduce his price in the competitive market, so that entry occurs in some ranges where it would have not in the presence of price discrimination. This result does not necessarily hold when the scale of entry is endogenised (in fact, price discrimination may promotes entry). <u>In any case, the price in the competitive market</u>

is (weakly) reduced as a consequence of price discrimination.

Even allowing for average revenue regulation, welfare effects still remain ambiguous. Price discrimination hurts consumers in the captive market (as higher prices are charged by the incumbent), while it favours the ones in the competitive market. However, *regulation makes entry less desirable*, by constraining price directly. In particular, the larger the scale of entry, the lower is the price in the competitive market, because higher prices can be charged in the captive market. Thus, pricing below marginal costs may become profitable, under certain conditions. Price cap regulation eliminates this anticompetitive behaviour.

The *desirability of allowing for more competition* (given the incumbent's optimal response) is another important question, whose answer remains ambiguous. In the unregulated context the benefits of customers in the competitive market may offset the incumbent's profit losses (due to a sort of business stealing effect) if the entrant's supply function is sufficiently elastic or if she is the most efficient producer. This latter case may still hold in the presence of regulation, even if entry becomes less desirable, because prices are already directly constrained.

While in Laffont and Tirole (1990b) with second degree price discrimination the incumbent's predatory behaviour arises in order to avoid the potential danger of cream skimming by inefficient competitors, in Armstrong and Vickers (1993) even more efficient competitors may be thwarted, due to the presence of average price regulation. In fact, this anticompetitive strategy chosen by the incumbent would not be optimal without regulation or with separate price caps. This latter type of regulation on the price *structure* has also the same advantageous features as optional pricing regulation. Some interesting issues related to regulation of the price structure (rather than the price level) are discussed by Ireland (1992) who shows that a welfare gain can be achieved in the presence of a restriction of the relative prices a monopolist can charge for a range of products in a model with linear demands and constant variable costs. His contribution can be also interpreted in terms of price discrimination, referring to Ireland (1991). In particular, the policy implication would be that bans to price discrimination can be welfare enhancing, a conlusion in line with the results of Armstrong and Vickers. Sappington and Sibley (1992) consider in a dynamic context the incentives for the firm to engage in strategic non-linear pricing under price cap regulation (arising because of the intertemporal linkage introduced in calculating the average revenue) and propose some implementations (in practice leading to lower levels for the price cap) able to eliminate such incentives.

At this point it is worth examining further complications which arise in vertically related markets, focusing on regulation of the terms of access in network industries.

1.4.3 Access pricing regulation

When regulated firms operate in a competitive environment an important issue is the determination of the access pricing, i.e. the pricing of an intermediate good that allows rivals to survive with the regulated firm in the final output market. This is not a merely theoretical question, as in many utility industries the feasibility of access constitutes a "conditio sine qua non" entry is blockaded. A complete treatment of this problem includes the possibility of *divestiture* of the integrated firm by the regulator. The regulator can in fact break up the integrated firm in order to avoid favourable treatment of internal consumers, and at the same time reduce a line of business, to prevent the supplier of the input from reintegrating once again with the competitive sector. In this case, it is essential to perform a cost-benefit analysis to ensure that the eventual coordination failure (and expropriation of specific investment) does not overcome the benefits due to greater opportunities of regulation and a reduction of the discriminatory policy towards previous competitors.

Ignoring this possibility and incentive issues, Laffont and Tirole's approach allows us to compute the pricing rules, using the standard Ramsey principles, interpreting the intermediate and the final output as substitutes.[22] In this framework it is easy to extend the standard analysis to examine predatory behaviour by the regulated firm, through high access pricing.

The regulator can perfectly check the presence of this sort of predatory behaviour only if the regulated firm and its competitors have the same technical requirement for the intermediate good, i.e. there is no need to create new infrastructure or facilities to connect the competitors with the existing network. When the regulated firm cannot falsely claim a higher access price, without being hurt in the final goods market, the input price exceeds the marginal cost of providing access even at the full information benchmark. Furthermore, with asymmetric information, it is optimal to increase access price, due to the incentive correction associated with the network activity. However, the same is true for the price of the final good produced by the regulated firm.

In the *network expansion* case, instead, less access is given than under the full information benchmark, a result (*too little* competition) that seems to clash with the one obtained in the analysis of the optimal bypass (*excessive* competition), discussed in section 1.3. However, the difference in these results is easily explained: here, the rise in the cost of the intermediate good (due to incentive issues) hurts competitors, whereas, in the bypass problem, it is directly transferred to the consumers, through higher prices in the final output, which in turn favours competitors.

Laffont and Tirole (1994a) extend the previous analysis considering limitations on regulatory instruments, showing how the separation of the regulatory functions from the taxation's ones creates difficulties in the presence of bypass. The determination of the access price involves a trade off between limiting inefficient bypass and contributing to budget balance. Therefore, it is clear that if the regulator is unable to tax in order to raise funds to cover the firm's fixed costs then only a

---

22 This interpretation can be justified given that the fringe makes no profit (so that the shadow social cost attached to their profit is irrelevant). The access price is not only greater than the marginal cost, but also of the one provided by the application of the Ramsey rule.

rough balance between these two conflicting objectives may be reached. They also examine the optimality of the efficiency component pricing rule, henceforth labeled as the Baumol-Willig rule. According to this rule, charges should be set equal to the marginal cost of access plus a term which reflects the opportunity cost of entry. Armstrong and Vickers (1995) define the concept of opportunity cost of entry under different assumptions about supply conditions, bypass and possibility of substitution in a model in which both prices and output are exogenously given -following the approach of Armstrong and Doyle (1994) and abstracting from information and incentive problems. They conclude that opportunity costs are often lower than those implied by the application of a simple price margin rule.

Laffont and Tirole (1994b) show how Ramsey pricing can be obtained by imposing a global price cap -global -i.e. a cap including access as a final good- on the incumbent, with weights exogenously determined and proportional to the goods' forecasted quantities. They criticise the use of a partial price cap together with the efficient component pricing rule, since it provides a form of subsidy to non competitive segments (in the sense that access charge are considered too high).

A more detailed discussion of optimal access pricing regulation is provided in chapter 3. In particular, we will explore within our framework: (a) the necessity that entry may be allowed only if an amount, which reflects the opportunity cost of entry incurred by the public incumbent, is paid, (b) the idea to exploit the merger's efficiency enhancing properties, or more precisely the more general idea of "internalisation" (of which the merger represents a specific application). As will be shown, the first point will prove quite relevant with all the welfare function specifications and with some welfare functions a strictly more efficient competitor (able to pay an amount greater than the opportunity cost of entry) may be required.

In the following sections we reconsider the scope for promoting competition, distinguishing competition *in* the product market from bidding *for* the market.

1.5 Interactions between competition in the market and regulation

1.5.1 The use of yardstick competition as a regulatory means

Competition, providing information about the firm's performance, may play an important role in reducing asymmetric information, as well as other types of *regulatory failures*. As the literature on optimal regulation under asymmetric information has shown, the regulated firm enjoys rents from private information about its technology. In other words, the firm can be seen as an *informational monopoly* opposed to a natural monopoly (arising for instance due to the presence of increasing returns to scale). An obvious way of collecting more precise information about the technology of the regulated firm is through the comparison with other firms. This is the concept that is at the basis of the formulation of so called "yardstick competition".

We will first consider yardstick competition as a regulatory system: as shown by Shleifer (1985) the regulator can force firms serving different markets to effectively compete in *cost reduction*.[23] The structure of the game is as follows: the regulator proposes his pricing rule; firms then invest in cost reduction and the regulator observes their cost level and relative effort to achieve such reductions. The regulator, by setting the price and the lump sum transfer respectively equal to the average marginal cost and the average effort level of all the other firms, can reach the **first best** solution. These average values are used as yardsticks against which each firm's performance is compared, so that each firm is assigned its own "shadow firm".

The promotion of competition through regulation offers in this way the opportunity of combining *internal* and *allocative efficiency*. The first version of the model makes the unrealistic hypothesis that firms operate in identical environments and in the absence of uncertainty. However, the model can be extended to accommodate the presence of heterogeneity, if the regulator correctly observes the

---

[23] If the regulator can credibly threaten to make inefficient firm incur losses, yardstick competition can lead to the first best solution.

diversity, through a statistical inference.

Shleifer's conclusions, generalised in the ambit of incentive theory, lead to a very strong result: under the hypotheses of *risk neutrality*, *no bankruptcy problems* and (appropriately) *correlated information* extracting the firm's rent is costless, so that fixed price contracts are optimal. For a rigorous proof of this result, the classical reference is Crémer and McLean (1985). Specifically, Crémer and McLean impose appropriate rank conditions on the conditional distribution functions, which determine a particular stochastic structure. In this context, the **first best** solution can be obtained as a Bayesian Nash equilibrium. More intuitively, any level of correlation of types between firms leads to a situation where *de facto* all information becomes common knowledge.

One possible way to escape from this extreme result is to assume the presence of two kinds of shocks, an aggregate and an idiosyncratic one (whose sum constitutes the technological parameter of each firm). In this framework it is the aggregate shock that can be treated as the common knowledge's factor, and can therefore be elicited costlessly by the regulator, as in Auriol and Laffont (1992). Lockwood (1995) points out that this mechanism which leads to a first best solution may cause the violation of the consumers' participation constraint (in particular when the correlation is low), due to the presence of transfers from consumers to the firm in some particular states of the world.

Even in this new context (in the presence of two kinds of shocks), the relative performance of the agents plays a useful role, since it can provide a better indicator of their individual efforts, while controlling for the effect of common aggregate shocks. Holmström's (1982) *sufficient statistic* condition implies that *relative performance evaluation* has a value whenever the agents face common uncertainties, so that the performance of one agent provides information about another agent's state of uncertainty. This is because relative performance evaluation provides *insurance* against random events that affect in a similar way the agents and are beyond their control.

The optimal incentive scheme with many agents, as derived in Mookherjee (1984), consists in the general case of a combination of schemes based on individualised and relative performance indicators. Naturally, in the presence of merely idiosyncratic shocks the regulator finds it optimal to regulate individually, while in the opposite case, in which only aggregate shocks are present, the first best solution can be achieved through a fixed-price contract based on relative performance comparisons.

Many of the insights in the literature of *tournaments*, such as Green and Stokey (1983) and Nalebuff and Stiglitz (1983), are just an application of Holmström's result. Potential collusion between firms can in theory constitute a relevant limitation to the use of yardstick competition. In particular, rank-order tournaments may suffer from this problem and may be unusable in some situations.

There are several possible extensions of the analysis. For instance, we can deal with organisational design issues, discussed by Auriol and Laffont (1992), Dana (1993) and Dana and Spier (1994). Auriol and Laffont (1992) examine the optimal industry structure, isolating some relevant effects -apart from the *yardstick* effect already considered in detail above- which may favour a duopoly structure compared with a monopoly:

1) the *sampling* effect, that is the achievement of a higher probability of small marginal costs, particularly valuable when the technologies are very risky;
2) the *complementary* effect, arising from a larger product space associated with the imperfect substitutability;
3) the *information* effect, which favours the duopoly structure only under particular circumstances. Specifically, it introduces an additional cost which favour duopoly if it grows in a linear way with production (if the market structure is chosen ex post, otherwise it must grow faster than production in order to favour duopoly).

The disadvantage of the duopoly structure is that it involves *duplication* of fixed costs.

Notice how if firms are identical the sampling effect has no value. Auriol and

Laffont (1992) model practically collapses to Shleifer's (1985) one. In fact, whenever the Crémer and McLean condition is fulfilled the regulator can capture all the information rents. On the other side side if firms' characteristics are stochastically independent monopoly is favoured under incomplete information, since the sampling effect is weakened.

Dana and Spier (1994) explore the optimal mechanism for auctioning production rights in a setting in which the market structure is endogenous. They conclude that duopoly is implemented less often under incomplete information (than under complete information). This is basically because incomplete information introduces a bias toward less competition, since the private information about the production rights is likely to be correlated with the benefits to increase the competitive structure of the market.

We must also be aware of the possibility of *breaking up* an integrated firm in order to enhance the advantages brought by the introduction of yardstick competition (if such advantages are not offset by the duplication of fixed costs). Furthermore, the *multi-task* issue introduces further distortions. We believe that an important development can be made following the lines of Holmström and Milgrom (1991) by introducing multidimensional efforts in the basic model of yardstick competition.

Finally, in a *dynamic* context matters become more complicated. With the introduction of the *ratchet* effect -a kind of *implicit* incentive, which implies the dampening of a firm's incentive to reduce costs, because of the anticipation of future price reductions- yardstick competition can be undesirable. Meyer and Vickers (1995) discuss this question, showing the relevance of the information structure of the model. Specifically, in a two periods set-up, an increase in the weight on the first period performance in forming expectation makes yardstick competition harmful. The intuition is that the larger is the agent's bargaining power in the second period the stronger becomes the ratchet effect.

We have already discussed the insurance effect brought by comparative performance information, which allows *explicit* incentives to be provided at lower costs (in term of risk). In a simple dynamic model where no explicit incentives can be addressed (the agent being risk neutral) the crucial variable to look is the correlation between firms' types relative to the correlation of transitory shocks affecting the firms' performance. Depending on the size of this variable, yardstick competition can either increase or decrease social welfare. In particular, the ratchet effect becomes less relevant if the correlation between firms' types is greater than the one relative to the shocks, so that the insurance effect outweighs it. In a more complex set up in which both implicit and explicit incentives can be designed Meyer and Vicker (1995) show that if risk aversion is not too important and the intertemporal linkage is not too strong yardstick competition, or more generally comparative performance information, is welfare enhancing.

1.5.2 The effect of product market competition on regulation

Apart from yardstick competition, product market competition interacts with regulation in many other ways. In particular, neglecting distortions in the pricing rule of the regulated firm, a positive question examined at the end of this section, we can first consider competitors as a price-taker fringe, assuming, for instance, the existence of small costs and/or delays in entry.

Under conditions of cost unobservability [following the approach pioneered by Baron and Myerson (1982)] Caillaud (1990) emphasises the gains that can be achieved through the threat of entry, used as an *endogenous regulatory mechanism.* The presence of this uncontrolled competitive fringe has a relevant influence either within the range of regulation, leading to the reduction in the rents of the regulated firm, or outside this range, with the shutdown of the regulated firm (if the fringe is more efficient than the regulated firm). In particular, it is the *correlation* between the unit cost of the regulated firm and the fringe that can be used to save on costly transfers, through the entry's threat. But, even in the case of *stochastic independence*, the ex post price which prevails will be lower than the implicit price chosen by the regulator.[24] Hence, not only do consumers benefit, but also the regulator, as he can reduce the amount of transfers.[25] Biglaiser and Ma's (1995) in a similar perspective, i.e. introducing uncertainty on the demand (rather than on the cost of the regulated firm), model the entrant as a Stackelberg follower (instead of a competitive fringe).

Although there are several analogies with the models presented by Anton and Yao (1987) and Demski et al. (1987), analysed in section 1.6.2, this approach is very different from the second sourcing literature. In fact, the regulator cannot contract with the fringe and cannot exercise any power of control on it. Nevertheless, the presence of a competitive fringe, that represents the role of *market forces*, is welfare improving. Therefore, it could be argued that not only has competition *direct effects*, that interact with the incentive of the firm [as already shown in the previous subsection], but also it may limit the consequences of some *regulatory failures*, such as asymmetric information, as well as problems of regulatory capture and credibility.

Let us finally tackle the distortions in the pricing rule arising from the introduction of competition. Assuming *cost observability* and *incentive-pricing dichotomy*,[26] we will follow Laffont and Tirole (1990a) in the derivation of the modified Ramsey elasticity formulas in the presence of competition (in the form of a competitive fringe). The well known Ramsey elasticity formula derived from welfare maximisation, [where social welfare is defined as is the Laffont and Tirole's (1986) basic model, i.e. $W = S - (1+\lambda)tr + U_m$] holding for each product is:

$$L = (\lambda/1+\lambda)\ 1/\eta$$

where $L = (p-c)/p$ is the Lerner index and $\eta = -(q/p)dq/dp$ is the elasticity of demand.

---

[24] However, the price distortion of the regulated firm may increase or decrease with respect to the unregulated case.

[25] Lockwood (1995) examines the case in which transfers are ruled out.

[26] For a one dimensional cost reducing activity (e), the incentive-pricing dichotomy holds if and only if the cost function is separable C=C(H(c, e), q); that is, changing the production output (q) does not affect the extent to which the firm can transform productivity increases into rents.

The interpretation of this formula is straightforward as the price mark up is inversely related to the elasticity of demand. The correction term $(\lambda/1+\lambda)$ measures the relative cost of transfer. Note also how the foundation of the shadow cost of public funds and taxation can be found in a *general equilibrium* analysis, in which regulation can be used as a substitute for a perfect taxation system (which does not exist in the real world). Two special cases can be easily derived when this factor does not appear:

1) *marginal cost pricing* ($p = c$) by setting λ equal to zero (that is, leaving aside distortionary taxation);
2) *monopoly pricing* ($L = 1/\eta$) by letting λ go to infinity (that is, assuming a very high shadow cost of public funds).

In order to simplify matters and ignore incentive issues, in contrast to the previous model, perfect information about competitors and *no correlation* between technologies are assumed. The competitor is assumed to produce an imperfect substitute (n+1) of the incumbent's product (n), her cost function being common knowledge.

We can distinguish two situations according to whether or not the fringe is regulated. In the presence of an unregulated fringe, the optimal price decision can be delegated to the regulated firm, who correctly reasons in terms of the net residual demand. The relevant *net* elasticity, is equal to the ordinary one in the case of *simultaneous* competition in prices:

$$L_n = (\lambda/1+\lambda)\ 1/\eta_n$$

It is instead lower for strategic complements (that is, for a positive value of the fringe's price reaction: $\varepsilon_f = (dp_{n+1}/dp_n)/(p_{n+1}/p_n) > 0$) and demand substitutes (i.e. when the cross elasticity between n and n+1 is positive: $\eta_{n,\ n+1} > 0$) in the case of *sequential* competition, as is evident from the following equation:

$$L_n = (\lambda/1+\lambda)\ [1/(\eta_n - \eta_{n,\ n+1}\ \varepsilon_f)]$$

The conclusions in the presence of regulated competition are completely different, because of the presence of an *externality* associated with a change in the consumption of the fringe's good. In particular, because of the general hypothesis that the shadow price of public funds is independent of firms' costs, the optimal pricing rule would be as if the two firms were *merged*. In this case the superelasticity formula of Boiteux will do the job:

$$L_n = (\lambda/1+\lambda)\ 1/\tilde{\eta}_n$$

where $\tilde{\eta}_n = 1/\eta_n\ [1+(p_{n+1}q_{n+1}/p_nq_n)\ (\eta_{n+1,\ n}/\eta_n)]/[1\ -\ (\eta_{n,\ n+1}\eta_{n+1,\ n}/\eta_n\eta_{n+1})]$ is the superelasticity of demand.

Finally, Laffont and Tirole (1990a) extend the analysis to the cases of a competitor with 'market power' and subsidised competitors. The regulated firm in both cases attempts to manipulate the demand for unregulated goods, raising its price above the Ramsey price in the first case, lowering it in the second. In fact, the presence of a competitor endowed with market power involves a sub-optimal level of production and enhances the consumption of the regulated firm's product if the goods are demand substitutes. So, under *simultaneous* competition (given the monopoly price) a higher price is optimal because it encourages demand for the monopolised product, raising its consumption.

$$L_n = (\lambda/1+\lambda)\ 1/\eta_n + (p_{n+1}q_{n+1})\eta_{n+1,\,n}\ /\ (p_nq_n)(1+\lambda)\ \eta_n\eta_{n+1}$$

In the case of *sequential* competition two strategic effects are introduced: first, assuming strategic complementarity in prices, there is a further distortion in the pricing of the competitor's products (an effect that tends to reduce the Ramsey index, being captured by the term $\eta_{n,\ n+1}\ \varepsilon$) and an increase in the revenue of the regulated firm (effect that goes the other way round, provided the first effect does not lead to charge a price below marginal cost) brought by the term $(1 - \eta_{n,\ n+1}\varepsilon/\eta_n)$ as shown by the following formula:

$$L_n = (\lambda/1+\lambda)1/\eta_n + [p_{n+1}q_{n+1}/(\eta_n - \eta_{n,\ n+1}\varepsilon)/(p_nq_n)(1+\lambda)\eta_n\eta_{n+1}]/(1 - \eta_{n,\ n+1}\ \varepsilon/\eta_n)$$

The case in which the competitors' pricing rule is distorted by subsidies is represented, for instance, by the absence of payments of social costs (due to pollution or congestion costs generated by competitors). Given these premises, it is obvious that the reduction in the price by the regulated firm is aimed at the reduction in the negative externality generated by competitors. Assuming that the previous effect does not lead to a fall in price below marginal cost, under sequential competition there is an opposite effect (a price's rise) which is due not only to the usual strategic effect, but also it depends on the reduction in the demand of the subsidised good (a positive externality).

So far, we have considered the interactions between product market competition and regulation. In the next section we will look at competition for natural monopoly.

1.6 Competition through auctioning and second sourcing

Since procurement is (and will probably remain) a major component of the public budget in many countries, its cost effectiveness is of the greatest practical importance. The dominance of monopolistic practices and of arrangements with a single producer, due to various reasons, may lead to costly rents and overpricing. Consequently, at least in a *static* context (section 1.6.1) it seems appealing to encourage the introduction of competition in order to improve such a situation and to realise savings through price reduction. In any case, as will be shown, competition cannot perfectly substitute regulation; the best that can be done is just to auction a regulatory incentive contract able to achieve the second best level of effort.[27] The selection of the producer (through competition) is just the first step in the organisation of a natural monopoly. It should be followed by the aspects related to the production stage (such as the determination of the production level and of the firm's rent). In this enlarged regulatory context, apart from the further stages of the monitoring and enforcement of the contract, in which potential or yardstick competition may play a role (at least mimicking the audit process) reducing

---

[27] The auction theory developed for procurement can be readily extended to the regulation framework, so that the same model can be used both for procurement and regulation.

informational asymmetries, for long term activities (such as utility industries) a complete contract with commitment is quite inconceivable. This opens the problem of underinvestment in specific assets, (considered in 1.6.3) because the added value, generated by the investment, may be expropriated during the subsequent bargaining processes.[28]

However, we will show how competition may be partially helpful also in a *dynamic* context, by providing incentives for investment or innovation (1.6.2). Even if in this ambit important factors may limit the efficacy of competition, giving the incumbent a big advantage (as investments in R&D), any possibility to somehow weaken the monopolist's positions should not be neglected.

In any case competition can hardly solve effectively all these problems and our analysis will show that a continuous process of regulation is needed so that the auctioning of a regulatory contract remains a theoretical extreme case.

1.6.1 Static competition and regulation

Before any regulation stage, the most important part of the procurer's task is to identify the most efficient potential firm. Following Vickrey (1961), theory suggests that by awarding the right to be the sole producer with a *second-price* auction (or Vickrey's auction), the regulator selects the most efficient firm and reduces rent (to the difference between his and the second-highest bidder's valuation of the award), since the firms' dominant strategy is to reveal through bidding their true intrinsic cost. In particular, in first best welfare analysis (when the social welfare function is defined by the expected value of consumers' surplus) a Bayesian regulator, setting the subsidy equal to consumers' surplus to induce marginal pricing [a rule proposed first by Loeb and Magat (1979)], eliminates all producer's rent with competitive bidding. However, as will be shown in what follows, this is true only with identical firms or with a great number of bidders.

---

28 Incentives to invest may be created by reimbursing a high part of the firm's total cost at the investment stage or through the existence of a long term contract or a credible substitute (i.e. developing a reputation for being fair).

The auctioning of a franchise is partly different, because the regulator can extract further rents from the winner (promoting more aggressive bidding) distorting prices and linking the contract rules to the winning bid. Thus, the optimal scheme may be seen as a *menu of franchise contracts*, defining prices and net transfer payments in function of revealed intrinsic costs.

Riordan and Sappington's (1987) analysis is built on the model of Baron and Myerson (1982) and allows for cost uncertainty at the time the monopoly franchise is auctioned (specifically, fixed costs are common knowledge, marginal ones are learned by the winning firm only after the auction and will be denoted by c_i) but there is no "winner's curse" (because of the independence of private technology signals τ). Each (risk neutral) bidding firm knows a private technology signal, τ, of the future value assumed by c (unknown to the regulator) independently drawn from the same uniform distribution on [0, 1] and the associated conditional density function $F_1(c|\tau)$ with strictly positive support on $[\underline{c}, \bar{c}]$. Using the revelation principle, the regulator can restrict attention to mechanisms which induce truthful reports: whoever reports the highest τ realisation is selected and pays a franchise fee. He incurs fixed costs and reports the realised parameter c. The regulator establishes the price and the producer receives both the sales revenues and the subsidy.

In general, it is optimal to auction a menu of contract types to induce self-selection and obtain the second best level of effort through production distortion, following a price-cost relationship which is independent of the number of bidders (n):

$$p(c|\tau) = c + (1-t)[F_2(c|\tau)/F_1(c|\tau)]$$

where F_1 and F_2 are the partial derivatives of the conditional distribution function with respect to its arguments (in particular F_2 associates the prospect of low production cost with high realisation of the signal τ).[29]

[29] The fulfilment of a regularity condition requires the adjusted marginal cost (the left hand side of the equation) to be a monotonically increasing (decreasing) function in $c(\tau)$.

In practice, the winner faces the same incentives (independent of n) and his rent is a nonincreasing function in c, as if there had been no bidding (the so called "separation property"). This can be explained by the fact that *after* the auction the winner must be prevented from overstating its realised production cost c.

Competition *benefits* the regulator because the winner's rent is also a nonincreasing function in the number of bidders (n), as the franchise fee is a nondecreasing function of n. In particular, the franchise fee is the sum of the winner's estimate of the second-highest bidder's valuation of the award (which is nondecreasing in n) and of the gains due to the fact that production and transfer are linked to the winning bid. This latter factor (the production distortion that reduces the incentive to misrepresent τ) increases the fee and represents an improvement on the rule proposed by Loeb and Magat with a small number of firms and informative signals [i.e. $F_2(c \mid \tau) > 0$]. If this is not the case, the regulator can extract without any distortion all the winner's profits as the fee is equal to the expected rents due to the transfer. As in the standard model of Baron and Myerson (1982), the fundamental *trade-off* is between reducing production distortion of a given firm type $(p-c)F_1(c \mid \tau)$ and raising rents (decreasing franchise fee) of the more efficient ones $(1-t)F_2(c \mid \tau)$.[30]

From our perspective, it may be also instructive to examine the way in which Laffont and Tirole (1987), generalising their monopoly regulation model, formalise Demsetz's idea to auction the right to operate a natural monopoly industry. They show how, with a very large number of (risk neutral) bidding firms with private information about their intrinsic costs (independently drawn from the same distribution), the winner's intrinsic cost is close to the lowest one and we are near to implement a fixed price contract with an optimal level of effort. Thus, as Laffont and Tirole (1993) conclude: "competition asymptotically solves the moral hazard problem by solving the adverse-selection problem" (p. 318).

---

[30] From the price-cost relationship we see that the distortion is greater the higher is the likelihood to have a lower c [i.e. for higher values of the informative technology signal $F_2(c \mid \tau)$ and of the range of higher bidders $(1-\tau)$] and the lower is the likelihood to have exactly c [for higher values of the probability to distort production $F_1(c \mid \tau)$].

Even in this case, competition may act more as a useful complement than as a perfect substitute of regulation. This can be shown if we stick to their model with a finite number of competing bids (without restricting to fixed-price contracts) and ignore the complexities of contract specification and transaction costs. In fact, in this simplified context, the *optimal Bayesian auction* is equivalent to asking firms how much they are willing to pay to be regulated as a monopolist (in a Vickrey auction with dominant strategy). "The auction selects the most efficient firm and awards the winner an incentive contract to induce a second best level of effort" (p. 322). The "separation property" tells us that with the optimal auction the winner faces the same incentives as if there had been no bidding. This result comes from the fact that in this model the incentive constraint is downward binding and hence the trade-off is between reducing distortion of a given type of firms and raising rents of the more efficient firms. This explains why in general it is still optimal to offer a menu of contracts to different types of firm and also why effort is below the optimal level and decreasing in the level of cost, just as in the monopoly regulation model.

We are back to the optimal monopoly contract with a *random upward truncation point*, given by the intrinsic cost revealed by the second lowest bid. Competition in the auction amounts just to reducing the interval of the possible intrinsic costs. This limits the information rent because now the second lowest bid (and not the lowest type) has a zero rent. In this way *asymptotical efficiency* is ensured: more competition, reducing the interval, leads to select a more efficient firm and to decrease distortion in effort (as optimal regulation implies no distortion with the most efficient type). Therefore, we move toward a fixed price contract.

The previous models, however, emphasise the benefits of selecting the most efficient firm and do not consider the *costs associated with the auction.* At the limit, if transaction costs are very high it may be better to regulate just a single competitor. It would be interesting to see how much the results are modified when a fixed transaction cost must be incurred (before cost realisations are known) to obtain a private estimate of intrinsic cost and when the number of competing firm becomes endogenous. The case of *insufficient entry* may prevail so that (in expected terms) an additional entrant would bring a positive externality.

1.6.2 Sequential competition (through second sourcing) and regulation

It is important to go beyond the previous static model, since the most relevant advantages of competition are perhaps to be found in a *dynamic* context. Like the auctioning of regulated monopoly, the threat of competition, *after* this position has been awarded, can play a useful role in order to mitigate monopolistic problems.

This introduces the issue of *franchise renewal* and of the usefulness of *reprocurement* through repeated auctions. In particular, this possibility may be used in order to exploit superior opportunities in future or to discipline the regulated firm by the threat of a *break-out*. Under asymmetric information the incumbent's advantage imposes the problem of the deviation from the bidding parity rule (which tells, very loosely, to replace the incumbent only with a firm characterised by lower intrinsic cost) to favour a *second source*. Therefore, we are in the presence of a basic trade-off between the entrant's production cost (allocative inefficiency) and the incumbent's rent extraction (saving in incentives or "control cost").

In the absence of investments, Demski, Sappington and Spiller (1987) show how it may be useful to make a regulated monopoly incumbent subject to the threat of break-out in favour of a second source. This makes it more costly for him to pretend to have higher intrinsic costs. In fact, producing at a cost that reflects a high intrinsic cost, the incumbent will face a higher probability of being replaced. Thus, a second source is not only valuable when the incumbent is inefficient, but also its threat may alleviate the incentive constraint and reduce the rent of efficient firms.

In practice, as the second supplier may be used to monitor the incumbent's cost, Demski et al.'s (1987) generalisation of Baron and Myerson's (1982) model allows a comparison between optimal *auditing* (intended as a means of verifying the incumbent's cost report) and *entry* policies, under the following hypotheses:

(a) the regulator (who maximises social welfare, defined as in the standard model) is a Stackelberg leader, able to commit and to enforce only limited penalties (so that he is endowed with a limited power);

(b) the products are identical and the cost functions are, as usual, linear in total output with two possible realisations (high and low) of intrinsic costs, respectively denoted as c^I and c^E for incumbent and entrant, positively *correlated*, and with a marginal cost advantage for the incumbent;

(c) it is common knowledge that, in order to learn their marginal costs, the incumbent and the entrant incur fixed costs that can be only partially recovered. Initially, the regulator offers an opportunity to produce to the incumbent firm asking it to sink a fixed cost and to report its marginal cost. Afterwards, on this basis the regulator decides whether to start production, to shut-down the incumbent, or to ask the entrant for a cost-projection before production (or shut-down) occurs.

The second source (as well as the audit) is considered only if expected welfare gains are greater than expected incremental costs. But, as by construction, entry can always mimic audit, the presence of an alternative supplier may be more powerful in limiting the incumbent's rent. Given these assumptions, this implies that <u>whenever it is not optimal to invite entry an audit should not been undertaken, but entry may be useful even when the regulator never conducts an audit</u>.

In general, the optimal policy mix differs systematically between a second source and an audit. For instance, one can compare the prices in the two regimes finding that entry can increase as well as decrease price distortion. That happens because prices constitute just a set of instruments that may perform more than one role. In general, under any specific condition, the interplay among the various policy instruments will establish in which regime a single one will be higher. Hence, it is possible that the optimal probability of entry is greater (or lower) than the optimal probability of audit, as the first instrument, being the most effective, may be used more extensively (or less extensively to impose the desired level of discipline).

In particular, it may be useful to increase the probability of replacing a high intrinsic cost firm to stop an efficient incumbent pretending to be an inefficient one. In practice, the stochastic dominance by the incumbent's intrinsic cost may call for favouring the second source. In this case, the bidding parity (which implies break-out when intrinsic costs are equal) will *not* hold and the entrant may be intrinsically *less* efficient. In their example, the extra production costs more than offset the gains due the reduced "control cost" (given by the rent of the efficient incumbent), since the incumbent is reluctant to falsely report high costs in order not to be shut-down.

A similar result may be expected to hold also in a dynamic model of price competition where the cost advantage for the incumbent is due to the fact that, being the developer, he may possess more accurate cost information and may have gained experience during the initial production stage. In such a *two stage procurement* process, the developer has a greater incentive to misrepresent cost, because for a low cost type it is useful to rise the second source bid (hiding efficiency to capture rents), while a high cost type finds it profitable to take full advantage of the initial production stage incentives and to lose the reprocurement auction (adopting the so called "take the money and run" strategy).[31] Given this informational handicap, the second source takes into account the "winner's curse" and bids less aggressively, rising in this way the winning auction price. Furthermore, when the second source has won, price competition problems occur in the form of additional cost, delay, or quality failure.

Analysing a sequential model for the acquisition of a newly developed product, Anton and Yao (1987) consider how, under asymmetric information and perfect cost correlation, the experience curve [taken from Spence (1981)] may affect the incentive contract design that induces revelation when the regulator is committed to the use of a reprocurement auction with a second source (and not to exploit the revealed cost of the developer) in order to reduce the monopoly advantages with the new system.[32] They focus on the sequential acquisition process (composed of an

---

[31] In a pooling equilibrium different developers' types submit the same report and, since the second source is uncertain about the production cost, we have an auction with asymmetric information and valuation ("winner's curse").

[32] The experience advantage over a second source (which simply consists in lower production cost of additional units) clearly reduces the benefits of competitive bidding. The auction commitment limits the ability of the government to dictate the

initial production and reprocurement stage with the new system or with the old technology) in order to simplify matters, avoiding problems of design preference and uncertainty about technological possibilities. Given the existence of a less efficient alternative system (an old technology that would never be used with full information), they show how the second source may be useful because:

(1) it establishes the price when the auction is used for the reprocurement, and

(2) it creates the possibility to cut-off the new system, taking as given the cost of the technology transfer.

The optimal cost minimising policy should start with an initial production contract, which reveals the private cost information, eliminating the information handicap of the second source, and should use the cut-off rule for the new system at the reprocurement stage (a threat that is made credible by the presence of a second source). In this way the optimal contract trades off between allocative inefficiencies and savings on the cost of ex-ante incentives. The developer is prevented from reporting falsely high costs, because above the cut-off level there will be no auction and the residual production will be conducted using the alternative system, notwithstanding the higher costs.

However, in this model the second source has a limited setting-price role, as the developer (who has a lower production cost) matches his bid and the learning process is fully automatic, with no place left for discretionary investments.

1.6.3 Regulation, competition and the underinvestment issue

Let us examine now a fully dynamic context where investments in infrastructure and in human capital produce benefits in favour of the incumbent and problems with break-out. Without commitment, in the context of repeated auctions, the major issue which emerges is that of *underinvestment* by the incumbent in long-lived industry specific (non-transferable) assets, which will be lost in case of break-out. However, even if the investment is transferable, it may be difficult to measure it

terms of reprocurement, using the cost information revealed by the developer during the initial production stage.

and to adequately compensate the incumbent for it. With a regulator able to commit, this would be the *only* case in which the incumbent underinvests, given the possibility of being replaced.

An interesting case of *re-auctioning* is examined by Rob (1986) who explicitly derives the R&D process (treating experimentation as sampling with perfect recall) and deals simultaneously with the choice of the most efficient developer and of the incentives needed to induce him to pursue a socially desirable R&D strategy.

First, the firm's R&D strategy is derived from search theory: experimentation continues till the expected gain from searching once more becomes less than its relative cost. Thus, the optimal cut-off cost level denoted by z, turns out to be an increasing function of learning (or searching) cost (specific to the firm and represented by s) and a decreasing function of output auctioned by the government (y). Then, the possibility of realising savings with a "learning" buy procedure (that is, only a fraction of the project is competitively purchased after an R&D stage) is examined. It is clear that we have a trade-off between competitive price reduction (reauctioning) and production cost reduction; in fact, the smaller is y, the quantity the regulator is committed to buy from the developer, the less are the incentives to reduce production costs and invest in R&D.

The analysis starts from the benchmark case in which even the most efficient firm can be undercut by competitors when it makes normal profits and it is assumed that the government and a developer initially make a fix-price agreement with technology disclosure. Under perfect competition profits are driven to zero, so that the price is equal to the most efficient cut-off cost level ($p=\underline{z}$), even in the absence of interface problems (which means that there is no efficiency loss when switching to another producer). Therefore, costs are minimised only in the case in which the single contractor used for R&D is also the only producer.

Instead, with a small number of bidders and a limited interface problem, Rob is able to show how entry by other suppliers, at a later stage of the project, may lead to a net welfare gain, notwithstanding the negative effects on the developer's incentives.

The results are only slightly modified when the quantity to be purchased is not predetermined but depends on the R&D performance.

In any case, Rob does not examine really in depth the incumbent underinvestment problem and if bidding parity should not hold in the re-auctioning, two fundamental issues tackled by Laffont and Tirole (1988).

More specifically, Laffont and Tirole consider the effect of a non-observable *monetary investment*, made in the first period by the incumbent, on the optimal contract and on the second period auction, under the following simplifying assumptions:

(a) the incumbent and the entrant have intrinsic costs c and c' independently drawn from the same regular distribution F and it is always optimal to realise the project (no shut-down);

(b) the regulator is able to commit and maximises welfare, i.e. the sum of expected utilities of consumers and firms (as usual, a shadow cost of public funds is present);

(c) monetary investment d(i) (with $d'>0$, $d''>0$) lowers the incumbent's cost by i and the entrant's cost by ki in the second period (where $0<k<1$);

(d) the discount factor δ (for second period utility) is identical for all agents.

In period 1 the incumbent facing an incentive contract chooses effort e_1 and investment d(i). In period 2, after reprocurement, the intrinsic cost of the regulated firm depends on first period investment (c-i for the incumbent, or c'-ki for the entrant). The regulator chooses a break-out rule $c^*(c)$ which allows the entry of a lower intrinsic cost firm [i.e. such that $c' < c^*(c)$].

It is worthwhile to start from the complete information benchmark. For the regulator it is optimal to set a breakout rule $c^*(c) = c-(1-k)i$ (so that bidding parity holds in the second period auction) and to equalise marginal disutility and benefit of effort. Furthermore, the socially optimal investment i*, chosen by the regulator, imposes the equality between marginal cost and marginal social benefit {in analytical terms: $d'(i^*) = \delta\,[(1-F(c-(1-k)i^*)+kF(c-(1-k)i^*)]$} and fully internalises the externality on the entrant cost [given by $\delta kF(c-(1-k)i)$].

With asymmetric information (i.e. nonobservable investment), instead, it is not always possible to separate incentive problems from the break-out rule and the nature of the optimal deviation from bidding parity crucially depends on the type of investment. If the investment is general ($k=1$) the regulated firm has a low incentive to invest, since it does not internalise part of the social benefit [namely the positive externality on entrant cost]. As a consequence of this "externality effect" the incumbent underinvests and should then be favoured to increase investment [i.e. $c^*(c) < c-(1-k)i$]. Furthermore, the optimal incentive scheme in period 1 relative to period 2 is low-powered, in order to reduce investment cost in the first period and to favour the capture of its benefits in the second one, through greater rents. In this way, the underinvestment problem is mitigated. It may be solved, at least in theory, if big delayed penalties are allowed for and the incumbent's cost depends also on the entrant's realised cost after a break-out. Then, the incumbent would not deviate from the socially desired investment i^*, as if it were observable.

On the other hand, given the regulator's commitment ability, if the investment is specific ($k=0$) the monopolist makes the socially optimal choice. In this case, since nothing should be internalised, the unobservability of non-transferable investment imposes no cost and the slope of incentive schedule is time invariant. Then, it is optimal to bias the break-out rule in favour of the entrant [$c^*(c) > c-i$] taking into account the "rent differential effect" due to the investment advantage (that is, the incumbent obtains a greater rent at the bidding parity point). In fact, given the trade-off between efficiency and rent, with bidding parity (i.e. equal intrinsic cost in period 2) the entrant (being $c' = c - i < c$) needs lower effort distortion because the moral hazard rate is non-decreasing [$F(c')/f(c') < F(c)/f(c)$]. A similar result was already found by Demski et al. (1987) and Anton and Yao (1987).

However, Laffont and Tirole do not allow for a choice of the transferable fraction of investment; in fact, k is only an exogenous variable. Alternatively, it would be of some interest to assume an increasing cost of hiding useful information away from competitors, transforming k into a discretionary variable.

In another similar and interesting extension of their basic model, Laffont and Tirole (1993) deal with *learning by doing* (in the absence of monetary investment) assuming that the incumbent's effort in the first period (e_1) reduces the incumbent's cost by the quantity $(g + h)\, e_1$ and the entrant's cost only by the amount $g\, e_1$ in the second period (so that g/(g+h) represents the ratio of transferable learning). With complete information the regulator imposes the equality of marginal private disutility of effort and its social benefit. In particular, the optimally social effort for the incumbent should internalise the transferable effect of learning by doing on the entrant [$\delta\, g\, F(c\text{-}he_1)$].

Under asymmetric information and unobservable fully transferable learning by doing (h=0) the incumbent is affected, as in the previous model, by the "externality effect" (he doesn't internalise the positive externality on entrant). Furthermore, we encounter a "learning by doing effect" (a reduction of the first period effort e_1 increases intrinsic cost and decreases rent in the second period) that calls for favouring the incumbent, reinforcing the previous effect. Hence, the incumbent is favoured [c*(c) being less than c] in order to increase the first period effort e_1. In fact, in this way the rent lost in the second period, by hiding efficiency in the first period, becomes more costly.

Like in the previous model, when learning is specific (g=0) the monopolist's rent obtained with bidding parity is higher. Due to this "rent differential effect" the break-out rule should favour the entrant [that is, c*(c) is greater than $c\text{-}he_1$]. However, in this model the final result is ambiguous, given the presence of the opposite incentive, due to the "learning by doing effect".

Let us examine now the *optimal incentive scheme*. When the shadow cost of public funds is near to one, the derivative of the marginal disutility of effort is constant, and g is big enough, the scheme should be high-powered in period 1 relative to period 2 in order to increase the first period effort e_1, while the opposite is true when learning by doing is non-transferable (g=0), in order to allow the regulator to reduce the first period rent. This result, in complete contrast with the conclusion reached with the previous monetary investment model, depends entirely on the fact that the investment is now embodied in the first period effort. In particular e_1 should be incentivized to encourage internalisation when g has a big value, while in the absence of externality (g=0) a steeper incentive scheme in period 2 favours rent extraction, making it more costly to hide efficiency in the first period (reducing e_1). Also in this model it could be interesting to introduce an increasing cost of hiding learning, transforming g into a discretionary variable.

1.7 Conclusion

From our overview it is easy to realise how topics such as competition (without regulation) natural monopoly regulation, have been analysed in great detail by the economic literature. As we have shown the "new economics of regulation and procurement" has proposed an incredible number of variants and extensions of the standard analysis, early formulated in the 80s.

However, it also emerges, especially from the last section, how, in most cases, the modelling of competition in the presence of price discrimination has not received a sufficient attention. We believe that is worthwhile to explore further the interactions between competition and regulation when non-linear pricing rules are allowed. This is the route that we are going to follow in the next chapter. A particular emphasis will be given to the telecommunication case and to vertical issues related to the regulation of the terms of access to the network. We believe that the answers to these questions has a particular relevance not only from a theoretical point of view but also from a practical perspective. Just in passing it is worth noticing how many of these issues have been (and are still) the object of several discussion both in the economic literature and in the political arena.

Chapter 2

NON-LINEAR PRICING AND COMPETITION IN VERTICALLY RELATED MARKETS*

* The modelling approach (section 2.2) follows the lines of Vagliasindi (1994). It went through major revisions and it has been extended to vertically related markets. I have received helpful comments from James Mirrlees, John Vickers, and participants of the Industrial Economist Workshop held on 7 October 1994 at the University of Warwick.

Chapter 2

NON-LINEAR PRICING AND COMPETITION IN VERTICALLY RELATED MARKETS

2.1 Introduction

As already noticed in the previous chapter, there is a fairly extensive literature on topics such as natural monopoly regulation, competition (without regulation), and the welfare effects of price discrimination (by a monopolist). However, the problem of the direct interactions between regulation (including in the definition of regulation also the policies toward price discrimination) *and* competition is still largely unexplored, especially in the non-linear pricing framework. Moreover, in the more realistic case in which markets are *vertically* related, even distinguished economists such as Laffont and Tirole (1993) recognise the need of developing a "general theory of access pricing" (p. 266). A renewed interest in the access pricing problem, a problem first introduced by Willig (1979) and Baumol (1983) with the formalisation of the so called Baumol-Willig rule, followed the contributions of Baumol and Sidak (1994), Armstrong and Doyle (1994) Laffont and Tirole (1994b), Armstrong and Vickers (1995) and Economides and White (1995). However, there has been no attempt to deal with non-linear access pricing.

An important question in this setting involves the determination of the range of circumstances in which *access* pricing can be used to bring about a welfare enhancing competitive solution to final goods supply, whilst a monopoly remains at one essential point in the chain of delivery. In fact, it is *not* always the case that the entry of a competitor (even if characterised by lower marginal costs) automatically increases social welfare in a general price discrimination setting. Notice how from a *normative* point of view it is relevant to determine the socially optimal access charge, that is the one (or the ones) that maximises social welfare. In chapter 3 we will specify different definition of social welfare, namely the unweighted and weighted sum of consumers' surplus and profit (with the introduction of distributional considerations or the cost of public funds). We will focus on *regulating access* directly, a kind of regulation which has been subject of controversies especially in the case of telecommunications.

Following a more *positive* approach in this chapter we want to examine the *private incentives* of the economic players (basically an incumbent and a potential entrant) in the absence of regulatory constraints in horizontal and vertical settings (where the rival enters the downstream level). In dealing with private incentives of both parties one interesting perspective to explore is to follow the incumbent's point of view in analysing the conditions under which he can maintain monopoly profits while entry occurs at one vertical level and act as if he were the only player. In doing this we want also to find out whether it is in his interest to oblige the (eventual) competitor to be efficient by the use of an appropriate access charge.

We will see that we are basically dealing with a simplified game in which the regulator does not appear -specifically the preliminary stage involving regulation is removed. This kind of approach (in which first we leave aside regulation) can be justified on economic grounds, because one important regulator's task involves, first of all, to determine what behaviour he could expect from the economic agents had they operated with no constraints. Only in this way he can find out the types of market failures, and, if necessary, he can design appropriate non market mechanisms that can remove, or at least to ameliorate, such failures.

Typical market failures, which occur in the network utilities, are usually related to the presence of natural monopoly, network externalities and the danger of potential cream-skimming competition, which is generally be thought as undesirable by itself. We analysed in detail what they effectively imply in the presence of vertically related markets and in which ways the regulator can intervene to improve social welfare.

It seems quite natural to start the analysis from an initial setting in which there is a natural monopoly in the network and the incumbent is vertically integrated. The regulator can intervene by directly setting the terms of interconnection between the incumbent and potential rivals, or more generally by imposing some constraints on firms' *conduct*. Various solution of the access pricing problem have been proposed in the UK with particular reference to telecommunications and gas industries. The regulator can more radically change the *structure* of the industry, for instance by adopting a divestiture approach. This seems to resemble the view taken by the US regulator in the telecommunication case. Problems brought by the presence of vertically separated structures will be examined in chapter 4.

Making reference to the existing literature, even in the simple context of *third degree price discrimination* only the model of Armstrong and Vickers (1993), which represents the extension of Armstrong and Vickers (1991) to endogenous scale of entry, somehow offers a possible framework to deal with some of the questions related to the interactions between competition and regulation. However, as shown before (see section 1.4.2), even in this case, the complications of non-linear tariffs and vertically related markets are not treated by the authors. In the ambit of *non-linear pricing*, there have been indeed very little attempts to tackle this kind of problems, which seem particularly relevant not only from a theoretical point of view, but also from a practical perspective.

The only model which introduces second degree price discrimination in a regulatory setting; that is, Laffont and Tirole (1990b), has already been analysed in detail in section 1.3. Regarding *competitive issues*, however, this model is quite restrictive, since it ignores the most interesting problems of strategic competition. In fact, as in the original (1986) model competitors are assumed to have an unlimited capacity, and they don't really take any type of economic decisions (competitors do not even choose which type of customer is more profitable to serve, or the optimal pricing strategy to apply). Basically, this occurs because their technology is exogenously given and, due to their hypothesis of two part tariffs (where the fixed part is simply given by the fixed per capita cost of access) with perfect competition, it automatically determines the surplus competitors offer to the most profitable customers. In practice, the problem of which type of customers will be served in equilibrium is already settled in advance by the authors, by appropriately choosing a relatively high fixed per capita cost in order to avoid competition for the low customers (the skimmed milk). In fact, Laffont and Tirole (1990b) concentrate their analysis on a very specific case; that is, the complete market invasion by a perfectly competitive fringe using a two-part tariff; this explains why high-demand customers (referred henceforth as the H type) are always the cream. Furthermore, they only deal with the case of an optimally regulated incumbent in an asymmetric information setting, or we may argue, following their convention on transfers, the case of a public firm once we remove asymmetries of information between the regulator and the firm.

However, in a more general situation, it is not clear at all that the H type should always be seen as the most profitable part of the market by any possible type of entrant. In other words, the entrant's choice of the type of customers to serve should in general depend also on the relative efficiency of the competitor and needs to be *endogenised* in a sequential multistage game. This more complicated setting is avoided by Laffont and Tirole (1990b) in assuming the existence of a competitive fringe with high fixed per capita costs that automatically eliminates any further possible stage in which strategic entry decisions take place. In reality we usually do not see a optimally regulated incumbent or a perfectly competitive fringe, and the typical initial situation entails a "big" incumbent who faces a "small" entrant not able to undertake a complete market invasion.

If we want to follow the original Laffont and Tirole model we may focus on the special case in which competition is able to attack only one type, serving all the population (that is market L or market H) we may need to introduce ad hoc assumptions that prevent entry for one type. Clearly, in Laffont and Tirole's original model one cannot even conceive the problem of what is the most profitable part of the market for the *entrant*, because both markets are equally profitable for the competitors, given that their profits are always zero. The only *structure* examined is the one implied by a competitive fringe. What matters for the outcomes of their model is only what the incumbent finds it more lucrative. Given the existence of a bypass regime, provided that their number is large enough, the H customers represent the most profitable part of the demand market served by the incumbent that is engaged in price discrimination. However, this does not imply that the profit per unit of customer obtained from the H type is higher than the one relative to the L type for any type of competitor which can enter the market.

On the other hand, when the competitor does not behave as a purely competitive market things become more complicated, and in this case what is the most lucrative part of the market for the entrant and for the incumbent remain open questions. The answer to them depends crucially on the respective marginal costs, the game's structure, the strategy of competition chosen by the entrant, her cost and scale of entry. Here, what is more lucrative for the entrant will determine which type of customer she will choose to serve, if she is able to attack both part of the market. Thus, we will show how cream skimming is no longer necessarily the most profitable strategy for a potential entrant (see 2.4.2). We also explore the question regarding the desirability of cream skimming, showing that this type of competition, contrary to the general point of view, is not necessarily more "harmful" than skimmed milk competition.

Within this chapter we will not deal with regulation, since we just limit ourselves to set up the basis of the model and solve it in the unregulated case. In order to develop in the next chapter a regulatory research topic (which will be fully specified in chapter 3) we try to extend the model proposed by Laffont and Tirole (1990a, 1994) introducing non-linear pricing as in Laffont and Tirole (1990b) and tackling the problem of entry as a crucial one. In Laffont and Tirole (1990a), a more complicated model, where vertically related markets are introduced, the price competitors set is just equal to the sum of the marginal cost and the access price. Furthermore, both competitors and the regulated firm make use only of linear pricing also for the intermediate good. Hence, it may be interesting to introduce the possibility of price discrimination. Since this pricing strategy is a more general one, always at least as good as the linear pricing strategy, it is not clear why it should be rejected a priori by the firms (and in particular the incumbent) and by the regulatory authority. That is, they are dealing with a regulation which is optimal only in a framework in which non-linear tariffs are prevented on an "a priori" ground, or due to an exogenous political constraint which has not been justified.

We will study a simpler model than the one proposed by Laffont and Tirole, in order to pay a particular attention to the entrant's behaviour and to the way of modelling it within a model in which non-linear pricing strategies are used by both the incumbent and the entrant. Specifically we decided not to introduce asymmetric information on the regulatory side. As shown in chapter 1 (section 1.2) moral hazard if introduced in the standard way (a la Laffont and Tirole) does not bring substantial changes in the outcomes relative to a pure adverse selection model. We also leave aside adverse selection between the regulator and the firm; this assumption is quite usual in the case of a *public firm* which directly maximises social welfare. Moreover we can also ignore asymmetric information on the demand side, since in a model where the marginal cost of the incumbent is fixed things would not changed, as proved by Lewis and Sappington (1989).

The only type of asymmetric information is on the consumers' side. Technically, we will deal with a special class of games of incomplete information, classified as games of *mechanism design*.[33] In the canonical situation a monopolist (the principal) has incomplete information about the agent's types, namely the consumers' willingness to pay for his goods. He has to design a tariff schedule which determines the price to be paid as a function of the quantity purchased.[34] We extend the canonical framework to the case of competition, the competitor being a small output constrained entrant, the idea being that competition should start somewhere.

---

[33] See for instance Fudenberg and Tirole (1991).

[34] A similar problem is the regulation of a natural monopoly. Here the regulator has incomplete information of the cost (generally only of the cost level, rather than its structure). His task is to design an incentive scheme, which related the transfers received by the regulated firm to its cost or price, or both of them, depending on the type of regulation considered.

If we want to correctly specify the structure of this type of game we should follow a three step procedure. In the first step the principal designs the mechanism, the mechanism itself being a game in which the agents send costless messages. It should then be clear that the subsequent allocation between types (that is, the decision on the quantity purchased) depends on the realisation of these messages. In the second step the agents simultaneously accept or reject the mechanism. If the agent rejects the mechanism he gets a reservation utility. In the final step, in which only the agents who accepted the mechanism can participate, they play the game as specified by the mechanism itself. However, the *revelation principle* allows us to restrict attention to mechanisms that are accepted by the agents (bypassing stage 2) and in which the agents' types are truthfully revealed (so that there is no need to specify the third stage).[35]

A basic difference between our approach and the one proposed by Laffont and Tirole is that here the *incumbent* is the *first mover* with respect to pricing policies, i.e. the first player who chooses the tariff for which he is committed to serve the customers. We think that this resembles more closely the typical situation in utilities industries, where the incumbent is more powerful and enjoys the first mover advantage in many aspects.

Given these premises, in the absence of vertical issues, our approach may also be interpreted as a transposition of Armstrong and Vickers (1993) to a second degree price discrimination model. However, the presence of non-linear pricing and of vertical relations remarkably modifies and complicates their basic model. For instance, we need to consider very carefully all incentive-compatibility problems (absent in the context of third degree price discrimination) and, as a direct consequence, the structure of the game itself must be modified.

[35] The perfection requirement involved here is simply that the agents have an incentive not to misreport their types, being not allowed to threaten to reject the mechanism. With this requirement the highest payoff of the principal can be obtained through a static game among the agent, so that mechanism design can be treated as a special case of static Bayesian game.

As the presence of non-linear pricing complicates the structure of the game, introducing incentive compatibility problems, when we consider vertical issues, we will slightly simplify the model proposed by Laffont and Tirole (1990a, 1994a), eliminating, for instance, the product differentiation issue, in order to derive clear-cut results and to pay a particular attention to the problems related to the entrant's behaviour and non-linear pricing. We will study the simplest framework, i.e. the one in which both the incumbent and his competitors have the same technical requirements for the intermediate good; i.e. there is no need to create new infrastructure or facilities to connect the competitors with the existing network. We can refer to this situation as the "common network case", following the definition by Laffont and Tirole.

Whilst in the basic model (ignoring vertical issues) we argued that cream skimming is not always the most profitable strategy for the entrant, except for very particular cases, this is no longer true for the common network case. In fact, when the incumbent remains the monopolist of an intermediate good which is consumed both internally and by any potential competitors *cream skimming* turns out to be the only strategy of competition allowed by the incumbent. We will see how, generally in the absence of regulation, the access charge determined by the incumbent may depend, apart from the type of network cost, on the entrant's production cost, the vertical game's structure, the strategy of competition chosen by the entrant and her scale of entry.

We will also show how the entry of an equally efficient competitor in the market of the non monopolised good 1, represents a limiting case as the incumbent is indifferent about entry. In fact, setting the per customer access charge equal to the monopoly variable profits, he maintains the monopoly pricing strategy and the related profits, *independently* of the entry scale. It is optimal for the incumbent to allow entry, if, maintaining the previous monopoly pricing, he can set a per customer access charge equal to the variable profits of the entrant and if the latters are equal or greater than his own. This happens only with an equally or more efficient competitor. Moreover, the incumbent finds it optimal to oblige the competitor to behave as a surplus taker (i.e. setting a new tariff which allows the same surplus determined by the incumbent for each type) and to allow only cream skimming competition. However, it would be preferable for the incumbent (and for the maximisation of social welfare) that the competitor sells all the produced good 1 to the incumbent (reaching only break-even profits), so that the latter is able to resell it to the consumers, applying the monopoly tariffs.

The chapter is organised as follows. Before providing the details and better specifying the hypotheses within which we will move both in the absence and in the presence of vertically related markets, we will first discuss the variety of approaches available to model the entrant's behaviour (section 2.2). The following sections provide the solution of the game in the absence of regulation first of all ignoring vertical markets. They focus on the private incentives of the incumbent (section 2.3) and the entrant (section 2.4) in the absence of regulatory constraints. We then extend the previous model in order to deal with vertical markets, providing the solution of the game in the absence of regulation first for the case in which the entrant is as efficient as the incumbent, and then in section 2.5 for the general case. Section 2.6 provides further refinements to the vertical game, endogenising the scale of entry (previously exogenously given) and considering more general cost functional forms for the entrant. A final section (2.7) summarises the main results of our analysis.

2.2 Modelling approaches

In our analysis we proceed step by step, disregarding at first the relevant issue of access pricing, limiting the analysis to an horizontal framework where only one final good is produced by the incumbent and the entrant. We then extend this framework by assuming that the incumbent has a natural monopoly on the network and introducing an intermediate good (access).

Even in the absence of vertically related markets there are several possible schemes, and correspondingly several games, that can be used to model the behaviour of the entrant and it is difficult (and also in part arbitrary) to choose between them, since many are quite plausible and seem to capture certain particular features of what happens in some public utilities industries.

For instance, we may assume that in stage (1) the entrant chooses her scale of entry in terms of output capacity (X) or in terms of the number of customers to be served (K). Furthermore, assuming that there are only two customers' types (as in Laffont and Tirole), the scale of entry can be fixed in each of the two markets (X_t, K_t, where t = L, H) to be served, or in global terms (X, K), postponing the determination of the split between the two markets to a subsequent stage. In this way, even if we stick to the simplest possible framework (i.e. we assume a *tariff taker* competitor, that is a competitor that takes as given the optimal pricing strategy fixed by the incumbent), we can obtain at least four possible games, each of which deserves a different and detailed analysis on its own. The number of games increases once we take into account a surplus taker competitor, who designes a new tariff, which allows each consumer type the same surplus fixed by the incumbent.

2.2.1 Our basic game and a possible alternative one

In this subsection we introduce two main categories of modelling approaches, which may be interesting to analyse, whose structure is here summarised.

Horizontal Game

(1) the entrant decides the number of customers $\{K_t\}$ to be served in each market;

(2) the incumbent chooses his pricing rule $\{T_t, q_t\}$, where $T_t = T(q_t)$ is a fully non-linear tariff;

(3) the entrant takes as given $\{T_t, q_t\}$ acting as a tariff-taker.

In this first type of model the entrant commits herself to the number of customers to be served in each market before the incumbent chooses his pricing rule (naturally we are referring only to a fully non-linear tariff). We believe that this approach might be used for example as a starting model, useful to describe telecommunication industries (where the crucial variable of competition is the number and the type of customers to be served) e.g. the case of "British Telecom" and "Mercury", especially if we focus on competition for large users.

Capacity Game

(1) the entrant determines her total output capacity X;

(2) the incumbent chooses his pricing rule $\{T_t, q_t\}$;

(3) the entrant decides how to allocate her output between the customers' types $\{X_t\}$, taking as given the incumbent's tariff structure.

In the second type of model, where the competitor decides her scale of entry in terms of output capacity, it makes much more sense to assume that she will allocate his output between the customers' types only after the incumbent has chosen his pricing rule.[36]

In what follows, a detailed analysis is provided only for the first type of game, which seems the most interesting and simple. In this type of game the competitor seems relatively more powerful since she commits herself to serve a given number of customers of given types, without taking into account the tariffs chosen by the incumbent. In other words, the crucial variables for the entrant $\{K_L, K_H\}$ are determined independently of the pricing policy chosen by the incumbent, so that there is less interaction between the two players.

The second approach, instead, allows for strategic interactions between the two players. The presence of strategic interactions crucially depends on the structure of the game, and in particular on the fact that the entrant determines the split of his total capacity between the two customers' types (X_t), only after the incumbent has fixed his pricing strategy. Therefore the incumbent, in determining the tariffs to apply to each type of customer, takes into account not only the value of the scale of entry already fixed in the previous stage, but also its expected split between the two types. In practice in this case, as the entrant serves the market with the highest average

---

[36] This second approach has emerged in a discussion with John Vickers. This particular approach seems to capture some of the typical situations that may arise in some of the gas and electricity industries.

price, we may expect that the result crucially depends on the entry scale and the numerical dimension of the types.

Focusing on *the decision of the type of customer to be served by the entrant* and making reference to our basic horizontal game, as we will see in what follows, the particular value assumed by the marginal cost of the incumbent turns out to be the crucial variable. Naturally, also the value assumed by the cost of entry affects the choice of the customers' type (since it matters in the preliminary decision taken by the competitor; that is, whether or not to enter the market). Instead, in the alternative capacity game the entrant chooses which type to serve only on the basis of the highest average tariff, which she is allowed to charge, given the tariff fixed by the incumbent.

Despite the several differences between the two approaches, it is worth noticing how in both of them the incumbent, who is undoubtedly the most powerful player, is the first mover, i.e. the first to fix the tariffs, and consequently, the quantities allocated to each type of customers. This is what really differentiates these approaches from the one proposed by Laffont and Tirole (1990b), where, instead, in reality it is the entrant who determines the pricing policy. The incumbent play the role of a surplus taker competitor, as the analysis of section 1.3 has shown.

2.2.2 Modelling approach to non-linear access pricing

If we want to model access pricing and vertical issues, the situation becomes even more complicated. For instance, we may assume that the interconnection costs are functions of output capacity or of the number of customers to be served or, more generally, that they are functions of both the variables. However, following this more general option, we have finally sorted out a modelling approach, which is built upon the basic horizontal game in the presence of vertical relationships. The structure of this multistage game is here summarised.

Vertical Game

(0) the authority sets up the regulatory system and decides the policy toward price discrimination;

(1) the incumbent fixes the access pricing function $F(K_H, K_L, Q^e)$, where K_t denotes the number of customers served by the entrant in each market and Q^e represents the total output produced by the entrant;

(2) the entrant decides her scale of entry in terms of the number of customers $\{K_t\}$ to be served in each market;

(3) the incumbent chooses his pricing rule $\{T_t, q_t\}$, where $T_t = T(q_t)$ is a fully non-linear tariff;

(4) the entrant chooses the strategy of pricing competition in the final good market.

Naturally, we may have different versions of this game, depending on the nature of the network cost function and the pricing strategy chosen by the entrant. On the whole, we believe that this approach might be quite appropriate to describe the case of telecommunication industries in the UK.

2.2.3 The main assumptions of the proposed games

Let us start introducing the following simplifying hypotheses within which we will initially move in our analysis of the horizontal and vertical games. We will consider only two final goods markets. The total output in the final goods' sector can be decomposed in Q^0, the output of a monopolised good, and Q^1 which is the one of a non-monopolised good, hereafter denoted also as good 1. In the horizontal game the total output is simply Q^1.

On the **consumers' side**, for simplicity's sake, we restrict the analysis to two types of customers: high and low-demand customers ($t = H, L$).

(i) each type is present in the same number (denoted by N);

(ii) as in Laffont and Tirole, for simplicity's sake, we will assume that the monopolised good is sold by the incumbent at the linear price p^0. We can in fact, for instance suppose that both the customers of type L and H receive the same utility $v(q^0)$ from consuming a unit of the monopolised good.

For the non-monopolised good the *gross surplus function* of type t is assumed to be a function of the relative output per unit of customer (q_t) and a taste parameter (θ_t)

which captures the willingness to pay for the bundle q_t.

To simplify notation, let θ_L=1 and consider the following simple functional form, such that for the same quantity, the high-demand utility becomes a fixed multiple of the low-demand's one, with $\theta > 1$:

[2.1] $$u_H = \theta\, u(q_H)$$

[2.2] $$u_L = u(q_L)$$

with $u'(q_t)>0$ and $u''(q_t)<0$. In what follows we examine in more detail the *quadratic utility function* case $[u_t = q_t - q_t^2/2]$ which has a nice geometrical interpretation.

On the **industry's side** both the incumbent and the entrant are allowed to make use of *fully non-linear tariff*:

$$T_t = T(q_t); \qquad T_t^e = T(q_t^e)$$

where T_t is the amount paid to the firm by the customer of type t.

(iii) In the absence of entry the incumbent acts as a monopolist and is characterised by the following *cost* and *revenue functions*:

$$C(Q) = c^* Q$$

$$R(Q, Q^0) = (T_L + T_H)N + 2N\, p^0 q^0$$

where c* is the incumbent's constant marginal cost, Q and Q^0 respectively denote his total output for the non-monopolised and monopolised good. For simplicity's sake good 0 has only network costs. Naturally, in the horizontal game being Q^0 equal to zero, the revenue function is simply R(Q). We are assuming that the incumbent's fixed costs are already sunk.

(iv) the competitor has a *fixed limited scale of entry* in terms of number of customers to be served:

$$K = K_L + K_H < N$$

Loosely speaking, she can't pre-empt either of the two markets. For simplicity's sake, we will keep this assumption in all the stages of the game. We assume that she must pay an access charge F fixed by the incumbent (which can be considered as a cost of entry) which may depend in general both on the scale of entry K and on her total output Q^e:

$$F(K_L, K_H, Q^e) \geq 0$$

Naturally in the horizontal game F is equal to zero.

Finally, she has a linear variable cost function:

$$CV^e = m\,Q^e$$

where m denotes the value of her marginal cost. Usually, to simplify matters we assume that the total cost function and the variable cost function do coincide, $C^e = CV^e$.

Notice how the competitor will enter the non monopolised good market if the usual participation constraint is satisfied:

$$[PC^e] \qquad \Pi^e(Q^e) = R^e - C^e = K_L(T_L^e - mq_L^e) + K_H(T_H^e - mq_H^e) - F(K_L, K_H, Q^e) \geq 0$$

We do not need to specify an incentive compatibility constraint, as we are dealing with a single competitor with complete information.

(v) the production cost of the two final goods depends on the network subcost function $\mathcal{NC}$. In general the latter may depend on the number of customers to be served (2N) and on the total quantity of commodities which flows through the network, $Q^1 = Q + Q^e$:

$$\mathcal{NC}(2N, Q^0, Q^1) = \mathcal{NC}(2N) + c^0Q^0 + c^1Q^1$$

To simplify matters, for some of the analysis, we will assume c^1 equal to zero.

In the presence of entry the incumbent is characterised by the following *cost* and *revenue functions*:

$$C(Q) = c^*\,Q = c^*\,(N_L\,q_L + N_H\,q_H)$$

$$R(Q, Q^0) = N_LT_L + N_H\,T_H + 2N\,p^0q^0$$

where N_t=N-K_t denote the residual number of customers of type t served by the incumbent, c^* is the incumbent's constant marginal cost and $Q = N_Lq_L + N_Hq_H$ denotes the total output. Notice how we are assuming that the incumbent's fixed cost depends on the total number of customers 2N.

For clarity of exposition, here is sketched the overall structure of the game in the absence of vertical markets:

(0) the authority sets up the regulatory system and decides the policy towards price discrimination;

(1) the entrant decides the number of customers $\{K_t\}$ to be served in each market;

(2) the incumbent chooses his pricing rule $\{T_t, q_t\}$, where T_t is a fully non-linear tariff;

(3) the entrant chooses the strategy of competition.

In the next sections we will solve the game, first in the absence of vertical issues and also making some simplifying assumption regarding the strategy of competition (stage 3), that will be later relaxed. Specifically assuming that the entrant is a *tariff taker* (i.e. she makes use of the same pricing rule of the incumbent) we can directly move to the second stage, determining the optimal pricing strategy of the incumbent. In the vertical case, we will solve the game, starting from the simple case of an equally efficient competitor (i.e. m=c*).

2.3 The incumbent's problem

2.3.1 The basic problem in the absence of vertical issues

We can directly tackle the incumbent's problem allowing for **fixed levels of entry**, by considering the decision of the incumbent in *residual terms*. Let N_t=N-K_t denote the residual number of customers of type t served by the incumbent. The solution of the problem in the absence of entry can be directly determined by imposing N_L=N_H=N. We are now dealing with stage 2, in order to tackle the incumbent's optimisation problem.

In the **absence of entry** (i.e. when $Q^1 = Q$) the incumbent acts as a monopolist; that is, he maximises his profit function $\Pi(Q)$ with respect to the tariff system $\{T_L$, q_L, T_H and $q_H\}$ subject to the individual rationality and incentive compatibility constraints:

$\max \Pi(Q) \equiv \; R(Q) - c^*Q$ subject to:

$[IR_L] \quad u(q_L) - T_L \geq 0$

$[IR_H] \quad \theta\, u(q_H) - T_H \geq 0$

$[IC_L] \quad u(q_L) - T_L \geq u(q_H) - T_H$

$[IC_H] \quad \theta\, u(q_H) - T_H \geq \theta\, u(q_L) - T_L$

and $\quad Q = N_L\, q_L + N_H\, q_H$

The first two constraints $[IR_L]$ and $[IR_H]$ represent the participation constraints for the two types of customers, when we assume a reservation price equal to zero, whereas $[IC_L]$ and $[IC_H]$ specify the incentive compatibility constraints. In particular, the upward binding incentive constraint $[IC_L]$ prevents the low-demand type from consuming the high-demand bundle (q_H), while the downward binding incentive constraint $[IC_H]$ prevents the high-demand consumer from mimicking the low-demand customer.

In this model as in Laffont and Tirole we assume that the incumbent serves both types of customers. We can simply say that this derives from an universal service obligation, but we will see how for him it is profitable to serve both types for a certain range of values of his marginal cost c and the taste parameter θ. If both types are served by the incumbent it is easy to demonstrate that only $[IR_L]$ and $[IC_H]$ are binding:

$[IR_L] \quad T_L = u(q_L)$

$[IC_H] \quad T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$

These expressions can be interpreted as follows: no surplus is allowed to the L type, whereas the H type enjoys a positive net surplus [given by $(\theta - 1)\, u(q_L)$].

In order to show that the only binding constraint are $[IR_L]$ and $[IC_H]$ we will proceed as follows. First of all, we will prove that $[IR_H]$ will be automatically satisfied, once the previous constraints are binding. From $[IC_H]$ we know that:

$$\theta\, u(q_H) \geq T_H + \theta\, u(q_L) - T_L$$

Making use of [IR_L], which tells us that T_L is less or equal to $u(q_L)$, keeping in mind that θ is greater than unity, we can rewrite the previous inequality as:

$$\theta\, u(q_H) \geq T_H + (\theta - 1)\, u(q_L) \geq T_H$$

This is exactly what we want to prove; i.e. [IR_H] is fulfilled. The *intuition* behind this result is that the H type can always mimic the L type, at a lower cost, since his willingness to pay for the same bundle is higher ($\theta>1$).

The last step of the procedure involve showing that [IC_L] is satisfied ex post by the solutions of our model. We are going to verify this at the end of the optimisation problem, solved below, applying a two stage procedure.

Since the incumbent's cost function depends only on the total level of output Q (and not separately on q_H and q_L), we can split up the maximisation problem in **two stages**. The first stage involves the determination of the *optimal revenue function* for all levels of the total output. The solution of this stage gives the optimal relationship between the shares of output allocated by the incumbent for each customer of type L and H, whereas only in the final stage we are able to determine the *optimal level of output* by solving the profit maximisation problem.

The two stages maximisation problem becomes:

STAGE A: Determination of the optimal R(Q) for all Q

max $R(Q) \equiv$ $T_L N_L + T_H N_H$ subject to:

[IR_L] $T_L = u(q_L)$

[IC_H] $T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$

and $Q = N_L q_L + N_H q_H$

The previous problem can be solved only with respect to q_L and q_H, once we substitute the two binding constraints [IR_L] and [IC_H] into the objective function:

max $R(Q) \equiv$ $[N_L - N_H(\theta - 1)]\, u(q_L) + N_H\, \theta\, u(q_H)$ subject to:

$$Q = N_L q_L + N_H q_H$$

We can then write down the Lagrangean function, which is:

$$\mathcal{L} = R(Q) - \mu\,[Q - (N_L\, q_L + N_H\, q_H)]$$

The first order conditions are given by:

$[q_L]$ $\quad [N_L - N_H(\theta - 1)]\, u'(q_L) - \mu\, N_L = 0$

$[q_H]$ $\quad N_H\, [\theta\, u'(q_H) - \mu] = 0$

Substituting the value of the Lagrangean multiplier $\mu = \theta u'(q_H)$ into the first order condition with respect to q_L we get the optimal relationship between q_L and q_H:

[2.3] $\quad \theta\, u'(q_H) = [1 - (\theta - 1)\, N_H/N_L]\, u'(q_L)$

The optimal relationship between the two marginal prices is very easily obtained from **[2.3]**, making use of the optimality conditions which hold for the consumers, i.e. $p_t = \theta_t\, u'(q_t)$:

[2.4] $\quad p_H = [1 - (\theta - 1)\, N_H/N_L]\, p_L$

STAGE B: Determination of the optimal level of output Q

max $\Pi(Q^i) \equiv$ $\quad R(Q) - c^*\, Q$

Since we already know, from the solution of the previous stage, the optimal revenue function, we can get the solution of the profit maximisation simply by imposing the equality between marginal revenue and marginal cost [i.e. $dR/dQ = c^*$]. Substituting in **[2.4]** the value of the marginal revenue [$dR/dQ = p_H = \theta u'(q_H)$], we immediately get the expression for the two marginal prices:

[2.5] $\quad p_L = u'(q_L) = c^* / [1 - (\theta - 1)\, N_H/N_L] > c^*$

[2.6] $\quad p_H = \theta\, u'(q_H) = c^*$

Equations **[2.5]** and **[2.6]** show how the low-demand marginal price p_L is always greater than p_H. Since the inverse demand function is a decreasing function of the price, the sign of the inequality in terms of quantities is clearly reversed, so that the quantity consumed per unit of customer q_H is always greater than q_L. Hence, we can easily verify that in equilibrium [IC_L] is satisfied:

$$T_H - T_L = \theta\, [u(q_H) - u(q_L)] \geq u(q_H) - u(q_L)$$

This concludes our proof that the only binding constraints are [IR_L] and [IC_H].

Let us now explore the admissible values of the model standardising, for simplicity's sake, the utility function [i.e. $u'(q_L)=1$ for $q_L=0$]. To be sure that the incumbent will find it profitable to serve both markets when non-linear pricing is allowed we require p_L to be less than unity; that is, making use of **[2.5]**:

$$p_L = c^* / [1 - (\theta - 1)N_H/N_L] < 1$$

which reduces to:

$$c^* < 1 - (\theta - 1)N_H/N_L$$

We must also check that this condition will hold when the ratio N_H/N_L reaches its maximum (that is, for the skimmed milk case which occurs when $K_H=0$, the ratio N_H/N_L being a decreasing function of K_H). Corresponding to K_H equal to zero N_H/N_L is equal to $1/(1\text{-}s)$, where $s = K/N$ denotes the proportional scale of entry:

$$c^* < 1 - (\theta - 1)/(1 - s) = \hat{c}^*(\theta, s)$$

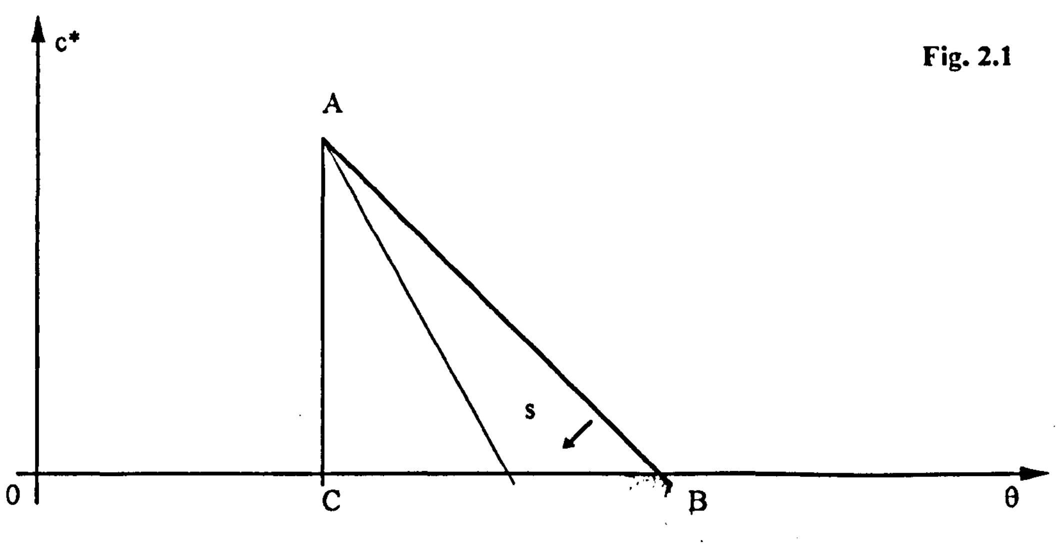
This means that the incumbent must be efficient enough; i.e. his marginal cost has an upper bound which is a function of θ and s, denoted by $\hat{c}^*(\theta, s)$. This value must be positive (hence, economically meaningful):

$$\hat{c}^*(\theta, s) = 1 - (\theta - 1)/ (1 - s) > 0$$

which is equivalent to say:

$$\theta < 2 - s$$

Fig. 2.1



In the space (θ, c) $\hat{c}^*(\theta, 0) = 2 - \theta$ is just a straight line which passes through the points A = (1, 1) and B = (2, 0). Graphically, the admissible values of c* and θ are confined to the region inside the triangle ABC [where C = (1,0)], as shown in fig. 2.1 above. This hinges on the fact that $\hat{c}^*(\theta, s)$ is a decreasing function of s; i.e. when s decreases the function $\hat{c}^*(\theta, s)$ simply rotates clockwise around the point A.

Let us pause now to interpret the results of the optimisation problem. It directly follows from [2.6] that the marginal price applied to the H type equals the marginal cost, so that there is **no distortion at the top**. The conclusion is that the *standard result on price discrimination* for pure monopoly can be directly applied allowing for exogenous scale of entry (K_L and K_H), once we take into account only residual customers (N_H and N_L).

This shows how pervasive is this result, which holds in several different frameworks. Mirrlees (1971) was the first to derive it in the ambit of optimal taxation, Mussa and Rosen (1979) and Spence (1980) confirm its validity respectively for a multi-quality and a multi-product monopolist who price discriminates. This result deserves a brief digression for its importance, in order to better specify its meaning and relevance.

The necessary and sufficient conditions in order to have "no distortion at the top" in a very stylised principal agent setting (where a monopolist is serving a population of customers but is unable to distinguish each type) are:

1) the existence of a "top customer". This condition will generally imply an upper bound on the value of the taste parameter. Here for simplicity's sake we consider only the two types case, that is we assume that the taste parameter takes only two values.
2) the relevant incentive compatibility constraint [IC_H] being binding, while the other [IC_L] being ex post satisfied by the solution, in the case of two types.

The intuition behind this result is easily explained. The monopolist would like to extract the high demand surplus; however, doing this he faces the threat of personal arbitrage by high demand customers. In fact each high demand customer can consume the low demand bundle if the latter generates more surplus (compared with his own bundle). The reduction of the quantity offered to the low demand consumers is needed in order to relax the arbitrage constraint. In fact, due to a technical sorting condition high demand customers suffer more from a reduction of the quantity than low demand customers do.[37] Very loosely, the monopolist, by reducing the quantity of low demand customers will reduce the "temptation" of high demand customers to exercise arbitrage (that is, to mimic low demand customers).

In what follows we are going to explore some interesting properties which can be derived from the previous analysis. Specifically, we focus on the determination of the optimal reply, in terms of tariff and bundle allocation, of the incumbent relative to the scale of entry of and the desirability to apply quantity discount in the presence of competition.

The optimal tariffs of the incumbent as a function of the scale of entry

It is interesting to analyse how the incumbent's tariffs and quantities vary with respect to the residual number of customers served by the incumbent (we will make reference to the ratio N_H/N_L) in the horizontal game. In particular, we will show that:

(a) q_L is a decreasing function of the ratio N_H/N_L;

(b) q_H is independent of the ratio N_H/N_L;

(c) $T_L = u(q_L)$ is a decreasing function of the ratio N_H/N_L;

(d) $T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$ is an increasing function of the ratio N_H/N_L.

Since it is straightforward to prove these results in what follow we just sketch the main arguments.

Result (a): From [2.5] the value of p_L is clearly an increasing function of the ratio N_H/N_L, since this term appears in the denominator with a negative coefficient (by assumption $\theta>1$). But q_L is a decreasing function of p_L (this follows directly from the fact that the utility function is concave), so that it must be a decreasing function of the ratio N_H/N_L.

[37] In the literature on incentives this condition is known as the "single crossing condition" or the "Spence-Mirrlees condition".

Result (b): From **[2.6]** the value of p_H is independent of the ratio N_H/N_L and the same is true of q_H. It is worth noticing that this result crucially hinges on the fact that we are assuming constant marginal cost for the incumbent.

Result (c): This result is just a corollary of (a), since we know that $T_L = u(q_L)$, where u is an increasing function of q_L.

Result (d): This result follows from (b) and (c). Since $T_H = \theta\, u(q_H)-(\theta-1)u(q_L)$, using (b) the first term is independent of the ratio N_H/N_L, whereas from (c) the latter term turns out to be an increasing function of q_L.

These result are very relevant if we want to analyse which kind of competition the incumbent would prefer to face: namely competition for the high or low type. We are going to make use of these properties later, after solving the entrant's problem (see section 2.4.3, where we explore the desirability of cream skimming in terms of productive and allocative efficiency).

The quantity discount result

Finally, we might want to check whether it is optimal for the incumbent to offer **quantity discounts** (i.e. in analytical term we need to verify that T_L/q_L is greater than T_H/q_H). A proof of this result is provided below for simplicity's sake in the quadratic utility functions case.

Let us then verify that T_L/q_L is greater T_H/q_H for quadratic utility functions. To simplify the comparison between T_L/q_L and T_H/q_H let us decompose the latter as the weighted average between T_L/q_L and T_a/q_a where $T_a/q_a = (T_H - T_L) / (q_H - q_L)$:

$$T_H/q_H = \{q_L/q_H\}\ T_L/q_L + \{(q_H - q_L)/q_H\}\ [T_H - T_L] / (q_H - q_L)$$

This means that in order to prove that $T_L/q_L > T_H/q_H$ we can limit ourselves to show that $T_L/q_L > T_a/q_a$. For quadratic utility functions T_L/q_L is simply the average between 1 and $u'(q_L)=p_L$, as it can be easily verified:

$$T_L/q_L = [q_L-(q_L)^2/2]/q_L = 1 - q_L/2 = [1+(1 - q_L)]/2 = [1+u'(q_L)]/2$$

Similarly T_a/q_a is the average between $p_H=c^*$ and $\theta u'(q_L)=\theta p_L$:

$$T_a/q_a = (T_H - T_L)/(q_H - q_L) = \theta[(q_H - q_L) - 1/2(q_H - q_L)(q_H + q_L)]/(q_H - q_L)$$
$$= \theta\,[(1-q_H) + (1-q_L)]/2 = (p_H + \theta p_L)/2$$

Substituting the value of p_H from [2.5] into the previous equations:

$$T_L/q_L = \{1 + c^*/[1-(\theta-1)N_H/N_L]\}/2 = \{1-(\theta-1)(N_H/N_L)+c^*\}/2[1-(\theta-1)N_H/N_L]$$

$$T_a/q_a = \{c^* + \theta c^*/[1-(\theta-1)N_H/N_L]\}/2 = c^*[(\theta+1)-(\theta-1)N_H/N_L]/2[1-(\theta-1)N_H/N_L]$$

We are now able to derive the condition which makes $T_L/q_L > T_a/q_a$ being fulfilled:

$$[1 - (\theta - 1)N_H/N_L + c^*] > c^*[(\theta + 1) - (\theta - 1)N_H/N_L]$$

Subtracting 2c* from both sides:

$$[1 - (\theta - 1)N_H/N_L - c^*] > c^*[(\theta - 1)(1 - N_H/N_L)]$$

Notice how the left hand side is always positive, that is $[1 - (\theta - 1)N_H/N_L - c^*] > 0$, since the marginal price p_L and the marginal cost c* are less than unity, both inequalities being required in order to ensure that the incumbent start off by serving both types of customers (see fig. 2.1). Let then examine the sign of the right hand side of the equation. For $N_H > N_L$ (that is, in the case of skimmed milk competition) the right hand side is negative; so that the required condition is always fulfilled. For $N_H < N_L$, that is in the cream skimming case, although the right hand side is now positive, the condition still holds. It is easy to check that if we reduce by the amount of z the ratio N_H/N_L the left hand side is increased by a factor given by $z(\theta - 1)$, while the right hand side rises only by $z\,c^*(\theta - 1)$ (where c* is less than unity).

This completes the proof that the incumbent finds it optimal to apply quantity discount in the presence of competition. In the next subsection we extend the game to vertically related market where the incumbent has a natural monopoly on the network.

2.3.2 Extension to the vertical case

Following the same procedure applied before, we first tackle the incumbent's optimisation problem. However, we first solve the game in the **absence of entry** (i.e. when $N = N_L = N_H$). In this case the incumbent acts as a monopolist; that is, he maximises his profit function $\Pi(Q, Q^0, Q^1)$ -now inclusive of the revenue and costs of the monopolised good and network costs NC- with respect to the tariff system $\{T_L, q_L, T_H$ and $q_H\}$ subject to the following constraints:

[Problem 1] $\max \Pi(Q, Q^0, Q^1) \equiv R(Q^0, Q^1) - NC(2N, Q^0, Q^1) - C(Q)$ subject to:

[MME] $v'(q^0) - p^0 \leq 0$

[IR_L] $u(q_L) - T_L \geq 0$

[IR_H] $\theta\, u(q_H) - T_H \geq 0$

[IC_L] $u(q_L) - T_L \geq u(q_H) - T_H$

[IC_H] $\theta\, u(q_H) - T_H \geq \theta\, u(q_L) - T_L$

The first constraint [MME] simply implies the monopolised market equilibrium on the consumers' side, imposing the price of the monopolised good to be less or equal to the marginal utility enjoyed by its consumption. As previously said, both customers of type L and H receive the same utility from consuming a unit of the monopolised good. The two following constraints [IR_L] [IR_H] are the usual participation constraints (with reservation prices equal to zero) for the two types of customers, whereas [IC_L] and [IC_H] represent the incentive compatibility constraints. As in the basic model, the upward binding incentive constraint [IC_L] prevents the low-demand type from consuming the high-demand bundle (q_H), while the downward binding incentive constraint [IC_H] prevents the high-demand consumer from mimicking the low-demand customer. As we have shown before the upward binding constraint [IC_L] and the participation constraint [IR_H] are automatically satisfied by the solution of the problem when [IR_L] and [IC_H] are binding (the same proof being provided above holds in this case). In practice, no surplus is allowed to the L type, whereas the H type enjoys a positive net surplus [given by $(\theta\text{-}1)\, u(q_L)$].

Notice how in the vertical game, apart from the additional constraint derived for the consumers in the intermediate demand market, the same constraints referred to the consumers' in the final demand market in the horizontal game are binding.

Once we substitute all the functional forms of the revenue and cost functions, as previously defined, **[Problem 1]** becomes:

[Problem 2] $\max \Pi(Q^i) \equiv 2Np^0q^0 + N(T_L+T_H) - NC(2N) - c^*Q^i$ subject to:

[MME] $p^0 = v'(q^0)$

[IR_L] $T_L = u(q_L)$

[IC_H] $T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$

This problem, hereafter called **[Problem 2]**, can be easily solved only with respect to q^0, q_L and q_H, once we substitute the three binding constraints into the objective function:

$$\max \Pi(Q^i) \equiv 2Nq^0v'(q^0)+N(2-\theta)u(q_L)+N\theta u(q_H) - NC(2N) - c^0(2Nq^0) - Nc^*(q_L+q_H)$$

Making use of the optimality conditions which hold for the consumers for the non monopolised good, i.e. $p_t = \theta_t\, u'(q_t)$ we can write the first order conditions as:

[q^0] $c^0 = v'(q^0)+ v''(q^0)\, q^0$

[q_L] $p_H = \theta\, u'(q_H) = c^*$

[q_H] $p_L = u'(q_L) = c^* / (2-\theta)$

The first equation simply states the equality between the marginal cost and the marginal revenue for the monopolised good. The following ones imply that there is **no distortion at the top** and **some distortion at the bottom**.

A comparison with the first order condition in the horizontal game (naturally for N_H/N_L equal to unity, since we are dealing with a monopolist) shows how the incumbent finds it optimal to fix in the final demand market the same tariffs as in the horizontal game.

In particular, from the two latter first order conditions we obtain the optimal relationship between the two marginal prices p_L and p_H:

[2.10] $p_H = (2 - \theta)\, p_L$

and between q_L and q_H:

[2.11] $\theta\, u'(q_H) = (2 - \theta)\, u'(q_L)$

What is more surprising, as will be shown later in section 2.5, is that the optimal relationship between the marginal prices **[2.11]** still hold in the presence of competition. That is, differently from the vertical game the incumbent can act as a monopolist, fixing his tariffs independently of the scale of entry. However, before tackling the entrant's problem let us provide a graphical representation of the incumbent's problem.

2.3.3 A graphical exposition

In order to clarify the exposition of the model, it may be of some interest to give a representation of our game in the monopoly case of *quadratic utility function* $u(q)=q - (q)^2/2$. In fig. 2.2 the linear marginal utility functions $u'(q)$ and $\theta u'(q)$ are shown, as well as the optimal discriminatory marginal prices $p_H=OG=c^*$, $p_L=OF>c^*$.

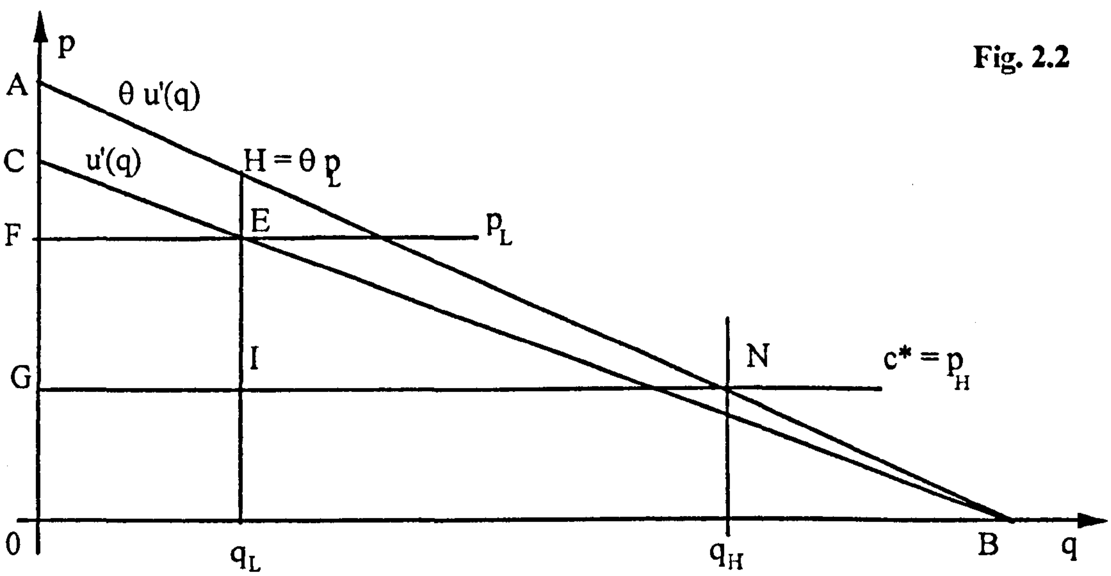


Fig. 2.2

It is immediate to derive graphically the quantities allocated per unit of customer (which are obtained in correspondence of the intersection between the curves representing the marginal utility and marginal price for each type) and to verify that q_H is greater than q_L (since q is a decreasing function of p, the sign of the inequality is clearly reversed). The tariffs T_L and T_H give a measure of the gross revenue per unit of customer. The tariff T_L can be measured by the integral from 0 to q_L of the area which lies below the L type marginal utility function $u'(q_L)$.

In the linear case T_L is simply given by the area COq_LE. To obtain the tariff T_H, since q_H is greater than q_L, the integral from q_L to q_H of the area which lies below the H type marginal utility function should be added. In the linear case this additional tariff is represented by the area Hq_Lq_HN. If we want to derive the net revenues per customer it suffices to subtract the marginal cost c*; graphically CGIE and CGIE + HIN measure the respective net revenue obtained from a customer of type L and H. It is also worth noting how, in order to respect the incentive compatibility constraint of the H type a net surplus which amounts to $(\theta - 1)\, T_L$ (the area ACEH) must be given to each customer of type H.

Notice how the representation of our *vertical* game in the case of monopoly when c^1 is equal to zero does not differ from the previous pictures. Therefore, no further explanations are required.

For the horizontal game we can make more detailed consideration, since we have already solved the game in the presence of entry, showing that the optimal prices and quantities allocated per unit of customer do not depend in any way on the scale of entry in the vertical game.[38]

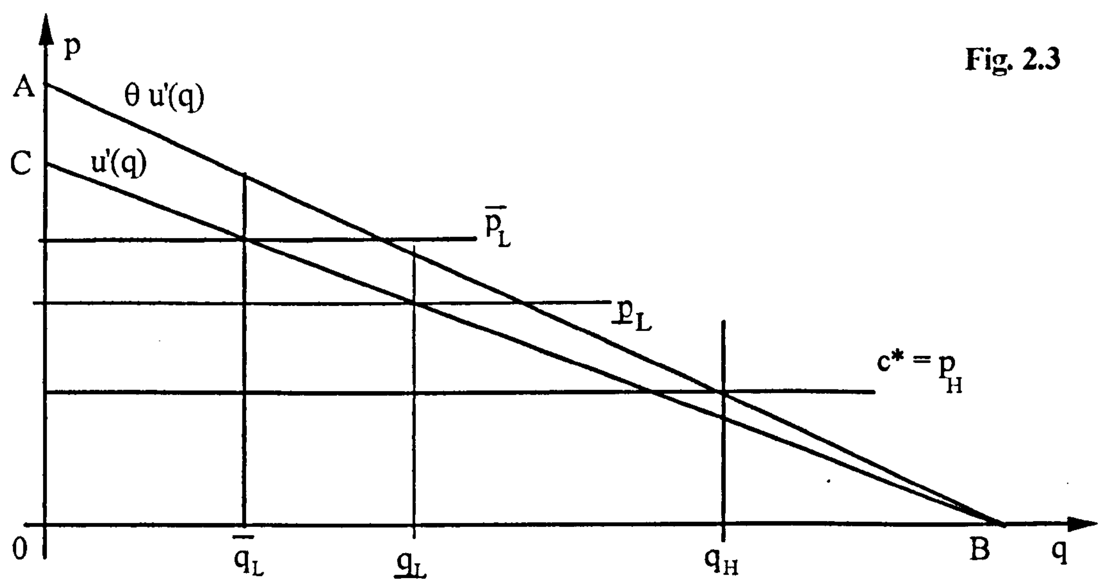


Fig. 2.3

We know that in the horizontal game the value of q_H remains the same as long

[38] This would be no longer true when the incumbent can fix charges to give access to his network, since in this case we will see how he can act strategically in a way of appropriating all the entrant's profits (by using an appropriate access charge) enjoying monopoly profits independently of the scale of entry.

as the marginal cost is constant. Instead, we have different values of q_L whenever the values of N_L and N_H vary. In fig. 2.3 above for a given value of K, the two extreme values assumed by the low-demand per customer bundle q_L are shown. From result **(a)** we know that q_L is a decreasing function of the ratio N_H/N_L. Since the total scale of entry is fixed (that is, we are allowing K_H to vary), then, the ratio N_H/N_L is clearly a decreasing function of the scale of entry for the high type K_H [since N_H/N_L is equal to $(N-K_H)/(N-K+K_H)$]. Therefore, q_L is an increasing function of K_H.

If for each variable x, we denote by $\underline{x}$ and $\overline{x}$ respectively the value assumed for $K_H = K$ and $K_H = 0$, it immediately follows that $\overline{q}_L$ is strictly greater than $\underline{q}_L$.This implies that the incumbent can enjoy monopoly profits for the residual number of customers' served in the horizontal game. In the next section we solve the maximisation problem for the entrant.

2.4 The entrant's problem

2.4.1 Strategies of competition

The previous results derived for the incumbent's optimal pricing strategy are useful to solve the entry game, that is the decision of the competitor to determine her scale of entry in terms of number of customers. In this section we are now going to solve stage 1 of the horizontal game, tackling the question of how the competitor's profits vary with the decision to serve the two types of customers in a given proportion.

However, before doing that, let us discuss in more detail the strategy of competition available to the entrant in stage 3. So far we assumed that the competitor acts as a *tariff-taker*; i.e. that she takes as given the tariff T_t determined by the incumbent. However, in principle there are several strategies of competition that the entrant can use; for instance, a *surplus-taker* competitor can propose a new tariff T_t^s that allows to type t the same surplus fixed by the incumbent. This kind of strategy is superior to the previous one and in Appendix 2 we are going to show that this is indeed the optimal pricing strategy for the entrant. Intuitively, it should be clear how the tariff-taking strategy is just a special case of the surplus taking behaviour. This consideration justifies our focus on the simple price taking behaviour on simplicity grounds. The point is that a priori we find out an incredible number of variants of the basic model. That is the reason why we limit ourselves to analyse in detail some of them, suggesting how the reasoning developed could be extended or modified in order to deal with more complicated cases following the geometric intuition.

For both these strategies of competition it is important to *endogenise* the scale of entry and determine which of the two types the entrant will choose to serve. In general, as will be shown, this will crucially depend on the level of the entrant's marginal costs m, or we may say, on the efficiency level of the competitor.

In what follows, first of all, we will provide a sketch of the entrant's problem in general terms, just in order to specify the crucial variables relevant in the choice of the customer's type to be served *at the margin*. Loosely, from the solution of the optimisation problem for the entrant we are able to determine two critical values assumed by the marginal cost as a function of the scale of entry. Outside the interval determined by these values, two corner solutions appear; an inefficient competitor will always find it more profitable to serve only low-demand customers, whereas a relatively efficient competitor will cream skim (i.e. she will serve only the H type).

The reason that drives this result is very simple: serving one customer of type H involves a *trade-off* between the benefits enjoyed because of the increase in the quantity sold and the costs due to the decrease in the average tariff (we are referring to the quantity discount result). It is then clear how a relatively efficient competitor is willing to accept the reduction in the average tariff, in order to enjoy a higher profit in absolute terms. For intermediate values of m the entrant will serve *both* types, being indifferent between one type and the other at the margin. This first result holds for the case in which there are relevant *infra-marginal effects*.

It is interesting to notice how only, ignoring infra-marginal effects, a competitor as efficient as the incumbent will always go for the H type. In a more general context (that is, allowing for infra-marginal effects) an entrant with the same marginal cost of the incumbent *may* increase his profit by substituting a customer of type H with a customer of type L (see Appendix 3). Basically, a decrease in the number of the H type served by the entrant allows a reduction in the net surplus to leave to the remaining high-demand customers. In fact, the gains achieved in this way might be greater than the losses due to the forgone net additional revenue (due to the decision of no longer serving a customer of type H).

2.4.2 Endogenising the choice of the customers' types

We are now solving the first stage of the horizontal game. The competitor decides whether or not to enter the market and in the first case she determines the number of customers to be served in each market. Since we are still keeping the assumption that the entrant has a fixed scale of entry (i.e. $K=K_H+K_L$ is given) there is only one variable left to his control, say K_H. Naturally, the value of K_L will be automatically determined as a residual $K_L=K-K_H$.

The optimisation problem for the entrant can be stated as:

$$\max \Pi^e \equiv K_H\,(T_H - mq_H) + (K - K_H)\,(T_L - mq_L) - F(K_H, K - K_H)$$

where T_H and T_L represent the incumbent's optimal tariffs, whose value will be determined in stage 2 (as described in section 2.3 above) and we considering the case in which the cost of entry depends on the scale of entry for both types of customers.

Let us give a brief account of all the terms that enter into the marginal profitability per unit of customer:

$$d\Pi^e/dK_H = (T_H\text{-}mq_H)\text{-}(T_L\text{-}mq_L)+K_H d(T_H\text{-}mq_H)/dK_H+(K\text{-}K_H)d(T_L\text{-}mq_L)/dK_H - (fe_H\text{-}fe_L)$$

1) $T = (T_H\text{-}mq_H)\text{-}(T_L\text{-}mq_L)$ represents the difference between the two net revenues enjoyed by serving a customer of type H and a customer of type L.

2) $\Omega = K_H\, d(T_H\text{-}mq_H)/dK_H + (K\text{-}K_H)\, d(T_L\text{-}mq_L)/dK_H$ represents the total net gain (or loss) on the infra-marginal customers of type H and L arising because of the variation in the incumbent's tariffs and quantities (as a response to the change in the value of K_H chosen by the entrant).

3) $fe = (fe_H - fe_L)$ represents the effect brought by a different marginal cost of entry for the two types of customer, where: $fe_H = dF/dK_H$ and $fe_L = dF/d(K - K_H)$.

In Appendix 3 we will consider in more detail the problem. In what follows we focus on a particular case.

The choice of an additional customer (type L or H) when the scale of entry is small

The simplest framework to examine is the one in which it is optimal for the entrant to go only for one type and her decision concerns just the choice of the additional customer to be served (i.e. the entrant is free to choose between serving a customer of type H or L).

We initially abstract from the preliminary decision whether or not to enter the market, postponing some considerations on the influence of the cost of entry on the choice of the type of customer to be served at the end of this section. Ignoring infra-marginal gains (or losses) on the residual number of customers (i.e. Ω=0) and assuming that the cost of entry depends only the total number of customers (fe=0) the crucial variable for the entrant is clearly the net revenue per customer (using the previous notation T). Infra-marginal effects can be surely ignored if we are dealing with infinitesimal changes in the value of the number of customers served by the entrant, i.e. when the percentage scale of entry (K/N) is small enough. A formal justification is provided in Appendix 3.

We will start from a *tariff-taker* competitor. The criterion which will guide her decision is simply the achievement of the highest profit per unit of customer *at the margin.*

The optimisation problem for the entrant can be stated as:

$$\max \Pi^e \equiv K_H (T_H - mq_H) + (K - K_H)(T_L - mq_L) - F(K)$$

The marginal profitability per unit of customer is given by:

$$d\Pi^e / dK_H = (T_H - mq_H) - (T_L - mq_L)$$

where T_H is a decreasing function of K_H, whereas both T_L and q_L are increasing functions of K_H and q_H is a constant (see section 2.3). The competitor obtains from each type the same net revenue per unit of customer if and only if $d\Pi^e/dK_H = 0$; that is, whenever:

$$m = (T_H - T_L) / (q_H - q_L)$$

Keeping the same notation as before, for each variable x let $\underline{x}$ and $\overline{x}$ represent respectively the value assumed for $K_H = K$ and $K_H = 0$. That is, the first value is associated to cream skimming the other to skimmed milk competition.

Given these premises, it is straightforward to conclude that an interior solution ($0<K_H<K$) arises whenever m assumes values within the interval determined by $\underline{m}$ and $\overline{m}$, whereas for all values outside this interval two corner solutions appear, as sketched below.

The entrant's choice of customers

1) for $m < \underline{m}$ the entrant will go only for the H type, since $d\Pi^e/dK_H > 0$;

2) for $m > \overline{m}$ the entrant will serve only the L type, since $d\Pi^e/dK_H < 0$;

3) for $\underline{m} \leq m \leq \overline{m}$ the entrant will be *indifferent* between serving one type or the other (as for all these values of m the equilibrium value of K_H will be set such that $d\Pi^e/dK_H = 0$).

In Appendix 3 we will prove that m is a decreasing function of the scale of entry in the high-demand market K_H; therefore, we have that $\underline{m}$ is less than $\overline{m}$. In economic terms this means that the level of efficiency required for the entrant in order to engage in cream skimming competition is strictly higher (the level of the marginal cost being strictly lower) than the one required to engage in skimmed milk competition. It can also be proved that $\underline{m}$ is always greater than the H type's marginal price p_H, which is simply equal to c*, the incumbent's marginal cost (see Appendix 3).

It is then immediate to deduce that in the case of a *competitor as efficient as the incumbent* case 1) will always occur (since $\underline{m} > c^*$). The same conclusion holds also for a surplus taker competitor (since in this case she can do no better than a tariff taker). As will be shown, the best that a surplus taker can do is to set marginal price equal to marginal cost. It is worth noticing that for the cream skimming case to occur the entrant does not need to be more efficient than the incumbent, as in the LT game or in the Laffont and Tirole approach.

Let us finally consider how the level of the cost of entry that are sunk in the *preliminary decision of entry* affects the choice of the type of customer to be served. First of all, the competitor will enter the market only if she gets a positive profit. By imposing $\Pi^e = 0$ we can determine the limiting value of the marginal cost (hereafter denoted by me) for which entry is profitable:

$$\mathbf{me} = [K_H T_H + (K - K_H) T_L - F(K)] / [K_H q_H - (K - K_H) q_L]$$

Depending on the value assumed by the fixed cost F **me** varies, so we can classify two cases:

I) for sufficiently high level of F **me** will be less than $\underline{m}$, so that the entrant will serve only the H type. Laffont and Tirole (1990b) limit the analysis to this case.

II) For sufficiently low level of F **me** will be greater than $\overline{m}$, so that the entrant will serve only the L type.

Let us conclude this section related to the entrant's problem by providing a graphical representation.

2.4.3 A graphical representation

It is useful to continue to represent the case of *quadratic utility functions*, in order to have a clear-cut picture of the possible situations.

The limiting levels $\underline{m}$ ($\overline{m}$)in the linear case are just given by the average between $\theta\, \underline{p}_L$ ($\theta\, \overline{p}_L$) and p_H (= c); that is:

$$\underline{m} = (\theta\, \underline{p}_L + c)/2$$

$$\overline{m} = (\theta\, \overline{p}_L + c)/2$$

Graphically, as shown in fig. 2.5 below, this occurs because the areas of two triangles STY and YDW (HVZ and ZDX) between θ u'(q) and the horizontal line $\underline{m}$ (or $\overline{m}$) [and the two boundaries $\overline{q}_L$ (or $\underline{q}_L$) and q_H] must be equal.

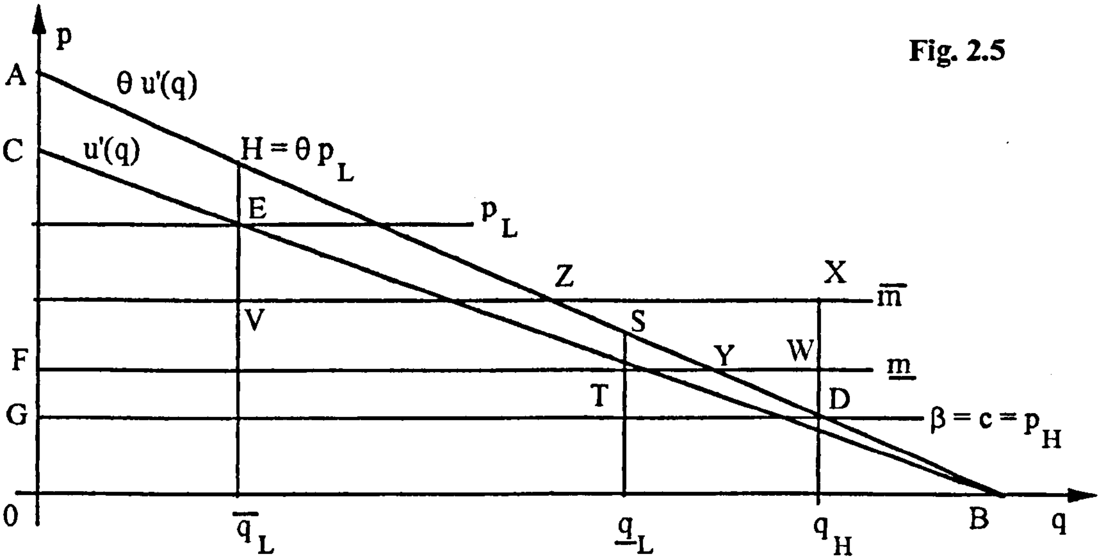


Fig. 2.5

Let us now extend the analysis to a *surplus-taker* entrant. The same reasoning can be applied and the same cases hold. The real difference hinges on the fact that a surplus taker competitor sets for each type the marginal price equal to her marginal cost [$p^e=\theta u'(q^e_H)=u'(q^e_L)=m$], so that the two limiting values of m are greater than the ones relative to a tariff-taker. In any case the surplus of high-demand customers is the same, no matter who they buys from. Another important fact to notice is that the entrant does not create any incentive compatibility problems to the incumbent (differently from Laffont and Tirole), because the L type consumer will always get a quantity q^e_L from the entrant which is less than the one offered by the incumbent (q_L). Instead of providing a detailed analysis of this case we will limit to represent the three possible situations for quadratic utility functions.

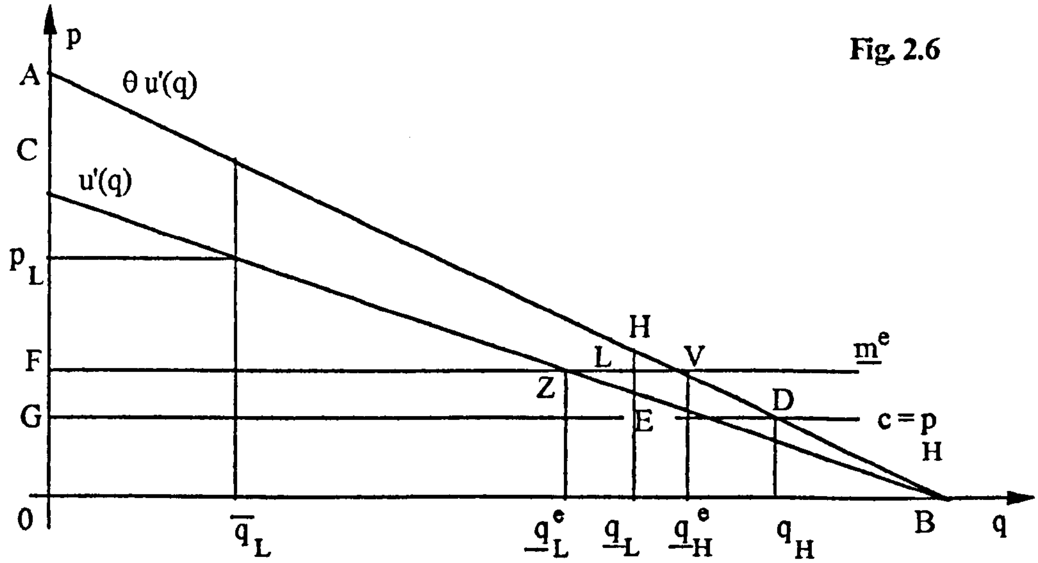


Fig. 2.6

To picture the cream skimming case (fig. 2.6 above) we should first of all find a price $\underline{m}^e$ (graphically a horizontal line) for which the consumer's surplus of the H type [the area AFV = $\theta u(q^e_H)$ - $q^e_H\,\underline{m}^e$] is equalised to the sum of the surplus of the L type [the area CFZ = $\theta u(q^e_L)$ - $q^e_L\,\underline{m}^e$] and the surplus allowed by the incumbent to the H customer [the area ACHE = $(\theta\text{-}1)T_L$]. This occurs in correspondence of the equality of the areas of the two shaded triangles (ZEL and LVH).

The second case in which only the L type is served is shown in fig. 2.7 below.

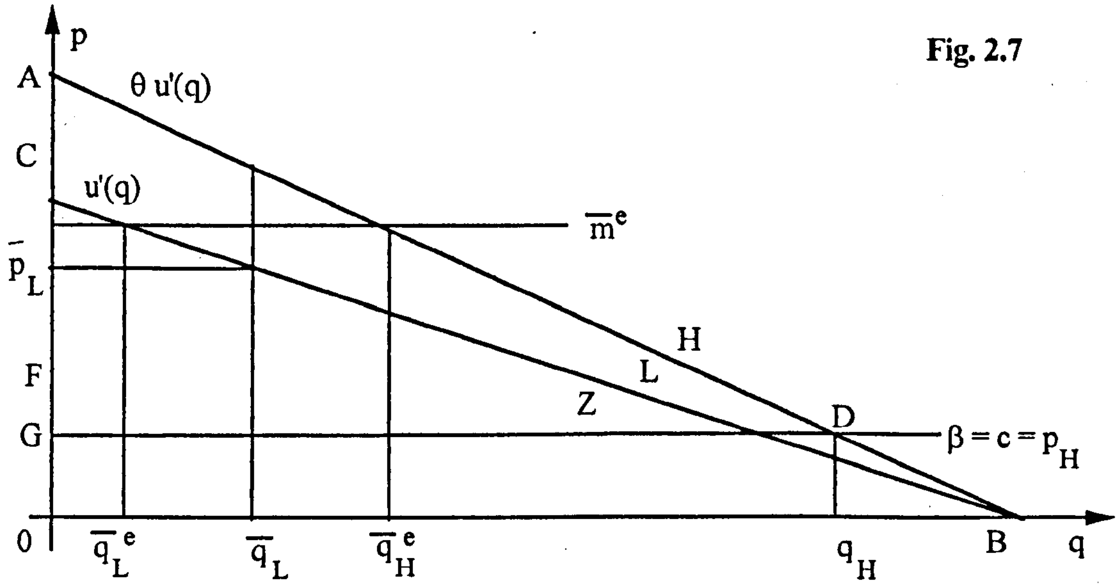


Fig. 2.7

As is geometrically intuitive from the previous figures, for all values of the scale of entry K we have:

$$\underline{m}^e > \underline{p}_L > \underline{m}; \qquad m^e > \overline{p}_L > \overline{m}$$

It goes without saying that the competitor will be indifferent about serving one type or the other for intermediate values of m (that is, for $\underline{m}^e \le m \le \overline{m}^e$).

At this point we can finally compare the two strategies of entry. It is important to notice how the surplus-taker behaviour is a more effective strategy for the entrant which allows her to appropriate a greater profit per customer. As has been shown, we have $\overline{m}^e > \overline{m}$ and $\underline{m}^e > \underline{m}$ for any value of K. Furthermore, for any value of the entry scale K and of the marginal cost m, the entrant is able to get a greater profit if she acts as a surplus-taker. The explanation of this is that, since a surplus-taker entrant tends to compete more for high-demand customers, she will allow them to get a greater net consumer surplus for any given value of K. Hence, this strategy improves the sum of the net consumers' surplus; that is, their welfare.

In the next subsection we will focus on the cream skimming case, in order to explore its implication in terms of productive and allocative efficiency and to compare to the case of skimmed milk competition.

2.4.4 The cream skimming and skimmed milk cases

It seems relevant at this point to compare the results derived in this section with the one of Laffont and Tirole (1990b). A comparison with the cream skimming paradigm carried out within an unregulated framework is perfectly legitimate, since our analysis have proved that no major changes of the outcomes of Laffont and Tirole (1990b) will take place without regulation (cf. section 1.3.2). In fact, paradoxically, this paradigm proves to be so *robust* that the very same distortion arises as in an unregulated monopoly. Only regarding regulation the cream skimming paradigm seems not to be so robust. In fact, we have shown how when social welfare is defined as the unweighted sum of consumers' surplus and profits (a particular case which occurs in Laffont and Tirole's framework when the shadow cost of public funds is equal to zero) there can be at most a sequence of only three regimes with no distortion (instead of the original sequence of five regimes). Therefore, in order to derive all the sequence of the five regimes present in Laffont and Tirole (1990b) with optimal regulation we need to introduce the distortion due to the shadow cost of public funds in the social welfare function, or to impose a binding budget constraint.

Our point in doing a comparison with Laffont and Tirole's cream skimming model is not merely theoretical. Whether natural monopolies should be protected from cream skimming entry is a question that has been raised by the recent theoretical and applied literature. In the previous section we developed an alternative modelling of competitive issues, which allows us to compare the welfare implications of different strategy of competition -namely, competition for the low and high type- and to derive some policy implications. Here we want to find out whether cream skimming competition is necessarily more harmful than skimmed milk competition, from an efficiency point of view.

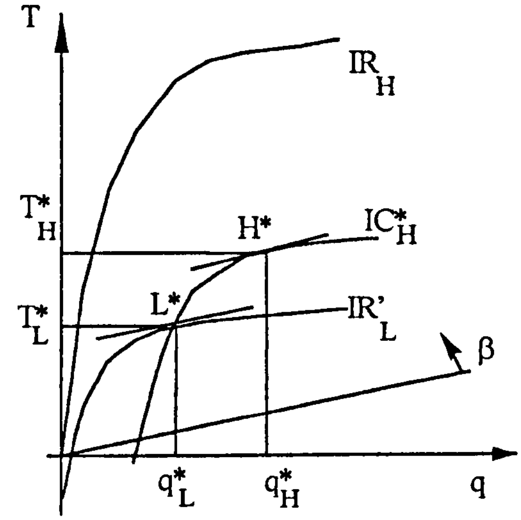
In order to carry out a comparison with the cream skimming model of Laffont and Tirole, let us briefly recap what happens in terms of productive and allocative efficiency when bypass from a cream skimming competitive fringe takes place. We clearly satisfy *productive* efficiency being the entrant more efficient than the incumbent, we do also get *allocative* efficiency having no distortion at the bottom (the price for low-demand customers being equal to the marginal cost). Furthermore bypass is also optimal from the point of view of the regulator, as shown in section 1.3. However, we must also acknowledge that the profit of the incumbent are clearly reduced, as he is left with the less lucrative part of the market. Moreover, high-demand customers will enjoy a higher surplus, compared with the one they would enjoy if they had been served by the incumbent and low-demand customers are given a positive surplus, lower than the one they would have got in regime 5, where their incentive compatibility constraint was binding.

It is also quite interesting to consider the case in which the bypass technology is appropriate only for the low type (the skimmed milk), that is the exactly the opposite case of cream skimming, following the Laffont and Tirole approach. Under the hypothesis of perfect competition what we may call the "skimmed milk" case implies that the surplus left to the H type is lower than the one given only by the regulated firm. This case may occur when we have low or zero fixed costs (in our notation for a low value of the parameter fe_L) and high marginal costs (i.e. a relative high value of m). What happens is that the L type will enjoy a positive surplus, causing a reduction in the tariffs of the regulated firm.

We will not get any change of regimes, since this time the incentive compatibility constraint of the low type IC_L is not affected, whereas its individual rationality constraint IR_L' will not pass through the origin, allowing him to enjoy a positive surplus (offered by the competitors). Hence, under different specifications of the social welfare function (as well as in the unregulated case) regime 1 remains the only regime, apart from bypass (for a very high value of fe_H).

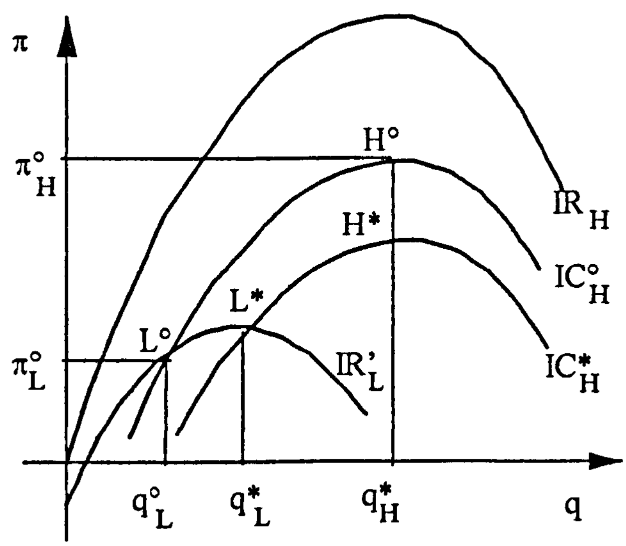
In fact, if we apply the same procedure as in the cream skimming case, carried out in section 1.3.1 increasing the marginal cost β of the regulated firm, the surplus the incumbent must leave to the L customer is given (and as a consequence also the additional net surplus of the H type). Henceafter, the superscripts * and ° refer to the optimal bundles derived assuming respectively the social welfare function as the unweighted sum of consumers' surplus and profits (i.e. for λ equal to zero) and as a weighted sum in the presence of public funds costs (i.e. when λ is greater than zero).

Fig. 2.8



In fig. 2.8 above the only non-bypass regime is illustrated in the space (q, T). It is immediate to derive the graphical representation in the space (q, π), as shown in fig. 2.9 below.

Fig. 2.9



As β increases, the bundle (q_L, q_H) corresponding to the separating equilibrium is reduced, till the firm finds it no longer profitable to serve the L type and to allow an additional surplus to the H type, leaving way to bypass. Even if there is no change of regime, in this only possible regime social welfare is reduced by the threat of skimmed milk competition, since the profit of the regulated firm decreases (compared with the monopoly case). Low-demand customers enjoy a positive surplus, so that they are better off than in the standard monopoly case (in which they would enjoy no surplus, their rational individual constraint being binding). The transfer of surplus from the regulated producer (or public budget) to customers reduces social welfare, if transfers are costly (for instance, if we acknowledge the presence of a shadow cost of public funds, i.e. we assume that λ is greater than zero). It is worth noticing how, even in the bypass case, following considerations of *productive* efficiency, social welfare is reduced compared with the cream skimming case, because the marginal cost of the entrant is higher (compared with the incumbent's one). Let see the consequences of these types of competition on consumers' surplus. Low-demand customers would be better off in the presence of skimmed milk competition, but at the expenses of *allocative* efficiency (since there would be a distortion in the tariff for the low type). On the other hand, high-demand customers would rather prefer cream skimming, as in this case they enjoy a greater surplus offered by competitors. Therefore, from an allocative and productive efficiency point of view cream skimming would be better than skimmed milk competition.

Laffont and Tirole's analysis clearly does not endogenise the competitor's choice of customers' types, since rivals are able to serve the whole population of one type, pre-empting only one market (that is, market L in the case of skimmed milk competition). Moreover, competitive issues are not really tackled, as we were still assuming perfect competition with two part tariffs. The crucial differences of our alternative approach, with Laffont and Tirole (1990b), which are basically all related to the modelling of competition, are summarised below:

(i) the entrant is "small" and cannot totally *pre-empt* either of the two market. Instead, in Laffont and Tirole (1990b), the incumbent risks losing all the H market;

(ii) the entrant is *not perfectly competitive*, since she can enjoy a positive profit;

(iii) the entrant maximises her profit acting a *surplus taker* (that is offering to each type a surplus equal to the one determined by the incumbent, and not an exogenous one, technologically given in perfect competition);

(iv) the choice of which type of customers is *endogenised* so that we do not have to assume exogenously what is the cream. This differentiates the present model from the original one where only high-demand customers are served by the entrant, since the L type would get a negative surplus;

(v) entry may take place even if the competitor is less *efficient* than the incumbent

(vi) the incumbent is the *first mover*, that is the first player who determines the optimal pricing strategy.

In what follows we allow the entrant to compete for both types and to make use of fully non-linear pricing. Notice that we replace the perfect competitive fringe with a single entrant with limited capacity. As already derived in section 2.4, focusing on the decision of the type of customer to be served, when the competitor does not behave as a perfect competitor things become more complicated, and in this case what is the most profitable part of the market for the entrant and for the incumbent remains an open question. We have in fact shown how either cream skimming or skimmed milk competition can occur depending on the relative efficiency of the entrant.

Let us first of all describe the two main cases of competition for the high and low type. We first consider the *cream skimming* case. We have seen that this kind of competition can take place only when the entrant's marginal cost m is lower than a given threshold ($\underline{m}$). We have shown that the incumbent in the presence of entry can enjoy monopoly profits for the *residual* number of customers served: that is, N_H customers of high type and N_L customers of low type. Notice how in the absence of infra-marginal effects the incumbent loses the most profitable part of his customers. He would rather prefer to face the same scale of entry in the low-demand market. However, we should keep in mind that in the presence of optimal regulation this consideration does not matter, as the profits of the public firm are appropriated by the regulator. Examining issues related to productive efficiency we can conclude that clearly cream skimming competition is good in terms of *productive* efficiency, as long as the entrant is at least as efficient as the incumbent. Cream skimming is also beneficial for low-demand consumers. In fact regarding *allocative efficiency* considerations, we have seen that the marginal price of low-demand customers decreases in the presence of cream skimming. This derives from the fact that the marginal price p_L is an decreasing function of the scale of entry in the high demand market K_H -a direct consequence of result **c)** examined in section 2.3. There is no distortion at the top, so that the marginal price for high-demand customer is equal to the marginal cost (as in the standard monopoly case).

Let us now focus on the *skimmed milk* case. We have seen that this case occurs when the entrant is less efficient than the incumbent (and his marginal cost m is greater than a given threshold $\overline{m}$). Similarly to the cream skimming case, the incumbent can enjoy monopoly profits for the residual number of customers served, but this time is better off since he loses customers which gives him a lower variable profits.

Skimmed milk competition is clearly bad both for *productive efficiency* -since the competitor is surely less efficient than the incumbent- and *allocative efficiency* hurting low-demand customers, since the marginal price of low-demand customers rises as a consequence of entry. This derives from the fact that the price p_L is an increasing function of the scale of entry in the low demand market.

Therefore, even in this enlarged framework the same conclusion as before holds; namely cream skimming is not very likely to be less harmful than competition for low-demand consumers, even in the absence of regulation, at least as long as it allows the entry of more efficient rivals. An additional reason that makes us prefer cream skimming (compared to skimmed milk competition) is that the marginal price of low-demand customers will decrease, bringing a welfare increase of consumers' surplus.

Let us sum up the most relevant conclusions. In this subsection we first analysed competition for the skimmed milk carried out by a perfectly competitive fringe, a case which is the opposite of cream skimming. When bypass takes place, social welfare is reduced, because the marginal cost of the entrant are higher (compared to the incumbent's one) and both from a productive and allocative point of view cream skimming would be preferable. Later, when we removed the hypotheses of perfectly competitive entrant serving only the H type also in this context emerges that cream skimming is not necessarily the most "harmful" type of competition. This time in the case of skimmed milk competition a reduction in allocative efficiency will surely take place, whereas in the opposite case of cream skimming allocative efficiency would be satisfied. This is due to the fact that in this the marginal price in the low-demand market is an increasing function of the scale of entry in the high demand market.

These results contrast on the standard idea of cream skimming, always being considered an undesirable regime by itself or one of the major causes of the unsustainability of regulated monopolies. In particular, cream skimming *may* involve a lower reduction of social welfare compared to the one implied by competition for the skimmed milk. In our model cream skimming is welfare enhancing, at least from the point of view of allocative efficiency. Moreover, society can benefit from the entry of a more efficient player who serve the high demand market, and the incumbent can keep his monopoly tariff for the residual number of customer served.

We must of course acknowledge many other drawbacks (ignored in our analysis for simplicity's sake) against restrictions to entry (i.e. the provision of insufficient incentives for the incumbent to be efficient). In fact, in the absence of other types of regulation the incumbent may be tempted to overprice his products or to engage in predatory behaviour. These caveats lead us to be against justified restrictions on competitive entry based only on the claim that cream skimming would be undesirable and inefficient by itself, without providing any evidence for the specific case analysed. In the next section we will provide introduce competition in the vertical game, and examine which strategies of competition will survive in this enlarged framework.

2.5 The solution of the vertical game

2.5.1 The case of an equally efficient competitor

Let us postpone the solution of the vertical game in general terms, focusing, first on the case of an equally efficient competitor (i.e. m is equal to c*). We solve the game in the absence of regulation, that is ignoring stage (0). In particular, we tackle the question on how the incumbent's profits vary, depending on the competitor's entry decision. In this case, if the competitor acts as a tariff taker [i.e. she takes as given the tariffs T_t determined by the incumbent in stage (3)] she will maximise her gross profits (abstracting from the access charges). That is, as will be shown later, a surplus taker competitor and a tariff taker competitor will end up proposing the same tariffs. Hence, for simplicity's sake we will consider only the case of a tariff taker competitor and demonstrate the following proposition.

Proposition 1

When an equally efficient competitor enters the market of good 1 it is optimal for the incumbent to set the per customer access charge equal to the monopoly variable profits; that is, T_H - c^*q_H *and to maintain the previous monopoly pricing strategy, independently of the scale of entry K. In other words, the incumbent is indifferent between facing a duopoly or a monopoly in the market of good 1.*

In what follows we will provide just a sketch of the proof of Proposition 1, as what matters at this stage is to develop the crucial reasoning relevant in the solution of the game.

Since in the absence of regulation the incumbent can fix at his discretion the level of the access charge, his profit can never be reduced in the presence of entry (as he can always deter entry by setting a very high access price). Furthermore, as the incumbent knows the competitor's marginal cost he can extract all the entrant's profits, simply by setting the per customer access charge equal to the entrant's per capita gross profits; that is, T^{*}_{H} - $c^{*}q^{*}_{H}$. Notice how we are considering a simplified framework (i.e. in the absence of infra-marginal effects, or alternatively assuming a small scale of entry) the entrant will cream skim, since her variable profits in this case are strictly greater than the ones she had serving low-demand customers. A lower access charge would not be levied, since in this case the incumbent would be worse off, getting lower profits. Therefore, assuming the presence of a tariff taker competitor, stage (3) is automatically solved, given the incumbent's choice of T_H.

In particular, if the per customer access charge is set equal to the gross profits of a tariff taker competitor, the incumbent will expropriate all the entrant's profits. In this case, we can write the access pricing condition [AP] as:

[AP] $\quad F(K_H) / K_H = T_H - c^{*}q_H$

In order to deal with stage (3) we may write down the incumbent's maximisation problem, in presence of entry, subject to the access pricing condition:

[Problem 3] $\max \Pi(Q^{i}) \equiv$

$2Np^{0}q^{0}+N_LT_L+N_HT_H+F(K_H) - NC(2N, Q^{0} + Q) - c^{0}(2Nq^{0})$

$- c^{*}(N_Lq_L+N_Hq_H)$ subject to:

[MME] $\quad p^{0} = v'(q^{0})$

[IR_L] $\quad T_L = u(q_L)$

[IC_H] $\quad T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$

[AP] $\quad F(K_H) = K_H\, (T_H - c^{*}q_H)$

It is easy to show how substituting the constraint [AP] the problem becomes exactly the same as the one examined in the monopoly case. This allows us to solve stage (3) as we get the values of q^{0}, q_L and q_H. Given the [AP] condition, clearly the entrant is indifferent about entry for p_H=c*=m, only if she serves the H type; otherwise, she will always incur in losses. Thus, we can assume that she enters and goes for the H type.

The proposed solution is not necessarily the unique solution of the game; however, this is the only economically relevant solution, since it allows the incumbent to get the same results as if he were the *only* player. Notice how, in practice, the incumbent has a sort of control on the scale of entry, since by setting the appropriate access charge to the entrant (extracting all her profits) he can oblige the entrant to take the action he wanted to; that is in this case to go only for the high-demand type (setting $K= K_H$). The previous reasoning isolates the relevant points that will allow us to solve the general game.

It is immediate to derive graphically the quantities allocated per unit of customer (obtained in correspondence of the intersection between the curves representing the marginal utility and marginal price for each type) exactly as we did in section 2.3.3 (cf. fig. 2.2) and to verify that q_H is greater than q_L.

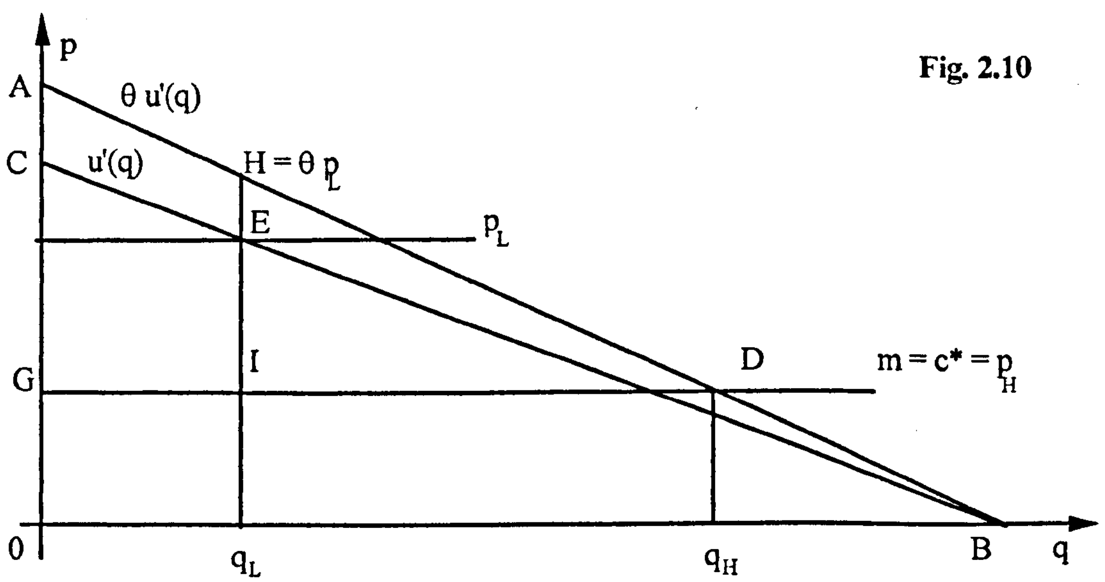


Fig. 2.10

The tariffs T_L and T_H (i.e. the gross revenue per unit of customer) are given by the integral from 0 to q_L and from 0 to q_H of the area which lies below the L type and the H type marginal utility function $u'(q_L)$ and $\theta\ u'(q_H)$. In the linear case T_L and T_H are respectively given by COq_LE and Hq_Lq_HD. The net revenues per customers' type are derived by subtracting the variable costs and are given by CGIE and CGIE +

HID. The H type enjoys a positive net surplus which amounts to $(\theta - 1)\ T_L$ (the area ACEH).

Let us first consider a competitor with the same marginal cost of the incumbent ($m=c^*$). If the non-linear monopoly tariffs T_L and T_H are maintained unchanged by the incumbent, the best the entrant can do, for any scale of entry, is to go for the H type and to get a gross profit per customer given by CGIE + HID (always greater than CGIE). Hence, if the competitor enters she will certainly decide to do cream skimming, setting $p^e_H=m=c^*$.

Furthermore, she will enter only if her implicit participation constraint is satisfied; that is, the per customer access charge she must pay to the incumbent is lower than her gross profits per unit of customer (CGIE + HID).

To complete the argument, we just need to show that the optimal pricing for the incumbent is indeed to choose the monopoly tariffs T_L and T_H for any scale of entry. This seems obvious if we realise that, given an equally efficient competitor, the incumbent who expropriates all her gross profit is exactly in the same position as a *monopolist.*

In fact, if he increases q_L by a marginal amount he will gain p_L-c^* (that is, segment EI) from his N low type customers, but he will lose $(\theta\text{-}1)\ p_L$ (that is, segment HE) from his N_H high type customers, and similarly from the $K=K_H$ per customer access charge paid by the entrant. Hence, in practice his net marginal losses are $N(\theta\text{-}1)\ p_L$, as if entry had not occurred. Basically, we have an internal optimum when marginal losses equate marginal gains; that is, $N(\theta\text{-}1)\ p_L=N\ (p_L\text{-}c^*)$ which gives $p_L=c^*/(2\text{-}\theta)$, that is independently of the scale of entry. Thus, the monopoly pricing T_L and T_H, derived analytically in the section 2.3.2, still represents the optimal pricing strategy for the incumbent for any scale of entry, which is exactly what we needed to show. As before, in practice, the incumbent has a sort of control of the scale of entry, since by extracting all her profits by means of the access charge he can oblige the entrant to set $K= K_H$.

Notice in fact how the crucial variable is the *total* population served by the two player, composed by N customers of type L entirely served by the incumbent and N customers of type H served by both players ($2N=N+N_H+K_H$). This explains why the pricing schedule and the bundle allocated per each customer type do not depend on the scale of entry, in contrast with the horizontal game, as noticed in section 2.3.2.

So far we have examined just the case of the entry of an equally efficient competitor. Let us now generalise the model.

2.5.2 The general solution

Let us tackle the general game introducing a surplus taker competitor, for simplicity's sake but without any loss of generality, in order to show the general solution and how the incumbent's profit will vary, depending on the competitor's type. To achieve this aim we can simply add in the previous fig. 2.8 a line parallel to p_H in correspondence of m (which is less or equal to c*). This line intersects the vertical lines ($q=0$, q_L) and the marginal utility function θ u'(q) of the H type respectively in L, M and N, as shown in fig. 2.11 below.

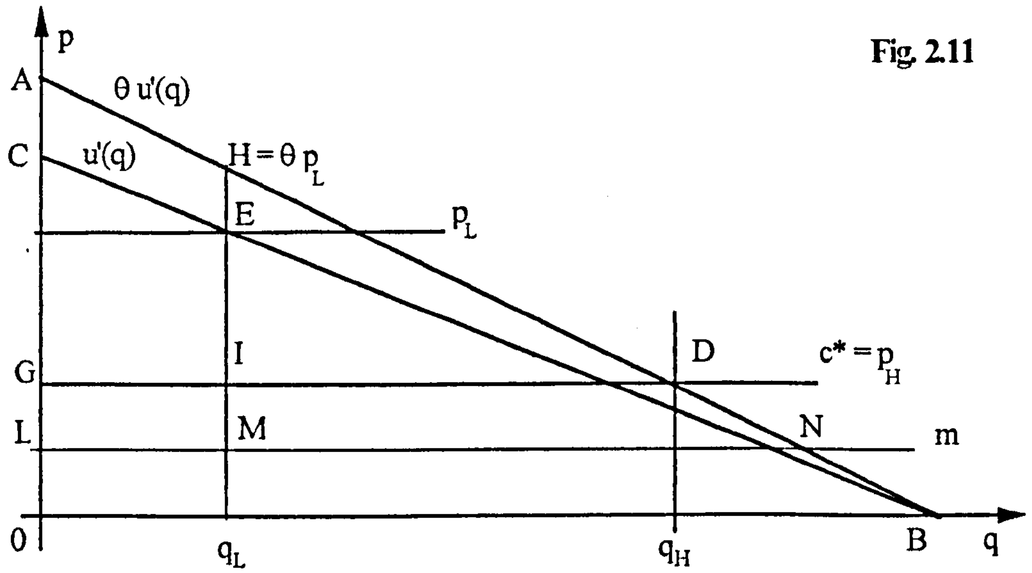


Fig. 2.11

With a more efficient competitor (with m less than c*) the incumbent can increase the per customer access charge asking the entrant exactly her gross profits per unit of customer (CLME+HMN). This is in fact, on the basis of the previous reasoning, the new optimal per customer access charge, for any scale of entry (given the monopoly price $p_H=c^*>m$). In fact, as before, the competitor will enter only for non negative profits and will decide to do cream skimming, setting $p_H=m<c^*$.

As before, the incumbent will maintain his optimal non-linear tariffs T_L and T_H for any scale of entry, because the first order condition relative to q_H does not depend on the value of N_H and the additional K_H per customer profit (GLND) is fixed and does not affect the optimal level of q_L. This happens because, given p_H the additional *access profit* (GLND) remains untouched for any given level of p_L greater than p_H, so that the incumbent remains in the position of a monopolist, with an additional profit which depends only on the scale of entry $K= K_H$. A higher access charge would not be levied, since in that case entry would not occur and the incumbent would just get his monopoly profits, losing any additional access profits. On the other hand a lower access charge would not be levied either. In fact, if the incumbent increases q_L by a marginal amount he will gain p_L-c* (that is, segment EI) from his N low type customers, but he will lose $(\theta-1)\ p_L$ (that is, segment HE) from his N_H high type customers, and similarly from the $K=K_H$ per customer access charge paid by the entrant. Hence, as before, his marginal losses are $N(\theta-1)p_L$, as if entry had not occurred and the monopoly tariffs T_L and T_H still represents the optimal pricing strategy for the incumbent for any scale of entry, which is exactly what we needed to show.

To complete the reasoning, we just need to consider the case of a less efficient competitor and of a positive network marginal subcost associated with the delivery of good 1 (that is, considering the case in which c^1 is greater than zero).

Entry is not optimal for a less efficient competitor. In fact, if we assume c* less than m, clearly the per customer access charge the entrant can afford to pay is less than the gross profits the incumbent can obtain by making use of the monopoly tariffs T_L and T_H for any scale of entry. It is obvious that knowing this outcome, the incumbent will set the per customer access charge greater than the competitor's gross profits, so that no entry would occur.

A positive level of c^1 creates no major problems. In fact, with a single competitor we can just limit to analyse two part tariffs, setting:

$$F(K_H, Q^e) = c^1\ K_H\ q^e_H + H(K_H)$$

where $H(K_H)$ plays the same role of $F(K_H)$ in the previous analysis. In practice, the presence of an additional variable access cost just shifts the previous marginal costs curves upward from c* to c^1+c^* for the incumbent and from m to $m+c^1$ for the entrant. However, once this has been taken into account, we can apply the previous reasoning simply by setting $p_H=c^*+c^1$ and $m^1=m+c^1$. No other changes affect fig. 2.10 and 2.11.[39]

The previous results referred to the entrant's and the incumbent's pricing strategies are useful to solve the general game and to prove the following proposition.

Proposition 2

In the general model, without regulation, there is no competition for the L type in the market of good 1 (the non monopolised commodity). It is optimal for the incumbent to allow the entry of a cream skimming competitor, only if she is at least as efficient as the incumbent ($c^* \geq m$), and to set a per customer access charge equal to the monopoly variable profits; that is, T^*_H-c^*q_H (for $c^1=0$). In this way he can maintain the previous monopoly pricing independently of the scale of entry K (<N). When a more efficient competitor enters the market the incumbent prefers to have a duopoly rather than a monopoly in the market of good 1. He also finds it optimal to oblige the competitor to behave as a surplus taker. However, it would even be preferable for the incumbent (and for the maximisation of social welfare) that the competitor sells all the produced good 1 to the incumbent (without any losses) at the marginal cost, so that the latter is able to resell it to the consumers at the monopoly tariffs.

---

[39] It might be interesting, before concluding the graphical analysis to notice how the fact that a more efficient competitor sets a marginal price less than p_H implies a surplus loss for the monopolist and the collectivity. In fact, if the marginal units q^e_H - q_H were sold as infra-marginal units to other H customers, or to L customers, they would produce (in this linear case) a surplus which is exactly the double of the previous one. In practice, very loosely instead of having a triangle we would have a rectangle; more precisely Pareto optimality implies equal marginal rates of substitution).

Let us sketch of the proof of Proposition 2. Since we know that the incumbent fixes the access charge in order to maximise his profits, on the basis of the previous analysis we can ignore the presence of a less efficient competitor. In fact, allowing entry the incumbent will always get a lower per customer gross profit. Namely by asking the entrant a per customer access charge equal to her variable per customer profit he would be worse off, incurring in losses with respect to the monopoly profits, being $T_t - c^* q_t > T_t^e - m q_t^e$ (for $c^* < m$). On the other hand, for a more efficient competitor, the per customer gross profit advantage [that is, $(T_t^e - m q_t^e) - (T_t - c^* q_t)$] assumes the highest value if we have cream skimming competition. In practice, $(T_H^e - m q_H^e) - (T_H - c^* q_H)$ is greater than $(T_L^e - m q_L^e) - (T_L - c^* q_L)$ for c^* greater than m. Consequently, it is clear that, as the entrant has a fixed scale of entry, we can set $K = K_H$ (i.e. K_H is now given being no longer a control variable). The competitor's optimisation problem can therefore be stated as:

[Problem 4] $\max \Pi(Q^e) \equiv K_H (T_H^e - m q_H^e) - F(K_H)$ subject to:

$[IC_H^e]$ $\quad T_H^e = \theta\, u(q_H^e) - (\theta - 1)\, T_L$

where T_L is the incumbent's optimal tariff determined in stage (3).

Solving the previous problem, we get the following first order conditions:

$[q_H^e]$ $\quad p_H^e = \theta\, u'(q_H^e) = m$

Knowing m, the incumbent will set the per customer access charge equal to the entrant's per capita gross profits; that is $T_H^e - m q_H^e$. Furthermore, as the incumbent expropriates all the entrant's profits, it will be in his interest to oblige the competitor (by appropriately setting the access charge tariff) to behave as a surplus taker. From the previous reasoning we can write the access pricing condition [AP] as:

[AP] $\quad F(K_H) = K_H (T_H^e - m q_H^e) = K_H [\theta\, u(q_H^e) - (\theta - 1)\, u(q_L) - m\, q_H^e]$

which is needed in order to solve the incumbent's maximisation problem.

[Problem 5] $\max \Pi(Q^i) \equiv$

$2Np^o q^o + N_L T_L + N_H T_H + F(K_H) - NC(2N) - 2Nc^o q^o - c^*(N_L q_L + N_H q_H)$

subject to:

[MME] $p^0 = v'(q^0)$

[IR_L] $T_L = u(q_L)$

[IC_H] $T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$

[AP] $F(K_H) = K_H\, [\theta\, u(q_H^\circ) - (\theta - 1)\, u(q_L) - m\, q_H^\circ]$

Solving **[Problem 5]** with respect to q^0, q_L and q_H we obtain the very same first order conditions of the monopoly case and in particular: $p_L = c^* / (2-\theta) = p_H / (2-\theta)$. This can easily be checked by substituting the four binding constraints into the objective function. Thus, solving the incumbent's problems we get the optimal values of q^0, q_L and q_H. Given the previous result, the competitor is indifferent to enter or not, knowing that the best outcome is to break even, serving the H type. Thus, we can assume that she enters and cream skim the market. On the other hand, the incumbent cannot do better than to oblige the competitor to maximise her gross profits (behaving as a surplus taker) setting the per customer access charge equal to the entrant's per capita gross profits.

2.6 Further refinements of the model

One may feel that the previous model has two major drawbacks. On one side, the scale of entry is exogenously given and the entry decision is limited to a binary choice. On the other side, the cost function of the entrant may seem "ad hoc", as in practice she has a fixed marginal cost m, till she serves a number of customers less than K and an infinite marginal cost from there onwards.

Finally, it may seem undesirable that the purchasing solution (that is, the one in which the incumbent buys all the output of the entrant to resell it to the customers) should be preferable to allowing entry and direct selling of the competitor. Hence, further refinements are appropriate to satisfy any potential reader.

In what follows we will examine first separately and then jointly these two problems. In particular, we will show how it is possible to refine the previous model in order to have a single optimal scale of entry chosen by the potential competitor at stage (2) of our game, without any substantial change in our assumptions.

At the same time we feel that this will in reality just complicate the model, without enriching it with further essential economic content. In practice, as the optimal purchasing solution depends only on the indivisibility of the number of customers' problem it will always appear in one way or another. Even if it is not possible in the real world to buy say a third of an egg, economists still continue to assume fully divisible goods, for simplicity's sake and they keep on believing that taking into account this indivisibility problem would not sensibly change the understanding of the basic economic problems.

Clearly, the first drawback -that is, an exogenous given scale of entry- is more apparent than real. In fact, we can easily demonstrate the following proposition.

Proposition 3

If we leave unchanged all the other assumptions and allow the potential competitor to choose any finite number of customers as his scale of entry -that is, $0 \leq k \leq K$ $(<N)$*- in stage (2) the solution of the game will not change.*

Entry is still not optimal for a less efficient competitor. In fact for any possible scale of entry (given $c^*<m$) the per customer access charge that allows entry is less than the gross profits the incumbent can obtain from the monopolist tariffs T_L and T_H. Thus, the per customer access charge will be greater than the competitor's gross profits and consequently entry would not take place.

On the other hand, for a more efficient competitor the entrant has a per customer gross profit advantage, which is given, and reaches its maximum value if she serves the H type. Hence, for the incumbent it is optimal not only to have cream skimming, but also to oblige an efficient competitor to maximise the scale of entry, setting k equal to K. This can be done decreasing (by a small amount ε, very close to zero) the per customer access charge associated with the maximum scale of entry. Hence, as before we end up with $k_H = k = K$.

Let us now tackle the two drawbacks jointly. It can be shown that if we consider an increasing marginal cost for the competitor and allow her to choose any finite number of customers as her scale of entry -that is, $0 \leq k \leq K$- at stage (2) of the game, we may have a unique solution of the game, but the purchasing solution due to the indivisibility of the number of customers will not disappear.

Let us change the initial assumption (iv) considering the specific case of a competitor that is more efficient than the incumbent only if she keep himself relatively small; that is, she doesn't pre-empt either of the two types' markets for good 1.

(iv)' the competitor can choose any *scale of entry* $0 \leq k \leq K$ in terms of number of customers:

$$k = K_L + K_H < N$$

She must pay an access charge F fixed by the incumbent dependent both on the scale of entry K and on the total output Q^e:

$$F(K_L, K_H, Q^e) \geq 0$$

Finally, she has an increasing production cost function:

$$C^e(k, Q^e) = m(k)\, Q^e$$

where m(k) represents a continuously differentiable marginal cost function (with respect to the scale of entry) that satisfies the following properties:

1) it is increasing with the scale of entry;
2) its initial value is lower than the incumbent's marginal cost;
3) it intersects the incumbent's marginal cost when the scale of entry is less than the number of customer of any type;

The competitor will enter the non monopolised good market if the usual participation constraint is satisfied:

$$[IR^e] \qquad \Pi^e(Q^e) = K_L(T^e_L - mq^e_L) + K_H(T^e_H - mq^e_H) - F(K_L, K_H, Q^e) \geq 0$$

With assumption (iv)' we have thus removed the two drawbacks and we can now state the following proposition.

Proposition 4

Under assumptions (i), (ii), (iii), (iv)' and (v) allowing any finite discrete number of customers as scale of entry there is only cream skimming competition in the market of good 1 (the non monopolised commodity). It is optimal for the incumbent to allow entry, to set a per customer access charge equal to the monopoly variable profits and to maintain the previous monopoly pricing. The incumbent also finds it optimal to oblige the competitor to behave as a surplus taker. However, it would be preferable for the incumbent (and for the maximisation of social welfare) that the competitor sells all the produced good 1 to the incumbent (reaching only break-even profits), so that the latter is able to resell it to the consumers, applying the monopoly tariffs.

Entry is optimal only when the competitor is more efficient -that is, for low marginal costs m(k) less than c*- being possible for the incumbent to set the per customer access charge greater than the gross profits he had obtained from the monopolist tariffs T_L and T_H. Thus, for any possible scale of entry, it is also in the incumbent's interest that the entrant has the greatest per customer gross profits advantage and this implies cream skimming competition and surplus taking behaviour.

Thus, the incumbent in order to maximise his profits subject to the entrant's participation constraint will ask her to pay exactly his gross profit and to set K in such a way to maximise her gross profits K $[T^e_H - m(K)q^e_H]$ minus the net surplus $K[T_H - c^* q_H]$ that he would have extracted from the K customers in the absence of competition. Consequently we may set $K=K_H$ and state the competitor's optimisation problem as:

$$\max \Pi^e(Q) \equiv K(T^e_H - m(K)q^e_H) - [H + K(T_H - c^* q_H)] \text{ subject to:}$$

$$T^e_H = \theta\, u(q^e_H) - (\theta - 1)\, u(q^e_L) = \theta\, u(q_H) - u(q_L) = \theta\, u(q_H) - T_L$$

where T_H and T_L are the incumbent's optimal tariffs and q_H and q_L the related quantities obtained by the solution of stage (3). Assuming that K is a continuous variable we get the following first order conditions:

$[q^e_H]$ $\qquad p^e_H = \theta\, u'(q^e_H) = m$

$[K_H]$ $\qquad T^e_H - m(K)\, q^e_H - T - c^*\, q_H = Km'(K)\, q^e_H$

In particular from the last condition that states the equality between the incumbent's marginal losses and gains connected to a change in the number of customers served we can derive the optimal scale of entry K* for the entrant given the access charge $F(K^*)=H+K(T_H-c^*q_H)$ paid to the incumbent; that is, the number of customers the incumbent finds it optimal that the entrant should serve. In reality for K as a discrete variable we should require that the marginal gains connected to an unitary change of K starting from the optimal one K* be greater than the marginal losses.

Just in order to clarify the exposition let us provide a graphical representation of the problem, limiting ourselves to the case of linear marginal utility functions.

The incumbent is clearly interested in maximising only his own profits, not the entrant's ones. If the entrant decides not to serve an additional high-demand customer the incumbent doesn't lose all the area HMN which represents T^e_H-$m(K^*)q^e_H$, but only the area between p_H and m that is, IMDN equal to T^e_H-$m(K^*)q^e_H$ - T_H-c^*q_H, as it can be easily seen from fig. 2.10.

The optimal internal solution for the entrant in the case in which the variable k could assume all continuous values would be reached when the losses due to the fact that he is no longer serving an additional H type customer equal the gains derived by the decrease in the marginal cost m(K) upon the K residual customers; that is, $K^*m'(K^*)q^e_H$. In practice, $m'(K^*)$ represents just the reduction in the entrant's marginal cost due to the decrease in one unit of customer $m(K^*)$-$m(K^*$-$1)$, and is simply multiplied for the quantity sold; that is, $(K^*- 1)\, q^e_H = Q^e$. To represent this graphically we can simply add to fig. 2.10 another horizontal line $m(K^*- 1)$. The area between the two lines $m(K^*)c$ and $m(K^*$-$1)$ represents just the losses due to the fact that an H type customer is no longer served.

In any case, if (as it should be) the access charge F is equal to the entrant's gross profits $K^*[T^e_H - m(K^*)q^e_H]$ the access pricing condition (needed in order to solve the incumbent's maximisation problem) is practically the same as [AP] used in **[Problem 4]**, once we set K and m respectively equal to K* and m(K*):

[AP]' $F(K^*) = K^*[T^e_H - m\, q^e_H] = K^*[\theta\, u(q^e_H) - (\theta-1)\, u(q^e_L) - m(K^*)\, q^e_H]$

Thus, the incumbent, solving the maximisation problem **[Problem 4]** will obtain the very same monopoly first order conditions with respect to q^o, q_L and q_H, as derived in the previous analysis. As before, the competitor will enter in stage 2 knowing she will break even by serving the H type only. Thus, we can assume that she will enter and do cream skimming. In stage (1) of the game the incumbent, as already argued, cannot do better than to oblige the competitor to maximise her gross profits (behaving as a surplus taker) by setting the access charge tariff equal to the entrant's gross profits $K^*[T^e_H - m(K^*)\, q^e_H]$. As in the previous analysis, the fact that the competitor's marginal price p^e_H is less than the incumbent's one p_H implies a surplus loss -which amounts to the integral of $K^*[\theta u'(q) - u'(q_H)]$ for q that goes from q_H to q^e_H- for the incumbent and the society. If the marginal units $K^*(q^e_H - q_H)$ were purchased by the incumbent at their marginal cost m(K*) and then sold as infra-marginal units to the high-demand customers (or even to the low-demand ones) there would be no longer any surplus loss. We believe that no further comments are needed, as these refinements add more complications than fundamental economic contents.

2.7 Conclusion

The main conclusions that we can derive from this chapter is how pervasive is the *standard result on price discrimination* for pure monopoly (with **no distortion at the top)** in the absence of price regulation. We have shown how it can be directly applied in the presence of competition within different frameworks. In a very stylised model (which ignores vertically related markets) the incumbent can keep his monopoly profits and apply the relative tariffs in *residual* terms (in our model taking into account only residual customers). Extending the game to vertically related markets, the incumbent can entirely enjoy monopoly profits, as if he were the only player allowing the entry of an equally efficient competitor. In the presence of a more efficient rival he can enjoy an *additional* profit, so that he would be better off by allowing entry, rather than deterring it. There is a very clear intuition beyond this result. That is, the use of non-linear access charges allows the incumbent to control entry, in the sense of obliging the competitor to cream skim the market (for productive efficiency reason). Cream skimming turns out to be the only strategy of competition allowed by the incumbent.

In particular, when an equally efficient competitor may enter the market of the non monopolised good 1, it is optimal for the incumbent to set the per customer access charge equal to the monopoly variable profits and to maintain the previous monopoly pricing strategy, independently of the scale of entry. In other words, for the incumbent it is indifferent to face a duopoly or a monopoly situation in the market of good 1.

Moreover, allowing any finite discrete number of customers as scale of entry, there is only cream skimming competition in the market of the non monopolised commodity. It is optimal for the incumbent to allow entry, to set a per customer access charge equal to the monopoly variable profits and to maintain the previous monopoly pricing strategy. The incumbent also finds it optimal to oblige the competitor to behave as a surplus taker. However, it would be preferable for the incumbent (and for the maximisation of social welfare) that the competitor sells all the produced good 1 to the incumbent (reaching only break-even profits), so that the latter is able to resell it to the consumers, applying the monopoly tariffs.

It is worth noticing how, in this general framework, *cream skimming* arises naturally because of the vertical structure of the model and therefore does not depend on ad hoc hypotheses regarding the entrant, as in Laffont and Tirole (1990b). We have also shown how in many aspect it is wrong to consider this kind of competition as the most harmful, both in horizontally and vertically related markets.

This chapter, however, was just meant to provide a framework, starting from which we could analyse the problem of the direct interactions between regulation (including the policies toward price discrimination) and competition. We have spent quite a lot of time in modelling competition and non-linear pricing, since a priori there was no obvious way of approaching the problem of entry and in particular the strategies of competition.

We have departed from the standard approach (that is, modelling competition as simple bypass), using a more satisfactory way of dealing with the problem of competition in a second degree price discrimination model. In the next chapter socially optimal regulation will be introduced in the analysis, so that we can then derive some policy implications, following a more normative approach.

Chapter 3

SOCIALLY OPTIMAL REGULATION: NON-LINEAR PRICING, COMPETITION AND ACCESS PRICING*

* Vagliasindi and Waterson (1995b) summarised and partly developed the analysis of section 3.3. This paper has been presented at the Annual Conference of Game Theory and Applications (3-4 March 1995, Siena, Italy) and at the 22nd EARIE Conference (3-6 September 1995, Sophia Antipolis, France). I am grateful to Jonathan Cave, Gianni de Fraja, Morten Hviid, Norman Ireland, John Vickers and participants to the above conferences for helpful comments.

Chapter 3

SOCIALLY OPTIMAL REGULATION: NON-LINEAR PRICING, COMPETITION AND ACCESS PRICING

3.1 Introduction.

Using the results of the previous chapter, where we analysed the private incentives of the incumbent and the entrant in the absence of regulatory constraints, we are now able to explore the direct interactions between regulation *and* competition. Following a *normative* perspective we are going to determine the socially optimal access pricing rules, that is the ones that maximise social welfare. We will focus on *regulating access* directly, a kind of regulation recently subject of several controversies in the UK and other countries, especially in the case of telecommunications.

The access pricing problem has been lately object of a debate in the economic literature and in the political arena. The optimality of the application of a simple rule, such as the Baumol-Willig rule (recently applied to telecommunications in New Zealand) has been re-discussed, among others, by Baumol and Sidak (1994), Armstrong and Doyle (1994) Laffont and Tirole (1994b), Armstrong and Vickers (1995) and Economides and White (1995).

Naturally, the first question that should be addressed before discussing the design of socially optimal tariffs (and related issues) is whether government intervention is needed at all (i.e. if competition would solve the whole matter). Traditional economic views assume that *competition* is a desirable regime in itself and that it is always feasible. In this world there would be no space for regulators and public firms. The predilection for competition as a process automatically generating efficiency arises from the two well known fundamental theorems of welfare economics. Theoretical models of perfect competition have several attributes (among others, the fact that the market is composed by a large population of passive price takers firms) that make them non applicable to questions of economic regulation of public utilities (naturally apart from the treatment of externalities and environmental issues). Regulation instead generally addresses phenomena such as monopoly and market power, phenomena that clash with the hypothesis of an infinite number of small firms with no market power. This explains our particular care in modelling the strategies of competition (cf. section 2.4) and the decision to replace the assumption of a perfectly competitive fringe with a single entrant. Moreover, economic efficiency requires prices equal to marginal costs. However, even in a first best setting (in the absence of any redistributional issues) we will show how, only on efficiency grounds, the public firm should *not* set marginal access pricing equal to his marginal cost if doing so he allowed the entry of a less efficient competitor.

Clearly, the aims and objectives of regulation are several and can be partly contradictory. They include not only allocative and productive efficiency, but also the promotion of competition and the respect of universal service obligations (or more generally of a criterion of fairness). Moreover, information and incentives issues complicate the intervention of regulators. However, the amount of private information on the regulatory side built into the standard models seems too limited by comparison with the problems actual regulators face. There is no reason why the regulator should be merely uncertain about the level of the cost function, as in Laffont and Tirole (1994) and Vickers (1995), and not about its shape, or the demand condition of the industry. Furthermore, as Vickers (1995) says "It would be nice to examine a model in which the regulator was uncertain about both upstream and downstream cost levels" (p. 4), but till now no model assumes that the entrants' costs are private information. Also the usual way in which moral hazard has been introduced is not fully satisfactory (as seen in section 1.2). Consequently, we decide to leave aside problems of moral hazard and adverse selection on the regulatory side to make the problem more manageable and to reveal its fundamental structure. These assumptions are quite usual in the case of a *public firm* which directly maximises social welfare, as well as in the large part of the debate on access pricing and vertical issues. In chapter 1 (section 1.2.4) we have also shown that the accounting convention of Laffont and Tirole, which we will continue to follow, perfectly fits a public firm.

We consider instead only asymmetric information between the firm and its consumers, focusing on second degree price discrimination. In fact, despite the extensive literature on this field theoretical models fail to incorporate the use of *non-linear pricing*. Since price discrimination by regulated firms is common in practice this omission can lead to misleading results, as our analysis will show. Specifically, arguments against the use of regulation might not be justified and simple pricing rules might no longer be socially optimal when the incumbent is allowed to use price discrimination.

Therefore in what follows, we will exclusively deal with socially optimal non-linear pricing regulation of the incumbent under full information, taking into account three different social welfare specifications which may represent the objective of the regulator. Dealing directly with a public firm the regulator's maximisation problem coincides with the incumbent's one.

Even within this simplified framework and postulating at first the absence of vertical relationships -examining the horizontal game (proposed in section 2.2.1)- regulation is needed. Our analysis will in fact show that socially optimal non-linear tariffs do not coincide with the ones private firms would choose, even in the presence of competition. Regulation is needed to deter the entry of a less efficient competitor and to ban competition for low-demand customers. Specifically, it is always welfare improving to create an eventual budget deficit of the inefficient competitor with negative lump sum transfers, ensuring that she will not be able to enter even the high type market.

It is then relevant to examine the impact of cream skimming competition by a more efficient entrant on social welfare. We have already shown in chapter 2 (section 2.4.3) that, starting from a private monopoly, cream skimming is likely to be less harmful than skimmed milk competition. Specifically, in a private setting cream skimming is socially desirable both from the productive and allocative perspectives. Basically, this occurs because profit maximising firms, giving no weight to consumers' surplus, tend to distort prices by a greater amount than it would be socially optimal. Here, starting from a public monopoly (or equivalently, a private regulated firm), the only benefits cream skimming can bring are related to the productive side. In fact, it will be shown that a public firm maximising social welfare (defined a la Loeb Magat or a la Baron Myerson) would indeed create no distortion both at the top and at the bottom. Consequently, cream skimming competition would not bring any improvement in allocative efficiency. It may instead cause a departure from socially optimal tariffs introducing the cost of public funds (a la Laffont and Tirole). This last problem can be solved through specific forms of intervention such as the institution of *entry taxes* or the regulation of the entrant. In both of these scenarios the public monopoly tariffs would be kept, so that cream skimming may be used by the regulator in order to enhance social welfare, through its positive impact on productive efficiency.

Also within the vertical game (introduced in section 2.2.2) it is relevant to determine the socially optimal tariffs for final and intermediate markets, and the way they can be eventually implemented. An important question we address introducing vertically related markets is indeed the determination of the range of circumstances in which *access* pricing can be used to bring about a welfare enhancing competitive solution to final goods supply, whilst a monopoly remains at one essential point in the chain of delivery. In fact, it is *not* always the case that the entry of a competitor (even if more efficient than the incumbent) automatically increases welfare in a general price discrimination setting. Moreover, we will also briefly discuss various other forms of interventions which may enhance social welfare, such as the institution of an entry tax or the regulation of the entrant.

The previous considerations remind us the original Baumol's subtle pricing regulation issues of 1983, where he looked for efficiency improvements in vertically related markets, showing that paradoxically a collusive result may be preferable to a Cournot-like competitive one. Baumol proposes a rule that attains some of the efficiency improvements implicit in a potential vertical merger. According to this rule, charges should be set equal to the marginal cost of access plus a term which reflects the opportunity cost of entry. The validity of this rule will be examined in detail in his original framework, as well as in the cream skimming model proposed in chapter 2, where we also introduce non-linear access charges.

Specifically, two issues which we are interested to explore in both games are: (a) the necessity that entry may be allowed only if an amount, that reflects the opportunity cost of entry incurred by the public incumbent, is paid, (b) the idea to exploit the merger's efficiency enhancing properties, or more precisely the more general idea of "internalisation" (of which the vertical merger represents just a specific application). The first point will prove quite relevant with all the welfare function specifications and with some welfare functions a strictly more efficient competitor (able to pay an amount greater than the opportunity cost of entry) would be required. Only with an entry tax or a regulated entrant, the entry of an equally efficient competitor will always be optimal.

The critiques which has been recently moved to the Baumol-Willig rule are quite specific and crucially derive from the relaxation of some of the basic assumptions made in the original framework. In this regard our approach is different, since we move in a setting which is quite close to the spirit of Baumol's model. Specifically, we will assume that the public firm is not operated inefficiently, both the vertically integrated incumbent and the rival produce homogeneous final goods, there are fixed coefficients in production and there is no possibility of bypassing the access provider. Baumol himself pointed out the need for corrections in this simple rule, once we move away from these assumptions. As we won't deal with these issues, let us just provide a brief account of the most relevant extensions of the Baumol-Willig rule, which we will discuss in section 3.3.

Both Cave (1994) and Armstrong and Doyle (1994) consider a correction of the parity rule related to the introduction of differentiated products. This relatively straightforward generalisation is in substance implicit in the definition of the components of the Baumol-Willig rule, if correctly specified. Armstrong and Vickers (1995) introduce alternatively substitution possibilities, variable proportion coefficients and bypass in a framework in which the entrant is characterised by a finite elasticity of supply (rather than an infinite one). They analyse the relevant assessment of opportunity costs in this setting, showing the need to reduce this component of the Baumol-Willig rule. As in Armstrong and Doyle (1994)'s model in practice prices and quantities are exogenously given. Moreover, there seems to be an artificial asymmetry between the entrant and the incumbent; in particular the elasticity of supply of the incumbent (in contrast to the one of the entrant) does not appear in any point of the analysis. Economides and White (1995) pointed out that it can be socially beneficial to allow the entry of an inefficient rival in the presence of complementary components. Specifically, even the entry of an inefficient competitor may bring about a reduction in the price of the complementary component .

Finally, let us consider a general criticism of the Baumol-Willig rule, which would apply to any other access pricing rule. Basically it abstracts from the environment in which access is provided, and, consequently it doesn't consider altogether the regulation of final products' prices. Baumol and Sidak (1994) recommend to impose a price ceiling -based on the stand-alone cost- and a price floor in order to prevent predation on entrants. Laffont and Tirole (1994b) show how Ramsey pricing can be obtained by imposing a global price cap -including access as a final good- on the incumbent, with weights exogenously determined and proportional to the goods' forecasted quantities. They criticise the use of a partial price cap together with the efficient component pricing rule, since it provides a form of subsidy to non competitive segments (in the sense that access charge are considered too high).

Finally, notice how the majority of these contributions, like ours, abstracted from information and incentive questions, with the only exception of Laffont and Tirole (1994a) where, however, thanks to the dichotomy property pricing and incentives issues are separated. Moreover, as noticed before, these papers entirely fail to consider the use of non-linear access charges. The aim of our approach is instead to offer a comprehensive analysis of socially optimal access pricing within a second degree price discrimination setting.

In the cream skimming model we will show that the Baumol-Willig rule allows entry of an equally efficient competitor, who faces no fixed costs, apart from the access charge, and it represents the minimum charge to apply to avoid inefficient entry. Starting from a private equilibrium, welfare is enhanced applying the Baumol-Willig rule when the retail average price and the access charge are decreased with all the welfare function specifications.

In a *Loeb and Magat*'s setting, where distributional considerations are left aside (and social welfare is simply the unweighted sum of profits and consumers' surplus), we will see how "parity" unit access charges are socially optimal, differently from what implicitly assumed by Baumol and explicitly by Armstrong and Doyle (1994). Hence, under socially optimal non-linear pricing, the presence of lump sum transfers may be compatible with the Baumol-Willig rule.

In this framework Pareto efficiency will be achieved (when the entrant behaves as a surplus taker, maximising her own profits) since the rule would not allow entry by an inefficient competitor and the marginal prices equate the marginal costs. However, in reality, considering the welfare function as the unweighted sum of the consumers' and the producer's surplus it does not matter if the incumbent would instead set an higher access charge and expropriate (partially or totally) the more efficient competitor's profits.

When the public (or private regulated) firm's profits are given an additional weight λ being costly to get public revenue from taxes, following the definition of social welfare of Laffont and Tirole, the Baumol-Willig rule will not always represent the socially optimal access pricing rule. In fact in this case it is socially optimal to totally expropriate the most efficient competitor's profits, by setting a *higher* access charge than the one it would derive from the application of the parity rule. More precisely, the per customer profits of the incumbent, having a weight greater than unity, are strictly preferred to the ones made by an efficient competitor, even if he pays corporate taxes on them (being the tax rate normally less than unity).

Similarly to the previous case, following the Baron and Myerson's approach, it will be always socially optimal that the incumbent expropriates all the entrant's profits, as the incumbent's profits are now given a lower weight (in the welfare function). But, like in the ultra-liberal case, now there will be no distortion at the bottom; hence, Pareto efficiency will be achieved with marginal prices equal to marginal costs. In these two cases the application of the Baumol-Willig rule alone will lead to sub-optimal results because, differently from what implicitly assumed by Baumol in his framework, the merged monopolist's optimal pricing criteria are *not* socially optimal in these two contexts.

To sum up, following uniquely an efficiency point of view, the Baumol-Willig rule remains a useful benchmark for all institutional settings characterised by the different functional specifications of social welfare mentioned before. However, as we will show, the welfare enhancing role played by competition should complement rather than replace regulation. Specifically, in vertically related markets, allowing cream skimming competition, we have only an improvement in productive efficiency, as the incumbent fully takes advantage of the greater efficiency of the competitor, through the use of non-linear access charges. But competition by itself will not reduce the high monopoly pricing distortion in the market for final goods and the public (or regulated) firm will continue to behave as he were the only player, independently of the scale of entry.

Therefore, it will be evident that the Baumol-Willig rule, built upon the vertical merger's efficiency properties, does not provide general socially optimal pricing policies, differently from the more general "internalisation criterion" which should instead be considered more carefully.

Just in passing, notice how the regulator can intervene more radically changing the *structure* of the industry, for instance by imposing a divestiture between the upstream and the downstream sector. This is exactly the case of AT&T in the USA in the telecommunication case. Also in the UK this approach has been adopted in the electricity which has been separated into three vertical levels; business separation is also taking place in gas. Oftel has recently reviewed the whole framework of interconnection to British Telecom's network and accounting separation of network activities has been achieved. Problems and distortions brought by the presence of vertically separated structures will be postponed to the next chapter.

The structure of this chapter is as follows. Before tackling vertically related markets, we derive some preliminary welfare considerations in the horizontal game (partly introduced in chapter 2) in order to analyse the optimal behaviour of a regulated or of a public incumbent firm in the absence of vertical issues (section 3.2). In section 3.3 we analyse the socially optimality of the Baumol-Willig rule in the original Baumol's setting and in our vertical game. A final section (3.4) summarises the main results achieved by our analysis.

3.2 Preliminary welfare implications in the horizontal game

Before tackling vertically related markets and non-linear access pricing regulation, let us derive some preliminary welfare considerations in the horizontal game. The aim of this section is to briefly analyse the optimal behaviour of a regulated or of a public incumbent firm in the absence of vertical issues. In particular, we will exclusively deal with socially optimal non-linear pricing regulation of the incumbent under full information, taking into consideration three different social welfare specifications which may represent the objective functions of the public incumbent firm.

A general implication that will emerge is that competition should complement rather than replace regulation. Recent ultra-liberal views assume that competition is a desirable regime in itself, and that it represents always a feasible solution. The general idea of competition as a process automatically generating efficiency has been justified on the basis of the two well known fundamental theorems of welfare economics. However, even in this first best setting, competition cannot be considered as an end in itself, but just only possible means in the hand of the public policy maker (whose general objective is to reach the so called "Pareto efficiency" and "interpersonal equity").

Moreover when efficiency consideration entirely dominates the scene, not only should the public firm set marginal pricing equal to his marginal cost, but also deter the entry of a less efficient competitor. In other words, it is always welfare improving to create an eventual budget deficits of the inefficient competitor with negative lump sum transfers, ensuring that she will not be able to enter even the low-demand market.

3.2.1 Preliminary remarks on regulation

In the previous analysis we have already noticed how competition by an inefficient rival is welfare reducing; this consideration emerges clearly in the horizontal model without access pricing (introduced in section 2.2.1 and re-examined in section 2.3.1 and 2.4). In fact, even when both the incumbent and the entrant maximise profits, the total consumers' surplus is reduced because some consumers will face the competitor's higher costs and also because, as she will serve the L type, price distortion will be increased by the incumbent in order to maximise his profits.

To show this, here we will first refer to efficiency problems in a first best setting, considering what it may be called an ultra-liberal welfare function. That is, we simply define welfare as the unweighted sum of consumers' and producer's surplus a la *Loeb and Magat* (1979) adopting quasi linear utility functions. In this way, we go beyond the basic utilitarist (or benthamite) welfare function, because we directly assume equal marginal utility of income for any consumers, eliminating at its root any scope for redistribution.

Even under full information on the cost and effort of firms, other types of concerns, apart from productive and allocative efficiency considerations, should be taken into account. For instance, second best considerations may emerge if we impose a binding break even budget constraint when public subsidies are not available. In the previous reasoning we have completely neglected the fundamental problems determined by the fact that taxes normally change the behaviour of economic agents, creating unavoidable distortions (due to the presence of suboptimal efficiency). This issue may be roughly considered introducing the cost of public funds, as notably done by *Laffont and Tirole* (1986). Moreover, adopting the approach of *Baron and Myerson* (1982), specific redistribution concerns emerge as the policy maker puts a higher weight on consumers' surplus, differently from the previous welfare functions.

In the analysis of the horizontal game we first referred to the negative consequences in terms of *productive* and/or *allocative efficiency* which arise in correspondence of: (1) the entry of an inefficient competitor -that takes place whenever the entrant's marginal costs are higher than the incumbent's ones, (2) skimmed milk competition, even if undertaken by an equally or more efficient competitor.

The negative impact on consumers' surplus directly follows from the analysis of the optimal relationship between the two marginal prices (and utilities) derived in the horizontal game:

$$p_H = \theta\, u'(q_H) = [1 - (\theta - 1)\, N_H/N_L]\, u'(q_L) = [1 - (\theta - 1)\, N_H/N_L]\, p_L = c^*$$

In fact, in these cases the negative effect on social welfare is not only due to the fact that probably the competitor faces higher costs, but also specifically to the increase in the marginal price p_L. In this way, the consumers' surplus is reduced, because the quantity sold to the L type decreases and the price distortion is increased by the incumbent in order to maximise his profits.

Consequently, it is clear that only cream skimming competition by more efficient entrants should be favoured, because it can be welfare enhancing, since: (1)

it implies that some of the high customers are served by the competitor at lower costs, (2) it brings a reduction in the tariffs applied to the low type of customers. In fact, cream skimming competition decreases the marginal price distortion automatically, since the relative weight of low types increases, reducing the distortionary factor (N_H/N_L). Notice how it is in the incumbent's own interest (in order to maximise his profits) to decrease the L type's marginal price p_L and it is not a consequence of regulatory constraints.

However, entry in the skimmed milk market *may* be welfare enhancing in particular cases in which *competition* is *regulated* and the rival's efficiency is sufficiently close to the incumbent's one. There may be in fact a trade-off between allocative and productive efficiency. Consider the case of a less efficient competitor (characterised by a value of her marginal cost intermediate between the incumbent's marginal cost and his optimal tariff for the low type, that is, $c^*<m< p_L$) obliged to serve only the low type, because of, say, a ban on cream skimming. Acting optimally as a surplus taker, the rival will set her tariff p_L^c below the one fixed by the incumbent p_L. In this way the incumbent, if his tariffs are not regulated, in order not to lose high-demand customers (who find it convenient to mimic low-demands and buy the bundle offered by the rival), must reduce the low customers' price distortion. In a different framework, with linear access prices (and with a standard intermediate demand curve) Economides and White (1995) show that competition by a less efficient rival could yield net social gains. As in our case, the improvement in terms of allocative efficiency may outweigh the losses in productive efficiency.

Apart from these particular cases, in what follows we will show how, in the absence of access pricing issues, regulation should not allow competition for the low-demand customers or by inefficient entrants. This result holds for all different social welfare specifications, as summarised in the table (p. 276).

3.2.2 Social optimality with Loeb and Magat's welfare function

In what follows, as a first approximation, following Loeb and Magat (1979) we consider a social welfare function simply given by the unweighted sum of consumers' surplus $U(Q^i, Q^e)$ and profits Π^i - tr + Π^e (specifically, the profit of the incumbent are net of transfers tr), disregarding any eventual distributional and excess burden issues.

[LM.1] $$W = U(Q^i, Q^e) + (\Pi^i - tr + \Pi^e) = S(Q^i, Q^e) - (C^i + tr + C^e) =$$
$$= N_H\theta u(q_H) + N_L u(q_L) + K_H\theta u(q^e_H) + K_L u(q^e_L) - C(Q^i) - tr - C^e(Q^e)$$

where as usual the two superscripts i and e refer respectively to the incumbent and the entrant.

Applying Laffont and Tirole's accounting notation, social welfare can be expressed as the sum of consumers' surplus less the total cost of the project (including the transfers from the authority to the public firm). In fact, remembering that the consumers' net surplus is equal to the consumers' gross surplus minus firms' revenues, that is, $U(Q^i, Q^e) = S(Q^i, Q^e) - R(Q^i) - R(Q^e)$, we see how in the formula the revenues drop out. The consumers' gross surplus function is specified as:

[LM.2] $$S(Q^i, Q^e) = N_H\theta u(q_H) + N_L u(q_L) + K_H\theta u(q^e_H) + K_L u(q^e_L)$$

where $N_t = N - K_t$ represent the incumbent's residual customer of type t (K_t is in fact the scale of entry in market t, in terms of number of customers).

Let us now consider the behaviour of a social welfare maximising public incumbent firm in this setting. A public monopolist would automatically equate the low type marginal price p_L and his marginal cost. This can be easily demonstrated in analytical terms, solving the public firm's social welfare maximisation problem with respect to good 1, the non-monopolised final product. In the monopoly case ($N_H = N_L = N$), neglecting fixed costs, the social welfare maximisation problem can be written as:

[LM.3] $$\max W = N[\theta u(q_H) + u(q_L)] - c^*N[q_H + q_L]$$

$[IR_L]$ $$u(q_L) - T_L \geq 0$$

$[IR_H]$ $$\theta\, u(q_H) - T_H \geq 0$$

$[IC_L]$ $u(q_L) - T_L \geq u(q_H) - T_H$

$[IC_H]$ $\theta\, u(q_H) - T_H \geq \theta\, u(q_L) - T_L$

Assuming that the public firm can keep positive profits, appropriating customers' surplus (to keep our analysis as close as possible to Loeb and Magat, who deal with a privately regulated firm), as in the standard model the upward binding constraint $[IC_L]$ and the participation constraint $[IR_H]$ are automatically satisfied by the solution of the problem when $[IR_L]$ and $[IC_H]$ are binding. Hence, no surplus is allowed to the L type, whereas the H type should always enjoy a positive net surplus [given by $(\theta - 1)u(q_L)$].

The welfare maximisation problem can be easily solved with respect to q_L and q_H, once we substitute the two binding constraints $[IR_L]$ and $[IC_H]$ into the objective function. Making use of consumers' optimality conditions, i.e. $p_t=\theta_t u'(q_t)$ from the first order conditions we can see that an equality between the two marginal price must hold:

$[q_H^w]$	$p_H^w = \theta\, u'(q_H) = c^*$	No distortion at the top
$[q_L^w]$	$p_L^w = u'(q_L) = c^*$	No distortion at the bottom

Hence, the public firm finds it optimal to equate marginal prices to marginal costs. Note how this result would also hold if customers of type H and L served by the public firm were present in a different number. It is also important to notice that, since lump sum transfers are allowed, this condition does not necessarily imply that the L type customers enjoy no surplus at all (as assumed instead just for simplicity's sake). On the other hand, since the usual downward incentive compatibility constraint $[IC_H]$ is binding any positive surplus given to customers of type L must also be given to high type customers.

Let us now tackle competitive issues, considering as before a public firm. In the presence of a tariff taker rival, it is easy to show that social welfare a la Loeb and Magat reaches the maximum in the case of cream skimming competition undertaken by a more efficient rival. In fact, social welfare derived in the presence of entry [i.e.

$N_H\theta u(q_H) + N_L u(q_L) - c^*(N_H q_H + N_L q_L)] + [K_H\theta u(q_H^e) + K_L u(q_L^e) - m(K_H q_H^e + K_L q_L^e)]$: (1) decreases if the entrant is less efficient ($m>c^*$), (2) stays the same if the entrant is equally efficient ($m=c^*$) and (3) increases only if the entrant is more efficient ($m<c^*$).

Social welfare is reduced if an inefficient competitor ($m > c^*$) enters the market. In fact, if entry takes place, social welfare per unit of customers decreases from the level $[\theta_t\, u(q_t) - c^*\, q_t]$ to $[\theta_t\, u(q_t^e) - m\, q_t^e]$ for the K customers served by the rival. With an efficient competitor ($m < c^*$) the maximum additional gain for society per unit of customers, that is $[\theta_t\, u(q^e_t) - m\, q^e_t] - [\theta_t\, u(q_t) - c^*\, q_t]$, will be achieved with cream skimming competition (i.e. for $t = H$).[40]

Hence, if the regulator does not impose any ban on competition for low-demand customers or by inefficient entrants, the public firm himself must make these socially undesirable forms of entry unprofitable for the entrant. This can be done only if he distorts his tariffs; specifically if the L type's tariff is reduced by an amount S_L, so that the entrant would not break even, being the maximum variable profits earned by a surplus taker competitor serving the low type $K[u(q_L^e) - S_L - mq_L^e]$ negative. However, since customers of type L enjoy a surplus S_L also customers of type H must be given the same additional positive surplus S_L in order to satisfy their binding incentive compatibility constraint $[IC_H]$. In the case in which the public firm does not break even by adopting these lower tariffs for both types of customers, he must be given adequate lump sum transfers.

Alternatively, the public authority may directly impose an *entry tax* equal to the foregone variable profits of the public firm (if K_H customers of type H and K_L customers of type L were taken away by the competitor). This payment amounts to $(K_H\, \pi_H^* + K_L\, \pi_L^*)$, where $\pi_t^* = T_t - c^* q_t$ represents the variable profit enjoyed by the public firm per unit of customer of type t. In this way, a less efficient competitor

---

[40] Notice how this reasoning applies perfectly to a private firm maximising profits the only difference being the presence of additional benefits, arising from the reduction of distortion on the low consumers associated with cream skimming, as already seen in section 2.4.3.

(m>c*) is prevented from entering the market and a more efficient competitor (m<c*) will find it optimal to enter and cream skim the market. In fact, taking the point of view of the entrant, a more efficient competitor (m<c*) will maximise her variable profits $K_H T_H^e + K_L T_L^e - m(K_H q_H^e + K_L q_L^e) = K_H \theta u(q_H^e) - K_H(\theta-1)u(q_L^i) + K_L u(q_L^e) - m(K_H q_H^e + K_L q_L^e)$ only when she cream skims the market. She will also find it optimal to behave as a surplus taker, setting marginal price equal to her marginal cost.

In order to clarify the reasoning, let us represent graphically the case of the entry of a more efficient rival. As represented in fig. 3.1 the public firm's tariffs are set equal to the marginal cost c* for both types of customer: T_L is the area COq_L^wI and to obtain the tariff T_H we must add the area $Eq_L^wq_H^wD$.

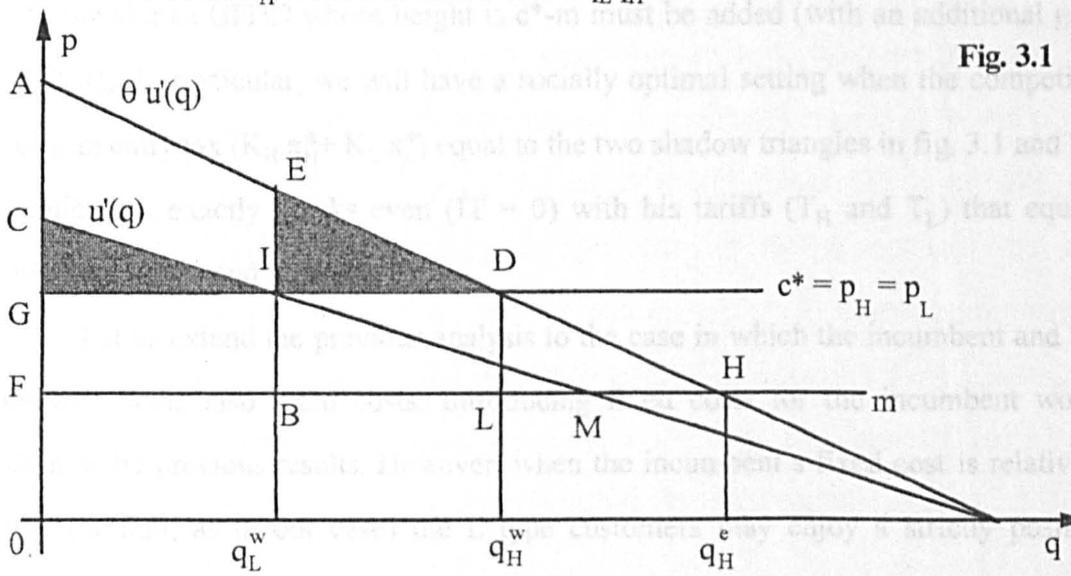


Fig. 3.1

Since the non-linear tariffs T_L and T_H are maintained unchanged by the public firm, independently of the scale of entry, the best a tariff taker entrant can do is to go for the H type and to get a gross profit per customer given by CFBI + EBLD (always greater than CFBI, namely the gross profit she would get serving the L type). It can also be shown that the optimal pricing strategy for the entrant is reached when she behaves as a surplus taker, that is she leaves to the high type the same surplus of the incumbent (the area ACIE, the minimum net surplus necessary in order to respect the incentive compatibility constraint). Doing this she maximises her profits by setting the marginal price equal to the marginal cost $p_t^e = m < c^*$. A lower marginal price would cause her losses, while a higher one will make her forego some profits (as in the tariff taker case). Also a surplus taker competitor finds it optimal to go for the H type and to get a gross profit per customer given by CFBI + EBH (always greater than CFM, namely the gross profit she would get serving the L type). It is self evident that the gross profit per unit of customer of type H enjoyed by a surplus taker competitor is greater than the one of a tariff taker by an amount given by the area DLH. Hence, if the competitor is more efficient she will enter and cream skim, behaving as a surplus taker (i.e. setting $p_H^e = m < c^*$).

This type of entry is socially optimal since social welfare per unit of customers $\theta u(q_H^w)$ - c^* q_H^w is less than $\theta u(q_H^e)$ - m q_H^e. As the previous variable profits per H customer are given by the shadow area of the two triangles CGI and EID the additional area GFHD whose height is c^*-m must be added (with an additional gain of DLH). In particular, we will have a socially optimal setting when the competitor pays an entry tax ($K_H \pi_H^* + K_L \pi_L^*$) equal to the two shadow triangles in fig. 3.1 and the public firm exactly breaks even ($\Pi^i = 0$) with his tariffs (T_H and T_L) that equate marginal prices and marginal costs.

Let us extend the previous analysis to the case in which the incumbent and the entrant incurs also fixed costs. Introducing fixed costs for the incumbent won't change the previous results. However, when the incumbent's fixed cost is relatively low (or null, as in our case) the L type customers may enjoy a strictly positive surplus, differently from what assumed above. Anyway, from an allocative point of view matters won't change, as the firm's profit and the consumers' surplus are given the same weight. On the other hand, when the incumbent's fixed cost is quite high the incumbent might need a lump sum transfer; a transfer which should always be less than the net surplus of high customers (ACIE) times the number of high-demand customers.

Instead, in the case of the entrant the presence of high fixed costs may penalise her, not allow to enter the market. Also from the social welfare perspective her entry may not be optimal, even if her efficiency was high enough to make entry profitable from her private point of view. Specifically, even if the entrant's marginal cost are lower than the incumbent's one, entry would not be welfare enhancing if the fixed cost is greater than the additional total variable profit (that is, the area GFHD times the number of high demand customers served).

Finally, let see how socially optimal tariffs can be obtained if we deal with a private firm which maximises profit. Specifically, in Loeb and Magat's setting (allowing for lump sum transfers) once inefficient entry is prevented, price regulation of the incumbent (here a private regulated firm) with a single cap on the average price of the L type may allow to reach even the first best. In fact, Pareto efficiency can be achieved setting the average price for low-demand customers in a way such that the marginal price p_L is equal to the incumbent's marginal cost.

In the specific case of linear marginal utility functions, the required price cap for low-demand customers should be equal to the average between u'(0) and c*, as this will lead to set the marginal price equal to c*. Under this constraint, the incumbent finds it optimal, in order to maximise his profits, not to create any distortion at the bottom, while keeping no distortion at the top. Hence, customers of type L will be offered a bundle such that the marginal price equals the marginal cost (that is, p_L=c*). Eventually, apart from the price cap, also positive lump sum transfers to the incumbent would be needed if the firm will not break even.

3.2.3 The welfare consequence of the cost of public funds in the horizontal game

It is possible to consider efficiency issues more in depth, introducing the cost of public funds and considering the managerial effort e and its disutility $\psi(e)$, following Laffont and Tirole. Within their approach, however, only efficiency issues are considered (as welfare remains the unweighted sum of consumers' and producer's surplus in the ambit of quasi linear utility functions) and transfers to the public firms and revenues will be given a weight equal to $(1+\lambda)$ and greater than unity. The moral hazard issue, introduced only in the form of managerial effort with risk neutral agents, has not a great relevance in a full information setting, and, furthermore, it does not imply any substantial change even in the ambit of their adverse selection model, as proved in section 1.2. Consequently, we will carry on the analysis disregarding the issue of moral hazard, represented by the effort e and its disutility $\psi(e)$.

Following Laffont and Tirole's assumptions within our public firm's setting, the collectivity will pay costs and transfers and get the revenues, which as a direct consequence, will be given a weight equal to $(1+\lambda)$ greater than unity. Taking into account distributional and excess burden issues (deriving from distortionary taxation) we consider a social welfare function given by the unweighted sum of consumers' surplus $U(Q^i, Q^e)$ and profits (specifically, the profit of the incumbent are net of transfers tr, Π^i - tr + Π^e).

[LT.1] $$W^{\lambda} = U(Q^i, Q^e)+(1+\lambda)(\Pi^i\text{-tr})+\Pi^e = S(Q^i, Q^e)+\lambda R(Q^i)\text{-}(1+\lambda)(C^i\text{+tr})\text{-}C^e$$
$$= N_H(1+\lambda)\theta u(q_H) + [N_L(1+\lambda)\text{-}\lambda N_H(\theta\text{-}1)]u(q_L) \text{-}(1+\lambda)c^*(N_L q_L+N_H q_H)$$
$$- (1+\lambda)tr + K_H\theta u(q_H^e) + K_L u(q_L^e) - C^e(Q^e)$$

Applying Laffont and Tirole's accounting notation, social welfare can be expressed as the sum of consumers' surplus less the total cost of the project (including the transfers from the authority to the public firm) times the shadow cost of public funds $(1+\lambda)$. In fact, since the consumers' net surplus is equal to the consumers' gross surplus minus firms' revenues, that is, $U(Q^i, Q^e) = S(Q^i, Q^e) - R(Q^i) - R(Q^e)$, we see how in the formula the revenues of the entrant drop out.

Hence, in the monopoly case, the consumers' net surplus may be written as $S(Q) = U(Q)+\lambda R(Q)$. The public firm will maximise the social welfare function $W(Q) = S(Q)+\lambda R(Q)\text{-}(1+\lambda)(c^*Q\text{+tr})$ with respect to the tariff system $\{T_H, q_H, T_L, q_L\}$ subject to the individual rationality and the incentive compatibility constraints:

[LT.2] $\max W^{\lambda} = N[\theta u(q_H)+u(q_L)] + \lambda N(T_L+T_H)\text{-}(1+\lambda)c^*N(q_L+q_H)$ subject to:

$[IR_L]$ $u(q_L) - T_L \geq 0$

$[IR_H]$ $\theta u(q_H) - T_H \geq 0$

$[IC_L]$ $u(q_L) - T_L \geq u(q_H) - T_H$

$[IC_H]$ $\theta u(q_H) - T_H \geq \theta u(q_L) - T_L$

As with the previous Loeb and Magat welfare function, the constraints $[IC_L]$ and $[IR_H]$ are automatically satisfied by the solution of the problem when $[IR_L]$ and $[IC_H]$ are binding. Hence, even in this case, the only relevant constraints are:

$[IR_L]$ $\qquad T_L = u(q_L)$

$[IC_H]$ $\qquad T_H = \theta\, u(q_H) - (\theta - 1)\, u(q_L)$

Therefore, in this second best setting, no surplus should be allowed to the L type. Welfare can be easily maximised only with respect to q_H and q_L once we substitute the individual rationality and the incentive compatibility constraints in the objective function:

[LT.3] $\qquad \max W^{\lambda} = N[\theta u(q_H)+u(q_L)]+\lambda N[(2-\theta)u(q_L)+\theta u(q_H)]-(1+\lambda)c^*N(q_L+q_H)$

Making use of the optimality conditions which hold for the customers we can easily write down the first order conditions as:

$[q^{\lambda}_H]$ $\qquad p^{\lambda}_H = \theta\, u'(q_H) = c^*$ $\qquad$ No distortion at the top

$[q^{\lambda}_L]$ $\qquad p^{\lambda}_L = u'(q_L) = c^*(1+\lambda)/[1+\lambda(2-\theta)]$ $\qquad$ Distortion at the bottom

Clearly, we have some distortion at the bottom, as in the case of a private monopolist examined in the previous chapter; the two cases coincide only for λ which tends to infinite. The distortion here arises from the fact that profits are now given an additional weight λ, being costly to get public revenues from taxes. From the previous first order conditions we obtain the optimal relationship between q_L and q_H, similar to the one holding in the monopoly case:

[LT.4] $\qquad \theta\, u'(q_H) = [1+\lambda(1-\theta)/(1+\lambda)]\, u'(q_L)$

Notice how in this case it is always optimal that customers of type L enjoy no surplus at all, since the public (or regulated) firm's revenues have a weight equal to $(1+\lambda)$, which is strictly greater than unity (as with Loeb and Magat's welfare function), which is the weight of the low demand customers' net surplus .

We can now reformulate the basic Laffont and Tirole welfare maximisation problem **[LT.2]** introducing competition following the same procedure as in the previous chapter (i.e. examining the public firm's problem with respect to the residual number of customers served by him).

[LT.5] $\max W^{\lambda} = N_H(1+\lambda)\,\theta u(q_H) + [N_L(1+\lambda) - \lambda N_H(\theta-1)]\, u(q_L) +$

$$- (1+\lambda)c^*(N_L q_L + N_H q_H) - (1+\lambda)tr + K_H \theta u(q_H^e) + K_L u(q_L^e) - C^e(Q^e)$$

The first order conditions are:

$[q_H^{\lambda}]$ $\quad p_H^{\lambda} = \theta\, u'(q_H) = c^*$ — No distortion at the top

$[q_L^{\lambda}]$ $\quad p_L^{\lambda} = u'(q_L) = c^*/[1 + (N_H/N_L)\,(1-\theta)\lambda/(1+\lambda)]$ — Distortion at the bottom

While the "no distortion at the top" result still holds, the socially optimal "distortion at the bottom" depends crucially on the scale of entry (as testified by the presence of the ratio N_H/N_L). This conclusion is opposite to the one reached in the previous subsection adopting Loeb and Magat's welfare function, where instead the marginal prices were invariant with respect to a change in the ratio N_H/N_L.

In particular, it is interesting to notice how the socially optimal distortion turns out to be exactly the same as the optimal one derived for the first regime in Laffont and Tirole (1990b)'s cream skimming model, which holds only in the pure monopoly case. It should however be reminded that in their paper the ratio N_H/N_L should be interpreted as the ratio between the total number of customers of different type.

The optimal relationship between q_L and q_H is:

[LT.6] $\qquad \theta\, u'(q_H) = [1 + (N_H/N_L)\,(1-\theta)\lambda/(1+\lambda)]\, u'(q_L) = c^*$

Let us examine the impact of competition on the socially optimal tariffs fixed by the public firm. A reduction in the number of high demand customers served by the public firm (due to cream skimming) leads him to reduce the efficient level of pricing distortion, as the ratio $N_H/N_L = (N-K_H)/N$ becomes less than unitary, for any positive value of K_H. In the limiting case, in which K_H tends towards N, there will be no distortion at the bottom, exactly as in Loeb and Magat's first best setting or in the bypass regime of Laffont and Tirole (1990b). On the other hand, a reduction in the number of low demand customers served by the public firm (due to skimmed milk competition) leads instead to an increase in the pricing distortion, as the ratio $N_H/N_L = N/(N-K_L)$ becomes greater than unity.

From the previous considerations it follows that if the entrant is less efficient and/or serves the L type the direct consumers' net surplus $U(Q^i, Q^e) = N_H\theta u(q_H) + N_L u(q_L) + K_H\theta u(q_H^e) + K_L u(q_L^e) - R^i - R^e$ is reduced, as the level of pricing distortion increases. Therefore, from the allocative point of view, the maximum level of total consumers' net surplus is reached with a more efficient entrant engaged in cream skimming competition. However, in what follows we will see that the decreased price distortion generated by cream skimming competition is a negative feature of entry (differently from we have seen in section 2.4.3 where we were dealing with a private monopolist maximising profits), as the public monopoly tariffs were socially optimal, due to the presence of a cost of public funds. Specifically, either the imposition of an entry tax or regulation of the entrant allow to maintain the public tariffs to their socially optimal public monopoly level.

Let us now analyse the impact of different kind of competition on productive efficiency. The total producers' profits $(1+\lambda)\Pi^i + \Pi^e = (1+\lambda)R(Q^i) + R^e(Q^e) - (1+\lambda)c^*Q^i - mQ^e$ in the absence of distortionary taxes -setting $\lambda = 0$ as in the Loeb and Magat case- decrease if the entrant is less efficient ($m > c^*$) and stay the same if the entrant is equally efficient ($m=c^*$), as showed in the previous subsection. For λ greater than zero, the incumbent's profits Π^i (that have attached a greater welfare weight) decrease in the presence of competition, reducing social welfare. However, what we are really interested is the overall effect on total profits $(1+\lambda)\Pi^i + \Pi^e$; in this regard, the reduction of the public firm's profits may be outweighed by the increase in the entrant's profit if she is sufficiently more efficient than him. Hence, when the entrant is strictly more efficient than the public firm ($m<c^*$) social welfare is likely to increase, because cream skimming competition surely increases consumers' net surplus as in the Loeb and Magat setting. Skimmed milk competition not only decreases the consumers' net surplus but also producers' total profits in the case of an inefficient rival.

Consequently, as before, the regulator should not allow competition for the low-demand customers or by inefficient entrants, because they are socially undesirable forms of entry. In order to deter inefficient entrants, the policy maker may use a direct intervention. He may directly impose an *entry tax* equal to the foregone variable profits of the public firm (if K_H customers of type H and K_L customers of type L were taken away by the competitor). In practice, the tax payment to any entrant should be set equal to **ET** = $(K_H \pi^*_H + K_L \pi^*_L)$ where $\pi^*_t = (T_t - c^* q_t)$ is, in fact, the monopoly variable profit enjoyed by the public firm from a customer of type t (for $N_H=N_L=N$). In this way, government revenues (accrued by the entry taxes and the public firm's net profits) are kept unchanged independently of the scale of entry; this occurs since the public firm has no incentive to react to entry modifying his monopoly tariffs. In fact, once the public authority receives this entry tax, it becomes part of the public budget; consequently, the social value of the payment becomes $(1+\lambda)$ **ET** $= (1+\lambda)\ (K_H\ \pi^*_H + K_L\ \pi^*_L) = (K_H\ \omega^*_H + K_L\ \omega^*_L)$ and it compensated the public firm's for the foregone variable profits due to the presence of entry. This is so because $\omega^*_t = (1+\lambda)(T_t - c\ q_t)$ represents just the value an additional customer of type t would bring to the public firm, if entry had not occurred.

As in the previous case, a more efficient competitor ($m < c^*$) will maximise her variable profits $K_H\theta u(q^e_H) + K_L u(q^e_L) - m(K_H q^e_H + K_L q^e_L) -$ **ET** serving the H type and behaving as a surplus taker. The same graphical reasoning of the Loeb and Magat case applies, even if marginal prices are no longer equal to marginal costs.

Furthermore, abstracting from fixed costs, social welfare would be additionally improved if also the entrant is regulated. In this case the social value of letting the entrant serving a customer of type H $\theta\ u(q^e_H) - m\ q^e_H$ would receive a higher weight (since her profits are now appropriated by the public authority) becoming $\theta\ u(q^e_H) + \lambda(T^e_H - m q^e_H) - m q^e_H$. This value is clearly greater than the one the society would get with an entry tax $\theta\ u(q^e_H) + \lambda(T_H - c^* q_H) - m q^e_H$ when the entrant is strictly more efficient compared with the public firm.

It should be reminded how, in these two specific cases (entry tax and regulated entrant), a reduction in the number of high demand customers served by the public incumbent will no longer lead to a reduction in the low customers' price distortion. In practice, the public firm's non-linear tariffs are kept the same as the public monopoly one, i.e. $\theta u'(q_H)=[1+(2-\theta)\lambda/(1+\lambda)]u'(q_L)=c^*$, since he does not find it optimal to respond to the scale of entry of a more efficient competitor by distorting his tariffs. This happens because in the presence of entry taxes, or if the entrant's profits are expropriated (representing a tax revenue for the public budget) the term $(1+\lambda)K_H(1-\theta)u(q_L)$ in the social welfare function **[LT.5]** replaces $K_H(1-\theta)u(q_L)$. Consequently, all the losses in the public firm's revenues, due to the net consumer surplus $(1-\theta)u(q_L)$ to allow to the H type, are given a weight $(1+\lambda)$, as in the public monopoly case. In the modified first order conditions **[LT.6]** the ratio N_H/N_L does not appear, being always equal to unity, for any positive K_H scale of entry.

It is important to notice how in the absence of an entry tax (or of the entrant's regulation), the competitor should be strictly more efficient than the public firm for entry to be optimal. In fact, even in the absence of fixed costs, the social welfare associated with a high demand customer being served by the public firm $\theta\ u(q_L) + \lambda(T_L-c^*q_L) - c^*\ q_L$, is strictly greater than $\theta u(q_L) - c^*\ q_L$ the welfare associated with a high demand customer being served by an equally efficient entrant ($m = c^*$). This happens because the public firm's profits are given a weight $(1+\lambda)$ greater than unity and are consequently strictly preferable to the ones made by an equally efficient competitor, even if he pays corporate taxes on them (since the tax rate is normally less than unity). The decreased price distortion generated by cream skimming competition becomes a negative feature of entry, as the public monopoly tariffs were socially optimal. This is due to the presence of a trade-off between efficiency in production and decreased public revenues.

3.2.4 Optimal non-linear pricing when the regulator cares of consumers' surplus

Let us now briefly examine Baron and Myerson's (1982) case, in which the producer's surplus has a weight ϕ less than unity. In this final setting, efficiency issues are no longer the only ones the public firm deals with (in first or in second best). In fact, the public authority is more concerned with consumers' surplus in a sort of partial first best setting, where lump sum transfers are still available but only towards public or private regulated firms. This quite peculiar feature of this model differentiates it from the Laffont and Tirole's one and should be kept in mind when we examine it in a full information setting.

Since the public firm's profits Π^i represent public revenues, being entirely redistributed to consumers, are given a unitary weight like consumers' surplus and differently from the entrant's profits $\phi\Pi^e$ Baron and Myerson's social welfare function W^ϕ can be written as:

[BM.1]
$$W^\phi = U(Q^i, Q^e) + \Pi^i - tr + \phi\Pi^e = S(Q^i, Q^e) - C(Q^i) - tr - C^e(Q^e)$$
$$+(\phi-1)[R^e(Q^e)-C^e(Q^e)] = N_H\theta u(q_H)+N_L u(q_L)+$$
$$K_H\theta u(q^e_H)+K_L u(q^e_L)-c^*(N_H q_H+N_L q_L)-m(K_H q^e_H+K_L q^e_L) -(1-\phi)\Pi^e$$

Maintaining Laffont and Tirole's accounting notation, social welfare can be expressed as the sum of consumers' surplus less the total cost of the project (including the transfers from the authority to the public firm). In fact, remembering that the consumers' net surplus is equal to the consumers' gross surplus minus firms' revenues, that is, $U(Q^i, Q^e) = S(Q^i, Q^e) - R(Q^i) - R^e(Q^e)$, as in the Loeb Magat case, we see how in the formula the revenues drop out.

In the public monopoly, neglecting the presence of fixed costs, the social welfare function collapses to the Loeb Magat one, so that we do not need to solve again the maximisation problem.

[BM.2]
$$W^\phi = U(Q^i) - c^*Q^i - tr = N\,[\theta u(q_H)+u(q_L)] - c^*N(q_H+q_L) - tr$$

Notice that we will assume, as we did in Loeb Magat's setting, that the low type's participation constraint $[IR_L]$ is binding, even if this may not always be the case, due to the presence of lump sum transfers towards the public firm. In practice, we assume that the public firm can keep positive profits, appropriating customers' surplus (to keep our analysis as close as possible to Baron and Myerson, who deal with a privately regulated firm), so that the two standard constraints are binding. Hence, with a public monopoly there is no distortion both at the top and at the bottom:

[BM.3] $p_H^\phi = \theta\, u'(q_H) = p_L^w = u'(q_L) = c^*$

Furthermore, the public firm finds it optimal to equate marginal prices to marginal costs, independently of the number of customers served.

Once we consider the entry problem, like with the Laffont and Tirole's welfare function, the public (or regulated) firm's objective function W^ϕ gives a lower weight, ϕ less than unity, to the per customer profits of the entrant. Hence, when the public (or regulated) firm's objective function is W^ϕ, i.e. Baron and Myerson's one, in this peculiar redistributive setting consumers' net surplus becomes $U(Q^i, Q^e)-(1-\phi)R^e(Q^e)$.

In fact, as incumbent's profits Π^i and consumers' surplus have a unit weight, differently from the entrant's ones $\phi\Pi^e$, allowing for entry the Baron and Myerson's social welfare function W^ϕ may be written as:

[BM.4]
$$\max W^\phi = U(Q^i, Q^e) - C(Q^i) - C^e(Q^e) + (\phi-1)\,[R^e(Q^e) - C^e(Q^e)] =$$
$$N_H\theta u(q_H)+N_L u(q_L)+K_H\theta u(q_H^e)+K_L u(q_L^e)-c^*(N_H q_H+N_L q_L)$$
$$-\, m(K_H q_H^e+K_L q_L^e)-(1-\phi)\Pi^e$$

The first order conditions remains unchanged compared to the public monopoly case:

$[q_H^\phi]$	$p_H^\phi = \theta\, u'(q_H) = c^*$	No distortion at the top
$[q_L^\phi]$	$p_L^\phi = u'(q_L) = c^*$	No distortion at the bottom

This clearly shows that, even in the absence of fixed costs and access pricing, social welfare W^ϕ will increase only if the entrant is strictly more efficient than the incumbent.

Consequently, in general, it is not necessary that a positive net surplus $S_L > 0$ should be allowed to the L type, whereas the H type enjoys always a positive net surplus [given by $(\theta - 1)u(q_L)+ S_L$]. In fact, being the downward binding constraint

$[IC_H]$ binding, any positive surplus $S_L > 0$ given to the L type customer must also be given to the H type customer. This may happen if the incumbent has to prevent the entry of an inefficient competitor.

As in Loeb and Magat's setting social welfare: (1) decreases if the entrant is less efficient ($m > c^*$), and (2) increases only if the entrant is more efficient ($m < c^*$). However, it decreases also if the entrant is equally efficient ($m=c^*$) instead of staying the same , because the entrant's profits are valued less than the public firm's one.

From an allocative point of view, the consumers' net surplus specified below

[BM.4] $$N_H[\theta u(q_H)-c^*q_H]+N_L[u(q_L)-c^*q_L] +K_H[\theta u(q^e_H)-mq_H] +K_L[u(q^e_L)-mq_L]$$

is reduced when the relative efficiency of the entrant is lower.

On the other hand, considerations of productive efficiency tell us that the social value of producers' profits $\Pi^i + \phi\Pi^e$ surely decrease if the entrant is less or equally efficient ($m \geq c^*$) and can increase only when the entrant is strictly more efficient ($m < c^*$).

Considering cream skimming competition, for each additional customer served by the entrant, social welfare varies by the amount $\{\theta u(q^e_H) - mq^e_H -(1-\phi)\pi^e - [\theta u(q_H) - c^*q_H]\}$ which is positive only for a more efficient competitor ($m < c^*$).

Thus, in the absence of specific public interventions, competition for low demand customers or by inefficient or equally efficient rivals must be faced by the public firm, who can avoid them making those forms of entry unprofitable. This can be done, as in both of the previous settings, reducing, by a sufficient amount S_L, the low type customers' tariff, so that the entrant's total variable profits $K\pi^e_L = K[u(q^e_L) - S_L - m\, q^e_L]$ are negative. But this means that also the high type customers' tariff is reduced by S_L and the public firm may not break even applying these lower tariffs, and may need lump sum transfers.

Alternatively, we can make use of the same regulatory tools described in Laffont and Tirole's setting to enhance social welfare. In particular, the entry of less efficient competitors ($m>c^*$) is prevented by a tax payment equal to the variable profits enjoyed by the public firm if entry had not occurred; i.e. $(K_H \pi_H^* + K_L \pi_L^*)$. In this case, since the entry tax represents a component of the public budget its social weight is unitary and is not reduced to ϕ as the entrant's profits are. Consequently, an equally efficient competitor may enter the market without reducing social welfare, as her variable profits are totally expropriated by the tax.

It should at this point be clear how, even in this setting, a more efficient competitor ($m<c^*$) finds it optimal to behave as a surplus taker and to serve the high type in order to maximise her own variable profits $K_H T_H^e + K_L T_L^e - m(K_H q_H^e + K_L q_L^e)$. Hence, a socially optimal pricing is reached when the competitor pays an entry tax equal to $(K_H \pi_H^* + K_L \pi_L^*)$ and the public firm equates marginal prices and marginal costs, causing no distortion for both type of customers.

The other alternative which allow to further improve social welfare, as before in Laffont and Tirole's case, is to regulate the entrant. In fact, in this case the social value of letting the entrant serve a high demand customer $\theta\, u(q_H^e) - m\, q_H^e$ would be attached an unitary weight. Moreover, this value is greater than the one obtained with an entry tax $\theta\, u(q_H^e) + (T_H - c^* q_H) - T_H^e - (1-\phi)(T_H^e - T_H - m q_H^e + c^* q_H)$, when the entrant is strictly more efficient.

In sum, similarly to Laffont and Tirole's setting, in the absence of entrant's regulation and/or an adequate entry tax, the rival should be strictly more efficient than the public firm for entry to be socially optimal. In fact, if a high customer is served by an equally efficient entrant ($m = c^*$) instead of the public firm, social welfare would be reduced, because the profits of the entrant are given a lower weight ($\phi < 1$).

3.2.5 Final remarks on regulatory issues in the horizontal game

Let us now reconsider in general terms the efficiency role played by competition with non-linear tariffs and in presence of a public firm who maximises social welfare. The previous results strengthened the initial reasoning according to which competition should complement rather than replace regulation. Furthermore, it is better from a social point of view that competition is directly regulated by the authority or is subject to an entry tax. In particular, the public incumbent or the regulator should not allow competition for the low-demand type or by a less efficient entrant, and should impose socially optimal non linear tariffs.

We have seen how in the horizontal game with purely profit maximising firms, the presence of cream skimming competition by a more efficient entrant increases social welfare, both from a productive and allocative point of view (since the distortion at the bottom tends to disappear).

However, even under what might seem at first sight to be the most favourable conditions, the private tariffs (T_L and T_H) are far from optimal. This occurs since private firms, giving no weight to consumers' surplus, tend to distort prices by a greater amount than it would be socially optimal, even in the absence of lump sum transfers and of redistributive issues.

Some of the considerations that have been previously made reminds us vaguely of the original Baumol's subtle pricing issues of 1983. Apparently, our analysis is not related to Baumol's one, because he was looking at ways to obtain efficiency improvements in vertically related market, showing that paradoxically a collusive result may be preferable to a Cournot-like competitive one and that access charges should be set equal to the marginal cost of access plus a term which reflects the opportunity cost of entry. We will deal with these issues in the section 3.3 in which we examine the welfare implications of the vertical game.

However, even in the previous analysis of the horizontal game two points emerged and it is worthwhile to discuss them a little bit further, especially because they are closely related to the welfare analysis that we are going to develop dealing with access pricing. The points in which we are specifically interested are: (a) the necessity that competitors' entry may be allowed if they are able to pay an amount that reflects the opportunity cost of entry incurred by the public incumbent, (b) the idea to exploit the efficiency enhancing properties, implicit in a potential horizontal merger, or more precisely the more general idea of "internalisation" (of which the merger may represent a specific private application).

The first point has already proved quite relevant with all the welfare function specifications. In particular, adopting the social welfare function of Laffont and Tirole and of Baron and Myerson in the absence of entrant's regulation and of an entry tax, entry is socially optimal if and only if the rival is strictly more efficient than the public or the private regulated firm (even in absence of fixed costs). This condition ensures that the competitor is able to pay an amount greater than the opportunity cost of entry, that is the foregone profits of the public firm. Only if we have an entry tax or the entrant is regulated, the entry of an equally efficient competitor is again optimal.

Furthermore, the ability to pay a transfer equal to the opportunity cost may be seen as a sort of Hicks-Kaldor welfare improving criterion. Once the criterion is satisfied there would be no need for the transfers to take place, as with the Loeb and Magat social welfare function. Instead, with the welfare specifications of Laffont and Tirole and of Baron and Myerson, it is always optimal that the compensatory transfer towards the public firm or better toward the public budget takes effectively place.

However, as the previous criterion of the "opportunity cost of entry payment" follows exactly from the efficiency properties of a horizontal merger, the fact that the entrant should be strictly more efficient shows that the merged monopolist's optimal pricing criteria are not always socially optimal with two distinct firms (a public and a private one) which behave following different objectives (welfare and profit maximisation).

Nevertheless, the efficiency properties implicit in a more general criterion that we may call "internalisation principle" suggest regulatory mechanisms able to enhance social welfare. In fact, ceteris paribus, the use of a regulatory instrument such as the entry tax (that we devised in the previous sections) improves a previously socially optimal setting adopting the welfare function proposed by Laffont and Tirole and by Baron and Myerson (while social welfare would not change with Loeb and Magat's function). In these two cases, further improvements on this new socially optimal setting (reached introducing the entry tax) can be achieved if we can also regulate the entrant.

Finally, with all the social welfare functions previously examined, we can make use of an additional instrument, able to better apply the "internalisation principle", to increase further social welfare. In our model this mechanism is given by the *purchasing solution* (i.e. the solution in which the incumbent buys the output of the entrant to resell it to customers) as already envisaged in chapter 2. In fact, as in the case already examined of private market equilibrium, the purchasing solution is strictly socially preferable to allowing entry and direct selling of the competitor.

We can easily show that social welfare increases when the competitor sells all the additional units $K(q^e_H - q^*_H)$ to the public incumbent firm, so that the latter is able to resell it to consumers at the socially optimal monopoly tariffs (T^*_L and T^*_H).

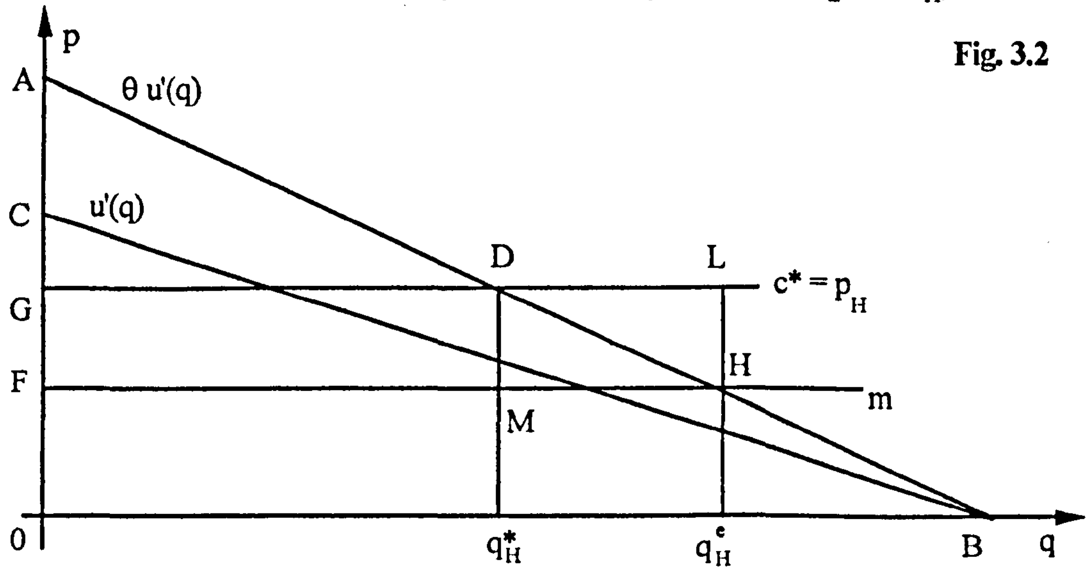


Fig. 3.2

As shown in fig. 3.2 above, the presence of a competitor whose marginal price p^e_H (equal to his marginal cost m) is lower than the incumbent's one p^*_H implies a surplus loss for the collectivity, which amounts to the integral of $K[c^* - \theta u'(q)]$ for q that goes from q^*_H to q^e_H. Hence, social welfare increases when the marginal units $K(q^e_H - q^*_H)$ are purchased by the incumbent at their marginal cost m (or at their average price, always less than c*) and then sold at a higher price as inframarginal units (for instance to high-demand customers) as the surplus loss fully disappears. The consumer surplus will increase by an amount equal in practice to the area DHL represented in fig. 3.2, and this will increase social welfare by $(1+\lambda)$ times the previous amount in Laffont and Tirole's case.

Before introducing vertical issues in the following section, let us now briefly reconsider the role of competition in the presence of access pricing in a private setting, as derived in section 2.5. It seems that the previous welfare implications are strengthened, even if some of the previous problems are no longer present. In particular, if he can fix access pricing the incumbent will not allow competition for the low-demand type or by an inefficient entrant, since it would not be profitable for him to set the per customer access charge less than the surplus he would get from the high type applying monopoly tariffs. Even under these most favourable conditions - that is, in the presence of cream skimming competition by a more efficient entrant- we will see that social welfare increases by a lower amount than it would do in the absence of vertical issues. Intuitively, this occurs since the pricing distortion is not reduced by entry, since the marginal price p_L and the tariff T_H remain unchanged, instead of decreasing.

3.3 Access pricing regulation and the optimality of the Baumol-Willig rule

The aim of this section is to consider in detail the implications of the presence of a regulated or of a public (vertically integrated) incumbent firm in vertically related markets. The following analysis will mainly focus on socially optimal non-linear price regulation under full information, taking into consideration the three different social welfare specifications.

A particular emphasis will be put on access pricing regulation, examining in detail the usefulness of the so called Baumol-Willig rule not only in the original linear access pricing framework devised by Baumol, but also in the cream skimming model previously analysed, restating its possible efficiency role in socially optimal pricing settings, in contrast with some recent interpretations.

From the society point of view, in a first best setting without price discrimination and redistributional concerns, it is clear that the regulator should set access pricing equal to the marginal cost of providing access. In fact, it is always possible to cover eventual budget deficits of the incumbent with lump sum transfers, ensuring that he will be able to continue to efficiently operate the network.

Unfortunately, even with full information on the cost and effort of the network operator, such first best outcomes are not normally available, due, for instance, to the presence of a binding break even budget constraint when public subsidies are not feasible or because public funds are costly, given the impossibility to recur to lump sum taxes, as Laffont and Tirole and many other economists have forcefully argued. Therefore, much of the debate should explicitly involve the determination of optimal departures from marginal cost pricing. We will see how both the powers that the public regulatory authority can use and the possibility to allow transfers to the firm (possibility that may be endogenised in the model, explicitly allowing for regulatory capture, as done by Laffont and Tirole) are quite relevant issues to analyse in this case.

3.3.1 The optimality of the Baumol-Willig pricing rule in the original framework

Once the original Baumol's model of 1983 is carefully examined, it is quite evident that the author implicitly considers a special case, where the budget constraint is binding, but transfers are not allowed, and proposes a rule, which is also known as the Baumol-Willig rule or the efficient component pricing rule, that can be used in order to attain some efficiency improvements. According to it, within non perfectly contestable industries, access charges should be set equal to the marginal cost of access plus a term which reflects the opportunity cost of entry. Moreover, he shows that, when dealing with linear pricing in a market characterised by vertical issues, the collusive result is normally preferable to the competitive one (leading to a Cournot-like equilibrium) as prices are closer to marginal costs. To fully understand the author's aims and the real relevance of his analysis, it is fundamental to realise that both of these results exploit the very same efficiency enhancing properties, implicit in a potential vertical merger.

In the first case where there are two different firms operating two segments of a given service (let say a rail service) and acting as profit maximisers, Baumol does not search for the optimal pricing (for each segment or the entire track) but simply considers "what price setting *process* will best serve the interests of consumers". His comparison between a complete collusion between the two firms with Cournot like competition shows the superiority of the first setting. Clearly, the same type of setting (within a profit maximisation model) should be preferable also in the second case in which there is a vertically integrated firm who is a monopolist of the first segment and the other firm still operates just in the competitive segment; in fact, Baumol does not even spend one more word in order to further specify matters. However, in the second case, differently from the first one, he seems to implicitly assume constant marginal costs of providing train space (ic_e) in his reasoning. As we will see, in this particular case, his result will also hold with socially optimal regulation.

In order to clarify matters, let us consider the original example due to Baumol (1983) in which he focuses on what we called the second case. A vertically integrated incumbent operates a rail service connecting R, S and T. An entrant offers a perfect substitute on the competitive route connecting R and S and has access to the incumbent's track ST. It should also be pointed out that, in this setting, he just deals with the efficiency issue (implicit in consumers' choice) of not distorting "the relative competitive advantages", even if the "optimal price" and "socially acceptable compensation" labels are used by Baumol himself. Now a non-discriminatory access price cap should be set by an outside regulator to enforce the "principle of parity" of implicit prices. In practice, the cap should be low enough to let efficient competitors enter the market, eliminating the possibility that the access charge may result in a disadvantage for one of the two firms competing for customers. Baumol believes that a low access charge may be achieved even as the result of a private bargaining process.

For simplicity's sake, by assumption we may consider, as in Cave (1994), the case where the incumbent's retail price p for the service over RST is equal to average costs, so that the incumbent just breaks even. However, in dealing with this case, it should be made fully clear how Baumol does not believe that firms would adopt a Ramsey pricing policy voluntarily (as he stated in detail in his reply to Tye's comment).

Let the incumbent's per unit incremental costs of providing track-space over ST be ic_A. Let ic_i and ic_e respectively denote the per unit incremental costs of providing track-space over RS and train space on RT if the transportation service is performed by the incumbent or by the entrant (subscript i and e). In this case the equitable access charge F* that the *incumbent* should charge is:

[BW] $$F^* = p - ic_i = ic_A + (p - ic_A - ic_i)$$

Let us show that F* is the only fixed charge that guarantees minimum costs of production. Assume that F ($= \varepsilon + p - ic_i$) be higher than $p - ic_i$. Then, some entrants with an incremental cost less than the incumbent's one ($ic_i > ic_e > ic_i - \varepsilon$) will find entry unprofitable. If entry occurred output would be produced at minimum costs, which would clearly improve efficiency and welfare. In practice, F* is "economically compensatory", since it is the implicit transfer price a merged monopolist would choose (for providing track-space over ST) when he faced the question of which production process should be used for the segment RS. From this it follows that the minimum costs of production property will hold immediately, as clearly the merged firm would always minimise costs. Thus, the rule just increases the industry's efficiency, as in the case of a vertical merger.

In sum this rule just "offers Pareto optimal criteria for pricing efficiency" related to the "intermediate input" (the track-space from S to T); the very same one that the merged monopolist would follow. However, not all the merged monopolist's optimal pricing criteria need to be followed and to be taken as socially optimal in this situation, as we continue to have two distinct firms, whose profits in general may be given different weights. Moreover, even in the presence of two profit maximising firms, probably this criterion may not necessarily hold if the two firms specialise on different types of customers having different demand elasticity. But this point is already implicitly recognised by Baumol (see. p. 352). If we look at the second formulation of the Baumol-Willig rule **[BW]**, we notice how the optimal charge is set equal to the marginal cost of access (ic_A) plus the opportunity cost of access ($p - ic_A - ic_i$), i.e. the profits foregone by the incumbent by not serving the traffic. This clearly shows how the incumbent enjoys a fully compensatory pricing.

Moreover, for practical purposes, Baumol refers to the price charged by the incumbent on his monopolised segment ST (p_{ST}) and to the incremental cost of providing train space service from S to T (ic_{ST}). In this setting his efficiency improving rule can be reformulated as:

[BW]' $$F^* = p_{ST} - ic_{ST} = ic_A + (p_{ST} - ic_A - ic_{ST})$$

and can be applied with any arbitrary pricing rule and not only as a part of a socially optimal pricing policy. According to Baumol, $F^* = p_{ST} - ic_{ST}$ imposes "all of the contribution toward common costs of shipments from S to T upon the track space rental component". But, as the train space will be offered at its marginal cost only in perfectly contestable markets (see. n. 5 p. 355) nothing ensures us that the access charge $p_{ST} - ic_{ST}$ (or $p - ic_i$) just covers effectively incurred costs (and not inefficiencies or monopoly returns, which can be included only in a quite broad definition of opportunity costs).

In this regard, when the network is earning excessive returns, or is being operated inefficiently, the application of the Baumol-Willig rule may lead to a situation in which the incumbent keeps on earning super-normal profits or imposing the costs of his inefficiency to competitors. Instead, one of the regulatory purposes is to eliminate super-normal profits and inefficiency costs from the access charge; our analysis in setting the incumbent's price equal to average costs aims to point out this. In any case, later (see for instance Baumol and Sidak (1994)) Baumol eliminates these potential problems, assuming that efficient regulation has already solved them, before dealing with access pricing regulation.

Nevertheless, even in Baumol's original case, due to the presence of a positive externality or due to dynamic considerations (see Cave (1994), among others), the regulator may wish to promote entry in the final services market, departing from the parity rule.

In an extended framework Cave (1994) notices how the Baumol-Willig rule should explicitly incorporate components related to the variety of products (i.e. the demand for the services) provided over the network. However, this perhaps minor generalisation is already implicit in the definition of the marginal and the opportunity cost of access (the profits foregone by the incumbent). Moreover, it was also recognised by Baumol himself in the explicit assumption that parity holds only when the intermediate good "will face the same [final] demand conditions and incur the same marginal costs" (p. 352).

Cave also points out that, with perfectly competitive entrants, social welfare may be enhanced by lowering p (the incumbent's retail price) and contemporaneously increasing F (the fixed access charge). This consideration leads him to conclude that the rule "is applicable only when the incumbent's retail prices are non-optimal". We will show that this is not always the case.

In claiming this Cave refers to quite specific models, such as Armstrong and Doyle (1994) and Armstrong and Vickers (1995), in which transfers are not allowed and the only access pricing policy allowed for is the linear one. The question analysed by the previous authors within this framework is which type of access pricing policy is socially optimal (mainly considering welfare as the unweighted sum of the consumers' and the producer's surplus) in their partial equilibrium setting. Proposition 1 (closely related to Laffont and Tirole 1990a) simply states the optimality to set access charge in excess of the marginal cost of providing access, on the basis of a simplified Ramsey pricing formula: $(p - F - ic_i)/(p - F) = -k/\eta_E$. However, even in their particular setting, their proposition is easily falsified in the plausible case in which the elasticity of supply of the entrant (denoted in their model η_E) is infinite because she has constant marginal costs. It should be also noticed how there is no reason which justifies the asymmetry between the incumbent and the entrant; in particular, their result may probably be modified, once one also considers the elasticity of supply for the incumbent. Thus it is arguable that the Baumol-Willig rule can be socially optimal pricing in the original context assumed by Baumol, so that we can now focus on our cream skimming model.

3.3.2 The social optimality of the Baumol-Willig pricing rule in the cream skimming model

Before dealing with socially optimal access pricing rules, let us first examine the impact of regulating access directly (through the application of the Baumol-Willig rule) starting from the purely private setting analysed in section 2.5. In vertically related markets we have seen how the incumbent will always enjoy monopoly profits, as if he were the only player, allowing the entry of an equally efficient competitor, while deterring less efficient rivals through high enough access charges. Moreover, in the presence of a more efficient rival the incumbent can enjoy an *additional* profit. In fact, the use of non-linear access charges allows him to control entry maximising at the same time productive efficiency.

In what follows we show that in our cream skimming model the Baumol-Willig rule allows entry to an equally efficient competitor who faces no entry costs (apart from the access charge) and serves either low or high-demand customers. The "no fixed costs assumption" is clear in what we called Baumol's second case, because for the entrant there are no additional fixed costs associated to offering train space in the segment ST (which we assume to be the high-demand market). However, if we introduce non-linear pricing whenever the average retail price p is greater than average costs this parity rule alone would no longer attain global social optimality, though it would still represent the minimum charge that should be applied to avoid the entry of inefficient competitors.

Let us then consider in detail the conditions under which in our model the Baumol-Willig rule is optimal for the *incumbent* and more generally for achieving *productive efficiency.* According to Baumol, in order to avoid inefficient entry, all potential competitors should pay one of the two charges specified below in order to enter the high or low demand market.

The per customer access charge which should be imposed according to the parity rule for the high-demand customers is

[BW$_H$] $\quad f_H^o = F(K_H) / K_H = T_H - c^*q_H$

It allows an equally efficient competitor, who serves the H type (in the absence of fixed costs), to break even and to enter the market. Similarly, the per customer access charge for the low-demand customers

[BW$_L$] $\quad f_L^o = F(K_L) / K_L = T_L - c^*q_L$

allows the entry of an equally efficient competitor serving the L type (in the absence of fixed costs).

Naturally, the competitor (in the absence of inframarginal costs) will always prefer to serve only high-demand customers, and is anyway obliged by the incumbent to cream skim, as our analysis has proved (cf. section 2.5). Nevertheless, in general there may exist particular specifications of the cost function for which the competitor will instead maximise her profits by serving the L type.

As we are not yet dealing with socially optimal regulation, one may put the usual objections on the use of this access pricing policy. However, in the previous setting, these prices are clearly the optimal ones for the incumbent (who faces an equally efficient competitor), as we have proved, and, given the presence of monopoly tariffs, they also represent the minimum fixed access charge which avoids the entry of inefficient competitors. In fact, in the case of an equally efficient rival, it is optimal for the incumbent to set the per customer access charge equal to the monopoly variable profits and to maintain his monopoly pricing strategy. In other words, the incumbent is indifferent to allow or deter entry (cf. proposition 1 of section 2.5).

In the presence of a more efficient rival the incumbent finds it optimal to set the per customer access charge greater than the monopoly variable profits in order to expropriate all the entrant's profits. Hence, he is better off in the presence of entry and productive efficiency is guaranteed, since the incumbent can get all the relative advantages. **[AP]'** represents the per customer access charge fixed by the incumbent, which is clearly greater than **[BW$_H$]**, the one derived from the parity rule:

[AP]' $F(K_H) / K_H = T^e_H - m\, q^e_H > f^0_H$

However, the application of the Baumol-Willig rule too, would still succeed in maximising productive efficiency, while leaving the entrant the additional profits (due to her greater efficiency), avoiding their complete appropriation by the incumbent.

3.3.3 The optimality of the Baumol-Willig rule in a Loeb and Magat's setting

So far we have seen how the Baumol Willig is optimal from the productive efficiency point of view. As a direct consequence, it might be expected its socially optimality if we stick to the very simple ultra-liberal social welfare function given by the unweighted sum of consumer surplus and profits a la Loeb and Magat (1979), focusing only on efficiency issues and ignoring any eventual distributional and excess burden considerations. However, our analysis will show that the tariffs induced by the Baumol-Willig rule do not represent overall the unique socially optimal prices for the society, even in this naive setting.

Let us then consider which access pricing rule would be the socially optimal one and in which direction we must move in order to achieve Pareto improvements. As in section 3.2.2, the collectivity will pay costs and transfers and get the public (or regulated) firm's revenues, including access charges. The entrant's individual rationality must be satisfied for entry to take place respectively in the high and low-demand market. These conditions, specified below as [AP$_H$] and [AP$_L$], imply the entrant to break even after paying the access charges:

[AP$_H$] $f_H K_H \leq (T^e_H - mq^e_H) K_H$

[AP$_L$] $f_L K_L \leq (T^e_L - mq^e_L) K_L$

All costs and transfers are attached weights equal to unity, so that the consumers' surplus can be simply expressed as $S(q^0, q^i_H, q^i_L, q^e_H, q^e_L, N_H, N_L, K_H, K_L)$. Adopting Loeb and Magat's social welfare function the public firm's general problem is to maximise social welfare W (with respect to his tariffs) and can be written as:

$$\max W = S - C(Q^i) - NC(2N, Q^0, Q) - C^e(Q^e) =$$

$$N_H\theta u(q_H)+N_L u(q_L)+K_H\theta u(q_H^e)+K_L u(q_L^e)+2Nv(q^0)-c^*(N_H q_H+N_L q_L)$$

$$- c^0(2Nq^0) - NC(2N) - m(K_H q_H^e+K_L q_L^e) \qquad \text{subject to:}$$

[MME] $p^0 \leq v'(q^0)$

[IR_L] $u(q_L) - T_L \geq 0$

[IR_H] $\theta\, u(q_H) - T_H \geq 0$

[IC_L] $u(q_L) - T_L \geq u(q_H) - T_H$

[IC_H] $\theta\, u(q_H) - T_H \geq \theta\, u(q_L) - T_L$

To verify whether the Baumol-Willig rule is socially optimal in this setting we assume that the regulator imposes the parity rule for the high and low-demand customers, in order to avoid inefficient entry:

[BW_H] $f_H = F(K_H) / K_H = T_H - c^* q_H$

[BW_L] $f_L = F(K_L) / K_L = T_L - c^* q_L$

Taking advantage of the previous analysis, all may be reduced to the maximisation of the following social welfare function, in which the entrant, being equally or more efficient than the public firm, will cream skim the market:

$$\max W = N_H\theta u(q_H)+Nu(q_L)+K_H\theta u(q_H^e)-c^*(N_H q_H+N_L q_L)-m(K_H q_H^e)+2Nv(q^0)-$$

$$c^0(2Nq^0) - NC(2N) \qquad \text{subject to:}$$

[MME] $p^0 = v'(q^0)$

[IR_L] $u(q_L) - T_L = 0$

[IC_H] $\theta\, u(q_H) - T_H = \theta\, u(q_L) - T_L$

The welfare maximisation problem can be easily solved with respect to q°, q_L and q_H, once we substitute the binding constraints [MME], [IR_L] and [IC_H] into the objective function. Making use of consumers' optimality conditions for the non-monopolised good, i.e. $p_t=\theta_t u'(q_t)$ from the first order conditions we can see that an equality between the two marginal price must hold:

$[q^0]$	$p^w = v'(q^0) = c^0$	Marginal cost pricing
$[q_H^w]$	$p_H^w = \theta\, u'(q_H) = c^*$	No distortion at the top
$[q_L^w]$	$p_L^w = u'(q_L) = c^*$	No distortion at the bottom

The public firm finds it optimal to equate price to marginal cost for the monopolised good 0 and to equate marginal prices (for both types of customers) to marginal costs in the market of good 1.

Therefore, in this setting the answer to the socially optimal access pricing question seems relatively easy. In fact, starting from a monopoly price setting with the Baumol-Willig access charges, we can always reduce the access charge f_H^0 and increase f_L^0 till we reach tariffs corresponding to the equality of marginal prices and marginal costs of the public firm; that is, the couple T_H^W and T_L^W for which we have $p_H^w = p_L^w = c^*$. Here, as usual the superscript W indicates a Loeb and Magat's socially optimal pricing, in contrast to the superscript B which would follow from the application of the Baumol Willig rule.

Once we consider non-linear pricing, differently from Cave's prescriptions, the retail average price is lowered, but simultaneously the access charge is decreased too (at least when we maximise an ultra-liberal welfare function and allow for lump sum transfers).

Under a full information setting whenever $p_H = p_L = c^*$ and $c^1 = 0$, if the regulator sets the unit access charges f_H^B and f_L^B following the Baumol-Willig rule, he will leave the public firm's profits unchanged.

$$f_H^B = T_H^W - c^* q_H^W$$

$$f_L^B = T_L^W - c^* q_L^W$$

These "parity" unit access charges are socially optimal being compatible with the previous first order conditions. Naturally, in this particular setting, differently from what implicitly assumed by Baumol and explicitly by Armstrong and Doyle (1994), lump sum transfers could be needed if the public firm will not break even when the regulator imposes the optimal tariffs T_H^W and T_L^W which equate marginal prices to marginal costs. Hence, differently from the original linear pricing setting, under socially optimal non-linear pricing the presence of lump sum transfers is fully compatible with the original Baumol-Willig rule.

Hence, in Loeb and Magat's setting, once the socially optimal tariffs (T_H^w and T_L^w) are reached we may impose the Baumol-Willig rule, as it represents the socially optimal access pricing policy for the whole society. In fact, in the previous model Pareto efficiency will be achieved (when the entrant behaves as a surplus taker, maximising her own profits) since the rule would not allow entry by an inefficient competitor and at the same time the marginal prices equate the marginal costs.

However, from the society point of view, as in the welfare function the entrant and the public firm's profits are given the same weight, it doesn't matter if the incumbent would instead set a higher access charge, expropriating (partially or totally) the more efficient competitor's profits. In practice, denoting by the parameter $1 > \gamma > 0$ the proportion of entrant's per customer variable profits (net of the Baumol-Willig charge) the public firm may take away from her- the family of unit access charges specified below will be socially optimal too:

$$f_H^W = f_H^B + \gamma\,(T_H^e - mq_H^e - f_H^B)$$

$$f_L^W = f_L^B + \gamma\,(T_L^e - mq_L^e - f_L^B)$$

Notice that even if the competitor's profits under the Baumol-Willig pricing rule may be negative (for eventual less efficient entrants) it will remain in the competitor's interest not to enter when she is less efficient than the public firm. Consequently, the Baumol-Willig pricing rule is just one of the possible socially optimal pricing rules. In any case, if we totally disregard redistributional concerns, the "parity principle" is fully applicable with socially optimal retail prices, differently from what stated by Cave, even if lump sum transfers are available.

3.3.4 Optimal access pricing regulation and the cost of public funds' approach

We may instead consider the fact that taxes change the behaviour of economic agents creating unavoidable distortion costs, following for instance the cost of public funds' approach due to Laffont and Tirole, or alternatively imposing the presence of a binding break even budget constraint and the impossibility to finance through subsidies. With this approach we still considers only efficiency issues (as welfare is the unweighted sum of consumers' and producer's surplus in the ambit of quasi linear utility functions) taking explicitly into consideration the cost of public funds. The collectivity will pay costs and transfers and get the public (or regulated) firm's revenues, including access charges. As in the previous subsection the entrant's individual rationality must be satisfied for entry to take place respectively in the high and low-demand market:

[AP_H] $\quad F_H \leq (T^e_H - mq^e_H)\, K_H$

[AP_L] $\quad F_L \leq (T^e_L - mq^e_L)\, K_L$

Allowing for costly public transfers and taking λ as given and adopting Laffont and Tirole's social welfare function, the consumers' surplus (now inclusive of public revenues) is $S(q^0, q^i_H, q^i_L, q^e_H, q^e_L, N_L, N_H, K_L, K_H,)+\lambda R(Q^0, Q^i)+\lambda F$ and the public firm maximises the new social welfare W^λ (with respect to his tariffs), where the superscript λ refers to Laffont and Tirole's formulation.

$$\max W^\lambda = S+\lambda R(Q^0, Q^i)-C(Q^i)-\mathcal{NC}(2N, Q^0+Q)+\lambda F-C^e(Q^e) =$$

$$N_H\theta u(q_H)+ N_L u(q_L) + \lambda(N_L T_L+N_H T_H+F_L+F_H) +2Nv(q^0)+\lambda 2Np^0q^0$$

$$+K_H\theta u(q^e_H)+K_L u(q^e_L)+\lambda(F_L+F_H)(1+\lambda)[c^*(N_H q_H+N_L q_L)$$

$$+c^0(2Nq^0)+\mathcal{NC}(2N)]-m(K_H q^e_H+K_L q^e_L) \quad \text{subject to:}$$

[MME] $\quad v'(q^0) - p^0 \geq 0$

[IR_L] $\quad u(q_L) - T_L \geq 0$

[IR_H] $\quad \theta\, u(q_H) - T_H \geq 0$

[IC_L] $\quad u(q_L) - T_L \geq u(q_H) - T_H$

[IC_H] $\quad \theta\, u(q_H) - T_H \geq \theta\, u(q_L) - T_L$

In order to verify whether the Baumol-Willig rule can be socially optimal also in this setting we impose the parity rule for the high and low-demand customers, which avoids inefficient entry:

[BW_H] $\quad f_H = F_H / K_H = T_H - c^*q_H$

[BW_L] $\quad f_L = F_L / K_L = T_L - c^*q_L$

Taking advantage of the previous analysis, all may be reduced to the maximisation of the following social welfare function, in which the entrant, being equally or more efficient than the public firm, will cream skim the market:

$$\max W^{\lambda} = N_H\theta u(q_H)+Nu(q_L)+2Nv(q^0)+\lambda(NT_L+N_HT_H+f_HK_H)+\lambda 2Np^0q^0+K_H\theta u(q_H^e) +\lambda f_HK_H-(1+\lambda)c^*(N_Hq_H+N_Lq_L)-(1+\lambda)[c^0(2Nq^0)+NC(2N)]-m(K_Hq_H^e)$$

subject to:

[MME] $v'(q^0) - p^0 = 0$

[IR_L] $u(q_L) - T_L = 0$

[IC_H] $\theta\, u(q_H) - T_H = \theta\, u(q_L) - T_L$

Making use of the optimality conditions which hold for the customers we can easily write down the first order conditions as:

[q^{λ}]	$p^{\lambda} = v'(q^0) = c^0\,[1 + \lambda/(1+\lambda)\eta^0]$	Public Ramsey pricing
[q_H^{λ}]	$p_H^{\lambda} = \theta\, u'(q_H) = c^*$	No distortion at the top
[q_L^{λ}]	$p_L^{\lambda} = u'(q_L) = c^*(1+\lambda)/[1+\lambda(2-\theta)]$	Distortion at the bottom

The public firm finds it optimal to apply Ramsey pricing (above marginal cost) in the monopolised market of good 0 and to equate marginal prices (for high types of customers) to marginal costs while leaving some distortion at the bottom. The departures from marginal pricing in [q^{λ}] and [q_L^{λ}] arise because revenues are given an additional weight λ, being costly to get public revenues from taxes.

The distortion at the bottom is lower than in the unregulated setting of section 3.3.1, apart from the case in which the shadow cost of public funds λ goes to infinity. Hence, Pareto improvements would also be achieved by reducing the access charge f_H^0 and increasing f_L^0 till we reach the second best tariffs T_H^{λ} and T_L^{λ} corresponding to $p_H^{\lambda}=[1+\lambda(2-\theta)/(1+\lambda)]p_L^{\lambda}$.

However, in this setting, the Baumol-Willig rule cannot be socially optimal, apart from particular cases, since applying the parity rule for high and low-demand customers **[BW_H]** and **[BW_L]** we impose stricter constraints than the access pricing condition. In fact, in this case, the access pricing conditions **[AP_H]** and **[AP_L]** are now always binding (differently from Loeb and Magat's maximisation problem) being always socially optimal that the public firm expropriates all the entrant's profits, as his profits are now given an additional weight λ (as F_H and F_L in the welfare function), due to the excess burden of the revenues coming from taxes.

If we consider now an equally efficient competitor (and $c^1 = 0$) the Baumol Willig rule turns out again to be optimal. In fact, it will prescribe to set the unit access charges:

$$f_H^B = T_H^\lambda - c^* q_H^\lambda$$
$$f_L^B = T_L^\lambda - c^* q_L^\lambda$$

so that the public firm's profits remain the same as all the entrant's profits are appropriated.

This result, stated in other words, says that the Baumol Willig rule is socially optimal when the rival is as efficient as the public firm and is in contrast with the one derived in the absence of vertical relations, which requires the entrant to be strictly more efficient than the incumbent (see section 3.2.3).

In sum, since the public (or private regulated) firm's profits are now given an additional weight λ, being costly to get public revenue from taxes, the Baumol-Willig rule no longer represents the socially optimal access pricing rule, apart from the case of an equally or less efficient entrant. In fact, the parity rule prevents the public firm from totally expropriating more efficient competitors' profits by setting a higher access charge. Hence a departure from the Baumol-Willig rule, increasing the access charge will be welfare improving. Specifically, total expropriation will be optimal as profits will be transferred from the competitor to the public firm.

Consequently, with full information -while second best tariffs (corresponding to $p_H^\lambda = [1+\lambda(2-\theta)/(1+\lambda)]\, p_L^\lambda$) will hold- socially optimal access pricing turns out to be the one according to which the incumbent sets the following unit access charges:

$$f_H^\lambda = \max\{T_H^e - m q_H^e;\ T_H^\lambda - c^* q_H^\lambda\}$$

$$f_L^\lambda = \max\{T_L^e - mq_L^e;\ T_L^\lambda - c^*q_L^\lambda\}$$

so that the public firm's profits are always maximised (when transfers are available).

We may rewrite the previous socially optimal unit access charges as follows, in order to show the differences with the one prescribed by the Baumol-Willig rule.

$$f_H^\lambda = f_H^B + \max\{(T_H^e - T_H^\lambda) + (c^*q_H^\lambda - mq_H^e);\ 0\}$$
$$f_L^\lambda = f_L^B + \max\{(T_L^e - T_L^\lambda) + (c^*q_L^\lambda - mq_L^e);\ 0\}$$

Here, the additional terms $(T_t^e - T_t^\lambda) + (c^*q_t^\lambda - mq_t^e)$ [with t = H, L] simply represent the additional profits that, under full information, the incumbent public firm can make, taking full advantage of the efficiency of the entrant (when $c^* > m$).

The case in which costly public transfers are available and the incumbent's budget is binding (in the absence of entry) and remains binding afterwards (in the presence of entry) is not substantially different from the previous one, apart from the fact that the value assumed by λ is in this case endogenously determined by the maximisation problem and does not reflect the marginal excess burden of public funds (financed by an optimal tax system).

3.3.5 Optimal access charges and pricing when the regulator cares about consumers' surplus

Finally, we may also follow the approach due to Baron and Myerson (1982) which puts more weight on consumers' rather than on producer's surplus, obtaining similar results. Adopting this approach, apart from efficiency issues, we are more concerned of consumers' surplus and we assume that lump sum transfers are available only towards the public firm. The public or regulated firm's profits represent public revenues which have unit weight being redistributed to consumers.

When the public (or regulated) firm's objective function is W^ϕ, i.e. Baron and Myerson's one, his profits, as consumers' surplus, have a unit weight, differently from the entrant's ones $\phi\pi^e$, consequently the consumers' surplus becomes $S(q^0, q_H, q_L, q_H^e, q_L^e, N_H, N_L, K_H, K_L)$. Therefore, the social welfare maximisation problem can be written as:

$$\max W^{\phi} = U+(1-\phi)R^{e}(Q^{e}) - C(Q^{i}) - \mathcal{NC}(2N, Q^{0}+Q) + (1-\phi)F - \phi C^{e}_{v}(Q^{e}) =$$

$$N_H\theta u(q_H)+N_L u(q_L)+2Nv(q^0) + K_H\theta u(q^e_H)+K_L u(q^e_L)+(1-\phi)(K_H T^e_H+K_L T^e_L)$$

$$-\mathcal{NC}(2N) - c^*(N_H q_H+N_L q_L) - 2Nc^0 q^0 + (1-\phi)(F_L + F_H) - \phi m(K_H q^e_H+K_L q^e_L)$$

subject to:

[MME] $p^0 \le v'(q^0)$

[IR_L] $u(q_L) - T_L \ge 0$

[IR_H] $\theta\, u(q_H) - T_H \ge 0$

[IC_L] $u(q_L) - T_L \ge u(q_H) - T_H$

[IC_H] $\theta\, u(q_H) - T_H \ge \theta\, u(q_L) - T_L$

[AP_H] $F_H \le (T^e_H - mq^e_H)\, K_H$

[AP_L] $F_L \le (T^e_L - mq^e_L)\, K_L$

Let us proceed considering first the parity rule the regulator may impose for the high and low-demand customers, in order to avoid inefficient entry:

[BW_H] $f_H = F(K_H) / K_H = T_H - c^* q_H$

[BW_L] $f_L = F(K_L) / K_L = T_L - c^* q_L$

Taking advantage of the previous analysis, all may be reduced to the maximisation of the following social welfare function, in which the entrant, being equally or more efficient than the public firm, will cream skim the market:

$$\max W^{\phi} = N_H\theta u(q_H)+Nu(q_L)+2Nv(q^0) + K_H\theta u(q^e_H) + (1-\phi)K_H T^e_H - c^*(N_H q_H+N_L q_L)$$

$$-2Nc^0 q^0 - \mathcal{NC}(2N) + (1-\phi)F_H - m\phi K_H q^e_H$$

subject to:

[MME] $v'(q^0) - p^0 = 0$

[IR_L] $u(q_L) - T_L = 0$

[IC_H] $\theta\, u(q_H) - T_H = \theta\, u(q_L) - T_L$

Making use of the optimality conditions which hold for the customers we can easily write down the first order conditions as:

[q^{ϕ}] $p^{\phi} = v'(q^0) = c^0$ Marginal pricing

$[q_H^\phi]$ $p_H^\phi = \theta\, u'(q_H) = c^*$ No distortion at the top

$[q_L^\phi]$ $p_L^\phi = u'(q_L) = c^*$ No distortion at the bottom

The public firm finds it optimal to apply marginal pricing in the monopolised market of good 0 and to equate marginal prices to marginal costs for both type of customers of good 1.

Here, as in Laffont and Tirole's setting, the terms $[F_H]$ (which represents the public or regulated firm's revenues) and $[-T_H^e K_H]$ (representing the consumers' costs paid to the entrant) do not disappear from the welfare function, because a lower weight ϕ (less than unity) is given to the corresponding costs and revenues of the entrant. Moreover, as in the previous case of the social welfare function with costly public transfers W^λ, but differently from Loeb and Magat's one W, the access pricing condition [AP] is binding, holding as an equality. In fact, it is always socially optimal that the incumbent expropriates all the entrant's profits, as the incumbent's profits are now given a lower weight ϕ (in the welfare function).

However, as already seen in the section 3.2.4, following Baron and Myerson's approach, like in the ultra-liberal case, there is no distortion at the bottom; hence, Pareto efficiency will be achieved when marginal prices equal marginal costs. But, similarly to Laffont and Tirole's welfare function, a lower weight ϕ (less than unity) is now given to the per customer profits of the entrant. Thus, social welfare will increase when the entrant's profits are transferred to the public firm (and to consumers through the public budget) and the Baumol-Willig rule will be no longer a socially optimal access pricing policy in this framework.

Hence, under full information and following the optimal pricing $p_H^\phi = p_L^\phi = c^*$ (with $c^1 = 0$) the Baumol-Willig rule, that prescribes to set the unit access charges

$$f_H^B = T_H^\phi - c^* q_H^\phi$$

$$f_L^B = T_L^\phi - c^* q_L^\phi$$

which leave the public firm's profits unchanged is socially optimal with an equally (or less) efficient competitor. However, analogously to the previous case, the socially optimal access pricing policy will set the unit access charges as:

$$f_H^{\bullet} = \max\{T_H^e - mq_H^e; T_H^{\bullet} - c^*q_H^{\bullet}\}$$
$$f_L^{\bullet} = \max\{T_L^e - mq_L^e; T_L^{\bullet} - c^*q_L^{\bullet}\}$$

As before, we may rewrite the optimal access charges in order to point out the additional terms $(T_i^e - T_i^{\bullet}) + (c^*q_i^{\bullet} - mq_i^e)$, representing the profits of the efficiency of the entrant (under the Baumol Willig rule), the incumbent should expropriate.

$$f_H^{\bullet} = f_H^B + \max\{(T_H^e - T_H^{\bullet}) + (c^*q_H^{\bullet} - mq_H^e); 0\}$$
$$f_L^{\bullet} = f_L^B + \max\{(T_L^e - T_L^{\bullet}) + (c^*q_L^{\bullet} - mq_L^e); 0\}$$

However, if we consider a case in which lump sum transfers from or to the public (or private regulated) firm are no longer allowed and the incumbent's budget constraint is not binding (and/or the entrants profits are directly extracted by the policy maker through lump sum transfers) the Baumol-Willig rule would be again a socially optimal access pricing policy.

3.3.6 Some considerations on the Baumol-Willig rule and the OFTEL rule

It should be noticed how in the previous analysis we have always assumed that there is *no product differentiation* and that (given full information) *the public (or private regulated) firm is not operated inefficiently*. However, we have shown how, even within this context which is very close to Baumol's original one, the application of the Baumol-Willig rule may lead to sub-optimal results following Laffont and Tirole's approach or even adopting the Baron and Myerson social welfare function. This happens because, differently from what implicitly assumed by Baumol in his framework, not necessarily the access pricing fixed by the Baumol Willig rule is socially optimal, as there still remains two distinct firms (a public incumbent and a private competitor) which behave differently and whose profit are given a different social weight.

Nevertheless, the parity principle is fully applicable and the Baumol-Willig pricing rule may be a socially optimal access pricing, when it is possible to avoid those particular redistributional concerns who lead to value less the entrant's profits, differently from what found in Cave (1994) and in Armstrong and Doyle (1994).

This proves that in reality, and even in quite simple models (with perfect regulation of the vertically integrated incumbent in a full information framework and without any product differentiation issue), the Baumol-Willig rule represent a quite useful efficiency reference point (as explicitly claimed by Baumol), even though it will not generally be the optimal access pricing policy for all institutional setting with different functional specifications of social welfare.

Reconsidering, in a private setting, the twofold efficiency role played (in consumption and production) by competition allowing non-linear tariffs within vertically related markets, it clearly emerges how regulation would still be needed. Specifically, while in this setting the regulator should not bother with the problem posed by competition for the low-demand type or by a less efficient entrant (not allowed by the incumbent itself), he should regulate access and impose socially optimal non-linear pricing tariffs in the final good markets (otherwise monopoly pricing would dominate). In fact, cream skimming competition by more efficient entrants (the only type of competition allowed by the incumbent) increases social welfare by a lower amount than it would do in the absence of vertical issues, as the incumbent can keep monopoly tariffs (so that the only gain is in terms of productive efficiency). Basically, this occurs because, while the incumbent fully takes advantage of the greater efficiency of the competitor, through the use of non-linear access charge, he will not reduce the monopoly pricing distortion in the market for final goods. Consequently, both the low type marginal price p_L and the high type tariff T_H remain unchanged notwithstanding competition. Both of them would instead decrease in the horizontal game, where, as the scale of entry increases welfare is enhanced and the distortion at the bottom is progressively reduced. That is, the beneficial efficiency role played in terms of allocative efficiency by cream skimming competition of a more efficient entrant fully disappears with vertically related markets.

Taking advantage of these results we have finally being able to test the

Baumol-Willig rule in the original Baumol's setting and in our vertical game. Naturally, notwithstanding the similar basic hypotheses, there are many relevant differences between these two settings, due to the fact that in our vertical game firms adopt fully non-linear pricing tariffs and that we search for socially optimal non-linear pricing in the final and intermediate goods (allowing for transfers).

The Baumol-Willig rule prescribes that access charges should be set equal to the marginal cost of access, plus a term which reflects the opportunity cost of entry, the idea being to exploit the efficiency enhancing properties of a vertical merger. Therefore it can be a useful point of reference in terms of productive efficiency, deterring the entry of less efficient competition. In what follows we will show how the recently proposed OFTEL rule has the same aim and can be examined in our framework taking the particular case in which prices are linear and socially optimal price cap regulation applies.

A relevant issue which should be consider in more detail is related to the presence of the incumbent's fixed costs. In our framework we already noticed how in many cases the imposition of socially optimal pricing or access rules may lead to problems of deficits, solved by appropriate transfers from the public budget to the firm. No additional distortion would follow from the deficit itself, but only from the additional cost of transfers needed to make the firm break even.

The more appropriate setting in which examine these issues is Laffont and Tirole's one. In fact, the parameter λ characterising their social welfare function is exogenous and can be interpreted as the shadow cost of public funds when transfers are permitted. When transfers are not allowed this parameter is endogenised and represent the multiplier associated to the binding budget constraint of the incumbent. In this different framework the incumbent's deficit increases the value of λ and relevant distortion of the marginal tariffs emerge in connection to higher fixed costs, even adopting non-linear tariffs.

Our model is particularly appropriate to describe telecommunication industries, where these issues have been addressed, through the adoption of specific rules designed to cover access deficit. According to our notation $\mathcal{NC}(2N)$ represent the access deficit, since we separated the incumbent's (British Telecom) costs due to the local network

$$\mathcal{NC}(2N, Q^0, Q^1) = \mathcal{NC}(2N) + c^0Q^0 + c^1Q^1$$

from the ones due to long distance calls

$$C(Q^i) = c^* Q^i$$

The entrant's variable costs, instead, are all related to long distance calls

$$CV^e = m\, Q^e$$

In the UK at the start of competition Mercury was allowed to interconnect with British Telecom's network without making a contribution to his local costs. This helped Mercury to offset her disadvantage from the lack of economies of scale. Since British Telecom was practically constrained to charge uniform tariffs p, Mercury could concentrate her effort on the densest telephone traffic, without having to face aggressive responses.

In Autumn 1993 OFTEL determined of interconnection charges between British Telecom and Mercury and indicated how to deal with the Access Deficit Contribution (ADC) waivers. The ADC system aims to compensate British Telecom for the deficit it incurs (compared with fully allocated costs) in the provision of access.

In the presence of linear prices p on long distance calls (regulated by a price cap) the OFTEL rule imposes on Mercury a tax ADC on a call proportional to the profitability of that call for British Telecom (defined as the ratio between the variable profits in the local network and the total variable profits, including the ones derived from access charges, i.e. fQ^e):

$$ADC = [\mathcal{NC}(2N)/Q^i]\ \{(p^0 - c^0)Q^0/[(p^0 - c^0)Q^0 + (p - c^*)Q^i + fQ^e]$$

Hence the per call access charge can be recursively determined as the sum of the network marginal cost and the access deficit contribution:

$$f = c^1 + ADC$$

In the case in which the price cap is successful in keeping the budget in balance $NC(2N) = AD = (p^0 - c^0)Q^0 + (p - c^*)Q^i + fQ^e$ this rule collapses to the Baumol Willig one, since:

$$f = p - c^1 = c^1 + (p - c^* - c^1)$$

It is then clear how the main purpose of ADC is indeed to discourage inefficient entry.

Major changes occurred in the market itself. Since mid 1993 competition has come to all market segments. There has also been a series of price reduction by British Telecom and the settlement of new prices for new or existing mobile services. The OFTEL rule in fact equally applies in the presence of multiple competitive segments (and therefore access charges). Naturally also in this case if the budget is balanced the OFTEL rule is nothing else than the Baumol-Willig rule examined in the previous sections.

3.4 Final remarks

From the previous analysis we have already noticed how competition can be even welfare reducing and this is clearly true even in the horizontal game without access pricing. In fact, when both firms maximise profits, the total collective surplus may be reduced because some consumers faces the competitor's higher costs and price distortion is increased by the incumbent in order to maximise his profits. A general implication that has emerged is that competition should complement rather than replace regulation.

Let us reconsider now in general terms the role played by competition with non linear tariffs and in presence of a public firm who maximises social welfare. The public incumbent or the regulator should not allow competition for the low-demand type or by less efficient entrant, and should impose the adoption of optimal non-linear tariffs. It is true that in the horizontal game with purely profit maximising firms, the presence of cream skimming competition by more efficient entrants increases welfare, since the distortion at the bottom tends to disappear.

However, even under the most favourable conditions, the private tariffs are not optimal, even adopting an approach that considers only efficiency issues and takes into account the cost of public funds, implicit in the use of public transfers. Basically, this occurs because profit maximising firms give no weight to consumers' surplus and distort prices too much, even in the absence of redistributive issues.

In this chapter we have taken advantage of the Baumol-Willig rule and put it to the test in the original Baumol's setting and in the vertical game. Naturally, in the vertical game both the public incumbent firm and the entrant are allowed to adopt fully non-linear pricing tariffs and we deal with a normative approach, aiming to find socially optimal pricing for access and final goods. However, the Baumol-Willig efficiency rule which prescribes that access charges should be set equal to the marginal cost of access plus a term which reflects the opportunity cost of entry and the idea to exploit the efficiency enhancing properties of a vertical merger, which may be seen as a narrow application of the more general idea of "internalisation".

Notice how we already envisaged within the horizontal game the use of welfare enhancing instruments, such as the adoption of entry taxes, the regulation of the entrant, and finally the application of the purchasing solution (instead of allowing entry and direct selling by the more efficient competitor), which would increase social welfare to a greater extent. Specifically the use of the simplest instrument given by optimal entry taxes avoids socially undesirable entry.

Within vertically related market the parity rule is nothing else than a way to impose entry taxes, while keeping the incumbent's position unchanged, as entry had not occurred. But clearly, adopting the social welfare function of Laffont and Tirole and of Baron and Myerson, it is socially optimal to set even higher access charge tariffs (compared with the ones implied by the parity rule) because the entrant' profits are given a lower weight than the public incumbent's ones. Hence, it is evident that the parity rule, which is built upon the efficiency properties of an vertical merger, does not provide general socially optimal pricing criteria, as it is not always socially optimal under different functional specifications and when firms pursue different objectives. Nevertheless, the enhancing welfare properties implicit in the more general "internalisation principle" may provide in all settings useful additional mechanisms available to the public authority. In fact, in general, ceteris paribus, the use of entrant regulation and of the purchasing solution (instead of allowing entry and direct selling by the more efficient competitor) can increase welfare even more.

In sum, in our analysis we have always assumed a very simplified model in which there is *no product differentiation* and (given full information) *the public (or private regulated) firm is not operated inefficiently*. However, we have shown how, even within this context, the application of the Baumol-Willig rule may lead to sub-optimal results following Laffont and Tirole's approach or even adopting Baron and Myerson's social welfare function. This happens because, differently from what implicitly assumed by Baumol in his framework, the merged monopolist's optimal pricing criteria are not necessarily always socially optimal in the situation where there are two distinct firms which behave differently and whose profits are given a different social weight. Further distortions will emerge when we introduce vertically separated structures, as we will see in the next chapter.

Chapter 4

NON-LINEAR ACCESS PRICING AND COMPETITION IN VERTICALLY SEPARATED MARKETS*

* I am grateful to Michael Waterson for his most valuable advices and his patience. Section 4.2 is partly related to joint work with him (Vagliasindi and Waterson, 1995b).

Chapter 4

NON-LINEAR ACCESS PRICING, COMPETITION AND OPTIMAL REGULATION: THE CASE OF VERTICAL SEPARATION

4.1 Introduction

There are several policy options that need to be analysed in a vertical setting, the most important of which are related to vertical structure and vertical conduct. Referring to the first question a network industry may be vertically integrated or vertically separated. If the incumbent starts off as a vertically integrated firm then a policy of vertical separation involves divestiture. Whether a network industry should be vertically integrated or vertically separated depends on the extent of vertical economies and on the cost of regulation.

An intermediate approach may be to allow the incumbent to own the network but to oblige him to keep separate accounts for his network (setting the transfer price paid by his retail organisation equal to the access charge imposed on any potential entrant). This approach allows to combine the benefits of economies of scope with the ban on discrimination by the network.

As we have already dealt with the case of a vertically integrated monopoly in the previous chapter, here we will focus on vertically separated structures in which the network is owned by an upstream monopolist. We will not be concerned with any transition effect and with the issue of the eventual welfare loss connected to the separation of the vertically integrated incumbent in two distinct firms (the upstream monopolist and the downstream incumbent) following the regulatory authority's decision. In fact here we will just examine the implication of this new setting under different hypotheses on the cost function of downstream producers.

There is a very little literature on the regulation of vertically separated structures. Yet for a long time that regulatory reforms have moved in the direction of divestitures and deregulated structures. Vickers (1995) offers some insights on vertical arrangements when there is regulation of monopoly pricing. Regulation is imperfect, due to the presence of asymmetric information. Also downstream competition is imperfect, as firms behave a la Cournot. The main question analysed is how information and competition imperfections can balance. Some conclusions on the optimal setting of access pricing are derived. In the case in which access price influences the number of the firm welfare losses associated with competition imperfections are greater than the ones associated with the other asymmetric information failure. Consequently access price is set above marginal cost basically to avoid duplication of fixed costs. Access price is instead set below marginal cost when it does not affect the number of firms in the industry.

Our approach differs in several aspects. First, we completely ignore asymmetric information between the regulator and the regulated firm. The reason of this choice should be clear at this point. Basically we believe that in this more complex framework a complete treatment of agency problems would require to introduce asymmetric information in the downstream sectors, not only between the regulator and the new agents, but also between the downstream producers'. Another difference is that we allow for price discrimination for the intermediate good as well as for final good, since we believe that in practice the use of non-linear tariffs is an important feature in regulated industries.

It seems quite natural to start the analysis from an initial setting in which there is a natural monopoly in the network and the regulator adopted a divestiture approach. This seems to resemble the view taken by the US regulator in the telecommunication case, as well as the approach of the UK policy makers in the electricity and gas industries. Following at first a *positive* approach here we want to examine the *private incentives* of the economic players (basically an upstream incumbent and more firms producing the final good in the downstream sector) in the absence of regulatory constraints, apart from the initial divestiture. In dealing with private incentives of both parties we will follow the upstream incumbent's point of view in analysing the conditions under which he can maintain the same profits as a fully integrated firm while entry occurs at one vertical level. In doing this we want also to find out whether it is in his interest to oblige the downstream incumbent and the (eventual) competitor by the use of an appropriate access charge to modify their tariffs in order to maximise joint profits rather than to act myopically in their own interest.

Our focus is still the determination of the range of circumstances in which access pricing regulation can be used to bring about a competitive solution to final goods supply, whilst a monopoly remains at one essential point in the chain of delivery. To develop this regulatory research topic, we will expand the model proposed in chapter 2, following the lines of chapter 3, tackling the problem of entry and price discrimination as crucial ones. We will limit the analysis within the simple framework in which both the incumbent and his competitors have the same technical requirements for the intermediate good, defined as the "common network case" as we did in the case of vertical integration examined in chapter 2. We have already shown how the access charge determined by the incumbent depends apart from the type of network cost, the game's structure, the strategy of competition chosen by the entrant, her cost and scale of entry.

In particular, in this chapter we will show how, under complete information, in the absence of regulation and following all the previous cost assumptions, the upstream monopolist, if allowed to price discriminate, will be able to expropriate all the downstream industry's profits. It is worth noticing how in this simplified setting, vertical separation will not introduce major changes in the vertical game proposed in chapter 2, so that the same results derived for a vertically integrated industry still hold.

When perfect discrimination between downstream producers is forbidden (and only non-linear pricing is allowed) by the regulatory authority, we are dealing with an incomplete information setting at the downstream level; however, as long as it is possible to resell access rights, or the access charge can be expressed as a function of the different consumption bundles that characterise each type of customers, nothing changes with respect to the complete information benchmark.

Even in this setting, as under vertical integration, *cream skimming* will be the only strategy of competition allowed by the upstream incumbent, who obliges the competitor to act as a surplus taker, as long as he faces a fixed scale of entry. However, we will also show how the standard result of 'no distortion at the top' does no longer hold once we endogenise the scale of entry. In fact, in this case, because of the constraint imposed on the quantity produced, a new trade off arises between rent extraction and profit maximisation for the downstream incumbent, so that a tariff *distortion at the top* is introduced.

The previous considerations seem to lead us to conclude that the vertical structure of a network industry may not matter a great deal adopting the previous cost assumptions, if a general non-linear pricing tariff is allowed and no other specific regulatory measures are taken. However, before deriving this strong conclusion, which holds for more general frameworks, we will analyse a more standard case of vertical separation, in which the marginal costs of the downstream producers do differ, as in the standard non-linear pricing models.

Specifically, the downstream incumbent (type "i") is strictly more efficient than the downstream entrant (type "e") for any equal number of customers served. Downstream producers face one or two types of consumers and the upstream monopolist can use non-linear access charge tariffs. In practice, the basic cream skimming model continues to hold, even if the two roles are interchanged: in fact, in the absence of any capacity constraint, it is likely that the downstream entrant will be obliged by the upstream monopolist's access charge tariffs to serve only low-demand customers, since she is less efficient and in this way the rent enjoyed by the downstream incumbent is reduced.

Furthermore, the previous type of welfare analysis with socially optimal regulation under full information, will be pursued in this case.

It will be shown how in general cream skimming turns out to be banned by the upstream monopolist (whether he is a private or a public firm) and going for the low consumers is the most likely strategy of competition for the entrant, who, differently from before, will no longer act as a surplus taker towards the H customers. This may eventually imply that the upstream monopolist now faces a new incentive compatibility constraint.

This is not the case for a public (or regulated) upstream monopolist, who maximises the Loeb Magat welfare function, as in first degree price discrimination, since there are not strong incentives to allow the downstream entrant to serve only low-demand customers.

The main general conclusion that we can derive from this analysis is that there is **no distortion at the top** for the downstream entrant's customers and **some distortion at the bottom, for low customers' allocations and bundles**. This means that, in practice, a modified monopoly result applies also in this framework, with quite relevant *exceptions*. In particular, when the upstream monopolist is allowed by the regulator to use the more general access charge tariff structure, the private monopoly solution implies a *tridimensional* distortion with respect to Loeb Magat's welfare maximising solution as: (a) on the number of customers served by the downstream entrant, (b) on the low customers' bundle, and (c) on the high customers' bundle offered by the downstream entrant. A similar result applies in Laffont and Tirole's case, as welfare maximisation distorts the marginal access charges, but to a lower extent. Instead, with Baron and Myerson's social welfare function there is a *twofold* distortion only with respect to the downstream entrant, as only the number of customers and his customers' bundle are reduced.

The structure of this chapter is the following. We will first discuss the structure of the approach chosen to model vertical separation (section 4.2), specifying the main assumptions of the proposed game. In sections 4.3 we will provide a more standard case of vertical separation; i.e. when the downstream producers are characterised by different levels of efficiency in the presence of only one type of customers. The same vertical structures will be discussed in section 4.4, extended to the presence of two types of customers. A final section (4.5) summarises the main results achieved by our analysis.

4.2 Modelling approach to vertical separation

Let us now consider the problem of vertical separation in the context of the vertical game, introduced in chapter 2 (section 2.5). To examine this case, we need to change assumption **(v)**, as specified below, focusing on a specific example of vertical separation; that is, the separation between the production of good 0 and of the intermediate good from the production of good 1. Apart from that, we will maintain the same assumptions and the basic rules as in the vertical game.

4.2.1 Vertical Separation

(v') In practice, with vertical separation instead of dealing with a single incumbent we are in the presence of two new firms. The first, the **upstream monopolist**, produces only Q^0 (good 0) and provides the intermediate good network's access. In fact, he owns and operates the network sustaining the related costs *NC*, which depend, as before, on the number of customers to be served (2N) and on the total quantity of commodities (Q^0, Q^1) which flows through the network. To simplify matters we may assume (setting for simplicity's sake c^1 equal to zero):

$$NC(2N, Q^0) = NC(2N) + c^0 Q^0$$

The upstream monopolist will also set an access charge tariff $F^i \geq 0$ for the second firm, i.e. the downstream incumbent (as well as for any potential downstream entrant $F^e \geq 0$).

The **downstream incumbent** is characterised by the same *cost* and *revenue functions* in the final goods market as the vertically integrated incumbent:

$$C(Q^i) = c^* Q^i = c^* (N_L q^i_L + N_H q^i_H)$$

$$R^i = N_L T^i_L + N_H T^i_H$$

where $N_t = N - K_t$ denotes the residual number of customers of type t served by the downstream incumbent, c* is the downstream incumbent's constant marginal cost and $Q^i = N_L q^i_L + N_H q^i_H$ denotes the total output. The downstream incumbent will produce a positive amount of good 1 (Q^i) only if his participation constraint PC (with reservation net profits equal to zero, for simplicity's sake) is satisfied:

[PC^i] $\Pi^i = R^i - c^*Q^i - F^i = N_L T^i_L + N_H T^i_H - c^*(N_L q^i_L + N_H q^i_H) - F^i(N_L, N_H, Q^i) \geq 0$

The **downstream entrant** is defined by the same assumption as the entrant in our basic game (cf. section 2.3) and will produce a positive amount of good 1 (Q^e) only if her participation constraint is satisfied:

[PC^e] $\Pi^e = K_L T^e_L + K_H T^e_H - m(K_L q^e_L + K_H q^e_H) - F^e(K_L, K_H, Q^e) \geq 0$

4.2.2 Price discrimination and vertical separation

Allowing for *perfect price discrimination*, under complete information and in the absence of regulation, the upstream monopolist does not need to bother with any incentive compatibility constraint in the downstream sector. In this setting, following the upstream monopolist's point of view it is evident that he can maintain the same profits as a fully integrated firm while entry occurs at the downstream level.

In fact, in this case, the upstream incumbent can appropriate the entire gross profits of the downstream incumbent and of the downstream entrant π^j (with $\pi^j = R^j - C^j$ for j = i, e) simply by setting the two access tariffs $F^j = \pi^j$. In this way, joint gross profits $\Sigma\pi^j = \Sigma[\Pi^j(Q^j)+F^j]$(for j = i, e) are maximised, so that the previous production levels (as in the absence of vertical separation) are reached.

In practice, the upstream monopolist will pretend from the downstream incumbent a per customer access charge equal to his variable profits: that is, $\pi^i_H = T^i_H - c^*q^i_H$ for each high-demand customer and $\pi^i_L = T^i_L - c^*q^i_L$ to give access to each low-demand customer. The downstream incumbent facing this access charge in order to break even must set his tariffs so as if he owned the network. In the case in which he does not set the same tariffs and quantities as an integrated monopolist, the perfectly discriminatory access charge he faces is greater than his gross profits. In this way the downstream incumbent is prevented from acting myopically in his own interest at the expenses of joint profits, that is decreasing the marginal price of the L type increasing his own profits (but reducing the total profits).

Given these premises, the upstream monopolist can enjoy the same profits as an integrated firm. Hence, he will set the same access price for the entrant as if he were a vertically integrated firm. More generally, we can take advantage of all the results of the analysis carried out in chapter 2 to describe his behaviour towards the entrant. Specifically, only the entry of a more efficient rival can take place. The downstream competitor's optimal response is to cream skim the market and behave as a surplus taker. All these results are summarised in the following proposition, whose proof can be immediately derived from the previous analysis (cf. the proof of proposition 3).

Proposition 5

Under assumptions (i) (ii) (iii) (iv) and (v'), allowing the upstream monopolist to make use of any type of access price discrimination (between downstream producers), there is only cream skimming competition for good 1 and the upstream monopolist is able to reach the same result as he would without vertical separation. It is optimal for the upstream incumbent to set the per customer access charge equal to the downstream firm's profits and in particular to oblige the downstream incumbent to maintain the same pricing strategy as in the vertical integration case, and the downstream entrant to cream skim and to behave as a surplus taker.

In this way (i.e. allowing first degree access price discrimination) vertical separation does not introduce major changes in the model; the only difference being the fact that it is now the upstream monopolist who gets the entire profits of the three firms, since the downstream incumbent and the entrant just break even.

Allowing instead only for *non-linear pricing*, the upstream monopolist will need to take into account an incentive compatibility constraint and to deal with the two downstream producers as if he were under an incomplete information setting. Let as assume, as often happens in practice, that regulation simply forbids to discriminate between different downstream producers of good 1. In order to examine a more interesting setting, we will now reintroduce assumption (iv') allowing for endogenous scale of entry of the downstream competitor.

Proposition 6

Under assumptions (i) (ii) (iii) (iv') and (v'), allowing for any finite number of customers as scale of entry but no first degree price discrimination between downstream producers (by the upstream monopolist), there is only cream skimming competition for good 1. Furthermore, the upstream monopolist is able to reach the same result as with perfect price discrimination only if it is possible to re-sell access rights, or the access charge is a function of the consumption bundles of the different customers. In these cases, as before, the upstream incumbent will set the per customer access charge equal to the downstream firm's profits and will oblige the downstream incumbent to maintain the previous monopoly pricing strategy, and the competitor to behave as a surplus taker.

Given our simplified assumptions about the cost functions, we will simply verify that at the end the downstream incumbent's and the entrant's incentive compatibility constraints are satisfied by the solution of the game. Moreover, in order to simplify matters and to allow for the comparison of the two cases with or without vertical separation, we will just proceed step by step, making use of the solution of the incumbent's problem, as derived from the proof of proposition 4.

The best that the upstream incumbent can do (according to what stated in proposition 5 above) is to set the access charge tariff F^e equal to the entrant's total gross profits $k[T^e_H - m(k)q^e_H]$ and the per customer access tariff for the downstream incumbent equal to his gross profits $T^i_t - c^* q^i_t$ for each customer of type t. In fact, in this way the total profit of the industry is maximised and fully expropriated by the upstream incumbent. This is possible in two special settings:

1) when it is possible for a downstream producer to resell access rights;

2) when the general access charge tariff $F(N^j_L, N^j_H, q^j_L, q^j_H)$ -where $N^j_t = K_t, N_t$ for $j = i$, e respectively- depends in a separate way upon the consumption bundles (q^j_L, q^j_H) of the two customers' types and not just on the aggregate consumption Q^j.

In what follows, we will examine the two cases in a separate way. Let us start from the first one. Here, the upstream monopolist can sell all the access rights to one of the downstream producers in order to impose total price discrimination. In fact, denoting by π^j the downstream producer's gross profits (as before) he can simply set the following tariff function:

$$F(N^j_L, N^j_H, Q^j) = \max \Sigma_j \pi^j \qquad \text{for } N^j_L=N;\ N^j_H=N \text{ and } Q^j=Nq^m_L+(N-K)q^m_H+Kq^e_H$$
$$= \max_j (\pi^j+\varepsilon) \qquad \text{elsewhere for } j = i, e \text{ and } \varepsilon>0$$

Clearly, then neither downstream producer can break even if he picks up separately his own offer. In fact, one of the two downstream producers must buy the access rights for all the customers (2N), retaining the ones that it would have served in the vertically integrated setting, while selling the rest to the other player. Assuming that the downstream incumbent buys all the access rights, he can break even only if he produces the monopoly bundles q^m_L and q^m_H for the residual number of customers N_L=N and N_H=N-K and resells the remainder of the access rights (K_H=K) to the downstream entrant fully appropriating her profits. In this way the downstream entrant breaks even only if she cream skims the market (serving K_H=K customers) and behaves as a surplus taker.

Let us turn now to case **(2)** where we have a general access charge tariff $F(N^j_L, N^j_H, q^j_L, q^j_H)$ expressed as a function of the consumption bundles in a separate way. In this case, the upstream monopolist will adopt the following tariff function:

$$F(N^j_L, N^j_H, q^j_L, q^j_H) = \pi^m \qquad \text{for } N^j_L = N,\ N^j_H = N - K,\ q^j_L = q^m_L,\ q^j_H = q^m_H$$
$$= \pi^e \qquad \text{for } N^j_L = 0,\ N^j_H = K,\ q^j_H = q^e_H$$
$$= \max_j (\pi^j + \varepsilon) \qquad \text{elsewhere, for } j = i, e \text{ and } \varepsilon > 0$$

Clearly, each one of the two downstream producers can break even if he (or she) picks up the offer designed in order to make him (her) maximise joint profits.

As before, the incumbent breaks even only if he produces the monopoly bundles q_L^m and q_H^m for $N_L = N$ and $N_H = N - K$ customers and the downstream entrant breaks even only if she serves only high-demand customers ($K_H = K$) behaving as a surplus-taker.

4.2.3 The downstream incumbent's maximisation problem: a reflection

It may be interesting to show that the downstream incumbent will not set the vertically integrated pricing when he faces the access tariff function specified below:

$$F(N_L^j, N_H^j, Q^j) = \pi^m \quad \text{for } N_L^j=N;\ N_H^j=N\text{-}K \text{ and } Q^i=Nq_L^m+(N\text{-}K)q_H^m$$
$$= \pi^e \quad \text{for } N_L^j=0;\ N_H^j=K \text{ and } Q^e=Kq_H^e$$
$$= \max_j(\pi^j+\varepsilon) \quad \text{elsewhere for } j=i, e \text{ and } \varepsilon>0$$

Clearly, once he chooses the access charge $F(N_L^j, N_H^j, Q) = \pi^m$ (because otherwise he will incur in losses) he can always do strictly better than break even by modifying his pricing away from pricing which would be set in the vertically integrated case. In order to deal with this case, we may write down the downstream incumbent's maximisation problem, in the presence of entry with the additional quantity and participation constraint, respectively [Q^m] and [PC^i], derived from the access pricing constraint:

[DI.1] $\max \Pi^i = NT_L^i+(N\text{-}K)T_H^i - \pi^m - c^*[Nq_L^i + (N\text{-}K)q_H^i]$ subject to:

[IR_L] $u(q_L^i) - T_L^i \geq 0$

[IC_H] $\theta\, u(q_H^i) - T_H^i \geq \theta\, u(q_L^i) - T_L^i$

[Q^m] $Q^i \leq Q^m=Nq_L^m+(N\text{-}K)q_H^m$

[PC^i] $\Pi^i = R(Q^i) - c^*Q^i - \pi^m = NT_L^i + (N\text{-}K)T_H^i - c^*(N_Lq_L^i + (N\text{-}K)q_H^i) - \pi^m \geq 0$

It is easy to notice how, substituting [IR_L] and [IC_H] into the objective function, ignoring the other constraints, the problem **[DI.1]** becomes exactly the same as the incumbent's problem in the absence of the access price issue in the presence of cream skimming competition (cf. section 2.3.1). Hence, the output produced in this case will be greater than the one the upstream monopolist would like in order to maximise joint profits (i.e. Q^m). In fact, following the results derived in chapter 2, while the per customer bundle offered to the high type q^i_H is the same, the one addressed to the low type q^i_L is always higher in the presence of cream skimming (since N_H=N-K and N_L=N).

In fact, solving the maximisation problem ignoring the additional constraints, we end up with:

$[q^i_H]$ $p^i_H = \theta\, u'(q^i_H) = [1 - (\theta - 1)\, N_H/N_L]\, u'(q^i_L) = c^* = p^m_H$ No distortion at the top

$[q^i_L]$ $p^i_L = u'(q^i_L) = c^*/[1 - (\theta-1)N_H/N_L] < p^m_L = c^*/(2-\theta)$ Distortion at the bottom

Thus, we would be back to the equations **[2.3]**, **[2.4]** and **[2.5]** previously derived for the horizontal game in chapter 2 in the case of cream skimming, but this time the downstream incumbent's choice is constrained by the fact that he cannot produce a quantity greater than $Q^m = Nq^m_L + (N\text{-}K)q^m_H$ and he must serve all low-demand customers (i.e. $N_L = N$) and the residual high-demand customers [N_H = (N-K)< N]. Clearly, as shown by the first order conditions, since the marginal price p^i_L is strictly less than the one derived in the vertically integrated case p^m_L we have that the bundle q^i_L is strictly greater than q^m_L. Consequently, in the absence of [Q^m] the output $Q^i = Nq^i_L + (N\text{-}K)q^i_H$ would be strictly greater than Q^m so that this constraint binds. On the other hand, since the downstream incumbent breaks even applying the monopoly tariff, his participation constraint [PC^i] is automatically satisfied by the solution of the maximisation problem.

Therefore, the solution of the maximisation problem **[DI.1]** becomes exactly the same as the revenue maximisation problem of the incumbent (for a given Q equal to Q^m) in the absence of the access price issue. In fact, from equation [Q^m] we know that the produced quantity is fixed, and so are all costs, so that the downstream incumbent will maximise profits simply by maximising his revenue subject to [Q^m]:

[DI.2] $\max R^i = NT^i_L + (N\text{-}K)T^i_H$ subject to:

$[IR_L]$ $u(q^i_L) - T^i_L = 0$

$[IC_H]$ $T^i_H = \theta\, [u(q^i_H) - u(q^i_L)] - T^i_L$

$[Q^m]$ $Q = Q^m = Nq^m_L + (N\text{-}K)q^m_H$

Hence, in the presence of vertical separation the general equation $\theta u'(q_H)=(2-\theta) u'(q_L)$ no longer holds as is shown by the solution of the previous maximisation problem **[DI.2]**. In particular, from the first order conditions that characterise an interior solution we obtain the usual relationship between the two marginal prices:

$$p^i_H = [1 - (\theta - 1) N_H/N_L] \, p^i_L$$

and between q^i_L and q^i_H:

$$\theta \, u'(q^i_H) = [1 - (\theta - 1) N_H/N_L] \, u'(q^i_L)$$

where the values $N_H = (N\text{-}K) < N$ and $N = N_L$ are fixed.

Furthermore, the usual 'no distortion at the top' condition $p^i_H = c^*$ no longer holds. In fact, as the production quantity is constrained, the relevant marginal cost for the downstream incumbent includes a positive shadow price and is greater than c^*. Consequently, while the bundle of the high-demand customers q^i_H is decreased, the bundle of the low-demand customers q^i_L is increased. Hence, in this case the upstream monopolist can no longer impose the monopoly tariffs to the downstream incumbent, so that the maximisation solution of the latter creates a *tariff distortion* which affects (in the sense of reducing) the profits of the downstream entrant. In practice, the type of tariff previously proposed collapses, as the surplus taker downstream entrant does no longer necessarily break even under the new tariffs fixed by the downstream incumbent. This poses a major problem for the upstream monopolist, since there is a trade off not only between rent extraction and efficiency, but also between rent extraction and profit maximisation. The upstream monopolist anticipates these problems but has no longer the instrument to impose (through the access charge function) the monopoly tariffs to the downstream incumbent.

4.3 A more standard case of vertical separation with one type of customers

In the previous section we have followed the original hypotheses of chapter 2 examining the case of more efficient ($m < c^*$) entrant in the ambit of the vertical game. We have concluded that the results of the unregulated vertical game do hold even under the drastic regulatory measure that imposes complete vertical separation if: a) the upstream monopolist cannot use complete discrimination; b) when it is possible to re-sell access rights; or c) the access charge can be expressed as a function of the consumption bundles of the different customer.

For completeness' sake, let us examine a more standard framework of price discrimination, in which the downstream sector is portrayed quite similarly to the final demand market. This implies that the upstream monopolist cannot perfectly discriminate between downstream incumbent and entrant. In what follows we will assume the downstream incumbent (type "i") to be strictly more efficient of the downstream entrant (type "e") for any equal number of customers served. Naturally the opposite case (in which the entrant is strictly more efficient than the incumbent) can be obtained simply by re-labelling the two types. In fact, notice how in this new setting the rival in determining the number of customers to be served is not constrained by a limited scale of entry, nor a fixed capacity (differently from the previous analysis). The reason why we consider the case of a less efficient entrant is to explore the benefits of entry, apart from the ones brought by productive efficiency.

In order to examine this more standard case of vertical separation, we will modify the assumption regarding the cost function of the downstream producers. Furthermore, to examine carefully this new situation, in this section we will just tackle the simplest case with only one type of customer (i.e. when the parameter θ is equal to unity). Therefore we have the same total population (i.e. 2N), but composed only of low-demand customers.

It may be useful to summarise the general structure of the vertical separation game which represents just a minor modification of the previous vertical game and will be maintained in the following sections where we deal within this framework.

Vertical Separation Game

(0) the authority sets up the regulatory system (e.g. the admissible terms of access);

(1) the upstream monopolist proposes the general access charge tariff function F(.) which allows the downstream producers to gain access to a number of customers in each market, and to serve them a given total output already produced;

(2) on the basis of the public access charge tariff function F(.) the downstream entrant decides her scale of entry in terms of the number of customers {K} to be served in each market;

(3) the downstream incumbent chooses the pricing rule {T, q} to apply to the remaining customers, where T = T(q) is a fully non-linear tariff;

(4) the entrant chooses her strategy of competition and the non-linear tariff $\{T^e, q^e\}$.[41]

Furthermore, we need to modify the initial assumptions (iv) and (v) considering a specific case when the downstream producers' marginal costs are a function of the number of customers served (n and k), the per customer output (q^i and q^e) and an additional parameter (β^i and β^e) which captures the efficiency of the firm. Naturally, apart from that, we will maintain the same assumptions and the basic rules of the vertical game.

(iv*) The downstream producers (the incumbent and the entrant) can freely choose the number of customers to be served:

$$n = N_L \leq 2N$$
$$k = K_L \leq 2N$$

They pay, as before, a non-linear access charge tariff F(.) fixed by the upstream incumbent (a cost of entry) which may depend in general both on the number of customers served ($N^j = n, k$) and on the per customer output ($q^j = q^i, q^e$) produced by a single firm:

$$F = F(N^j, q^j)$$

[41] Naturally, with a single type of customers the most profitable strategy is always to leave no surplus to the consumers. But, even in this simple framework, where the order of moves between the downstream incumbent and the downstream entrant loses all its meaning, we may have quite different versions of the game, depending on the nature of the network cost function and admissible access pricing strategies.

In substance, the upstream incumbent cannot discriminate between the two downstream producers, but may simply make use of non-linear pricing, since first and third degree price discrimination are no longer feasible.

They have a production cost which depends on the number of customers served, on the per customer output q^j produced by a single firm, and on a parameter β^j which captures the efficiency of the firm:

$$C^j = C(N^j, q^j, \beta^j)$$

In practice, in order to simplify notation, we set:

$$\beta^e = 1$$

$$\beta^i = \beta \qquad \text{with } 0 < \beta < 1$$

considering a simple functional form such that for the same scale of entry the fraction of production cost which varies with the scale of entry of the downstream competitor (i.e. the least efficient producer) is a fixed multiple $1/\beta$ of the downstream incumbent (i.e. the most efficient producer). In particular, we assume that costs are strictly increasing with respect to the number of customers served and total output, convex in their first argument and linear in the second one (the cross derivative has a positive sign):

$$C_N > 0;\ C_{NN} \geq 0;\ C_q > 0\ ;\ C_{Nq} \geq 0 \text{ and } C_{qq} = 0$$

Without loss of generality, in some parts of the analysis we will specify the function as:

$$C^j = [\hat{c} + \beta^j\, m(N^j)]\, Q^j$$

where $\hat{c}$ is the component of the cost function which is invariant with respect to the scale of entry (a constant which is often normalised to zero). An alternative way of representing this situation would be to assume c^j equal to the average cost C^j/Q^j (instead of a constant $\hat{c}$), with c^e greater than c^i, the incumbent's fixed cost being already sunk. Notice also how using this specification, that we will keep on using in the extension to the two customers' types, costs depend only on the total output and not on the bundle offered to each customer.

The downstream producers are characterised by the following *revenue* and *profit functions*:

$$R^j = R^j(N^j, q^j) = N^j\, T^j(q^j)$$

$$\Pi^j = \Pi(N^j, q^j, \beta^j) = R^j(N^j, q^j) - C(N^j, q^j, \beta^j) - F(N^j, q^j)$$

The downstream producers will enter the market for good 1 only if their participation constraints

$$[PC^e] \qquad \Pi^e = k\, T^e(q^e) - C(k, q^e, \beta^e) - F(k, q^e) \geq 0$$

$$[PC^i] \qquad \Pi^i = n\, T^i(q^i) - C(n, q^i, \beta^i) - F(n, q^i) \geq 0$$

(with reservation prices equal to zero) are satisfied.

(v*) Under vertical separation, as before, the upstream monopolist produces only Q^0 (good 0) and provides the intermediate good network's access, sustaining the related costs $\mathcal{NC}$, which depend, as before, both on the number of customers to be served (2N) and on the total quantity of commodities (Q^0 , Q^i , Q^e) which flows through the network. To simplify matters, we keep on assuming (setting c^1 equal to zero):

$$\mathcal{NC}(2N, Q^0) = \mathcal{NC}(2N) + c^0 Q^0$$

The upstream monopolist -who also sets an access charge tariffs $F(.) \geq 0$ for the two downstream producers- is characterised by the following *revenue* (net of the cost of the monopolised good 0) and *profit functions*:

$$\Pi^u(Q^0, k, n) = p^0 Q^0 - c^0 Q^0 + F(k, q^e) + F(n, q^i) - \mathcal{NC}(2N)$$

In what follows we will solve the vertical separation game by backward induction. Notice how, with one type of customer, since the order of moves between downstream producers in setting the tariffs will not affect the result we may solve their maximisation problems simultaneously. In fact, each of the two downstream producers will end up by proposing perfectly discriminatory tariffs, the ones that allow full surplus extraction.

The application of the revelation principle allows us to further simplify the game, restricting ourselves to the only two weakly separating access tariffs <k, F(k, q^e)> and <n, F(n, q^i)> optimally designed by the upstream monopolist for the two downstream players. In this way the game summarised above is equivalent to a static two stages game (apart from the preliminary regulatory decision). In the first stage the principal (the upstream monopolist) designs the contracts, and the second one in which the agents (the two downstream producers) reveal truthfully their types and choose their contracts, and act according to the game induced by the mechanism (that is, they produce respectively the quantities Q^e and Q^i serving k and n customers).

In what follows we proceed step by step, starting from first degree access charge discrimination specifying later the access charges as a function of the number of customers served and also of the bundle sold to them.

4.3.1 First degree access charge discrimination

Let us now examine the benchmark case of first degree access charge discrimination, in order to analyse later the ways in which the upstream incumbent creates distortions between the downstream producers, using the access charge tariff to maximise his total profit $\Pi^u(Q^0, n, k, q^i, q^e)$. In such a situation (in the absence of high-demand customers) the upstream monopolist can determine n and k and use lump-sum access charges F^i and F^e facing only the two participation constraints [PC^e] and [PC^i] of the downstream producers (with reservation prices equal to zero), since we do not need to take into account the incentive compatibility constraints [IC^e] and [IC^i]. Hence, both the efficient type "i" (the downstream incumbent) and the inefficient type "e" (the downstream entrant) enjoy no net surplus.

[U*]	$\max \Pi^u(Q^0, n, k, q^i, q^e)^* = p^0Q^0 + F^i + F^e - c^0Q^0 - \mathcal{N}C(2N)$	subject to:
[MME]	$p^0 = v'(q^0)$	
[PC^e]	$\pi^e = R^e - C^e = kT^e(q^e) - C(k, q^e, \beta^e) = F^e$	
[PC^i]	$\pi^i = R^i - C^i = nT^i(q^i) - C(n, q^i, \beta^i) = F^i$	

and $2N = k + n$

As usual, [U*] can be solved with respect to q^0, n, k, q^i and q^e, substituting the binding constraints into the objective function, obtaining:

$$\Pi^u(Q^0, n, k, q^i, q^e)^* = 2Nv'(q^0)q^0 - 2Nc^0q^0 + R^e(k, q^e) - C^e(k, q^e) + R^i(n, q^i) - C^i(n, q^i) -$$
$$\mathcal{NC}(2N) = 2Nv'(q^0)q^0 - c^0 2Nq^0 + k[u(q^e) - m(k)q^e] +$$
$$n[u(q^i) - \beta m(n)q^i] - \mathcal{NC}(2N) \qquad \text{subject to}$$
$$2N = k+n$$

where we replace the value of the two tariffs $T^e(q^e)$ and $T^i(q^i)$ with the consumers' surplus $u(q^e)$ and $u(q^i)$ fully expropriated by the downstream producers (allowed to make use of non-linear tariffs).

Hence we can write down the Lagrangean function.

$$\boldsymbol{L^*} = \Pi^u(Q^0, n, k, q^i, q^e)^* - \mu\,(k + n - 2N)$$

Hereafter we denote by $R^j_{Nj} = \partial R^j(N^j, q^j)/\partial N^j$ and $C^j_{Nj} = \partial C(N^j, q^j, \beta^j)/\partial N^j$ the partial derivative of the revenue and cost function with respect to the number of customers served by each downstream producers and by $R^j_L = \partial R^j(N^j, q^j)/\partial q^j$ and $C^j_L = \partial C(N^j, q^j, \beta^j)/\partial q^j = [\beta^j m(N^j)]N^j$ the partial derivative of the revenue and cost function with respect to the bundle q^j hereafter setting for simplicity's sake $\hat{c}$ equal to zero. The first order conditions of the Lagrangean function are specified below:

$[q^0]$	$c^0 = v'(q^0) + v''(q^0)q^0$		
$[{}^*N^i]$	$R^i_n - C^i_n = \mu = \pi^i_n$	$\Rightarrow$	$u(q^i) - \beta\, q^i\,[m'(n)n + m(n)] = \mu = F^i_n$
$[{}^*N^e]$	$R^e_k - C^e_k = \mu = \pi^e_k$	$\Rightarrow$	$u(q^e) - q^e\,[m'(k)k + m(k)] = \mu = F^e_k$
$[{}^*q^i]$	$R^i_L = C^i_L$	$\Rightarrow$	$u'(q^i) = \beta m(n)$
$[{}^*q^e]$	$R^e_L = C^e_L$	$\Rightarrow$	$u'(q^e) = m(k)$

Hence, considering the marginal access tariffs $F^j_L = \partial F(N^j, q^j)/\partial q^j$ with respect to the number of customers and the per customer bundle, as evident from $[{}^*q^i]$ and $[{}^*q^e]$, in a first degree price discrimination setting we have F^e_L and F^i_L equal to zero. Hence, the relation between the consumer's marginal prices $u'(q^j)$ served by different downstream producers is not distorted. Furthermore, the optimal per customer bundles q^j can be expressed only as functions of the number of customers served N^j; that is, $q^j = q^j(N^j)$. Specifically, they are both decreasing function of N^j; since $dq^j/dN^j = \partial(C^j_L/N^j)/\partial q^j = \beta^j dm/dN^j/d^2u/(dq^j)^2 < 0$. This property which will be useful in our graphical representation hinges basically on the fact that cost are increasing in the number of customers served.

Moreover, the other two marginal access tariffs with respect to the number of customers $F^j_{Nj} = \partial F(N^j, q^j)/\partial N^j$ are equal to each others and represent the marginal gross profit of the downstream producers (i.e. the value of the Lagrangean multiplier) $F^i_n = F^e_k = \mu$, as from [*N^i] and [*N^e]. Since F^e_n and F^e_k are equal, the profit $\Pi^u(Q^0, n, k, q^i_L, q^e_L)^*$ is maximised without any distortion between the marginal access prices. Hence, the Lagrangean multiplier μ is equal to the variable profits on the marginal customer of both downstream producers (from [N^i] and [N^e]).

As usual in a first degree price discrimination, the upstream monopolist takes away from downstream producers all variable profits $F^i = \pi^i$ and $F^e = \pi^e$. That is no net positive rent is allowed to the downstream incumbent.

The first order conditions referred to the final good can be summarised as:

[4.1] $$u(q^i) - u(q^e) = C^i_n - C^e_k = q^i\beta\, m(n)\,(1+\varepsilon(n)) - q^e m(k)(1+\varepsilon(k))$$

[4.2] $$u'(q^e) - u'(q^i) = m(k) - \beta\, m(n) = C^e_L - C^i_L$$

where $\varepsilon(n)$ and $\varepsilon(k)$ denote the elasticity of the cost function with respect to the number of customers, evaluated at the equilibrium points.

The efficient customer bundles (F^j_L =0) offered by different downstream producers (q^e and q^i), as well as downstream producers' marginal costs (C^e_L and C^i_L) are not generally equated. However, for isoelastic cost function (when $\varepsilon(n)$ equals $\varepsilon(k)$, for any n and k) equations **[4.1]** and **[4.2]** can be set equal to zero. Trivially, $q^e = q^i = q^*$ is an optimal solution of our problem. In this case also downstream producers' marginal costs per additional customer are equated ($C^e_k = C^i_n = C^*_N$). In fact, from **[4.2]** the per customer bundles are equated (i.e. $q^e = q^i = q^*$) if $m(k) - \beta\, m(n) = 0$. Replacing q^* in equation **[4.1]** we find that it is satisfied for the same optimal values of k and n, as derived from **[4.2]**: that is, u(q*) - u(q*) = q*β m(n) (1+ε)- q*m(k)(1+ε) = 0.

We may easily give a representation of the first degree discrimination in the vertical separation setting, assuming isoelastic cost function and drawing for simplicity's sake a linear functional form to represent the cost increase per marginal customer C^{j}_{Nj}.

In fig. 4.1 we choose as the length of the horizontal axis the total number of customers and call respectively 0 and 2N the origin for the downstream entrant and for the downstream incumbent. The linear marginal costs $C^{e}_{k}(k, q^{e}(k))$ and $C^{i}_{n}(n, q^{i}(n))$ (as functions of the number of customers served by each downstream producer) are shown, as well as the optimal level of consumer gross surplus u(q*). The downstream entrant serves k* customers, whereas the downstream incumbent serves n* customers as the marginal costs C^{e}_{k} and C^{i}_{n} intersect in point C determining the efficient marginal cost level C^{*}_{N}.

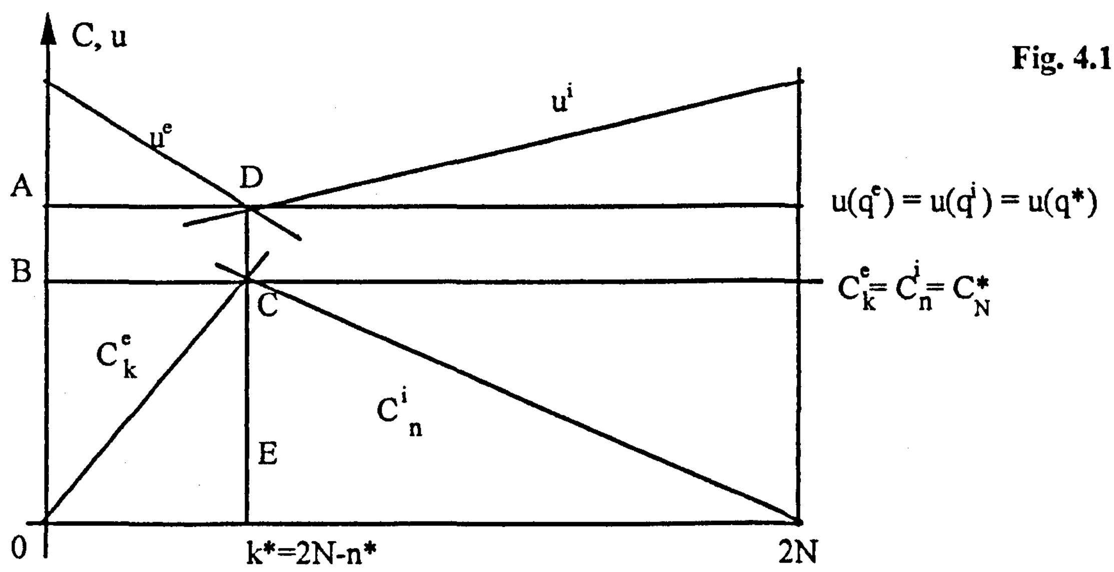


Fig. 4.1

We can also represent on the same axes the two decreasing gross consumer surpluses $u^{e}(k)$ and $u^{i}(n)$ (as functions of the number of customers k and n served by each downstream producer) q^{j} being a decreasing function of N^{j} (as $q^{j} = q^{j}(N^{j})$, with $dq^{j}/dN^{j} < 0$). In particular, gross consumer surpluses equate $[u^{e}(k) = u^{i}(n)]$ for k* (since marginal costs are equal, as evident from the first order conditions [$*q^e$] and [$*q^i$]) and hence intersect in point D determining the efficient gross consumer surplus level $u(q^*)$. The difference between $u(q^*)$ and C_N^* (given by the segment AB) represents the increase of variable profits for an additional customer ($\pi_n^i = \pi_k^e = u(q^*) - C_N^*$).

It should be noticed how the number of customers allocated to the less efficient downstream producers (i.e. the entrant) is less than the one served by the other producer. Specifically, in our isoelastic case, it directly follows from equation **[4.2]** the equality between $m(k^*)$ and $\beta\, m(n^*)$ which implies k^* lower than n^*, as stated above.

An additional question to explore would be to endogenise the *total number of customers*. We can ignore this issue, as we did so far, by assuming the level of customers 2N, fixed exogenously, to be the optimal one, in the sense that there is no reason to leave the market uncovered. Eventually, it could also be considered the alternative frameworks in which the number of customers (2N) is above the optimal level but there is a universal service obligation, or in which the upstream monopolist can endogenously determine the number of customers to be served (though maintaining the same population proportion between the two types, i.e. an equal number of each type of customers).

However, for completeness' sake let us briefly deal with the eventual determination of the optimal number of customers through the profit maximisation of the upstream monopolist specified below, the only difference with the usual one being the introduction of 2N as a choice variable:

$$\max \Pi^u(Q^0, n, k, q^i, q^e, 2N)^* = 2Nq^0(v'(q^0)-c^0)+R^e(k, q^e)-C^e(k, q^e)+R^i(n, q^i)-C^i(n, q^i)-NC(2N) \qquad \text{subject to:}$$

$$2N = k+n$$

We can then write down the modified Lagrangean function as:

$$L = \Pi^u(Q^0, k, n, 2N)^* - \mu\,(k + n - 2N)$$

The first order condition derived in this section remain unchanged; we need only to derive an additional condition with respect to 2N, which can be stated as $\mu = NC'$. From the first order conditions relative to [*N^j] we know that the value of the Lagrangean multiplier is equal to the marginal access charges with respect to the number of customers [$\mu = F^i_n = F^e_k$]; therefore, we immediately get: $F^i_n = F^e_k = NC'$. It directly follows that the marginal access prices applied to the downstream producers equals the marginal cost, when the number of customer is optimally chosen, so that there is **no distortion at the top** and **at the bottom** in the pricing of the intermediate good (i.e. the access to the network).

4.3.2 Distortions when the customers' bundle does not enter in the access charge

In this more standard case of downstream incumbent's cost function we have assumed with **(iv*)** that the upstream monopolist's charge may depend, in general, both on the number of customers served (N^j = n, k) and on the per customer output q^j. However, clearly as the upstream monopolist knows the number of customers served (k and n) he can also make use of the non-linear access charge tariff $F(N^j, q^j)$ which depends, instead, on the customers' bundles ($q^j = q^i$, q^e) to be offered by each downstream producer to customers.

Since in what follows we set c^1 equal to zero, for simplicity's sake and without loss of generality, it may seem useful to first examine a setting in which the access charge can be expressed only as a function of the number of customers served (N^j = n, k), as a consequence of a regulatory ban to discriminate with respect to the output, that is:

$$F = F(N^j)$$

In this regulated case, differently from the more general case, the upstream incumbent cannot discriminate between the two downstream producers using non-linear pricing, with respect to total output, because the latter, by assumption, does not affect the subcost function.

4.3.2.A The downstream producers' problem with one type of customers

Before tackling the upstream monopolist's problem we must explicitly solve the downstream producers' maximisation problem, proceeding by backward induction. In this case for good 1 first and second order price discrimination do coincide and the downstream producers can simply extract all consumers' surplus. In practice, each downstream producer solves his maximisation problem with respect to q^j and N^j.

$\max \Pi^j \equiv \quad R^j(N^j, q^j) - C(N^j, q^j, \beta^j) - F(N^j) \quad$ subject to:

$[IR_L] \quad u(q^j) - T^j(q^j) \geq 0$

Hence, as the participation constraint is binding for the optimal solution, substituting the revenues and cost functions, we get the following first order conditions:

$[q_L] \quad R^j_L = C^j_L \quad \Rightarrow \quad p^j = u'(q^j) = \beta^j m(N^j),$

In practice, as shown by equation $[q_L]$, consumers' marginal prices $N^j u'(q^j) = R^j_L = \partial R^j(N^j, q^j)/\partial q^j$ (which equate marginal revenues) are set equal to marginal costs $C^j_L = \partial C(N^j, q^j, \beta^j)/\partial q^j = \beta^j m(N^j)N^j$ by each downstream producers and no surplus is ever left to consumers. Moreover, the access charge tariff $F(N^j)$ does not directly influence consumers' marginal prices. In particular, from the first order condition we can derive the total output and the profit function specified below:

$$Q^j = N^j q^j; \qquad \Pi^j(Q^j) = N^j T^j - \beta^j m(N^j)Q^j - F(N^j)$$

From the previous solution we see how for the downstream entrant [as $T^e = u(q^e)$ and $u'(q^e) = m(N^e)$ the optimal per customer bundle q^e and the optimal tariff T^e can be expressed only as functions of the scale of entry N^e; that is, $q^e = q^e(N^e)$ and $T^e = T^e(q^e(N^e))$. Specifically, they are both decreasing function of N^e; since $dq^e/dN^e = \partial(C^e_L/N^e)/\partial q^e = dm/dN^e/d^2u/(dq^e)^2 < 0$ and $t^e = dT^e/dN^e < 0$. Similarly, for the downstream incumbent we have that the quantity and tariff offered to each customer are a decreasing function of the number of customers served; that is, $q^i = q^i(N^i)$ with $dq^i/dN^i = \partial(C^i_L/N^i)/\partial q^i = \beta dm/dN^i/d^2u/(dq^i)^2 < 0$ and $T^i = T(q^i(N^i))$ with $t^i = dT^i/dN^i < 0$.

These properties derive basically from the fact that cost are increasing in the number of customers served, due to the presence of diseconomies of scale.

The previous analysis holds in a quite general framework where the upstream monopolist offers a menu of access charges $F^i(N^i)$, summarised by a continuous function of the number of customers. Since we are dealing with the two type case we can further simplify the game, restricting ourselves to the only two weakly separating access tariffs $\langle k, F(k) \rangle$ and $\langle n, F(n) \rangle$ optimally designed by the upstream monopolist for the two downstream players. We are implicitly assuming that the upstream monopolist maximises his expected payoff by allowing entry, instead of offering a unique contract to the downstream incumbent.

4.3.2.B The upstream monopolist's problem with one type of customers

Let us now tackle the upstream monopolist's optimisation problem in the absence of high-demand customers (i.e. when $\theta = 1$ and second degree price discrimination coincides with perfect price discrimination). Regulation simply forbids to discriminate between the two downstream producers of good 1 using non-linear access charges as functions of their total output. The upstream incumbent continues to act as a monopolist, but, since he is allowed only to use non-linear pricing, he must also satisfy their incentive compatibility constraints:

[IC^e] $\quad \pi^e(k, q^e) - F(k) \geq \pi^e(n, q^i) - F(n)$

[IC^i] $\quad \pi^i(n, q^i) - F(n) \geq \pi^i(k, q^e) - F(k)$

In practice, he maximises his profit function $\Pi^u(Q^0, n, k)$ with respect to Q^0, n, k subject to the individual rationality and incentive compatibility constraints:

[U.1] $\quad \max \Pi^u(Q^0, n, k) = p^0Q^0 + F(n) + F(k) - NC(2N) - c^0Q^0$ subject to:

[MME] $\quad p^0 = v'(q^0)$

[PC^e] $\quad \Pi^e = kT^e - C(k, q^e, \beta^e) - F(k) \geq 0$

[PC^i] $\quad \Pi^i = nT^i - C^i(n, q^i, \beta^i) - F(n) \geq 0$

[IC^e] $\quad \pi^e(k, q^e) - F(k) \geq \pi^e(n, q^i) - F(n)$

[IC^i] $\pi^i(n, q^i) - F(n) \geq \pi^i(k, q^e) - F(k)$

and $2N = k + n$

The first constraint [MME] is the usual one, which simply states the monopolised market equilibrium on the consumers' side, imposing the price of the monopolised good to be equal to the marginal utility enjoyed by its consumption. The two following constraints [PC^e] and [PC^i] are the usual participation constraints (with reservation prices equal to zero) of the downstream producers, whereas [IC^e] and [IC^i] represent the incentive compatibility constraints.

As in section 2.3.1 with two types of customers, here the upward binding incentive constraint [IC^e] and the participation constraint of the most efficient producer [PC^i] are automatically satisfied by the solution of the problem when [PC^e] and [IC^i] are binding (the proof follows the lines of the one given in section 2.3.1).

[PC^e] $\Pi^e = kT^e - C(k, q^e, \beta^e) - F(k) = 0$

[IC^i] $\pi^i(n, q^i) - F(n) = \pi^i(k, q^e) - F(k)$

In practice, no surplus is allowed for the downstream entrant (i.e. the inefficient type), whereas the downstream incumbent (i.e. the efficient type) enjoys a positive net surplus.

The maximisation problem, keep on assuming for simplicity's sake $\hat{c}=c^1 = 0$, becomes:

[U.1A] $\max \Pi^u(Q^0, n, k) = p^0Q^0 - c^0Q^0 + F(k) + F(n) - \mathcal{N}C(2N)$ subject to:

[MME] $p^0 = v'(q^0)$

[PC^e] $F(k) = R^e - C^e = kT^e - m(k)Q^e$

[IC^i] $F(n) = \pi^i(n, q^i(n)) - [\pi^i(k, q^i(k)) - \pi^e(k, q^e(k))] = R^i(n, q^i(n)) - C^i(n, q^i(n))$
$- [R^i(k, q^i(k)) - C^i(k, q^i(k)) - R^e(k, q^e(k)) + C^e(k, q^e(k))]$

and $2N = n + k$

Before solving the problem let us introduce as before a simplified notation:

[4.3] $\pi^e_k = \partial\pi^e/\partial k = R^e_k - C^e_k$

[4.4] $\pi^i_n = \partial\pi^i/\partial n = R^i_n - C^i_n$

where $R^{j}_{Nj} = \partial R^{j}/\partial N^{j}$ and $C^{j}_{Nj} = \partial C^{j}/\partial N^{j}$ respectively denote the marginal revenue and cost with respect with the number of customer served.

The previous problem **[U.1A]** can be solved only with respect to k and n, once we substitute the binding constraints [PC^{e}], [IC^{i}] and [MME] into the objective function:

$$\max \Pi^{u}(Q^{0}, k, n) \equiv 2Nv'(q^{0})q^{0} - c^{0}2Nq^{0} + R^{i}(n, q^{i}(n)) - C^{i}(n, q^{i}(n)) - [R^{i}(k, q^{i}(k)) - C^{i}(k, q^{i}(k)) - 2R^{e}(k, q^{e}(k)) + 2C^{e}(k, q^{e}(k))] - \mathcal{N}C(2N) \quad \text{subject to:}$$

$$k + n - 2N = 0$$

We can then write down the Lagrangean function as:

$$\boldsymbol{L} = \Pi^{u}(Q^{0}, k, n) - \mu\,(k + n - 2N)$$

After few algebraic manipulations the first order conditions can be written as:

$$[q^{0}] \qquad c^{0} = v'(q^{0}) + v''(q^{0})q^{0}$$

$$[N^{e}] \qquad 2R^{e}_{k}(k, q^{e}(k)) - 2C^{e}_{k}(k, q^{e}(k)) - R^{i}_{k}(k, q^{i}(k)) + C^{i}_{k}(k, q^{i}(k)) = \mu$$

$$[N^{i}] \qquad R^{i}_{n}(n, q^{i}(n)) - C^{i}_{n}(n, q^{i}(n)) = \mu$$

Notice how the partial derivatives with respect to the per customer bundles q^{j} cancel out from the two final first order condition, due to the equality between the marginal cost and revenue with respect to the relative per customer bundles (as can be verified from [q_{L}] derived from the downstream producers' maximisation problem derived in subsection A). Substituting the value of the Lagrangean multiplier $\mu = R^{i}_{n}(n, q^{i}(n)) - C^{i}_{n}(n, q^{i}(n)) = \pi^{i}_{n}$ (i.e. the shadow profits on the marginal customer or the marginal access price of the most efficient firm) into the first order condition with respect to n we obtain the optimal relationship between n and k:

$$\textbf{[4.5]} \qquad \pi^{i}_{n} = \pi^{e}_{k} + (R^{e}_{k}(k, q^{e}(k)) - R^{i}_{k}(k, q^{i}(k))) - (C^{e}_{k}(k, q^{e}(k)) - C^{i}_{k}(k, q^{i}(k)))$$

The optimal relationship between the two marginal prices is very easily obtained from **[4.5]**, making use of the binding constraints which hold for the downstream producers, i.e. $F^{j}_{Nj} = \pi^{j}_{Nj} = R^{j}_{Nj} - C^{j}_{Nj}$:

$$\textbf{[4.6]} \qquad F^{i}_{n} = F^{e}_{k} - [R^{i}_{k}(k, q^{i}(k)) - R^{e}_{k}(k, q^{e}(k))] - [C^{e}_{k}(k, q^{e}(k)) - C^{i}_{k}(k, q^{i}(k))]$$

The previous relationship between the marginal access prices can be easily re-expressed for our functional form specification and interpreted as follows:

[4.6]' $$F^i_n = F^e_k - [u(q^i(k)) - u(q^e(k))] - [q^e(k) - \beta q^i(k)][m(k)+m'(k)k]$$

In practice, not only does the entrant (here the inefficient type) enjoy no surplus, but also we have a wedge between the marginal access prices, given by the net positive surplus the incumbent (that is the most efficient type) can enjoy, taking the access tariff of the entrant. This distortion of the marginal access charges (and consequently of the customers' marginal prices) is generated by the upstream incumbent, who cannot perfectly discriminate (through non-linear pricing) between the two downstream producers, in order to increase his profits. In fact, with a lower number of customers allocated to the "e" type (less efficient) downstream producer, the rent of the "i" type (more efficient) downstream producer is reduced.

Let us now represent in fig. 4.2, assuming isoelastic cost functions, the result of this section when the access charges can be expressed only as a function of the number of customers and the customers' bundles offered by different downstream producers are not yet directly distorted. In the same picture k* and n* represent the allocation derived in first degree price discrimination, the benchmark case analysed before. As the incentive compatibility constraint of the "i" type is now binding, the most efficient type (i.e. the downstream incumbent) should be allowed a net positive rent, which can be decomposed in two main components. The first rent $[q^e(k^*) - \beta q^i(k^*)]m(k^*)k^*$ (that is the area OCE) is attributed to the lower cost of the downstream incumbent due to his greater efficiency level; in our graphical representation we assume that this effect outweighs the cost increase associated with the quantity increase. The other rent $[u(q^i(k^*)) - u(q^e(k^*))]k^*$ (that is the area SUR) derives from the greater revenue obtained by the downstream incumbent due to the increase in the per customer bundle. Consequently, to enforce the perfectly discriminatory allocation of customers <k*, n*> the downstream incumbent's access charge must be reduced by the total rent. This situation would be no longer optimal for the upstream monopolist, since the marginal profits of a customer allocated to the less efficient type "e" is now reduced by a loss given by the increase in the downstream incumbent's rent $[u(q^i(k^*))-u(q^e(k^*))]+[q^e(k^*)-\beta q^i(k^*)][m(k^*)+m'(k^*)k^*]$ (i.e. segments RU and CE).

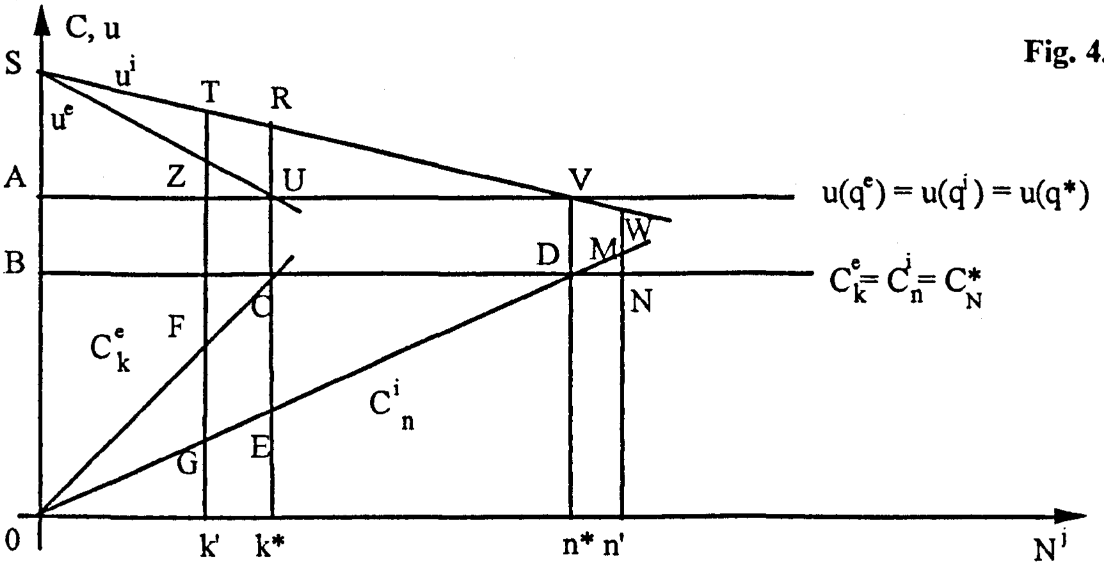


Fig. 4.2

In this case, there is no longer an efficient allocation of customers between downstream producers, because both the downstream incumbent's marginal access charge (F_n and F_k) per additional customer are no longer equal. In order to get the downstream incumbent's marginal access charge we must subtract from the segment ZF the two segments TZ and FG, as evident from equation [4.6].

In general the sum of the marginal rents (TZ + FG) is strictly positive. In this way, a number of customers is transferred from the downstream entrant to the downstream incumbent; i.e. $k^* - k' = n' - n^*$. The net positive rent of the downstream incumbent (that is, the most efficient type) is consequently reduced to the area OFG + SZT. With this new allocation of customers (k', n'), the downstream entrant's marginal cost (C^e_k) is reduced, whereas the downstream incumbent's marginal cost (C^i_k) is increased and there is a positive wedge ($F_k - F_n$) between them. Hence, π^i_n decreases from the level DV to MW and π^e_k increases from the level CU to FZ.

In graphical terms the left hand side of equation [4.6] MW is equal to the right hand side ZF - TZ - FG. In this way, the gains from the marginal customer allocated to each downstream producer are equated; that is, $\pi^i_n = \pi^e_k - [u(q^i(k')) - u(q^e(k'))] - [q^e(k') - \beta q^i(k')][m(k')+m'(k')k']$ (or in graphical terms MW = FZ - GF - ZT). In fact, now the marginal customer allocated to the less efficient type "e" brings a gain equal to the additional variable profit per marginal customer π^e_k (segment FZ) minus the increase in the rent (segment GF plus ZT) to be paid to the more efficient type "i". The upstream monopolist's profits are maximised when this gain (FZ - GF - ZT) is equated to the marginal loss due to the foregone profit of a customer no longer served by the more efficient type "i", i.e. his additional variable profit per marginal customer (segment MW).

An interesting variant of the game would be to endogenise the *total number of customers*. We can ignore this issue, by assuming the level of customers 2N, fixed exogenously, to be the optimal one, in the sense that there is no reason to leave the market uncovered. Eventually, it could also be considered the alternative frameworks, in which the number of customers (2N) is above the optimal level but there is a universal service obligation, or in which the upstream monopolist can endogenously determine the number of customers to be served (though maintaining the same population proportion between the two types, i.e. an equal number of each type of customers). However, let us briefly deal with the eventual determination of the optimal number of customers through the profit maximisation of the upstream monopolist specified below, the only difference with the usual one being the introduction of 2N as a choice variable:

$$\max \Pi^u(Q^0, k, n, 2N) \equiv \quad 2Nv'(q^0)q^0 - c^0 2Nq^0 + R^i(n, q^i(n)) - C^i(n, q^i(n))$$

$$-[R^i(k, q^i(k)) - C^i(k, q^i(k)) - 2R^e(k, q^e(k)) + 2C^e(k, q^e(k))]$$

$$-\mathcal{NC}(2N) \qquad \text{subject to:}$$

$$k + n - 2N = 0$$

We can then write down the modified Lagrangean function as:

$$\mathcal{L} = \Pi^u(Q^0, k, n, 2N) - \mu\,(k + n - 2N)$$

The first order condition derived in section remain unchanged; we need only to derived an additional condition with respect to 2N, which can be stated as $\mu = NC'$. Making use of **[4.6]** to express the value of the Lagrangean multiplier as a function of the marginal access charges $[\mu = F^i_n = R^i_n - C^i_n]$, we immediately get:

[4.7] $$F^i_n = F^e_k - [R^i_k(k, q^i(k)) - R^e_k(k, q^e(k))] - [C^e_k(k, q^e(k)) - C^i_k(k, q^i(k))] = NC''$$

It directly follows from **[4.7]** that the marginal access price applied to the incumbent (the most efficient type) equals the marginal cost, when the number of customer is optimal, so that there is **no distortion at the top**. On the other side, there emerges some distortion at the bottom, even when the upstream monopolist can optimally choose the number of customers to serve through the network.

The previous conclusion shows that the *standard result on price discrimination* for pure monopoly with no distortion at the top can be reinterpreted in this new vertical framework where there are two downstream producers characterised by different efficiency level (unknown to the upstream monopolist), allowing for an endogenous scale of entry (k) of the downstream entrant. The strict equality of equation **[4.7]** will follow when the upstream monopolist can choose the parameter 2N and consequently the total number of customers.

4.3.3 The distortionary solution when customers' bundles influence access charges

Let us now consider the more general assumption present in **(iv')** and let the upstream monopolist's non-linear access charge tariff $F(N^j, q^j)$ depend in general both on the number of customers served ($N^j = n, k$) and on the customer bundles ($q^j = q^i, q^e$) to be offered by each downstream producers to the customers:

[4.8] $$F = F(N^j, q^j)$$

In this case, as before, the upstream incumbent, using only non-linear pricing, cannot perfectly discriminate, in theory, between the two downstream producers, but even if we have only a consumers' type he may influence, as it will be clearer later, the customers' bundles ($q^j = q^i, q^e$) through the access charge, to further reduce the rent of the "i" type (more efficient) downstream producer.

In what follows, we will solve the previous game, with only one type of customers and considering non-linear access charge tariffs as $F(N^j, q^j)$. The upstream monopolist (using second order price discrimination) maximises the following profit function $\Pi^u(Q^0, n, k, q^i, q^e)$ with respect to q^0, n, k, q^i and q^e, subject to the individual rationality and incentive compatibility constraints:

[U.2] $\max \Pi^u(Q^0, n, k, q^i, q^e) = p^0Q^0 + F(n, q^i) + F(k, q^e) - \mathcal{N}C(2N) - c^0Q^0$

subject to:

[MME] $p^0 = v'(q^0)$

[PC^e] $\Pi^e = kT^e - C^e(k, Q^e, \beta^e) - F(k, q^e) \geq 0$

[PC^i] $\Pi^i = nT^i(q^i) - C(n, q^i, \beta^i) - F(n, q^i) \geq 0$

[IC^e] $\pi^e(k, T^e) - F(k, q^e) \geq \pi^e(n, T^i) - F(n, q^i)$

[IC^i] $\pi^i(n, T^i) - F(n, q^i) \geq \pi^i(k, T^e) - F(k, q^e)$

and $2N = k + n$

where the constraints have their usual interpretation, with [IC^e] and [PC^i] automatically satisfied by the solution and with [PC^e] and [IC^i] binding as in **[U.1]**. Hence, while the efficient type "i" (the downstream incumbent) enjoys a positive net surplus, no surplus is allowed to the inefficient type "e" (the downstream entrant).

We keep on assuming, just for simplicity's sake, $\hat{c} = c^1 = 0$, and specify the maximisation problem taking only into account the binding constraints:

[U.2A] $\max \Pi^u(Q^0, n, k, q^i, q^e) = p^0Q^0 - c^0Q^0 + F(k, q^e) + F(n, q^i) - \mathcal{N}C(2N)$

subject to:

[MME] $p^0 = v'(q^0)$

[PC^e] $F(k, q^e) = R^e - C^e = k\,T^e - m(k)Q^e$

[IC^i] $F(n, q^i) = \pi^i(n, q^i) - [\pi^i(k, q^e) - \pi^e(k, q^e)] = nT^i - \beta m(n)Q^i - (1-\beta)m(k)Q^e$

and $2N = n + k$

As usual, **[U.2A]** can be solved with respect to q^0, n, k, q^i and q^e, substituting the binding constraints [PC^e], [IC^i] and [MME] into the objective function, obtaining the new profit function:

$$\Pi^u(Q^0, n, k, q^i, q^e) = 2Nq^0(v'(q^0) - c^0) + R(k, q^e) - 2C^e(k, q^e) + C^i(k, q^e) - + R(n, q^i) - C(n, q^i) - \mathcal{NC}(2N)$$

$$= 2Nq^0(v'(q^0) - c^0) + kT^e - [(2-\beta)m(k)kq^e] + nT^i - \beta m(n)nq^i - \mathcal{NC}(2N)$$

and solving the Lagrangean function, specified as below:

$$\boldsymbol{L} = \Pi^u(Q^0, n, k, q^i, q^e) - \mu\,(k + n - 2N)$$

Substituting the value of T^j, equal to the consumers' surplus $u(q^j)$, and following the usual procedure, we get the new first order conditions:

[q^0]	$c^0 = v'(q^0) + v''(q^0)q^0$		
[N^i]	$R^i_n - C^i_n = \mu = \pi^i_n$		
[N^e]	$R^e_n - 2C^e_n + C^i_n = \mu$	$\Rightarrow$	$\pi^e_k - (1-\beta)C^e_n = R^e_n - (2-\beta)C^e_n$
[q^i]	$R^i_L - C^i_L = 0$	$\Rightarrow$	$u'(q^i) = \beta\, m(n)$
[q^e]	$R^e_L - C^e_L = 0$	$\Rightarrow$	$u'(q^e) = (2-\beta)m(k) = m(k) + (1-\beta)m(k) > m(k)$

This time, we find two additional first order conditions [q^i] and [q^e], while the first three are formally the same. Since from [N^i] the Lagrangean multiplier $\mu = R^i_n - C^i_n = \pi^i_n$ is equal to the variable profits on the marginal customer of the efficient type (i.e. the marginal access price of the most efficient firm), using equation [N^e] we get the optimal relationship between n and k; that is, the optimal distortion which holds between the two marginal access prices for the two downstream producers which is exactly the same as **[4.6]**.

[4.9] $$\pi^i_n = R^i_n - C^i_n = R^e_k - (2-\beta)C^e_k = \pi^e_k - (1-\beta)C^e_k$$

From [q^i] and [q^e] we obtain the new distortionary relation between the consumers' marginal prices $u'(q^j)$ served by different downstream producers, that the upstream monopolist sets, through the use of marginal access charges $F^j_L = \partial F(N^j, q^j)/\partial q^j$.

[4.10] $$\Delta F_L = F_L^e - F_L^i = (1-\beta)\, m(k)$$

In practice, there is a positive wedge between the differential marginal access charges $\Delta F_L = F_L^e - F_L^i = \partial F(k, q^e)/\partial q^e - \partial F(n, q^i)/\partial q^i$, which maximises the upstream monopolist's profits, reducing the net positive rent of the downstream incumbent (that is, the most efficient type), while the downstream entrant is left with no surplus. In fact, while the downstream incumbent's marginal access charge $F_L^i = c^i$ is set equal to zero (as from $[q^i]$ $u'(q^i) = \beta\, m(n) = \beta m(n) + F_L^i$), the downstream entrant's optimal marginal access charge $F_L^e = (1-\beta)m(k)$ is strictly positive (as from $[q^e]$ $u'(q^e) = m(k) + (1-\beta)\, m(k) = m(k) + F_L^e$).

[4.11] $$u'(q^e) - u'(q^i) = [m(k) - \beta\, m(n)] + (1-\beta)\, m(k) = (C_k^e - C_n^i) + (F_L^e - F_L^i)$$

Hence, using his general access charge tariff , the upstream incumbent is able to maximise his profit function $\Pi^u(Q^0, n, k, q^i, q^e)$ discriminating in two ways between the downstream producers. As we can see from **[4.11]** not only different marginal costs ($C_k^e - C_n^i < 0$) are imposed, through a distortionary allocation of the number of customers, but also the choice of customers' bundles (q^i and q^e) is distorted by charging different marginal access charges ($\Delta F_L > 0$).

The previous relationship between the consumers' marginal prices $u'(q^i)$, set by different downstream producers, means that the inefficient type enjoys no surplus and that a bidimensional wedge is set between the two downstream producers' marginal access prices. This aims to reduce the net positive rent the downstream incumbent could gain, taking the inefficient downstream producer's contract.

As before, when the number of customers is optimally chosen by the upstream monopolist the marginal revenue given by **[4.11]** must be greater or equal to marginal cost ***NC'***. In practice, the access charge per additional customer of the most efficient downstream producer is greater or equal to marginal cost ***NC'***

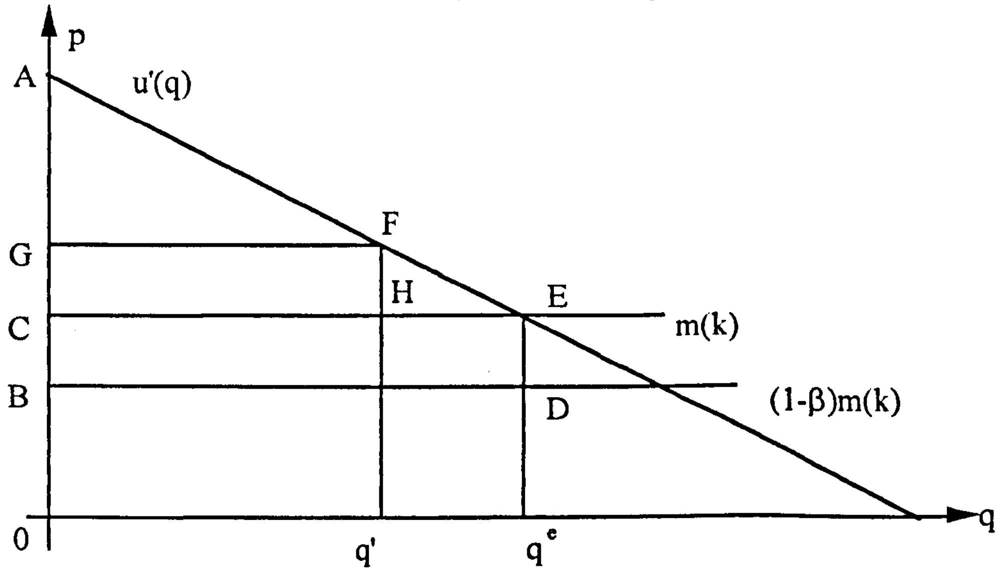
[4.12] $$\pi_n^i = R_n^i - C_n^i = \pi_k^e - (1-\beta)C_k^e = R_k^e - C_k^e - (1-\beta)C_k^e \geq NC'$$

Hence, the marginal access price applied to the incumbent (the most efficient type) equals the marginal cost and there is **no distortion at the top**. This conclusion is still quite close to the *standard result on price discrimination* for a pure monopoly.

Let us finally consider the more general case in which the non-linear access charges F(n, q^i) and F(k, q^e) are allowed (as section 4.3.2) and the customer bundles offered by different downstream producers may be directly distorted setting positive marginal access charges ($F^i_L > 0$). In relation to the allocation of customers between the downstream producers the previous condition $\pi^i_n = \pi^e_k - (1-\beta)\, C^e_k$ (and the relative graphical representation) still hold, but the new revenue function $R^u(q^0, n, k, q^i, q^e)$ offers a new opportunity for reducing the rent of the downstream incumbent.

As the upstream monopolist can also directly fix the bundles q^e and q^i offered by downstream producers, the previous situation ($F^i_L = 0$) is no longer optimal. To show this, in fig. 4.3 we represent the linear marginal utility functions u'(q) of a *quadratic utility function* $u(q)=q - (q)^2/2$. Consider now a non-distortionary marginal price $p^e = OC = C^e_L$, the corresponding bundle q^e and tariff T^e (given by the integral from 0 to q^e of the marginal utility function u'(q) the area AOq^eE).

Fig. 4.3



The variable profit per customer π^e/k, given by ACE, is derived by subtracting from the tariff T^e the variable costs $m(k)q^e$ (represented by the area COq^eE). With first degree access price discrimination this is optimal, as the upstream monopolist's marginal gain (with respect to the quantity set by the less efficient type "e") equates the marginal variable profit already set equal to zero ($dR^u/dq^e = \pi^e_L = 0$).

However, now that only second degree access price discrimination is allowed, the upstream monopolist's marginal gain with respect to the quantity set by type "e" dR^u/dq^e (differently from $\partial R^u/\partial q^i = \pi^i_L = 0$, as from [$q^i$] we have $u'(q^i) - \beta\, m(n) = 0$) is negative since it is equal to the marginal variable profit ($\pi^e_k=0$) minus the loss given by the increase in the downstream incumbent's rent $(1-\beta)C^e_L = (1-\beta)m(k)$ (segment DE).

Hence, while it is optimal that the downstream incumbent's marginal access charge $F^i_L = c^1$ remains set equal to zero, the downstream entrant's optimal marginal access charge is $F^e_L = (1-\beta)\ m(k)$ equal to segment DE. Adding CG (= DE) to the previous marginal cost OC the new distortionary marginal price is $p^e = OG = C^e_L + F^e_L$, and the corresponding tariff T^e is given by the area AOq^eE. The variable profit per customer π^e/k is reduced to ACHF, and $\partial R^u/\partial q^e = \pi^e_k - (1-\beta)C^e_L = 0$ since now the marginal variable profit (HF) is equal to the marginal increase in the downstream incumbent's rent (DE). Thus, in this final case different marginal prices (p^e and p^i) are the result of a distortionary allocation of the number of customers and of a positive marginal access charges $F^e_L > 0$ charged to the less efficient downstream producer.

4.3.4 Access charge discrimination with one consumers' type: welfare considerations

In the present standard framework (with one consumers' type) of the vertical separation model the upstream monopolist allows the entry of a less efficient competitor, who faces no entry costs (or more generally no fixed costs, apart from the access charge). He leaves her no rent and charges her (with non-linear access pricing) also the increase in the rent to be paid to the more efficient type "incumbent" (in correspondence of the marginal increase in the number of customers and in the quantity supplied). With non-linear access pricing, these distortion are clearly optimal for the upstream monopolist, as we have proved, but they also represent a way to avoid the choice of inefficient entry scales by a less efficient competitor. In order to avoid an inefficiently large scale of entry, the potential competitor must pay these two marginal charges, in order to be allowed to enter the low demand market. This leads

leads us to examine what kind of access charge tariff would be socially optimal.

In order to start dealing with optimal regulation in this new vertical framework, let us consider the conditions under which the first degree discrimination access pricing policy is optimal. Ignoring any eventual distributional and excess burden considerations, adopting the ultra-liberal social welfare function a la Loeb and Magat (given by the unweighted sum of consumer surplus and profits) the lump sum tariffs F^i and F^e (introduced in sub-section 4.3.3.A) represent overall the socially optimal access and final prices for the society. In this section, we will always assume that the upstream monopolist is a public (or regulated) firm i.e. the collectivity will pay costs and transfers and get the firm's revenues. But this is not strictly necessary adopting Loeb and Magat's social welfare function.

Since the consumers' net surplus is equal to consumers' gross surplus minus firms' revenues, that is $U(Q^0, Q^i, Q^e) = S(Q^0, Q^i, Q^e) - R^i(Q^i) - R^e(Q^e)\ R^u(Q^0, Q^i, Q^e)$ the public upstream monopolist's general problem is no longer to maximise his revenue $R^u(Q^0, n, k, q^i, q^e)$ but the social welfare W (with respect to q^0, n, k). Taking advantage of the previous analysis and assuming transfers equal to zero for simplicity's sake the whole problem reduces to the constrained maximisation specified below:

[U.LM] $\max W = S(Q^0, Q^i, Q^e) - \mathcal{NC}(2N, Q^0, Q) - C(n, q^i, \beta^i) - C(k, q^e, \beta^e) =$

$2Nv(q^0)+nu(q^i)+ku(q^e)-c^0(2Nq^0)-\mathcal{NC}(2N)-\beta m(n)nq^i-m(k)kq^e$ subject to:

[PC^i] $\Pi^i = nT^i - C(n, q^i, \beta^i) - F(n, q^i) \geq 0$

[PC^e] $\Pi^e = kT^e - C(k, q^e, \beta^e) - F(k, q^e) \geq 0$

[IC^i] $\pi^i(n, T^i) - F(n, q^i) \geq \pi^i(k, T^e) - F(k, q^e)$

[IC^e] $\pi^e(k, T^e) - F(k, q^e) \geq \pi^e(n, T^i) - F(n, q^i)$

and $2N = k + n$

Here, as usual the superscript W indicates optimal pricing a la Loeb and Magat. The problem **[U.LM]** can be solved with respect to q^0, k, q^i and q^e, solving down Loeb and Magat's social welfare function, rewritten (using [PC^e] to substitute the value of T^j) as:

$$W = 2Nv(q^0)q^0 - 2N c^0 q^0 - \mathcal{NC}(2N) + (2N-k)[u(q^i) - \beta m(2N-k)q^i] + k[u(q^e) - m(k)q^e]$$

and getting the new first order conditions.

[LM.1]* $c^0 = v'(q^0) = p^0$

[LM.2]* $\pi^i_n = u(q^i) - C^i_n = u(q^e) - C^e_k = \pi^e_k$

[LM.3]* $u'(q^i) = \beta\, m(n)$

[LM.4]* $u'(q^e) = m(k)$

Entry is optimal with $F(k°, q°)=k°T^e-m(k°)Q^e$ and $F(n°, q°)=n°T^i-\beta m(n°)Q^i-(1-\beta)m(k°)Q^e$ as constraints [PC^e] and [IC^i] are binding, while the others are satisfied. In this case, the downstream producer will automatically set the optimal tariffs $T^i=T^e=T^W$, which equate marginal prices to marginal costs. From the previous equations it emerges that while first degree access price discrimination is socially optimal under Loeb and Magat's social welfare function, this is not true with other types of non-linear access charge tariff function. In fact, with the upstream monopolist's access charges which take away from downstream producers all variable profits $F^i=\pi^i$ and $F^e=\pi^e$ we have the same final price setting in the final market for good 1. Notice how equation **[LM.2]*** is exactly equal to **[4.1]** while equations **[LM.3]*** and **[LM.4]*** are equivalent to **[4.2]**. Instead, in the monopolised final market for good 0 the price p^0 should be equal to marginal cost c^0 from **[LM.1]***, as it was without vertical separation.

But first degree access price discrimination is just a special case in which no net positive rent is allowed to the downstream producers. In general, there are many non-linear access charges which satisfy constraints [PC^e] and [IC^i] (not necessarily binding), leaving some rent to downstream producers, but do not produce any distortion, equating marginal access prices to zero ($F^i_n=F^e_k=F^e_L=F^i_L=0$) for $k=k°$, $n=n°$ and $q^e=q^i=q°$. Hence, Pareto improvements would be achieved by reducing the value of the marginal access charge the entrant is facing (F^e_k, F^e_L), till they are zero.

Let us now explicitly take into account, among the efficiency issues, the cost of public funds (following the approach of Laffont and Tirole) or let us, alternatively, impose the presence of a binding break even budget constraint in the absence of subsidies.

The public upstream monopolist maximises the Laffont and Tirole social welfare function W^{λ} (with respect to N, q^0, n, k, q^i, q^e), where for simplicity's sake we put the trasfer tr equal to zero.

[U.LT] $\max W^{\lambda} = S(Q^0, Q^i, Q^e) + \lambda R^u(Q^0,n,k,q^i,q^e) - (1+\lambda)\mathcal{NC}(2N,Q^0,Q) - C^i(n,q^i) - C^e(k, q^e) = 2Nv(q^0) + \lambda 2Np^0 + nu(q^i) + ku(q^e) + \lambda[F(n,q^i)+F(k,q^e)] - (1+\lambda)[c^0 2Nq^0 - \mathcal{NC}(2N)] - \beta m(n)nq^i - m(k)kq^e$ subject to:

[MME] $p^0 = v'(q^0)$

[PC^i] $\Pi^i = nu(q^i) - \beta m(n)nq^i - F(n, q^i) \geq 0$

[PC^e] $\Pi^e = ku(q^e) - m(k)kq^e - F(k, q^e) \geq 0$

[IC^i] $\pi^i(n, T^i) - F(n, q^i) \geq \pi^i(k, T^e) - F(k, q^e)$

[IC^e] $\pi^e(k, T^e) - F(k, q^e) \geq \pi^e(n, T^i) - F(n, q^i)$

and $2N = k + n$

Taking advantage of the previous analysis, we can substitute the binding constraints [MME], [PC^e] $F(k, q^e) = ku(q^e) - m(k)kq^e$ and [IC^i] $F(n, q^i) = nu(q^i) - \beta m(n)nq^i - (1-\beta)m(k)kq^e$ into the Laffont and Tirole social welfare function W^{λ} reducing the whole problem to the maximisation of the following function:

$$W^{\lambda} = 2Nv(q^0) + \lambda 2Nv'(q^0)q^0 + (1+\lambda)[(2N-k)u(q^i)+ku(q^e)] - (1+\lambda)[2Nc^0q^0 - \mathcal{NC}(2N)] - (1+\lambda)C^i - [1+\lambda(2-\beta)]C^e$$

In this case, we should have some distortion for the downstream entrant (the less efficient producer) but no distortion for the downstream incumbent (the more efficient producer), analogously to the unregulated case.

[LT.1]^s $(p^0 - c^0)/p^0 = -\lambda/(1+\lambda)\, v''(q^0)\, q^0/p^0 = +\lambda/(1+\lambda)\eta^0$

[LT.2]^s $\pi^i_n = u(q^i) - C^i_n = u(q^e) - \{[1+\lambda(2-\beta)]/(1+\lambda)\}C^e_k = \pi^e_k - [\lambda(1-\beta)/(1+\lambda)]C^e_k$

[LT.3]^s $\pi^i_L = 0 \quad \Rightarrow \quad u'(q^i) = \beta m(n)$

[LT.4]^s $\pi^e_L - [\lambda(1-\beta)/(1+\lambda)]C^e_L = 0 \quad \Rightarrow \quad u'(q^e) = \{[1+\lambda(2-\beta)]/(1+\lambda)\}m(k)$

However, the unregulated regime can be derived when the shadow cost of public funds λ goes to infinity. In fact, as λ goes to infinity, equation **[LT.2]^s** is exactly equal to **[4.9]**, while equations **[LT.3]^s** and **[LT.4]^s** combined are the analogues of **[4.10]** or

[4.11].

Hence, in this case, Pareto improvements would be also achieved by reducing the value of the downstream entrant's marginal access charges (F^e_k, F^e_L) till they get to the socially optimal values {i.e. $[\lambda(1-\beta)/(1+\lambda)]C^e_k$ and $[\lambda(1-\beta)/(1+\lambda)]C^e_L$}.

In fact, in this way, we just reach the second best tariffs $F^\lambda(n, q^i)$ and $F^\lambda(k, q^e)$ corresponding to the solution of the public upstream monopolist's problem **[U.LT]**. The participation condition [PC^e] is always binding, since it is always socially optimal that the upstream monopolist expropriates all the downstream entrant's profits, as profits Π^u have an additional weight λ, due to the excess burden of the revenues coming from taxes. Furthermore, there appears a trade-off between the expropriation of the downstream incumbent's rent and efficiency. Consequently, under full information, using the general access charge tariff $F^\lambda(N^j, q^j)$, the upstream incumbent is able to maximise total welfare W^λ causing a twofold distortion.

[LT.5]^e $\quad u'(q^e) - u'(q^i) = [m(k^\lambda) - \beta\, m(n^\lambda)] + [\lambda(1-\beta)/(1+\lambda)]\, m(k^\lambda)$

Not only are different marginal costs (C^e_k - C^i_n < 0) imposed, through a distortionary allocation of the number of customers ($k^\lambda < k^o$), but also the choice of customers' bundles (q^i and q^e) is distorted by charging a positive marginal access charges {$F^e_L = [\lambda(1-\beta)/(1+\lambda)]\, m(k^\lambda) > 0$} to the downstream entrant. Stated in other words, since the upstream monopolist's profits are now given an additional weight λ, the inefficient type enjoys no surplus and his choices are distorted by a bidimensional wedge in order to reduce the net positive rent the efficient type could gain, taking the inefficient type's contract.

To conclude, let us briefly follow the approach which puts more weight on consumers' surplus $U(Q^0, Q^i, Q^e) = S(Q^0, Q^i, Q^e) - R^i(Q^i) - R^e(Q^e)\, R^u(Q^0, Q^i, Q^e)$ and on the upstream monopolist's profits Π^u (which are redistributed to consumers as public revenues), rather than on the downstream producer's surplus $\phi\Pi^j$. Let then W^ϕ represent, as usual, the Baron and Myerson social welfare function, which is maximised by the upstream monopolist:

[U.BM] $\max W^{\phi} = S(Q^0, Q^i, Q^e) - \mathcal{NC}(2N, Q^0, Q) - (1-\phi)[R^i(Q^i)+R^e(Q^e)]$

$+(1-\phi)[F(n,q^i)+F(k,q^e)] - \phi[C^i(n, q^i)+C^e(k, q^e)] =$

$2Nv(q^0) + \phi nu(q^i) + \phi ku(q^e) + (1-\phi)[F(n,q^i)+F(k,q^e)] - 2Nc^0q^0 - \mathcal{NC}(2N)$

$- \phi[\beta m(n)nq^i + m(k)kq^e]$ subject to:

[PC^i] $\Pi^i = nu(q^i) - \beta m(n)nq^i - F(n, q^i) \geq 0$

[PC^e] $\Pi^e = ku(q^e) - m(k)kq^e - F(k, q^e) \geq 0$

[IC^i] $\pi^i(n, T^i) - F(n, q^i) \geq \pi^i(k, T^e) - F(k, q^e)$

[IC^e] $\pi^e(k, T^e) - F(k, q^e) \geq \pi^e(n, T^i) - F(n, q^i)$

and $2N = k + n$

To solve the problem, the binding constraints [PC^e] $F(k, q^e) = ku(q^e) - m(k)kq^e$ and [IC^i] $F(n, q^i) = nu(q^i) - \beta m(n)nq^i - (1-\beta)m(k)kq^e$ (since the others are satisfied by the solution) can be substituted into W^{ϕ}, obtaining:

$$W^{\phi} = 2Nv(q^0) + (2N-k)u(q^i) + ku(q^e) - 2Nc^0q^0 - \mathcal{NC}(2N) - C^i - [1+(1-\beta)(1-\phi)]C^e$$

In this case, we should have no distortion for the monopolised good 0, and for the downstream incumbent, but in order to reduce the latter rent some distortion may emerge for the less efficient downstream producer, as in the Laffont and Tirole case.

[BM.1]^{s} $c^0 = v'(q^0) = p^0$

[BM.2]^{s} $\pi^i_n = u(q^i) - C^i_n = u(q^e) - [1+(1-\beta)(1-\phi)]C^e_k = \pi^e_k - (1-\beta)(1-\phi)C^e_k$

[BM.3]^{s} $\pi^i_L = nu'(q^i) - C^i_L = 0; \quad \Rightarrow \quad u'(q^i) = \beta m(n)$

[BM.4]^{s} $\pi^e_L - (1-\beta)(1-\phi)C^e_L = 0 \quad \Rightarrow \quad u'(q^e) = [1+(1-\beta)(1-\phi)]m(k)$

Analogously to Laffont and Tirole's case, the access charge tariffs are $F(k^{\phi}, {}^{\phi}q^e) = k^{\phi}T^e - m(k^{\phi})Q^e$ and $F(n^{\phi}, q^i) = n^{\phi}T^i - \beta m(n^{\phi})Q^i - (1-\beta)m(k^{\phi})Q^e$. The participation condition [PC^e] is always binding and downstream entrant has no rent, since profits Π^u have a greater weight; moreover the constraint [IC^i] implies a trade-off between rent extraction (of the downstream incumbent) and efficiency.

While **[BM.1]^{s}** coincides with **[LM.1]^{s}** and Pareto efficiency is achieved in the monopolised market (where marginal prices equal marginal costs, like in the ultra-liberal case), the scale of entry and customers' bundles are again distorted, as $k^{\phi} < k^{\circ}$ and ${}^{\phi}q^e < q^{\circ}$.

[BM.5]* $u'(q^e) - u'(q^i) = [m(k^\phi) - \beta\, m(n^\phi)] + (1-\beta)(1-\phi)m(k^\phi)$

As in Laffont and Tirole's case, the maximisation of W^ϕ causes a twofold distortion: through the allocation of the number of customers ($k^\phi < k^\circ$) as ${}^\phi F^e_k = (1-\beta)(1-\phi)C^e_k > 0$, and the customers' bundle ${}^\phi q^e$, as ${}^\phi F^e_L = (1-\beta)(1-\phi)\, m(k^\lambda) > 0$. In particular, the number of customers of the downstream incumbent is relatively high, while the bundle offered to customers by the downstream entrant is relatively low, with respect to the optimal one with the Loeb and Magat social welfare function. However, the two regimes will clearly coincide as ϕ tends to one.

Access charges tend towards the ones holding under the private monopoly solution as the weight of profits ϕ approaches zero, since the tariffs F^ϕ become closer and closer to the monopoly ones $F(N^j, q^j)$. In fact, in the limiting case, equation **[BM.2]*** is exactly equal to **[4.9]**, while the combination of **[BM.3]*** and **[BM.4]*** is the analogue of **[4.10]** or **[4.11]**.

Hence, also in the Baron and Myerson case, a reduction of the downstream entrant's monopoly marginal access charges (F^e_k, F^e_L) would enhance social welfare, till we reach the second best access charges ${}^\phi F^e_k$ and ${}^\phi F^e_L$.

4.4 Vertical separation with two consumers' types in the standard framework

In order to deal with two consumers' types in the previous setting of vertical separation, we must modify some of the specific hypotheses contained in **(iv*)** and **(v*)** considering the more general assumptions specified below.

In particular, now in **(iv*)** the downstream producers (the incumbent and the entrant) can freely choose the number of customers to be served in each market (the high and low demand customers' markets). In each market we will set:

$$N^j_t = n_t \quad \text{for } j=i$$
$$ = k_t \quad \text{for } j=e$$

with $N^i = \Sigma_t N^i_t = n = n_L + n_H \leq 2N$ and $N^e = \Sigma_t N^i_t = k = k_L + k_H \leq 2N$, requiring:

$$n_L + k_L \leq N$$
$$n_H + k_H \leq N$$

Moreover, the upstream incumbent has a new specification of the access charge tariffs $F(N^j_H, N^j_L, q^j_H, q^j_L)$ which are a function of the number of customers served N^j_t (with j = i, e and t = H, L) in each market, and of the customer bundles ($q^j_t = q^i_t, q^e_t$) to be offered by each downstream producer to the different customers' types:

[4.13] $$F = F(N^j_H, N^j_L, q^j_H, q^j_L)$$

In this way, the upstream incumbent is allowed to discriminate, through the non-linear access charge tariff, between the two downstream producers choosing directly the customers' bundles (q^i_t and q^e_t) and consequently the pricing strategies of the downstream producers, who can either accept or reject the offer.

Downstream producers have the following *revenue* and *cost functions*:

$$R^j = N^j_L T^j_L + N^j_H T^j_H$$

$$C^j(N^j_H, N^j_L, q^j_H, q^j_L) = C(N^j, Q^j, \beta^j)$$

In particular, we set as before $\beta^e = 1$ and $\beta^i = \beta$ (with $0 < \beta < 1$) and assume that costs $C(N^j, Q^j, \beta^j)$ are strictly increasing with respect to the total number of customers served ($N^j = N^j_H + N^j_L$) and total output ($Q^j = N^j_H q^j_H + N^j_L q^j_L$), convex in their first argument and linear in the second (with positive cross derivatives). Consequently the above functional restrictions require:

$$C^j_{Nt} = \partial C^j/\partial N^j_t > 0;\ C^j_{NN} \geq 0;\ \ C^j_t = \partial C^j/\partial q^j_t > 0;\ C^j_{tt} = 0 \ \text{ and }\ \partial^2 C^j/\partial N^j_t \partial q^j_t \geq 0$$

with: $$C^j_H/N^j_H = C^j_L/N^j_L = \partial C(N^j,Q^j,\beta^j)/\partial Q^j \text{ and } C^j_{NH} - C^j_{NL} = (q^j_H - q^j_L)\, \partial C(N^j,Q^j,\beta^j)/\partial Q^j$$

Without great loss of generality, in most of the analysis we will consider the following simple functional form: $C^j = \beta^j\, m(N^j)\, Q^j$.

We may write downstream producers' *profit functions* as:

$$\Pi^j = R^j - C^j - F(N^j_H, N^j_L, q^j_H, q^j_L) \qquad \text{with} \qquad \pi^j(N^j_H, N^j_L, q^j_H, q^j_L) = R^j - C^j$$

The two downstream producers enter the market for good 1 only if their participation constraints are satisfied.

[PC^e] $$\Pi^e = k_L T^e_L + k_H T^e_H - C^e(k_H, k_L, q^e_H, q^e_L) - F(k_L, k_H, q^e_H, q^e_L) \geq 0$$

[PC^i] $$\Pi^i = n_L T^i_L + n_H T^i_H - C^i(n_H, n_L, q^i_H, q^i_L) - F(n_L, n_H, q^i_H, q^i_L) \geq 0$$

Furthermore, with second degree price discrimination, the downstream producers' incentive compatibility constraints should be satisfied to get true type revelation:

[IC^e] $\pi^e(k_L, k_H, q^e_L, q^e_H)$-$F(k_L, k_H, q^e_H, q^e_L) \geq \pi^e(n_L, n_H, q^i_L, q^i_H) - F(n_L, n_H, q^i_H, q^i_L)$

[IC^i] $\pi^i(n_L, n_H, q^i_L, q^i_H) - F(n_L, n_H, q^i_H, q^i_L) \geq \pi^i(k_L, k_H, q^e_L, q^e_H) - F(k_L, k_H, q^e_H, q^e_L)$

Moreover, in presence of two types, in **(v*)** the upstream monopolist is characterised by the following *profit* $\Pi^u(q^0, n_L, n_H, k_L, k_H, q^i_H, q^i_L, q^e_H, q^e_L)$ *function*:

$$\Pi^u = 2N(p^0 - c^0)q^0 + F(n_L, n_H, q^i_H, q^i_L) + F(k_L, k_H, q^e_H, q^e_L) - NC(2N)$$

Naturally, apart from these minor changes specified above, in what follows we will maintain the same assumptions and the basic rules of the vertical separation game. Moreover, as before, the order of moves between downstream producers will not affect the game's solution. In fact, the upstream monopolist directly chooses the customer bundles ($q^j_t = q^i_t, q^e_t$) and consequently the tariffs that the downstream producers offer to the different customers' types.

4.4.1 First degree access charge discrimination with two types of customers

Let us now analyse this more general setting starting from the benchmark case given by first degree access charge discrimination. This will show the ways in which the upstream incumbent can create distortions between the downstream producers, using the access charge tariff to maximise his total profit $\Pi^u(q^0, n_L, n_H, k_L, k_H, q^i_H, q^i_L, q^e_H, q^e_L)$. Under perfect discrimination both the efficient type "i" (the downstream incumbent) and the inefficient type "e" (the downstream entrant) enjoy no net surplus as the upstream monopolist faces only the two participation constraints [PC^e] and [PC^i] (using lump-sum access charges F^i and F^e) and does not need to take into account their incentive compatibility constraints [IC^e] and [IC^i].

[U*] $\max \Pi^u(q^0, n_L, n_H, k_L, k_H, q^i_H, q^i_L, q^e_H, q^e_L)^* = 2N(p^0\text{-}c^0)q^0+F^i+F^e - NC(2N)$

subject to:

[MME] $p^0 = v'(q^0)$

[PC^e] $\pi^e(k_L, k_H, q^e_L, q^e_H) = k_L T^e_L + k_H T^e_H - C^e(k_H, k_L, q^e_H, q^e_L) = F^e$

[PC^i] $\pi^i(n_L, n_H, q^i_L, q^i_H) = n_L T^i_L + n_H T^i_H - C^i(n_H, n_L, q^i_H, q^i_L) = F^i$

[N^j_L] $n_L + k_L = N$

[N^j_H] $n_H + k_H = N$

and $n_L, k_L, n_H, k_H \geq 0$

In order to solve his problem the upstream incumbent must consider the tariffs the downstream producers will set after accepting the contract offered. In fact, the contract involves the determination of the variables(N^j_H, N^j_L, q^j_H, q^j_L) and the corresponding access charges $F^j(N^j_H, N^j_L, q^j_H, q^j_L)$ specifically designed for each downstream producer. Hence, in what follows we will define the downstream producers' tariffs for the relevant values of q^e_H, q^e_L, q^i_H and q^i_L.

Considering the downstream entrant's pricing problem, it is easy to verify how given k_L, k_H, q^e_L and q^e_H the customers tariffs T^e_L and T^e_H, which satisfy the rationality and incentive compatibility constraints of the customers, are the highest ones.

In fact it is always optimal, that the "e" type producer always extracts low type consumers' surplus, setting:

[IR^e_L] $T^e_L = u(q^e_L)$

If the downstream entrant enters the H type's market ($k_H > 0$) her tariff T^e_H must satisfy the following *internal* incentive compatibility constraint

$$\theta\, u(q^e_H) - T^e_H \geq \theta\, u(q^e_L) - u(q^e_L)$$

She sets a tariff $T^e_H = \theta\, u(q^e_H) - (\theta - 1)\, u(q^e_L)$ for the H type when $q^e_L \geq q^i_L$ so that there are no external interference from the tariff T^i_L and T^i_H of the downstream incumbent and high demand customers have no incentive to buy from the downstream incumbent.

More generally she must also consider an *additional external* incentive compatibility constraint in maximising the tariff T^e_L.

$$\theta\, u(q^e_H) - T^e_H \geq \theta\, u(q^i_L) - T^i_L$$

The solution of the problem depends on which of the two incentive compatibility constraints is binding. We do not need to consider the high demand tariff T^i_H, since the downstream incumbent's too maximises the customers tariffs T^i_L and T^i_H (given n_H, n_L, q^i_H and q^i_L) Hence, the solution of the problem reduces to:

$[IC^e_H]$ $$T^e_H = \theta\ u(q^e_H) - (\theta - 1) \max [u(q^i_L), u(q^e_L)]$$

In practice no surplus is left to the L consumers, while given the customers' bundles (q^i_L and q^e_L) the rent of high-demand customers is minimised, since it is equal to $(\theta - 1) \max [u(q^i_L), u(q^e_L)]$ the maximum rent a consumer of type H can get mimicking the low type.

Let us now examine the downstream incumbent's tariffs problem, when he will serve both the H and L customers. Like the entrant he should extract all the low type consumers' surplus, setting:

$[IR^i_L]$ $$T^i_L = u(q^i_L)$$

If both downstream producers enter the H type's market (k_H, $n_H > 0$) his tariff T^i_H must satisfy at the same time the *internal* and *external* incentive compatibility constraints specified below:

$$\theta\ u(q^i_H) - T^i_H \geq \theta\ u(q^i_L) - T^i_L$$

$$\theta\ u(q^i_H) - T^i_H \geq \theta\ u(q^e_L) - T^e_L$$

Also this time the producer just wants to minimise the rent left to the H customers who are able to mimic the low type. This is achieved applying the same procedure as before by equating the high demand customers' rent to the amount $(\theta - 1) \max [u(q^i_L), u(q^e_L)]$.

$[IC^i_H]$ $$T^i_L = \theta\ u(q^i_H) - (\theta - 1) \max [u(q^i_L), u(q^e_L)]$$

Knowing the downstream producers' tariffs T^i_H, T^i_L, T^e_H and T^e_L in correspondence of q^e_H, q^e_L, q^i_H and q^i_L we can reformulate the upstream monopolist's profit maximisation problem as:

[U.1*] $\max \Pi^u = 2N[v'(q^0)-c^0]q^0+k_L u(q^e_L)+k_H\{\theta u(q^e_H)-(\theta-1)\max[u(q^i_L), u(q^e_L)]\}-$

$-C^e+(N-k_L)u(q^i_L) + (N-k_H)\{\theta u(q^i_H)-(\theta - 1)\max[u(q^i_L), u(q^e_L)]\}-C^i-NC(2N)$

with $k_L, k_H \geq 0$

In this case, as usual, we assume that the customers served by the incumbent in both markets (i.e. n_L and n_H), are non-negative numbers in the optimal contract. First of all, assuming the downstream incumbent to be sufficiently more efficient than the downstream entrant (i.e. β significantly smaller than one), we can argue that the downstream entrant (i.e. the less efficient producer) cannot pre-empt either of the two markets, since her average and marginal unit costs [m(k) and m'(k)] are both greater than the incumbent's ones [βm(k) and βm'(k)] when she serves N or more customers. In fact, in order to satisfy the upstream monopolist's profit maximisation problem one market is assigned to a downstream producer only if in that market his gross profits on the marginal customer are higher than the other downstream player's one, which in turn is possible only when he has the lowest costs associated with serving the marginal customer (in that market).

On the other hand, the upstream monopolist can find it optimal to oblige the downstream entrant to serve just one type of customers; while the downstream incumbent must always serve both types. Specifically, we will show how the problem of market pre-emption emerges just when the upstream monopolist obliges her the to serve only low demand customers. Hence, this time a particular attention should be paid to the possibility that the downstream entrant (the less efficient producer) specialises in serving just one of the two markets (i.e. to the possibility of either k_L or k_H being equal to zero). Moreover, as seen before from the downstream producers' tariffs problem we do not know in advance whether q^i_L is greater, equal or smaller than q^e_L.

Taking explicitly into account these problems the first order conditions can be specified as follows:

[q^0] $c^0 = v'(q^0) + v''(q^0)q^0$

[$^*k_H$] $\pi^e_{kH} - \pi^i_{nH} = 0$ for $k_H > 0$ $\Rightarrow$ $\theta u(q^i_H) - C^i_{nH} = \theta u(q^e_H) - C^e_{kH}$

	or: $\pi^e_{kH} - \pi^i_{nH} < 0$	for $k_H = 0$	$\Rightarrow$	$\theta u(q^i_H) - C^i_{nH} > \theta u(q^e_H) - C^e_{kH}$
[*k_L]	$\pi^e_{kL} - \pi^i_{nL} = 0$	for $k_L > 0$	$\Rightarrow$	$u(q^i_L) - C^i_{nL} = u(q^e_L) - C^e_{kL}$
	or: $\pi^e_{kL} - \pi^i_{nL} < 0$	for $k_L = 0$	$\Rightarrow$	$u(q^i_L) - C^i_{nL} > u(q^e_L) - C^e_{kL}$
[*q^i_H]	$R^i_H = C^i_H$		$\Rightarrow$	$\theta u'(q^i_H) = C^i_H/n_H = \beta m(n)$
[*q^e_H]	$R^e_H = C^e_H$		$\Rightarrow$	$\theta u'(q^e_H) = C^e_H/k_H = m(k)$
[*q^i_L]	$R^i_L = C^i_L - N\xi(\theta-1)u'(q^i_L)$		$\Rightarrow$	$u'(q^i_L) = \beta m(n) - N\xi(\theta-1)u'(q^i_L)/n_L$
[*q^e_L]	$R^e_L = C^e_L - N(1-\xi)(\theta-1)u'(q^e_L)$		$\Rightarrow$	$u'(q^e_L) = m(k) - N(1-\xi)(\theta-1)u'(q^e_L)/k_L$

where the parameter $0 \le \xi \le 1$ is equal to zero when $q^i_L < q^e_L$ and equal to one for $q^i_L > q^e_L$, and determined endogenously for $q^i_L = q^e_L$.

Let us explore first the case in which the downstream entrant serves both types (k_L, $k_H > 0$). In this case both equations [*k_H] and [*k_L] should hold as an equality and we must have

[*k] $$\pi^e_{kH} - \pi^i_{nH} = \pi^e_{kL} - \pi^i_{nL} = 0$$

This equality in turn implies that the additional gains from serving a high customer is equal for both downstream producers:

$$\theta u(q^e_H) - q^e_H C^e_H/k_H - [u(q^e_L) - q^e_L C^e_L/k_L] = \theta u(q^i_H) - q^i_H C^i_H/n_H - [u(q^i_L) - q^i_L C^i_L/n_L]$$

This equality is satisfied when the customers are distributed in a way such that average costs are equated $C^e_H/k_H = C^e_L/k_L = C^i_H/n_H = C^i_L/n_L$ [i.e. $\beta m(n) = m(k)$] and the customers' bundles offered by the downstream producers are equated (i.e. $q^i_L = q^e_L$ and $q^i_H = q^e_H$). Naturally, condition [*k_L] [i.e. $u(q^i_L) - C^i_{nL} = u(q^e_L) - C^e_{kL}$] must also hold for k_L, $k_H > 0$. This implies the equality between the two marginal costs C^i_{nL} and C^e_{kL}. Notice how, as we have already argued, the downstream entrant cannot serve a number of customers greater or equal to N. In fact, in these cases, her average cost would be greater than the incumbent's one, giving rise to a contradiction: $m(k) \ge m(N) > \beta m(N) \ge \beta m(n)$.

Isoelastic cost functions (for which $\varepsilon(n) = n\, m'(n)/m(n)$ equals $\varepsilon(k)$, for any n and k) meet all these requirements. In this case from the equality of unit costs [$\beta m(n) = m(k)$] we have also [$n\, \beta m'(n) = k\, m'(k)$] and consequently the equality between C^i_{nL}

$= \beta[m'(n)\, n\, (Q^i/n) + m(n)q^i_L]$ and $C^e_{kL} = m'(k)\, k\, (Q^e/k) + m(k)q^e_L$ when downstream producers serve the same number of the two types. In fact, (since $q^i_L = q^e_L$ and $q^i_H = q^e_H$) per capita output is equated ($Q^i/n = Q^e/k$) when $k_L = k_H = k/2$. It easy to show how in this case the marginal prices of the high and low demand customers set by the downstream producers are equal (i.e. $\theta u'(q^i_H) = \theta u'(q^e_H)$ and $u'(q^i_L) = u'(q^e_L)$). In fact, from the equality of unit costs [$\beta m(n) = m(k)$] and equations [$*q^i_H$] and [$*q^e_H$] the equality of the high demand customers' marginal prices $\theta u'(q^i_H) = \theta u'(q^e_H)$ follows. We can also get the optimal value of ξ (here given by n_H/N) directly from equations [$*q^i_L$] and [$*q^e_L$].

In practice, when the upstream monopolist perfectly discriminates between downstream producers, in the final demand market there is **no distortion at the top** if the unit cost m(k) functions are isoelastic, since the marginal access tariffs for the high demand customers F^e_H and F^i_H are set equal to zero. In fact, marginal prices $\theta u'(q^i_H)$ equate the per customer marginal costs $\partial C(\ ,\beta^j)/\partial Q^j$, i.e. $\theta u'(q^i_H) = \beta m(n) = m(k) = \theta\, u'(q^e_H)$. On the other hand, we get **distortion at the bottom,** $u'(q^i_L) = u'(q^e_L) < \beta m(n) = m(k)$, even if in this particular case the marginal access tariffs (for the low demand customers' bundle F^e_L and F^i_L) are equal to zero.

Let us briefly consider the other cases, in which the downstream entrant serve just one type of consumers, starting from the case in which the downstream entrant serves only the low-demand type ($k_H = 0$). In this case only equations [$*k_L$] should hold as an equality and we must have:

[*k'] $$\pi^e_{kH} - \pi^i_{nH} < \pi^e_{kL} - \pi^i_{nL} = 0$$

This inequality implies that the additional gains from serving a high customer are smaller for the downstream entrant. After some simplifications we get:

$$\theta u(q^e_H) - q^e_H\, C^e_H/k_H - [u(q^e_L) - q^e_L\, C^e_L/k_L] < \theta u(q^i_H) - q^i_H\, C^i_H/n_H - [u(q^i_L) - q^i_L\, C^i_L/n_L]$$

The previous inequality is satisfied for the relevant customers distributions (between downstream producers) when the downstream entrant's unit marginal costs are higher $C^e_H/k_H = C^e_L/k_L > C^i_H/n_H = C^i_L/n_L$ [i.e. $m(k) > \beta m(n)$], independently from the fact that the customers and the bundles offered by the downstream producers are equated (i.e. q^i_L and q^i_H are not necessarily equal to q^e_L and q^e_H). Naturally condition [$*k_L$] [i.e. $u(q^i_L) - C^i_{nL} = u(q^e_L) - C^e_{kL}$] must hold also for $k_H = 0$.

We can eventually also get condition [$*k_L$] holding as a strict inequality [i.e. $u(q^i_L) - C^i_{nL} < u(q^e_L) - C^e_{kL}$] in the case of low demand customers' market pre-emption by the downstream entrant , i.e. for $n_L = 0$. However, this case would not occur, provided that β is low enough. In fact, assuming $u(q^i_L) \geq u(q^e_L)$ as we will show later, the inequality $u(q^i_L) - C^i_{nL} < u(q^e_L) - C^e_{kL}$, for $n = k = N$, implies

$$\beta[\varepsilon(N)\, q^i_H + q^i_L] > \varepsilon(N)\, q^e_L + q^e_L$$

a condition which can be never satisfied when β (> 0) is low enough.

In this case condition [$*q^i_H$] is of no interest as the downstream entrant is specialised in serving only customers of type L. Analysing the low demand customers' bundles it turns out (as claimed before) that the marginal price offered by the downstream incumbent [which from [$*q^i_L$] is greater or equal to marginal costs $u'(q^i_L) \geq \beta m(n)$] should not be greater than the one offered by the downstream entrant, i.e. $u'(q^i_L) \leq u'(q^e_L)$.

In fact, otherwise (i.e. if $u'(q^i_L) > u'(q^e_L)$ were implied by the optimal solution) we incur in a contradiction. Since the downstream entrant's unit costs are greater [$\beta m(n) < m(k)$], the upstream monopolist can make positive gains increasing downstream incumbent's low demand customers' marginal price (e.g. setting $q^i_L = q^e_L$) while leaving $\xi = 0$.

This happens because, for $q^i_L \leq q^e_L$, the following statements hold:

(a) the high demand customers' rent $(\theta-1)u'(q^e_L)$ is determined by the level of q^e_L and does not decrease till q^i_L is smaller or equal to q^e_L

(b) the low demand customers' marginal price offered by the downstream entrant (from [$*q^e_L$]) is smaller or equal to his marginal costs (i.e. $u'(q^e_L) \geq m(k)$) and is clearly greater than the downstream incumbent's marginal cost [i.e. $u'(q^e_L) > \beta m(n)$].

On the other hand, it is instead possible that the downstream entrant's marginal cost $m(k)$ is greater than the low demand customers' marginal price offered by the downstream incumbent [i.e. $m(k) > u'(q^i_L)$] even for $\xi = 1$. Taking into account [$*q^e_L$], this implies that the downstream entrant must apply marginal cost pricing so that the inequality $u'(q^i_L) < u'(q^e_L)$ holds.

In practice, when the upstream monopolist obliges the downstream entrant to serve only low-demand customers, there is, as usual, **no distortion at the top** (since the marginal access tariffs for high demand customers F^e_H and F^i_H are set equal to zero) and **distortion at the bottom**, so that F^e_L may be positive for the downstream entrant. In fact, when $q^i_L = q^e_L$ and $u'(q^i_L) > m(k)$, from the first order condition [$*q^e_L$] $F^e_L = (N-k_L)(1-\xi)(\theta-1)u'(q^e_L)/k_L$ is greater than zero as ξ must be less than unity.

Let us turn now to the case in which the downstream entrant serves only the high type ($k_L = 0$). In this case only equations [$*k_H$] should hold as an equality and we must have:

[*k"] $$\pi^e_{kL} - \pi^i_{nL} < \pi^e_{kH} - \pi^i_{nH} = 0$$

The previous inequality implies that the additional gains from serving an high-demand customer are greater for the downstream entrant.

$$\theta u(q^e_H) - q^e_H C^e_H/k_H - [u(q^e_L) - q^e_L C^e_L/k_L] > \theta u(q^i_H) - q^i_H C^i_H/n_H - [u(q^i_L) - q^i_L C^i_L/n_L]$$

This inequality is satisfied when the downstream entrant's unit marginal costs are lower $C^e_H/k_H = C^e_L/k_L < C^i_H/n_H = C^i_L/n_L$ [i.e. $m(k) < \beta m(n)$], independently of the customers' bundles (q^i_L, q^e_L, q^i_H and q^e_H). Naturally condition [$*k_H$] [i.e. $u(q^i_H) - C^i_{nH} = u(q^e_H) - C^e_{kH}$] must hold for $k_L = 0$.

In particular, in this case it is easy to show that the downstream entrant can never serve a number of customers equal to N, pre-empting the high customer market. In fact, in these cases, her average cost would be greater than the incumbent's one, giving rise to a contradiction: $m(N) > \beta m(N)$. Consequently, the equality between $u(q^i_H) - C^i_{nH}$ and $u(q^e_H) - C^e_{kH}$ is met only when the downstream entrant serves a smaller number of customers.

As usual, we have **no distortion at the top** (the marginal access tariffs for the high demand customers F^e_H and F^i_H being equal to zero). In particular, high demand customers' marginal price offered by the downstream incumbent [which from [*q^i_H] is equal to marginal costs $u'(q^i_H) = \beta m(n)$] is greater than the one offered by the downstream entrant, i.e. $\theta u'(q^i_H) > \theta u'(q^e_H) = m(k)$.

On the other hand, if the downstream entrant will not serve low demand customers' and consequently $\xi = 1$, from the first order condition [*q^i_L] we see how the marginal price offered by the downstream incumbent is increased by $F^i_L = (N-n_H)(1-\xi)(\theta-1)u'(q^i_L)/n_L$, an amount which is greater than zero

Having examined the only three admissible regimes ($k_L, k_H > 0$; $k_L > k_H = 0$; $k_H > k_L = 0$) let us re-consider more in depth the marginal access tariffs for the high and low demand customers in first degree price discrimination setting. First of all, we should remember how the first order conditions [*q^e_H] and [*q^e_L] hold if and only if the downstream entrant is not specialised in serving either low-demand customers ($k_H = 0$) or high-demand customers ($k_L = 0$).

From the first order conditions [*q^i_H] and [*q^e_H] relative to the per customer bundle offered to high-demand consumers, the two marginal access tariffs F^e_H and F^i_H are always equal to zero. In practice, the relationship between the high demand consumer's marginal prices $\theta u'(q^i_H)$ served by different downstream producers is not distorted, but it is equated to the per customer marginal costs $C^j_H/n_H = \partial C(N^j, Q^j, \beta^j)/\partial Q^j$. Hence, equations [*$q^i_H$] and [*$q^e_H$] simply state that there is **no pricing distortion at the top** when the upstream monopolist is able to perfectly discriminate between the two downstream producers.

On the other hand, from the first order conditions [*q^i_L] and [*q^e_L], we see that we have **some pricing distortion at the bottom**, even in the particular case in which the two marginal access tariffs (for the low demand customers' bundle) F^e_L and F^i_L are equal to zero. In any case, no surplus is left to the L consumers; but their marginal prices $u'(q^i_L)$ and $u'(q^e_L)$ must be distorted at least by one downstream producers, in order to reduce the net surplus enjoyed by the customers of type H. Specifically, the value of the parameter $0 \le \xi \le 1$ in equations [*q^i_L] and [*q^e_L], shows how the distortion at the bottom for each downstream producer depends on which of the two constraints $[IC^i_H]$ or $[IC^e_H]$ is binding.

The most interesting case occurs in the non specialisation regime (k_L, $k_H > 0$), when the low demand customers' bundles -which equate the marginal rent of the high demand customers $N(\theta-1)u'(q^i_L)$- is split between the downstream incumbent and entrant respectively in quotas ξ and $(1-\xi)$ which do not need to be equal to the percentage of high customers served. F^e_L and F^i_L will be equal to zero only if the quotas ξ and $(1-\xi)$ are equal to the percentage of high customers served, as in the case of isoelastic curves. When this is not the case, one downstream producer incurs in a positive marginal access tariff, while the other faces a negative one.

Also when the downstream entrant is specialised in low demand customers, F^e_L and F^i_L can be equal to zero provided that q^i_L is greater than q^e_L the bundle offered by the downstream entrant. In this situation there is no point in departing from the marginal pricing rule $p^e_L = u'(q^e_L)$. In fact, if it were $p^e_L = C^e_H/k_H > p^i_L$ high-demand customers would be worse off taking the entrant's contract and the upstream monopolist would have no incentive to rise the marginal price $u'(q^e_L)$ of the low-demand consumers above that level. When this is not the case, the downstream entrant faces a positive marginal access tariff, while the downstream incumbent faces a negative one.

Finally, when the downstream entrant is specialised in high demand customers, F^i_L is always greater than zero, so that the downstream incumbent faces always a positive marginal access tariff.

4.4.2 Second degree access charge discrimination with two types of customers

We now examine, keeping all the initial assumptions on the players, the more general setting of second degree access charge discrimination in which lump sum access charges are no longer available. Hence, differently from before, the access charges $F(N^i_H, N^i_L, q^i_H, q^i_L)$ are no longer specifically designed for each downstream producer and their incentive compatibility constraints should also be satisfied to get true type revelation.

At the same time, even if we depart from the perfect discrimination case, the previous analysis of section 4.4.1 of the downstream producers' pricing problem, given the values N^i_H, N^i_L, q^i_H and q^i_L, still holds. In fact as before, the customers tariffs T^i_L and T^i_H, must be the highest ones satisfying the rationality and incentive compatibility constraints of the customers.

In practice, in order to maximise revenues, given the number of customers N^i_H, N^i_L served by each downstream producer and the related bundles q^i_H and q^i_L, no surplus is left to the L consumers, while the rent of the H customers (dependent on the low customers' bundles q^i_L and q^e_L) is minimised, being equated to $(\theta - 1) \max [u(q^i_L), u(q^e_L)]$ the maximum rent he can get mimicking the most profitable low type's contract. Specifically, we can still use equations $[IR^e_L]$, $[IC^e_H]$, $[IR^i_L]$ and $[IC^i_H]$ to determine the downstream producers' tariffs T^i_H, T^i_L, T^e_H and T^e_L in correspondence of the customers' bundles q^e_H, q^e_L, q^i_H and q^i_L.

Consequently, we can now analyse the ways in which the upstream incumbent creates additional distortions between the downstream producers, using the access charge tariff to maximise his total profit $\Pi^u = \Pi^u(q^0, n_L, n_H, k_L, k_H, q^i_H, q^i_L, q^e_H, q^e_L)$ under second degree access charge discrimination, taking into account their incentive compatibility constraints $[IC^e]$ and $[IC^i]$.

[U2.1] $\max \Pi^u = 2N(p^0-c^0)q^0 + F(n_L,n_H,q^i_H,q^i_L) + F(k_L,k_H,q^e_H,q^e_L) - NC(2N)$

subject to:

[MME] $p^0 = v'(q^0)$

$[PC^e]$ $\Pi^e = k_L T^e_L + k_H T^e_H - C^e(k_H, k_L, q^e_H, q^e_L) - F(k_L, k_H, q^e_H, q^e_L) \geq 0$

$[PC^i]$ $\Pi^i = n_L T^i_L + n_H T^i_H - C^i(n_H, n_L, q^i_H, q^i_L) - F(n_L, n_H, q^i_H, q^i_L) \geq 0$

$[IC^e]$ $\pi^e(k_L,k_H,q^e_H,q^e_L) - F(k_L,k_H,q^e_H,q^e_L) \geq \pi^e(n_L,n_H,q^i_H,q^i_L) - F(n_L,n_H,q^i_H,q^i_L)$

$[IC^i]$ $\pi^i(n_L,n_H,q^i_H,q^i_L) - F(n_L,n_H,q^i_H,q^i_L) \geq \pi^i(k_L,k_H,q^e_H,q^e_L) - F(k_L,k_H,q^e_H,q^e_L)$

$[N^j_L]$ $N - n_L - k_L \geq 0$

$[N^j_H]$ $N - n_H - k_H \geq 0$

and $n_L, k_L, n_H, k_H \geq 0$

From the previous analysis we can derive the downstream producers' tariffs T^i_H, T^i_L, T^e_H and T^e_L (in correspondence of q^e_H, q^e_L, q^i_H and q^i_L) and, as usual, we can verify that ex post the solution does automatically satisfy:

(i) non-negative value assumed by the constraints $[N^i_L]$, $[N^i_H]$ and by the incumbent's customers (i.e. n_L, $n_H \geq 0$),.

(ii) the incentive compatibility constraints $[IC^e]$ and $[PC^i]$.

In practice, it is profitable to serve all the customers in the two markets and, if the downstream incumbent is sufficiently efficient (i.e. β is significantly smaller than one), the upstream monopolist's profit maximisation implies that the downstream entrant (i.e. the less efficient producer) cannot pre-empt either of the two market. This is quite plausible, her average and marginal unit costs being sufficiently greater than the incumbent's ones (when she serves N customers). Moreover, since only the $[PC^e]$ and $[IC^i]$ are binding, the downstream incumbent may enjoy a positive rent while the downstream entrant enjoys no rent.

The upstream monopolist objective function becomes:

$$\Pi^u(q^0, k_L, k_H, q^e_H, q^e_L, q^i_H, q^i_L) = 2N[v'(q^0)-c^0]q^0 + 2\pi^e - \bar{\pi}^i + \pi^i - \lambda C(2N)$$

with:

$$\pi^e = \pi^e(k_L,k_H,q^e_H,q^e_L) = k_L u(q^e_L) + k_H\{\theta u(q^e_H)-(\theta-1)\max[u(q^i_L),u(q^e_L)]\} - C^e(k_H,k_L,q^e_H,q^e_L)$$

$$\bar{\pi}^i = \pi^i(k_L,k_H,q^e_H,q^e_L) = k_L u(q^e_L)+k_H\{\theta u(q^e_H)-(\theta-1)\max[u(q^i_L), u(q^e_L)]\} - C^i(k_H, k_L, q^e_H, q^e_L)$$

$$\pi^i = \pi^i(N-k_L,N-k_H,q^i_H,q^i_L) = (N-k_L)u(q^i_L) +(N-k_H)\{\theta u(q^i_H)-(\theta-1)\max[u(q^i_L),u(q^e_L)]\} - C^i(N-k_H,N-k_L,q^i_H,q^i_L)$$

Hereafter the superscript $^-$ denotes the values assumed by the downstream incumbent's functions when he mimics the downstream entrant. Therefore, the upstream monopolist profit maximisation problem can be reformulated as:

[U.2.2]
$$\max \Pi^u = 2N[v'(q^0)-c^0]q^0+k_L u(q^e_L)+k_H\{\theta u(q^e_H)-(\theta-1)\max[u(q^i_L), u(q^e_L)]\} -2C^e+\bar{C}^i+(N-k_L)u(q^i_L) +(N-k_H)\{\theta u(q^i_H) - (\theta-1)\max[u(q^i_L),u(q^e_L)]\} - C^i - \lambda C(2N)$$

subject to:

$$k_L, k_H \geq 0$$

Notwithstanding the previous assumption we can take explicitly into account the most relevant issues. This time it is more likely that the upstream monopolist obliges the downstream entrant to serve just one type of customers (and specifically the low-demand type, i.e. the possibility of k_H being equal to zero) while the downstream incumbent serves both types. The intuition behind this is that the downstream entrant is less efficient and in this way the rent enjoyed by the downstream incumbent is reduced. Hence, we will examine this case with a particular attention. As before, we cannot tell in advance whether q^i_L is greater, equal or smaller than q^e_L.

Using as before the superscript $^{-}$ to denote the values assumed by the downstream incumbent's functions when she mimics the downstream entrant, the first order conditions can be specified as follows:

$[q^0]$ $\quad c^0 = v'(q^0) + v''(q^0)q^0$

$[k_H]$ $\quad 2\pi^e_{kH} - \bar{\pi}^i_{nH} - \pi^i_{nH} = 0$ for $k_H>0$ $\quad \Rightarrow \quad \theta u(q^i_H) - C^i_{nH} = \theta u(q^e_H) - 2C^e_{kH} + \bar{C}^i_{nH}$

or: $\quad 2\pi^e_{kH} - \bar{\pi}^i_{nH} - \pi^i_{nH} < 0$ for $k_H=0$ $\quad \Rightarrow \quad \theta u(q^i_H) - C^i_{nH} > \theta u(q^e_H) - 2C^e_{kH} + \bar{C}^i_{nH}$

$[k_L]$ $\quad 2\pi^e_{kL} - \bar{\pi}^i_{nL} - \pi^i_{nL} = 0$ for $k_L>0$ $\quad \Rightarrow \quad u(q^i_L) - C^i_{nL} = u(q^e_L) - 2C^e_{kL} + \bar{C}^i_{nL}$

or: $\quad 2\pi^e_{kL} - \bar{\pi}^i_{nL} - \pi^i_{nL} < 0$ for $k_L=0$ $\quad \Rightarrow \quad u(q^i_L) - C^i_{nL} > u(q^e_L) - 2C^e_{kL} + \bar{C}^i_{nL}$

$[q^i_H]$ $\quad R^i_H = C^i_H$ $\quad \Rightarrow \quad \theta u'(q^i_H) = C^i_H/n_H = \beta m(n)$

$[q^e_H]$ $\quad R^e_H = 2C^e_H - \bar{C}^i_H$ $\quad \Rightarrow \quad \theta u'(q^e_H) = 2(C^e_H - \bar{C}^i_H)/k_H = 2m(k) - \beta m(k) > m(k)$

$[q^i_L]$ $\quad R^i_L = C^i_L - N\xi(\theta-1)u'(q^i_L)$ $\quad \Rightarrow \quad u'(q^i_L) = \beta m(n) - N\xi(\theta-1)u'(q^i_L)/n_L$

$[q^e_L]$ $\quad R^e_L = 2C^e_L - \bar{C}^i_L - N(1-\xi)(\theta-1)u'(q^e_L) \Rightarrow \quad u'(q^e_L) = 2m(k) - \beta m(k) - N(1-\xi)(\theta-1)u'(q^e_L)/k_L$

where, as in the previous section 4.4.1, the parameter $0 \leq \xi \leq 1$ is equal to zero when $q^i_L < q^e_L$ and equal to one for $q^i_L > q^e_L$, and determined endogenously for $q^i_L = q^e_L$.

Let us compare the previous conditions with the benchmark case given by first degree access charge discrimination. This shows the ways in which, in absence of lump sum access prices, the upstream incumbent creates distortions between the downstream producers in order to maximise his total profit Π^u.

Equations $[k_H]$ and $[k_L]$ express the optimal relationship between n_L, n_H, k_H and k_L; that is, the optimal distortion which holds between the two marginal access prices for the two downstream producers and are similar to $[{}^*k_H]$ and $[{}^*k_L]$.

In particular, condition $[k_H]$ can be rewritten as $\pi^e_{kH} - \pi^i_{nH} = \bar{\pi}^i_{nH} - \pi^e_{kH} > 0$ and tell us that, in order to serve the high-demand market, the downstream entrant's marginal profit on an additional customer of type H should be greater than the downstream incumbent's one. The additional profit -which is equal to zero in $[{}^*k_H]$- is just given by $(\bar{\pi}^i_{nH} - \pi^e_{kH})$ the marginal rent obtained by the downstream incumbent if he takes the contract designed for the downstream entrant. A similar reasoning can be made for condition $[k_L]$, which can be rewritten as $\pi^e_{kL}\ \pi^i_{nL} = \bar{\pi}^i_{nL} - \pi^e_{kL} > 0$, where it is easy too see how the downstream incumbent enjoys a rent whose marginal value for an additional low-demand customer is given by $(\bar{\pi}^i_{nH} - \pi^e_{kH})$.

Hence we have a positive differential marginal access price for the downstream entrant which reduces her scale of entry (in order to decrease the downstream incumbent's rent).

While the downstream incumbent pricing conditions, given by $[q^i_H]$ and $[q^i_L]$, stay the same the incentive compatibility constraint implies an additional distortion for the downstream entrant's pricing conditions $[q^e_H]$ and $[q^e_L]$. In particular, equation $[q^i_H]$ shows how the high customers' bundle which is independently chosen by the downstream incumbent is not distorted, in practice marginal price p^i_H equates marginal cost $\theta\ u'(q^i_H)$, and $F^i_H = 0$. As before, equation $[q^e_H]$ tells us that no surplus is ever left to the low-demand consumers. The marginal prices $u'(q^i_L)$ may be distorted, in order to reduce the net surplus of high-demand customers. As in equations $[{}^*q^i_L]$, the value of the parameter $0 \le \xi \le 1$ shows how the distortion at the bottom for the downstream incumbent depends only on which of the constraints $[IC^i_H]$ or $[IC^e_H]$ is binding.

On the other hand, condition $[q^e_H]$ can be rewritten as $R^e_H - C^e_H = C^e_H - \bar{C}^i_H > 0$ or equivalently $\theta u'(q^e_H) - C^e_H/k_H = (C^e_H - \bar{C}^i_H)/k_H = m(k)\text{-}\beta m(k) > 0$. Thus, differently from $[{}^*q^e_H]$, when the downstream entrant serves the high-demand market, her marginal price must be greater than the marginal cost and we have some **distortion at the top.** The difference (C^e_H - $\overline{C}^i_H$) is equal to the marginal rent obtained by the downstream incumbent if he takes the contract designed for the downstream entrant. The same consideration can be made for condition [q^e_L] since also in this case the downstream incumbent enjoys a rent whose marginal value for an additional unit in the low customers' bundle is given by $C^e_L - \overline{C}^i_L > 0$ or equivalently $k_L[m(k)-\beta m(k)] > 0$.

Hence, equations [q^e_H] and [q^e_L] show how the upstream monopolist finds it optimal to distort the bundle of the customers served by the downstream entrant. In particular, when the downstream entrant's marginal prices p^e_H and p^e_L are always higher than the ones which hold in the benchmark case given by first degree access charge discrimination. That is, we always have a positive marginal access price for the additional customer's quantity served by the downstream entrant.

In sum, we may notice how in the present framework (of non-linear access charges in vertical separation) the most relevant differences derive from the presence of this additional cost given by the marginal rent obtained by the downstream incumbent if he takes the contract designed for the downstream entrant. This makes more likely that the upstream monopolist prefers the "e" type to specialise in low consumers' type. In fact, in this way, the marginal rent (related to the additional customer and quantity) of the more efficient type is automatically reduced. Furthermore, for β low enough, it is also optimal that she will not serve all the customers of type L.

A natural question which arises is whether, in terms of social welfare, this kind of non-linear access distortions represent a minor price to be paid (in efficiency term) in order to prevent a less efficient competitor from an inefficiently high scale of entry or from undesirable forms of tariff competition. This leads us to the next subsection where we examine in which way monopoly access charge tariffs should be modified to improve social welfare.

4.4.3 Socially optimal access charge tariffs

In what follows, we always assume that the upstream firm is a public (or regulated) firm; i.e. his revenues represent public revenues, and also the firm's costs and transfers are components of the public budget. As usual, the superscript W, λ, ϕ, denote respectively Loeb and Magat's, Laffont and Tirole's and Baron and Myerson's optimal pricing.

Let us start ignoring any eventual distributional and excess burden consideration, when the public upstream monopolist's general problem is to maximise the ultra-liberal social welfare W (with respect to q^0, n_L, n_H, k_L, k_H, q^e_H, q^e_L, q^i_H and q^i_L) a la Loeb and Magat (i.e. the unweighted sum of surpluses and profits). Taking advantage of the previous analysis (setting constraints [PC^e] and [IC^i] as binding), we may write the following constrained maximisation:

[U2.LM]	$\max W = U(q^0, q^i_H, q^i_L, q^e_H, q^e_L) - NC(2N, Q^0, Q) - C^i - C^e$	subject to:
[PC^e]	$F(k_L, k_H, q^e_H, q^e_L) = \pi^e(k_L,k_H,q^e_H,q^e_L) = k_L T^e_L + k_H T^e_H - C^e(k_H, k_L, q^e_H, q^e_L) = 0$	
[IC^i]	$F(n_L,n_H,q^i_H,q^i_L) = \pi^i(n_L,n_H,q^i_H,q^i_L) - \pi^i(k_L,k_H,q^e_H,q^e_L) - \pi^e(k_L,k_H,q^e_H,q^e_L)$	
[N^i_L]	$n_L + k_L = N$	
[N^i_H]	$n_H + k_H = N$	
and	$k_L, k_H \geq 0$	

Using the constraints [N^i_L] and [N^i_H] and substituting the value of $U(q^0, q^i_H, q^i_L, q^e_H, q^e_L)$ we can reformulate Loeb and Magat's social welfare maximisation problem as:

$$\max W = 2N[v'(q^0)-c^0]q^0 - NC(2N) + (N-k_L)u(q^i_L) + (N-k_H)\theta u(q^i_H) - C^i + k_L u(q^e_L) + k_H \theta u(q^e_H) C^e$$

subject to: $k_L, k_H \geq 0$,

getting the following first order conditions:

[$^w q^0$]	$c^0 = v'(q^0) = p^0$	
[$^w k_H$]	$\theta u(q^i_H) - C^i_{nH} = \theta u(q^e_H) - C^e_{kH}$	for $k_H, n_H > 0$
	$\theta u(q^i_H) - C^i_{nH} > \theta u(q^e_H) - C^e_{kH}$	for $k_H = 0, n_H > 0$
[$^w k_L$]	$u(q^i_L) - C^i_{nL} = u(q^e_L) - C^e_{kL}$	for $k_L, n_L > 0$
	$u(q^i_L) - C^i_{nL} > u(q^e_L) - C^e_{kL}$	for $k_L = 0, n_L > 0$
[$^w q^i_H$]	$\theta u'(q^i_H) = C^i_H / n_H = \beta m(n)$	for $n_H > 0$

[$^w q_L^i$] $u'(q_L^i) = C_L^i/n_L = \beta m(n) \leq m(k)$ for $n_L > 0$

[$^w q_H^e$] $\theta u'(q_H^e) = C_H^e/k_H = m(k)$ for $k_H > 0$

[$^w q_L^e$] $u'(q_L^e) = C_L^e/k_L = m(k) \leq \beta m(n)$ for $k_L > 0$

From [$^w q^0$] we see how in the monopolised final market for good 0 price p^0 should equate marginal cost c^0, as in the previous chapter (section 3.3.3) and differently from first degree access charge discrimination (section 4.4.1).

Comparing the previous conditions [$^w k_H$] and [$^w k_L$] with the benchmark case (cf. equations [$^* k_H$] and [$^* k_L$]) we find out they are exactly the same. Hence, we can conclude that in order to maximise Loeb and Magat's social welfare, the upstream incumbent must not create distortions between the downstream producers. In practice, in order to serve either of the two markets, the downstream entrant's marginal profit on an additional customer of a given type should be equal to the downstream incumbent's one. The optimal relationship between n_L, n_H, k_H and k_L implies that no distortion should be created between the marginal access prices of downstream producers.

Hence, improvements with respect to the unregulated non-linear access charge case (where instead conditions [k_H] and [k_L] imply some distortion, since $F^e_{kH}(k_L,k_H,q^e_H,q^e_L) > F^i_{nH}(N-k_L,N-k_H,q^i_H,q^i_L)$ and $F^e_{kL}(k_L,k_H,q^e_H,q^e_L) > F^i_{nL}(N-k_L,N-k_H,q^i_H,q^i_L)$) would be achieved by reducing the value of the marginal access charges the downstream entrant (the less efficient competitor) face till they reach the values of the downstream incumbent's ones.

Notice how, when the constraints [PC^e] and [IC^i] are binding we consider only the tariffs which do not allow positive rents to the downstream entrant (as in the previous chapter, section 3.3.3). In general, other non-linear access charges, which satisfy constraints [PC^e] and [IC^i] (not necessarily binding), leaving to downstream producers some rent, would also be optimal.

We have explored the cases in which the downstream entrant serves both types or she must specialise in serving one market in the first degree access price discrimination setting. This analysis will continue to hold in this regulated setting, the only difference being that now the optimal customers' tariffs must equate marginal prices to marginal costs so that the downstream producers should be forced not to distort low consumers' bundles. In this regard, even the first degree access price discrimination would not be socially optimal adopting Loeb and Magat's social welfare function.

Let us then consider the marginal access tariffs set for the high and low demand customers and compare them with the previous price discrimination setting.

As before, the first order conditions [$^w q_H^e$] and [$^w q_L^e$] hold only if the downstream entrant is not specialised in serving respectively low demand customers ($k_H = 0$) or high demand customers ($k_L = 0$). From the first order conditions with respect to the low demand per customer bundle [$^w q_H^i$] and [$^w q_H^e$] (exactly the same as [$^* q_H^i$] and [$^* q_H^e$]), we see how the marginal access tariffs F_H^e and F_H^i are always equal to zero. Hence, differently from the second degree charge discrimination setting there is **no pricing distortion at the top.** The high demand consumer's marginal prices $\theta u'(q_H^i)$ set by the downstream producers are simply equated to the per customer marginal costs $C_H^j/n_H = \partial C(N^j, Q^j, \beta^j)/\partial Q^j$.

Referring to the low-demand per customer bundle the first order conditions [$^w q_L^i$] and [$^w q_L^e$], show that (differently from the first and second degree price discrimination setting) there is **no pricing distortion at the bottom**. No surplus is left to low-demand consumers; but marginal prices [$u'(q_L^i)$ and $u'(q_L^e)$] are not distorted by the downstream producers, as there is no interest in reducing the net surplus of the H customers. However, the value of the marginal prices in equations [$^w q_L^i$] and [$^w q_L^e$] should not be lower than the ones of the H customers, in order to prevent them from mimicking low demand customers.

Explicitly taking into account the cost of public funds (following Laffont and Tirole) the public upstream monopolist would give more weight to his profits in maximising the Laffont and Tirole social welfare function W^λ:

[U2.LT] $\max W^\lambda = U(q^0, q_H^i, q_L^i, q_H^e, q_L^e) + \lambda R^u - (1+\lambda)NC(2N, Q^0, Q) - C^i - C^e =$

$2Nv(q^0) + \lambda 2Np^0 + n_H\theta u(q_H^i) + n_L u(q_L^i) + k_H\theta u(q_H^i) + k_L u(q_L^e) +$

$$\lambda F(n_L,n_H,q^i_H,q^i_L)+\lambda F(k_L,k_H,q^e_H,q^e_L) - (1+\lambda)[c^0(2Nq^0)-\Lambda C(2N)] - C^i - C^e$$

subject to:

[PC^e] $\quad F(k_L, k_H, q^e_H, q^e_L) = \pi^e(k_L,k_H,q^e_H,q^e_L) = k_L T^e_L + k_H T^e_H - C^e(k_H, k_L, q^e_H, q^e_L) = 0$

[IC^i] $\quad F(n_L,n_H,q^i_H,q^i_L) = \pi^i(n_L,n_H,q^i_H,q^i_L) - \pi^i(k_L,k_H,q^e_H,q^e_L) - \pi^e(k_L,k_H,q^e_H,q^e_L)$

[N^j_L] $\quad n_L + k_L = N$

[N^j_H] $\quad n_H + k_H = N$

and $\quad k_L, k_H \geq 0$

Clearly, for $\lambda > 0$ the constraint [PC^e] is binding, since it is optimal to expropriate downstream producers' profits. In particular, [IC^i] being binding, there is a trade-off between the expropriation of the downstream incumbent's rent and efficiency.

We can substitute the binding constraints [PC^e] and [IC^i] into W^λ simplifying the problem which reduces to the maximisation of the following function (with $N_L = N - K_L$):

$$\max W^\lambda = 2Nv(q^0)+\lambda 2Nv'(q^0)q^0+(1+\lambda)[(N-k_H)\theta u(q^i_H)+(N-k_L)u(q^i_L) +k_H\theta u(q^e_H)+k_L u(q^e_L)]-\lambda N(\theta-1)\max[u(q^i_L),u(q^e_L)]- -(1+\lambda)[c^0(2Nq^0)-\Lambda C(2N)]-(1+\lambda)C^i -(1+\lambda)C^e -\lambda(C^e -\bar{C}^i)$$

subject to: $\quad k_L, k_H \geq 0,$

In this case there is some distortion for type "e" (the less efficient producer) and also for low-demand customers but no distortion for type "i" (the more efficient one), as in the unregulated case. Specifically, as the first order conditions below show, this regime coincides with the unregulated second degree price discrimination setting when the shadow cost of public funds λ is infinity and with the Loeb and Magat regime when λ is zero.

[$^\lambda q^0$] $\quad (p^0 - c^0)/p^0 = - \lambda/(1+\lambda)\, v''(q^0)\, q^0/p^0 = + \lambda/(1+\lambda)E^0$

[$^\lambda k_H$] $\quad \pi^i_{nH} = \theta u(q^i_H)-C^i_{nH} = \theta u(q^e_H)-C^e_{kH}-\lambda(C^e_{kH}-\bar{C}^i_{kH})/(1+\lambda) \qquad$ for $k_H, n_H > 0$

$\quad \pi^i_{nH} = \theta u(q^i_H) -C^i_{nH} > \theta u(q^e_H) -C^e_{kH} -\lambda(C^e_{kH}-\bar{C}^i_{kH})/(1+\lambda) \qquad$ for $k_H = 0, n_H > 0$

[$^\lambda k_L$] $\quad \pi^i_{nL} = u(q^i_L) -C^i_{nL} = u(q^e_L) -C^e_{kL}-\lambda(C^e_{kL}-\bar{C}^i_{kL})/(1+\lambda) \qquad$ for $k_L, n_L > 0$

$\pi^{i}_{nL} = u(q^{i}_{L}) - C^{i}_{nL} > u(q^{e}_{L}) - C^{e}_{kL} - \lambda(C^{e}_{kL} - \bar{C}^{i}_{kL})/(1+\lambda)$ for $k_L = 0, n_L > 0$

$[^{\lambda}q^{i}_{H}]$ $\theta u'(q^{i}_{H}) = C^{i}_{H}/n_H = \beta m(n)$ for $n_H > 0$

$[^{\lambda}q^{i}_{L}]$ $u'(q^{i}_{L}) = C^{i}_{L}/n_L - \lambda N\xi(\theta-1)u'(q^{i}_{L})/n_L(1+\lambda) \leq \beta m(n)$ for $n_L > 0$

$[^{\lambda}q^{e}_{H}]$ $\theta u'(q^{e}_{H}) = C^{e}_{H}/k_H + \lambda(C^{e}_{H} - \bar{C}^{i}_{H})/k_H(1+\lambda) \leq m(k)$ for $k_H > 0$

$[^{\lambda}q^{e}_{L}]$ $u'(q^{e}_{L}) = C^{e}_{L}/k_L + \lambda[(C^{e}_{H} - \bar{C}^{i}_{H}) + N(1-\xi)(\theta-1)u'(q^{e}_{L})]/k_L(1+\lambda) \leq m(k)$ for $k_L > 0$

where, as in the previous sections 4.4.1 and 4.4.2, the parameter $0 \leq \xi \leq 1$ is equal to zero when $q^{i}_{L} < q^{e}_{L}$ and equal to one for $q^{i}_{L} > q^{e}_{L}$, and determined endogenously for $q^{i}_{L} = q^{e}_{L}$.

For intermediate values ($\lambda > 0$ and $k_H > 0$), welfare would be enhanced by reducing the value of the downstream entrant's marginal access charges (related to the additional high demand customer and bundle), till they differ from the downstream incumbent's ones by the socially optimal values [i.e. respectively $\lambda(C^{e}_{kH} - \bar{C}^{i}_{kH})/(1+\lambda)$ and $\lambda(C^{e}_{H} - \bar{C}^{i}_{H})/k_H(1+\lambda)$]. The same type of reasoning applies to the low-demand customers served by the downstream entrant and their bundle (the additional issue in this case being the problem of reducing the high-demand customers' rent).

To reach the second best tariffs $F^{\lambda}(n_L, n_H, q^{i}_{H}, q^{i}_{L})$ and $F^{\lambda}(k_L, k_H, q^{e}_{H}, q^{e}_{L})$ which solve the public upstream monopolist's problem **[U2.LT]** it is also needed to enforce less distortion between the downstream producer's marginal prices (for type H and L).

Consequently, allowing for non-linear access charges, to maximise welfare W^{λ} the public firm causes a twofold distortion to the entrant imposing a distortionary allocation of the number of customers ($k^{\lambda} < k^{W}$) and a distortionary choice of (${}^{\lambda}q^{i}_{L} < {}^{W}q^{i}_{L}$) through positive marginal access charges.

Hence, similarly to the second degree charge discrimination setting, the presence of an additional cost given by the marginal rent obtained by the downstream incumbent leads to **some pricing distortion at the top**. In fact, the high demand consumer's marginal prices $\theta u'(q^{e}_{H})$ set by the downstream entrant are greater than the per customer marginal costs C^{e}_{H}/n_H.

With respect to the low demand per customer bundle the first order conditions $[{}^{\lambda}q^{i}_{L}]$ and $[{}^{\lambda}q^{e}_{L}]$, show that there is also **some pricing distortion at the bottom** (even if less than in a second degree price discrimination setting).

Finally, let us briefly follow the Baron and Myerson approach which puts more weight on consumers' surplus rather than on the downstream producer's profits $\phi\Pi^{i}$ and let W^{ϕ} represent the welfare function maximised by the upstream public monopolist:

[U2.BM] $\max W^{\phi} = U - N\mathcal{C}(2N, Q^{0}) - (1-\phi)(R^{i}+R^{e}) + (1-\phi)(F^{i}+F^{e}) - \phi(C^{i}+C^{e}) =$

$2Nv(q^{0}) + \phi[n_{H}\theta u(q^{i}_{H}) + n_{L}u(q^{i}_{L}) + k_{H}\theta u(q^{e}_{H}) + k_{L}u(q^{e}_{L})] +$

$N(1-\phi)\max[u(q^{i}_{L}), u(q^{e}_{L})] + (1-\phi)[F(k_{L}, k_{H}, q^{e}_{H}, q^{e}_{L}) + F(n_{L}, n_{H}, q^{i}_{H}, q^{i}_{L})]$

$-c^{0}(2Nq^{0}) - N\mathcal{C}(2N) - \phi(C^{i}+C^{e})$ subject to:

$[PC^{e}]$ $F(k_{L}, k_{H}, q^{e}_{H}, q^{e}_{L}) = \pi^{e}(k_{L}, k_{H}, q^{e}_{H}, q^{e}_{L}) = k_{L}T^{e}_{L} + k_{H}T^{e}_{H} - C^{e}(k_{H}, k_{L}, q^{e}_{H}, q^{e}_{L}) = 0$

$[IC^{i}]$ $F(n_{L}, n_{H}, q^{i}_{H}, q^{i}_{L}) = \pi^{i}(n_{L}, n_{H}, q^{i}_{H}, q^{i}_{L}) - \pi^{i}(k_{L}, k_{H}, q^{e}_{H}, q^{e}_{L}) - \pi^{e}(k_{L}, k_{H}, q^{e}_{H}, q^{e}_{L})$

$[N^{j}_{L}]$ $n_{L} + k_{L} = N$

$[N^{j}_{H}]$ $n_{H} + k_{H} = N$

and $k_{L}, k_{H} \geq 0$

This problem becomes easy to solve once the constraints are substituted into W^{ϕ}:

$\max W^{\phi} =$ $2N[v(q^{0}) - c^{0}q^{0}] + [(N-k_{H})\theta u(q^{i}_{H}) + (N-k_{L})u(q^{i}_{L}) + k_{H}\theta u(q^{e}_{H}) + k_{L}u(q^{e}_{L})] - N\mathcal{C} -$

$C^{i} - C^{e} - (1-\phi)(C^{e} - \bar{C}^{i})$

subject to: $k_{L}, k_{H} \geq 0,$

Maximising the previous function we obtain the following first order conditions.

$[{}^{\phi}q^{0}]$ $c^{0} = v'(q^{0}) = p^{0}$

$[{}^{\phi}k_{H}]$ $\pi^{i}_{nH} = \theta u(q^{i}_{H}) - C^{i}_{nH} = \theta u(q^{e}_{H}) - C^{e}_{kH} - (1-\phi)(C^{e}_{kH} - \bar{C}^{i}_{kH})$ for $k_{H}, n_{H} > 0$

$\pi^{i}_{nH} = \theta u(q^{i}_{H}) - C^{i}_{nH} > \theta u(q^{e}_{H}) - C^{e}_{kH} - (1-\phi)(C^{e}_{kH} - \bar{C}^{i}_{kH})$ for $k_{H} = 0, n_{H} > 0$

$[{}^{\phi}k_{L}]$ $\pi^{i}_{nL} = u(q^{i}_{L}) - C^{i}_{nL} = u(q^{e}_{L}) - C^{e}_{kL} - (1-\phi)(C^{e}_{kL} - \bar{C}^{i}_{kL})$ for $k_{L}, n_{L} > 0$

$\pi^{i}_{nL} = u(q^{i}_{L}) - C^{i}_{nL} > u(q^{e}_{L}) - C^{e}_{kL} - (1-\phi)(C^{e}_{kL} - \bar{C}^{i}_{kL})$ for $k_{L} = 0, n_{L} > 0$

$[{}^{\phi}q^{i}_{H}]$ $\theta u'(q^{i}_{H}) = C^{i}_{H}/n_{H} = \beta m(n)$ for $n_{H} > 0$

$[{}^{\phi}q^i_L]$ $u'(q^i_L) = C^i_L/n_L = \beta m(n)$ for $n_L > 0$

$[{}^{\phi}q^e{}_H]$ $\theta u'(q^e_H) = C^e_H/k_H + (1-\phi)(C^e_H - \bar{C}^i_H)/k_H \leq m(k)$ for $k_H > 0$

$[{}^{\phi}q^e{}_L]$ $u'(q^e_L) = C^e_L/k_L + (1-\phi)(C^e_L - \bar{C}^i_L)/k_L \leq m(k)$ for $k_L > 0$

Condition $[{}^{\phi}q^0]$ coincides with $[{}^{w}q^0]$ and Pareto efficiency is achieved, since marginal price equates marginal cost in the monopolised market. There is also no distortion between the types served by the downstream incumbent. Some distortion arises instead in the marginal access charges of the less efficient downstream producer, in order to reduce the incumbent's rent, as in the Laffont and Tirole case. In fact, the participation condition $[PC^e]$ is always binding so that the downstream entrant must have no rent; moreover $[IC^i]$ creates two trade-offs between rent extraction (of the downstream incumbent) and efficiency (i.e. $(1-\phi)(C^e_{kL} - \bar{C}^i_{kL}) > 0$ and $(1-\phi)(C^e_H - \bar{C}^i_H)/k_H > 0$).

As with one type of customers the entry's scale and the L customers' bundle are again distorted. In particular, the number of customers of the downstream incumbent is relatively high, while the bundle offered to customers by the downstream entrant is relatively low, with respect to the optimal one with the Loeb and Magat social welfare function.

Similarly to Laffont and Tirole's case, social welfare maximisation of W^{ϕ} causes this twofold distortion both on the number of customers ($k^{\phi} < k^{W}$) and on the customers' bundle (${}^{\phi}q^i_L < {}^{W}q^i_L$). However, the access charges are less distorted than they would be under the private monopoly solution, since they tend towards the monopoly ones as the weight of profits ϕ approaches to zero. Differently from the Laffont and Tirole regime there is no distortion between high and low customers served by the downstream incumbent. In order to equate the incumbent marginal prices p^i_H and p^i_L a negative marginal access charges is required.

Also as far as the access charge tariffs are concerned, in Baron and Myerson's case, a reduction of the entrant's monopoly marginal access charges will enhance social welfare, till we reach the second best access charges. This concludes our analysis of vertical separated markets and the design of socially optimal non-linear access charges. The next section summarised the results derived in this chapter.

4.5 Final remarks

In this chapter we first derived the modelling implications of introducing vertically separated structures where an upstream monopolist owns the network, within the previous cost assumption, without major changes with respect to original the vertical game proposed in chapter 2. One may ask himself whether, abstracting from transition and welfare costs of separation, in this simplified common network case, the upstream monopolist is able to retain the same profit of the vertically integrated incumbent when he faces two downstream producers.

We have explored this question in detail, showing how a discriminating upstream monopolist will continue to follow the monopoly pricing strategy expropriating all the downstream producers' profits, even when only non-linear pricing is allowed (provided that access rights can be resold, or the tariff is a function of the customers' bundles) and that cream skimming and surplus taking strategy turn out to be again the more profitable strategies of competition for the downstream entrant. However, when the access charge tariffs can no longer be determined by the upstream monopolist on the basis of the customers' bundles (due for instance to an additional regulatory constraint), the standard result of 'no distortion at the top' does no longer hold, because a new trade off arises between rent extraction and profit maximisation for the downstream incumbent, so that a tariff distortion at the top is introduced, with an additional constraint on the produced quantity.

In practice, in the context of our basic simplified common network case, the addition of vertical separation to competition seems to play the minor role of complicating the industrial framework without providing further efficiency gains. In order to verify if this result holds in a more standard set-up, we considered the case in which the downstream sector is portrayed (similarly to the two customer types' market) with one producer (type "i") strictly more efficient of the other (type "e") for any equal number of customers served.

Here, abstracting from many interesting variants and extensions, we focused mainly on the interactions between competition and regulation when the downstream producers face one or two types of consumers and the upstream monopolist can use non-linear access charge tariffs. Also in this generalised framework, we have shown how cream skimming may arise depending on the specification of the cost functions, given the vertical structure of the model. In fact, also in this framework we do not make any ad hoc hypothesis, such as excluding a priori the possibility for the entrant to serve low demand customers.

In practice, referring to a non-linear price setting, the basic cream skimming model continues to hold, even if the two roles are interchanged: in fact, in the absence of any capacity constraint, it is likely that the downstream entrant will be obliged by the upstream monopolist's access charge tariffs to serve only low-demand customers, since she is less efficient and in this way the rent enjoyed by the downstream incumbent is reduced. This is not the case for a public (or regulated) upstream monopolist, who maximises the Loeb Magat welfare function, as in first degree price discrimination, since there are no strong incentives to allow the downstream entrant to serve only low-demand customers.

The main general conclusion that we can derive from this analysis is that there is **no distortion at the top** for the downstream entrant's customers and **some distortion at the bottom, for low customers' allocations and bundles.** This means that, in practice, a modified monopoly result applies also in this framework, with quite relevant *exceptions*. In particular, when the upstream monopolist is allowed by the regulator to use the more general access charge tariff structure, the private monopoly solution implies a *tridimensional* distortion with respect to Loeb Magat's welfare maximising solution as: (a) on the number of customers served by the downstream entrant (which is reduced $k^m < k^W$), (b) on the low customers' bundle (which is decreased $q_L^i, q_L^e < q_L^W$), and (c) on the high customers' bundle offered by the downstream entrant (which is decreased since $q_H^e < q_H^W$). A similar result applies in Laffont and Tirole's case, as social welfare W^λ maximisation distorts the marginal access charges, but to a lower extent. Instead, with Baron and Myerson's social welfare W^{ϕ} function there is a *twofold* distortion only with respect to the downstream entrant, since only the number of customers ($k^{\phi} < k^{W}$) and his customers' bundle (${}^{\phi}q_L^e$ $< q_L^W$) are reduced.

In the next chapter we summarise the main results of our analysis, sketching the progress made in each chapter. To sum up the main policy implication of the case of vertical separation, analysed in this chapter, we can strongly argue that competition at the downstream level should be encouraged. In fact, in the alternative "more standard" framework also competition by a less efficient downstream entrant (compared to the downstream incumbent) turns out to be welfare enhancing in the presence of an upstream public firm. However, regulation is still needed to avoid inefficient scales of entry or distortions which are not socially optimal, but would emerge in the absence of any regulatory constraint.

Chapter 5

SUMMARY AND CONCLUSIONS

In broad terms what motivated this thesis was to gain a greater understanding of the particular features introduced in the standard regulatory set-up by competitive issues and vertically related markets. Specifically, we intended to explore their impact on the profitability of the market and the possibility for the incumbent to maintain monopoly profits under different regulatory regimes. Our focus has been on the access pricing problem which is becoming the key issue to the regulators. Despite the presence of a growing literature in these area, nevertheless models fail to incorporate the use of non-linear pricing. Since price discrimination is common in practice this omission can lead to misleading results.

In what follows we summarise the main results of our study, reconsidering in detail the whole structure of the thesis.

Chapter 1 provided an overview of the current state of the art of the literature on regulation (and related issues in procurement). We considered the recent developments of the literature on regulation, notably the substantial achievements of Laffont and Tirole, largely collected in their 1993 book. We explored in detail the relevance of the presence of asymmetric information between the regulator and the regulated firm. We showed that moral hazard as introduced in the standard way (referring for instance to Laffont and Tirole's canonical model) does not bring substantial changes in the outcomes relative to a pure adverse selection model. We then considered the ways in which competition and regulation interact in different scenarios. Namely, regarding regulation we took into account regulatory constraints on final goods, such as price caps, and on intermediate goods. Finally, competition within the market has been distinguished from competition for natural monopoly.

These are the reasons why we adopted a simplified framework in which we ignored asymmetric information between the regulator and the regulated firm, introducing the only type of asymmetric information on the consumers' side.

Technically, we dealt with a game of incomplete information, classified also as a game of *mechanism design,* sketched below. In the canonical framework a monopolist (the principal) has incomplete information about the agent's types, namely the consumers' willingness to pay for his goods. He has to design a tariff schedule, which determines the price to be paid as a function of the quantity purchased. We extended this canonical setting to the presence of competition and vertically related markets.

Making reference to the existing literature, the only model that introduces second degree price discrimination in a regulatory setting is Laffont and Tirole (1990b). Regarding *competitive issues*, however, this model is quite restrictive, since it ignores the most interesting problems of strategic competition. In fact, as in the original (1986) model competitors are assumed to have an unlimited capacity, and they don't really undertake any type of economic decisions (competitors do not even choose which type of customer is more profitable to serve, or the optimal pricing strategy to apply). Basically, this occurs because their technology is exogenously given and, due to their hypothesis of two part tariffs (where the fixed part is simply given by the fixed per capita cost of access) with perfect competition, it automatically determines the surplus competitors offer to the most profitable customers. In practice, the problem of which type of customers will be served in equilibrium is already settled in advance by the authors, by appropriately choosing a relatively high fixed per capita cost in order to avoid competition for the low customers (the skimmed milk).

However, in a more general situation, it is not clear at all that high-demand consumers should always be seen as the most profitable part of the market by the entrant. In other words, the entrant's choice of the type of customers to serve should in general depend also on the relative efficiency of the competitor and needs to be *endogenised* in a sequential multistage game.

In reality we usually do not see a perfectly regulated incumbent or a perfectly competitive fringe, since the typical initial situation entails a "big" incumbent who faces a "small" entrant not able to undertake a complete market invasion. We decided to consider a small output constrained entrant, the idea being that competition should start somewhere. Another basic difference between our approach and the one proposed by Laffont and Tirole is that here the *incumbent* is the *first mover* with respect to pricing policies, i.e. the first player who chooses the tariff for which he is committed to serve the customers. We think that this resembles more closely the typical situation in utilities industries, where the incumbent is more powerful and enjoys the first mover advantage in many respects.

When the competitor does not behave as a purely competitive fringe things become more complicated, and in this case what is the most lucrative part of the market for the entrant and for the incumbent remain open questions. The answer to them depends crucially on the respective marginal costs, the game's structure, the strategy of competition chosen by the entrant, her cost and scale of entry. Here, what is more lucrative for the entrant will determine which type of customer she will choose to serve, if she is able to attack both part of the market.

Chapter 2 examined the *private incentives* of the economic players (basically an incumbent and a potential entrant) in the absence of regulatory constraints in horizontal and vertical settings (where the rival enters the downstream level). Specifically, we explored the conditions under which the incumbent can maintain monopoly profits while entry occurs at one vertical level and act as if he were the only player.

This approach (in which first we leave aside regulation) allowed us to determine what behaviour the regulator could expect from the economic agents in the case in which they operated with no constraints. In this way we identified the persistence of natural monopoly in network and the presence of cream-skimming competition. We have shown that in the absence of vertically related markets cream skimming is not necessarily the most profitable strategy for a potential entrant except for very particular cases. We also examine the question regarding the desirability of cream skimming, showing that this type of competition, contrary to the general wisdom, is not necessarily as "harmful" as skimmed milk competition. Basically, productive efficiency benefits from the entry of a more efficient rival and from an allocative point of view we go towards socially optimal pricing since the distortion at the bottom is reduced.

Introducing vertically related markets we analysed a simple framework in which both the incumbent and his competitors have the same technical requirements for the intermediate good (i.e. there is no need to create new infrastructure or facilities to connect the competitors with the existing network). We referred to this situation as the "common network case", following the definition by Laffont and Tirole.

What we discovered is that, when the incumbent remains the monopolist of an intermediate good which is consumed both internally and by any potential competitors, *cream skimming* is the only strategy of competition allowed by the incumbent. Specifically we showed how the entry of an equally efficient competitor in the market of the non monopolised good 1, represents a limiting case as the incumbent is indifferent about entry. In fact, setting the per customer access charge equal to the monopoly variable profits, he maintains the monopoly pricing strategy and the related profits, *independently* of the entry scale. It is optimal for the incumbent to allow entry, if, maintaining the previous monopoly pricing, he can set a per customer access charge equal to the variable profits of the entrant and they are equal or greater than his own. This happens only with an equally or more efficient competitor. Moreover, the incumbent finds it optimal to oblige the competitor to behave as a surplus taker and to allow only cream skimming competition. However, it would be preferable for the incumbent (and for the maximisation of social welfare) that the competitor sells all the produced good 1 to the incumbent (reaching only break-even profits), so that the latter is able to resell it to the consumers, applying the monopoly tariffs. Further refinements of our vertical game including the endogenisation of the scale of entry (previously exogenously given) and the consideration of more general cost functional forms for the entrant are considered.

The outcomes of the models proved to be robust with respect to these extensions.

In chapter 3 we analysed further the market outcomes described above and the ways in which the regulator can intervene to improve social welfare. The regulator can intervene by directly setting the terms of interconnection between the incumbent and potential rivals, or more generally by imposing some constraints on firms' *conduct*.

Clearly, the aims and objectives of regulation are several and can be partly contradictory. They include not only allocative and productive efficiency, but also the promotion of competition and the respect of universal service obligations (or more generally of a criterion of fairness). Moreover, information and incentives issues complicate the intervention of regulators. However, the amount of private information on the regulatory side built into the standard models seems too limited by comparison with the problems actual regulators face. There is no reason why the regulator should be merely uncertain about the level of the cost function, as in Laffont and Tirole (1994a) and Vickers (1995), and not about its shape, or the demand condition of the industry. Furthermore, as Vickers (1995) says "It would be nice to examine a model in which the regulator was uncertain about both upstream and downstream cost levels" (p. 4), but till now no model assumes that the entrants' costs are private information. As already explained, we decide to leave aside problems of moral hazard and adverse selection on the regulatory side to make the problem more manageable and to reveal its fundamental structure. These assumptions are quite usual in the case of a *public firm* which directly maximises social welfare, as well as in the large part of the debate on access pricing and vertical issues.

Within the models of access pricing proposed by the recent economic literature both competitors and the regulated firm make use only of a linear access pricing policy. Hence, the results are somewhat special in their nature; that is, the proposed regulatory rules are optimal only in a framework in which non-linear tariffs are prevented on "a priori" grounds, or because of an exogenous political constraint.

Since this pricing strategy is a more general one, always superior to the linear pricing strategy, it is not clear why it should be rejected a priori by the firms (and in particular the incumbent) and by the regulatory authority. Hence, we introduced the possibility of making use of non-linear access charges.

The general conclusion we drawn are that the regulator should not allow competition for the low demand type or by a less efficient entrant and should impose the adoption of optimal non-linear pricing. In our horizontal setting the presence of cream skimming competition as shown in Chapter 2 is welfare enhancing since it reduces the distortion at the bottom. However, even in such a favourable case private tariffs are far from optimal; basically because profit maximising firms putting no weight to consumers' surplus distort tariffs too much (compared to a public firm, even if redistributive issues are ignored and we deal with a first best setting). Welfare enhancing instruments, such as the imposition of appropriate entry taxes or the regulation of the entrant have been considered in detail.

Introducing a vertically related market we examined optimal pricing both for access and final goods. We took advantage of the Baumol-Willig rule and tested its optimality both in Baumol (1983)'s original setting and in our vertical game where fully non-linear tariffs are allowed for. We have shown how the Baumol-Willig rule can also be interpreted as a way of imposing an entry tax, in order to avoid socially undesirable entry, while keeping the incumbent's position unchanged (as if entry had not occurred). Under this perspective this rule turned out not to be provide socially optimal pricing criteria, unless we confine ourselves to the Loeb and Magat setting, that is, ignoring redistributive considerations. The other welfare instruments envisaged before are instead still valuable, as they employ the welfare properties implicit in a (horizontal or vertical) merger.

In chapter 4 we introduced vertically separated markets. For a long time regulatory reforms have moved in the direction of divestitures and deregulated structures. Yet the theoretical and applied literature on this intriguing area is very little. We first expanded the model set out in Chapter 2 to see under which circumstances we can get the same outcomes where the regulatory stage involves divestiture. It turned out that if it is possible to resell access rights or if access charges can be expressed as a function of the consumption bundles that characterise each type of consumers nothing changes compared to the complete information benchmark. Basically the upstream incumbent through the use of an appropriate access charge tariff obliges the downstream entrant to cream skim the market and to act as a surplus taker. Doing so he can retain the same profit as entry had not occurred. Hence in such cases vertically separated structures do not introduce any major changes in our setting.

We then model vertical separation in an alternative framework that portrays access price discrimination by an upstream monopoly towards downstream producers of different types (that is, level of efficiency). We extended the model by introducing price discrimination also in the final demand market.

As should be expected, there are no clear cut results in terms of the optimal pricing strategy to be applied in the intermediate and final demand markets. We discover a variety of different combinations of cases that can arise depending on the specification of the functional form of downstream producers' costs and final customers' demand. In this setting downstream producers differ only in their efficiency level, their cost functions having the same functional form and being decreasing in the scale of entry, so that differently from before the entrant is not output constrained.

Nevertheless the basic cream skimming model continues to hold, even if the two roles are interchanged. In fact, in the absence of capacity constraints it is very likely (even if it not always the case) that the downstream entrant will be obliged by the upstream incumbent to serve only low-demand customers, the reason being that in this way the rent enjoyed by the downstream incumbent is reduced.

Another general lesson to be drawn is that distortion can arise not only in the tariff schedule offered to final customers but also in the customers' allocations between the downstream producers. Nevertheless, a modified monopoly result applies also in this framework, with quite relevant exceptions. This shows how pervasive is the standard no distortion at the top result.

To conclude, let us sketch some directions for further research. We believe it is important, building up on what has been already done, to set up a quite general framework which allows a comparison between what optimal regulation would suggest and the results achievable through the most common regulatory constraints on the price level and/or structure. In practice, in the UK supporters of the "price cap" or "RPI-X" argue that significant benefits have been achieved. Yet its architect (Littlechild, 1983) only claimed it to be a short to medium-term instrument, and there have been strident calls for its replacement. Nor was it designed for access pricing. Hence it is worth exploring whether price cap regulation (or other possible candidates) appropriate in a setting where there are vertically related markets and also for the purpose of regulating access pricing.

The advantage of the approach proposed in this thesis is in its simplicity, whilst capturing some of the flavour of Laffont and Tirole's (1990) and (1994a) papers. Yet clearly it is important to consider in some detail within our framework the theoretical implications of the introduction of agency problems between the regulator and the regulated firm. Following this perspective, with a particular emphasis on non-linear pricing in final and intermediate markets, it is possible to develop a model which fits better the needs of the regulatory authorities. Within this we can embed rules, such as RPI-X in order to examine their properties. An additional dynamic issue concerns the analysis of what elements in conjunction with RPI-X can lead the regulator to a closer understanding of the firm's cost structure. In this context, we would like to explore if RPI-X can be considered an appropriate instrument for the longer-term or are there superior alternatives a regulator might employ. Finally, we feel it is relevant to address specific questions related to vertical separation and divestitures. In a more general framework the downstream competitor can be selected through an auction and the optimal number of firms in the downstream sector can be endogenised depending on the industry's cost structure.

Appendix 1: Derivation of the regimes in the LT game

In what follows we will derive all the possible regimes of the LT game (an intuitive account of which has been given in section 1.3) maintaining the same notation and the same simplifying hypotheses sketched in section 1.3.

The incumbent maximises his profit with respect to T_L, q_L, T_H and q_H subject to the following constraints:

max Π	$\alpha_L T_L + \alpha_H T_H - c(\alpha_L q_L + \alpha_H q_H)$	subject to:
[IR_L]	$u(q_L) \geq T_L$	
[IR'_H]	$\theta\, u(q_H) - T_H \geq S_H$	
[IC_L]	$u(q_L) - T_L \geq u(q_H) - T_H$	
[IC_H]	$\theta\, u(q_H) - T_H \geq \theta\, u(q_L) - T_L$	

No comments are needed for [IR_L] [IC_L] and [IC_H], since they have the standard meaning. The modified individual rationality constraint [IR'_H] instead tells that a H consumer will not buy from the incumbent, unless he is allowed to enjoy a rent greater or equal to the exogenous surplus offered by the entrant S_H.

When both types are served by the incumbent, Laffont and Tirole (1993) demonstrate (Appendix A6.3) that there are five different possible no-bypass regimes with the following binding constraints: **(1)** IR_L and IC_H, **(2)** IR_L, IC_H and IR'_H, **(3)** IR_L and IR'_H, **(4)** IR_L, IR'_H and IC_L, **(5)** IR'_H and IC_L.

Here, we will derive all these regimes in a stylised version of the model (LT game), with no asymmetric information and no 'optimal' regulation.

In the LT game we obtain regime **(1)** when the value of S_H is less than the level $(\theta-1)u(q_L)$. Once we substitute IR_L and IC_H into Π, the profit function can be easily maximised only with respect to q_L and q_H:

$$\max \Pi_1 \qquad [\alpha_L - \alpha_H(\theta - 1)]\, u(q_L) + \alpha_H\, \theta\, u(q_H) - c(\alpha_L q_L + \alpha_H q_H)$$

Setting $p_i = \theta_i\, u'(q_i)$, we can verify that the first order conditions are:

$$p_L = c / [1 - (\theta - 1)\, \alpha_H/\alpha_L)] > c$$

$p_H = c$

In particular when the net surplus S_H is increased but is still less than the level $(\theta - 1)u(q_L'')$ (where q_L'' corresponds to the values which occurs for $u(q_L'')$ is equal to c) we are in regime **(2)**, with $p=p_1$ and IR_H' binding. Thus, taking also into account IR_H' we get that in equilibrium:

$$u(q_L) = S_H/(\theta - 1) = T_L;\ p_L' < p_L < c$$

$$p_H = c;\quad T_H = \theta\, u(q_H) - (\theta - 1)u(q_L)$$

Hence, as S_H increases, p_L decreases reaching c (and q_L increase till q_L'') when finally S_H is equal to $(\theta - 1)u(q_L'')$. In practice, while q_H remains constant, the distortion at the bottom is reduced till marginal price p_L becomes equal to marginal cost c.

Then we move to regime **(3)** when $(\theta - 1)u(q_L'') < S_H < (\theta - 1)u(q_H')$ and (due to a greater S_H) IC_H is no longer binding. Consequently, from IR_L and IR_H' we get: $T_H = \theta\, u(q_H) - S_H$ and the profit function becomes:

$$\Pi_3 = \alpha_L\, u(q_L) + \alpha_H\, [\theta\, u(q_H) - S_H] - c(\alpha_L\, q_L + \alpha_H\, q_H).$$

We can easily verify that in equilibrium marginal prices equal to marginal costs:

$$p_L = p_H = c;\quad T_L = u(q_L'');\quad T_H = \theta\, u(q_H) - S_H$$

Later, as S_H increases it becomes equal to $(\theta - 1)u(q_H)$ and we are in regime **(4)**, where IC_L becomes binding. Consequently taking into account this additional constraint we can easily get:

$$p_L'' = c;\quad T_L = u(q_L'')$$

$$u(q_H) = S_H/(\theta - 1) < c;\quad p_H < c$$

Hence, as S_H increases till S_H is equal to $(\theta - 1)u(q_H'')$, p_H decreases below c (and q_H increases to q_H''). In practice, the marginal price p_H is now below marginal cost c in order to retain the H customers and to prevent pooling.

Finally we get into regime (5) when S_H is greater than $(\theta - 1)u(q''_H)$ and IR_L is no longer binding. Consequently from IC_L and IR'_H we get the new profit function:

$$\Pi_5 = \alpha_L[u(q_L) + (\theta - 1)u(q_H) - S_H] + \alpha_H [\theta\, u(q_H) - S_H] - c(\alpha_L q_L + \alpha_H q_H)$$

and we obtain the following first order conditions:

$$p''_L = c; \quad T_L = u(q'_L) + (\theta - 1)\, u(q''_H) - S_H < u(q'_L)$$

$$p''_H = c/[1 + (\alpha_H/\alpha_L)(\theta\text{-}1)/\theta] < c; \qquad T_H = \theta\, u(q''_H) - S_H$$

Hence, as the individual rationality constraint of type L is no longer binding, the low-demand customers enjoy a positive net surplus $[T_L < u(q_L)]$.

Appendix 2: The surplus taker pricing strategy of the entrant

Here we will prove that to act as a surplus taker represents the optimal pricing strategy for the entrant. That is, marginal tariffs must be set equal to her marginal cost and customers must be allowed the same rent determined by the incumbent.

We are naturally dealing with the the sequential game in which the incumbent accomodates entry. From the formulation of our game the choice of the entrant is binary, in the sense that she can decide either to serve a fixed number of customers, or not to enter the market. In the case entry takes place, the rival must commit herself to serve a given number of customers of type high or low, after observing the incumbent's tariffs fixed for each type.

We have already determined the solution of the game for a tariff taker entrant. It is easy to show how the entrant may improve upon it, leaving unchanged the number of customers of each type served, if she acts instead as a surplus taker. In fact, in this way she will increase profits setting prices equal to her marginal cost and allowing to each type the same rent implied by the incumbent's tariffs. In practice, if she serves only the low type, no rent is left to the consumers and no marginal losses are incurred. However, the entrant must also respect the incentive compatibility constraint of the high type consumers if she chooses to serve them. The relevant incentive compatibility constraint is naturally the same one considered in the incumbent's problem, as the entrant allows the same surplus.

Notice how this constraint was trivially fulfilled in the case of a tariff taker competitor. In the surplus taker case the entrant proposes a new tariff the solution of her optimisation problem implies the equality between marginal prices and marginal costs. Specifically, in the cream skimming case by setting marginal price equal to marginal cost marginal profits will be equal to zero, so that profits cannot be increased any further.

Intuitively, on one hand if the entrant allows a lower level of surplus than the incumbent the consumer would rather buy the good from the incumbent. On the other hand, offering a higher surplus would entail a net loss of profit.

A similar reasoning carries over to the case of skimmed milk competition and for intermediate cases. The reasoning is also true changing the number of customers of each type served by the entrant, since (for given values of K_L and K_H) it is not possible to improve upon surplus taker pricing. Generally, apart from the case of an equally efficient competitor, the optimal values of the entry scale in each market (K_L and K_H) would be different. In fact, we have already seen (in section 2.4.3 considering the surplus taker case) that for quadratic utility functions we have $\overline{m}^e > \overline{m}$ and $\underline{m}^e > \underline{m}$ for any value of K.

Therefore, the optimal pricing strategy for the entrant is to behave as a surplus taker.

Appendix 3: Endogenising the entrant's choice of customers

The optimisation problem for a tariff-taker entrant, in the general case, can be stated as:

$$\max \Pi^e \qquad K_H (T_H - mq_H) + (K - K_H)(T_L - mq_L) - F(K_H, K - K_H)$$

The marginal profitability per unit of customer is:

$$d\Pi^e/dK_H = (T_H\text{-}mq_H)\text{-}(T_L\text{-}mq_L)+K_H d(T_H\text{-}mq_H)/dK_H+(K\text{-}K_H)d(T_L\text{-}mq_L)/dK_H - (fe_H - fe_L)$$

Let us examine all the terms that enter in $d\Pi^e / dK_H$:

1) $T = (T_H\text{-}mq_H) - (T_L\text{-}mq_L)$ represents the difference between the two net revenues enjoyed by serving a customer of type H and a customer of type L.

2) $\Omega = K_H\, d(T_H\text{-}mq_H)/dK_H + (K\text{-}K_H)\, d(T_L\text{-}mq_L)/dK_H$ represents the *total* net gain (or loss) on the inframarginal customers of type H and L. Let us examine the single components of this term:

$$\Omega_H = d(T_H\text{-}mq_H)/dK_H = -(\theta\text{-}1)\, u'(q_L)\, dq_L/dK_H \leq 0$$

The sign of Ω_H is negative (as $\theta>1$, $u'(q_L)>0$ and $dq_L/dK_H \geq 0$). In fact, W_H represents the *loss* incurred by the entrant for *each* customer of type H. Intuitively, as K_H rises the entrant must be willing to leave for each additional high demand customer a net marginal surplus $(\theta\text{-}1)u'(q_L)$ times the increase in the quantity sold to the L type (dq_L/dK_H).

$$\Omega_L = d(T_L\text{-}mq_L)/dK_H = [u'(q_L) - m]\, dq_L/dK_H \geq 0$$

The sign of Ω_L is positive (as long as $m < u'(q_L)$, since $dq_L/dK_H \geq 0$). In fact, W_L represents the *gain* enjoyed by the entrant for *each* customer of type L. As K_H rises the entrant for each additional high demand customer obtains a gain equal to the marginal revenue $[u'(q_L)\text{-}m]$ times the increase in the quantity sold to the L type (dq_L/dK_H).

3) $fe = (fe_H - fe_L)$ represents the effect brought by a different cost of entry per unit and type of customer, where $fe_H = dF/dK_H$ and $fe_L = dF/d(K\text{-}K_H)$. Its sign is determined by the values assumed by fe_H and fe_L. For simplicity's sake in what follows we will assume that fe=0, a condition is clearly fulfilled whenever the cost of entry depends only on the total scale of entry K.

The entrant will serve both type of customers if and only if $d\pi^E/dK_H=0$, which occurs whenever:

$$m_W = [(T_H-T_L) + (K-\theta K_H)u'(q_L)dq_L/dK_H] / [(q_H-q_L) + (K-K_H)dq_L/dK_H]$$

Proposition: m is a decreasing function of K_H.

Here we will prove that m is a decreasing function of K_H. First, let us deal with the case in which there are no infra-marginal effects (Ω=0). This will surely hold whenever $dq_L/dK_H=0$, as both the expressions Ω_H and Ω_L vanish. Let us then expand this term, applying the chain rule:

$$dq_L/dK_H = \{\partial q_L/\partial p_L\}\{\partial p_L/\partial R\}\{dR/dK_H\}$$

where $p_L=c/[1-(\theta-1)/R]$; $R=N_H/N_L=[N-K_H]/[N-K+K_H]$. The values of the single components are given by:

$$\partial q_L/\partial p_L = 1 / u''$$

$$\partial p_L/\partial R = c(\theta - 1) / [1 - (\theta - 1)/R]^2$$

$$dR/dK_H = - [2N-K] / [N - K + K_H]^2$$

It is self-evident that as N assumes very high values compared to K_L the last component vanish. The same is true when the percentage scale of entry (K/N) is small. Ignoring infra-marginal effects:

$$m = (T_H-T_L)/(q_H-q_L)$$

The denominator of m represents the additional tariff of the H type, which is the integral from q_L to q_H of the area which lies below the H type marginal utility function. This area can be decomposed in the integral from q_L to q_H of the area which lies between the H type marginal utility function and the marginal cost c and the total cost incurred by selling the quantity (q_H-q_L). Dividing the additional tariff by (q_H-q_L) the first component will vary only according to the relative height (θp_L) whereas the latter component is clearly independent of K_H. Since p_L is a decreasing function of K_H the same is true of m. Therefore, $\underline{m}$ is lower than $\overline{m}$.

Consider now the case in which Ω is not equal to 0. We will prove that the lower bound of m $\underline{m}(\Omega\neq0)$ is less than $\underline{m}(\Omega=0)$, where the upper bound $\overline{m}(\Omega\neq0)$ is greater than $\overline{m}(\Omega=0)$, so that also in this case m is a decreasing function of K_H.

Substituting the value of $\underline{m}(\Omega\neq0)$ into Π^e, we get:

$$d\Pi^e / dK_H(\underline{m}) = K\Omega_H = - K(\theta-1)\, u'(q_L)\, dq_L/dK_H < 0$$

To bring back $d\Pi^e / dK_H$ to zero, since the coefficient of m in the expression is negative, m must decrease, so that $\underline{m}$ is lower in the presence of infra-marginal effects.

Applying the same reasoning, we substitute the value of $\overline{m}(\Omega\neq0)$ into Π^e:

$$d\Pi^e / dK_H(\overline{m}) = -K\Omega_L = - K[u'(q_L) - \overline{m}]\, dq_L/dK_H > 0$$

To bring back $d\Pi^e/dK_H$ to zero $\overline{m}$ must decrease, since its coefficient in the expression is negative, so that $\overline{m}$ assumes a greater value in the presence of infra-marginal effects.

Let us provide a *counterexample* in which the entrant finds it more profitable to substitute one H type customer with one L type. Let us consider, for quadratic utility functions the extreme case in which the entrant in the initial setting serves all high-demand customers ($K=N=K_H$) and $p_H=p_L=\beta$. Compare this situation with the one in which the entrant is serving (N-1) customers of H type and only one customer of type L. Clearly, in the latter case $p_H=\beta$, whereas $p'_L > \beta$ (since q_L is a decreasing function of K_H we can also conclude that $q'_L < q_L$).

We need to prove that the gains obtained on the (N-1) customers of H type (hereafter denoted by G_H) are greater than the losses due to the forgone additional tariff on one customer of type H (hereafter G_L).

$$G_H = (N - 1)(\theta - 1)\, u'(q_L)\, (q_L-q'_L)$$
$$G_L = (q_H - q_L)(\theta - 1)\, \beta /2$$

To prove that G_H is greater than G_L, since $u'(q_L) > \beta$, it suffices to show that:

$$(N - 1)(\theta - 1)\, c\, (q_L - q_L') > (q_H - q_L)(\theta - 1)\, \beta /2$$

Dividing both sides by $(\theta-1)\,\beta$ we get:

$$(N - 1)(q_L - q_L') > (q_H - q_L)/2$$

For quadratic utility function we know that:

$$(q_L - q_L') = (1-\beta) - \{1- \beta / [1- (\theta -1)/ (N-1)] = (\theta - 1)/(N - \theta)\beta$$

$$(q_H - q_L) = (1 - \beta/\theta) - (1 - \beta) = (1 - 1/\theta)\, \beta$$

By substituting the values $(q_L - q_L')$ and $(q_H - q_L)$ in the previous expression, for G_H to be greater than G_L we just need to verify that:

$$(N-1)(\theta-1)/ (N-\theta)\, \beta > (\theta-1)\, \beta/2\theta$$

which reduces to:

$$(N-1) > (N-\theta)/2\theta$$

a condition which always holds, since θ is greater than unity.

Chapter 1's notation

Laffont and Tirole's model:

$C = cQ + \varepsilon$	is the general specification of the cost function for the regulated firm where:
$c = \beta - e$	is the marginal (and average) cost
β	is the adverse selection parameter (a technological parameter in the two type case $\beta = \beta_L, \beta_H$)
e	is the moral hazard parameter (the effort level)
Q	is the output
ε	is a noise parameter
W	is Laffont and Tirole's social welfare function in the basic (1986) monopoly model where:
S	is the consumers' gross surplus
λ	is the shadow cost of public funds
$U_m = tr - \psi(e)$	is the managerial utility function where:
tr	represents the transfers from the regulator to the regulated firm
$\psi(e)$	is the disutility of effort, where $\psi'>0$, $\psi''>0$ and $\psi'''>0$
ν	is the probability the regulated firm being efficient (i.e. $\beta=\beta_L$)
$\phi(\beta_H - c_H)$	is the differential rent of the more efficient firm
α_L and α_H	denote the proportion of low and high-demand consumers
S_H	is the surplus offered by the competitive fringe
L	is the Lerner index
η	is the elasticity of demand
$\eta_{n, n+1}$	is the cross elasticity of demand between n and n+1
ε_f	is the fringe's price reaction

The simplified model:

$v = c + \psi(e)$	is our newly defined cost function inclusive of managerial disutility
$\Pi = tr - vQ$	is the incumbent's profit
$u_H = \theta u(q_H)$	is the surplus of the high-demand consumer
$u_L = u(q_L)$	is the surplus of the low-demand type consumer
$\theta > 1$	is the taste parameter
$(T_H, q_H)\ (T_L, q_L)$	are the incumbent's non-linear tariffs chosen for each type of consumer t = H, L (T is the maximum willingness to pay for buying the bundle q)
$p_H = 1 - \theta q_H$	is the per customer inverse demand function for the H type (specified for a quadratic utility function)
$p_L = 1 - q_L$	is the per customer inverse demand function for the L type (specified for a quadratic utility function)
π_L and π_H	are the per customer profits of the incumbent
$R(Q) = [T_L + T_H]\ N$	is the revenue function of the incumbent, where N is the number of customers of each type.

Chapters 2's notation

N	is the number of consumers of each type
$u_t(\theta_t, q_t)$	is the surplus of the consumer of type t for good 1
θ_t	is the taste parameter
q_t	is the bundle of consumer t
T_t	is the incumbent's non-linear tariff chosen for each type of consumer t = H, L (T is the maximum willingness to pay for buying the bundle q)
$p_t(q_t)$	represents the per customer inverse demand function for type t for the final good (good 1)
$p^0(q^0)$	represents the per customer inverse demand function for the monopolised good (good 0)
$v(q^0)$	is the consumer surplus for good 0
Q^0	is the total output of good 0
$Q^1 = Q+Q^e$	is the total output of good 1, where Q and Q^e are the output of the incumbent and the entrant respectively
Π	denotes the profit function of the incumbent
$NC(2N, Q^0, Q^1)$	is the network subcost function of the incumbent
$K = \Sigma_t K_t$	is a fixed limited scale of entry, in terms of number of customers served of type t
$N_t = N - K_t$	is the residual number of customer of type t served by the incumbent
$F(K_H, K_L, Q^e)$	is the access tariff set by the incumbent
$CV = m\, Q^e$	is the variable cost of the entran
m	is the marginal cost of the entrant
c^*	is the unit production cost of the incumbent for good 1
c^1	is the marginal cost of access for good 1
c^0	is the marginal cost of the incumbent for good 0
$s = K/N$	is the proportional scale of entry

T_t^e and q_t^e	are respectively the tariff and the bundle offered to each customer of type t by the entrant
T	is the additional net revenue enjoyed by the incumbent serving a customer of type H in the place of a L type
Ω	is the inframarginal gain enjoyed by the incumbent serving a customer of type H in the place of a L type
fe	is the marginal fixed cost incurred by the incumbent in serving a customer of type H in the place of a L type
me	is the limiting value of the marginal cost for which entry is profitable
π_t	are the per customer profits of the incumbent for each type t
$\boldsymbol{L}$ and μ	are the Lagrangean function and the relative multiplier

Chapters 3's notation

w	as a superscript refers to Loeb and Magat's social welfare function
ϕ	denotes the weight attached to consumers surplus and to the entrant's profit (and as a superscript characterises Baron and Myerson social welfare function)
λ	denotes the shadow cost of public funds and as a superscript characterises Laffont and Tirole's social welfare function
ic	is the incremental cost of providing track space
f	is the per customer access charge
ADC	is the access deficit contribution
f	represents the per unit profit on the access charges

Social Welfare functions

	Loeb-Magat	Laffont-Tirole	Baron-Myerson
Weight on S	1	1	1
Weight on Π^i	1	>1	1
Weight on Π^e	1	1	<1

Chapters 4's notation

Vertical separation model (section 4.2):

Π^u	denotes the profit function of the upstream monopolist
$NC(2N, Q^0, Q^1)$	is the network subcost function of the upstream monopolist
R^j	is the revenue function of the downstream producer j
F^j	is the access tariff for the downstream producer j
N_t	is the number of customer of type t served by the downstream incumbent
K_t	is the number of customer of type t served by the downstream entrant
T^j_L and q^j_t	are respectively the tariff and the bundle offered to each customer of type t by the downstream producer j
$\pi^j = R^j - C^j$	is the gross profit of the downstream producer j (inclusive of access charges)
$\pi^j_t = \pi^j / N^j$	is the per customer gross profit of the downstream producer j
$\Pi^j = \pi^j - F^j$	is the net profit of the downstream producer j
π^m, p^m and Q^m	are the gross profit, the price and the output level in the vertically integrated case

Vertical separation model (section 4.3 and 4.4.):

$R^j = R(N^j, q^j)$	is the revenue function of the downstream producer j
$C^j = C(N^j, q^j, \beta^j)$	is the access tariff for the downstream producer j, where:
N^j	is the number of customer served by the downstream producer j ($N^j = n, k$ for j=i, e)
β^j	is the efficiency of the downstream producer j

$\hat{c}$	is the component of the cost function independent of the scale of entry
q^j	is the per customer output of the downstream producer j
R^j_{Nj} and C^j_{Nj}	are the partial derivatives of the revenue and cost function with respect to the number of customers served by each downstream producers
R^j_L and C^j_L	are the partial derivatives of the revenue and cost function with respect to the bundle q^j_L
$\varepsilon(k)$ and $\varepsilon(n)$	are the elasticity of the cost function with respect to the number of customers (k and n respectively)

References

Anton, J. and D. Yao (1987) "Second sourcing and the experience curve: price competition in defense procurement", Rand Journal of Economics, 18, 57-76.

Armstrong, M.; Cowan, S. and J. Vickers (1995) "Nonlinear pricing and price cap regulation", Journal of Public Economics, 58, 33-55.

Armstrong, M.; Cowan, S. and J. Vickers (1994) REGULATORY REFORM - ECONOMIC ANALYSIS AND BRITISH EXPERIENCE, MIT Press, Cambridge.

Armstrong, M. and C. Doyle (1994) "Access pricing, entry and the Baumol-Willig rule", Discussion Paper in Economics and Econometrics, University of Southampton, n. 9422.

Armstrong, M. and J. Vickers (1991) "Welfare effects of price discrimination by a regulated monopolist", Rand Journal of Economics, 22, 571-80.

Armstrong, M. and J. Vickers (1993) "Price discrimination, competition and regulation", Journal of Industrial Economics, 4, 335-59.

Armstrong, M. and J. Vickers (1995) "The access pricing problem", paper presented at the conference on "New developments in access pricing for network utilities", London, May 24, 1995.

Auriol, E. and J. Laffont (1992) "Regulation by duopoly" Journal of Economics and Management Strategy, 1, 507-33.

Averch, H. and L. Johnson (1962) "Behavior of the firm under regulatory constraint", American Economic Review, 52, 1053-69.

Baron, D. (1989) "Design of regulatory mechanisms and institutions" in Schmalensee, R. and R. Willig (eds.) HANDBOOK OF INDUSTRIAL ORGANIZATION, North Holland, Amsterdam, 1347-1447.

Baron, D. and D. Besanko (1987) "Monitoring, moral hazard, asymmetric information and risk sharing in procurement contracting", Rand Journal of Economics, 18, 509-32.

Baron, D. and R. Myerson (1982) "Regulating a monopolist with unknown costs", Econometrica, 50, 911-30.

Baumol, W. (1983) "Some subtle pricing issue in railroad regulation", International Journal of Transport Economics, 10, 341-55.

Baumol, W. and G. Sidak (1994) TOWARD COMPETITION IN LOCAL TELEPHONY, MIT Press, Cambridge.

Beesley, M and S. Littlechild (1989) "The regulation of privatized monopolies in the U.K." Rand Journal of Economics, 20, 454-72.

Biglaiser, G and Ma, C. (1995) "Regulating a dominant firm: unknown demand and industry structure", Rand Journal of Economics, 26, 1-19.

Bradley, I. and C. Price (1988) "The economic regulation of private industries by price constraint", Journal of Industrial Economics, 37, 99-106.

Braeutigam, R. (1993) "A regulatory bargain for diversified enterprise", International Journal of Industrial Organisation, 11, 1-20.

Brennan, T. (1991) "Entry and welfare loss in regulated industries" in Crew, M. (1991) COMPETITION AND THE REGULATION OF UTILITIES, Kluwer, Boston, 141-156.

Caillaud, B. (1990) "Regulation, competition and asymmetric information", Journal of Economic Theory, 52, 87-110.

Cave, M. (1994) "The role and effectiveness of network access regulation", mimeo, Brunel University.

Crémer, J. and R. McLean (1985) "Optimal selling strategies under uncertainty for a discriminatory monopolist when demands are interdependent", Econometrica, 53, 345-361.

Dana, J. (1993) "The organization and scope of agents: regulating multiproduct industries", Journal of Economic Theory, 59, 288-310.

Dana, J. and K. Spier (1994) "Designing a private industry government auctions with endogenous market structure", Journal of Public Economics, 53, 127-47.

De Fraja, G. and F. Del Bono (1990) "Game theoretic models of mixed oligopoly", Journal of Economic Surveys, 4, 205-9.

De Fraja, G. (1993), "Productive efficiency in public and private firms", Journal of Public Economics, 50, 15-30.

Demski, J.; Sappington, D. and P. Spiller (1987) "Managing supplier switching", Rand Journal of Economics, 18, 77-97.

Economides, N. and L. White (1995) "Access and interconnection pricing: "How efficient is the efficient component pricing rule"?", paper presented at the conference "Interconnection Pricing", Milan, April 1995.

Economides, N. and G. Woroch (1995) "Benefits and pitfalls of network interconnection", mimeo, Stern School of Business, New York University.

Einhorn, M. (1987) "Optimality and sustainability: regulation and intermodal competition in telecommunications", Rand Journal of Economics, 18, 550-563.

Fudenberg, D. and J. Tirole (1991) GAME THEORY, MIT Press, Cambridge.

Gale, I. (1990) "A multiple-object auction with superadditive value", Economic Letters, 34, 323-28.

Gilbert, R. and D. Newbery (1994) "The dynamic efficiency of regulatory constitutions", Rand Journal of Economics, , 538-54.

Green, J. and N. Stockey (1983) "A comparison of tournaments and contests", Journal of Political Economy, 91, 349-64.

Holmström, B. (1982) "Moral hazard in teams", Bell Journal of Economics, 13, 324-40.

Holmström, B. and P. Milgrom (1991) "Multitask Principal-Agent analyses: incentive contracts, asset ownership and job design", Journal of Law, Economics and Organization (Supplement).

Ireland, I. (1991) "Welfare and non-linear pricing in a Cournot oligopoly", Economic Journal, 101, 949-957.

Ireland, I. (1992) "On the welfare effects of regulating price discrimination", Journal of Industrial Economics, 3, 237-248.

Katz, M. (1983) "Non-uniform pricing, output and welfare under monopoly", Review of Economic Studies, 50, 37-56.

Laffont, J. and J. Tirole (1986) "Using cost observation to regulate firms", Journal of Political Economy, 94, 614-41.

Laffont, J. and J. Tirole (1987) "Auctioning incentive contracts", Journal of Political Economy, 95, 921-37.

Laffont, J. and J. Tirole (1988) "Repeated auctions of incentive contracts, investment, and bidding parity with an application to takeovers", Rand Journal of Economics, 19, 516-37.

Laffont, J. and J. Tirole (1990a) "The regulation of multiproduct firms I and II", Journal of Public Economics, 43, 1-66.

Laffont, J. and J. Tirole (1990b) "Optimal bypass and creamskimming", American Economic Review, 80, 1042-1061.

Laffont, J. and J. Tirole (1991) "The politics of government decision making: a theory of regulatory capture", Quarterly Journal of Economics, 106, 1089-1127.

Laffont, J. and J. Tirole (1993) A THEORY OF INCENTIVES IN PROCUREMENT AND REGULATION, MIT Press, Cambridge.

Laffont, J. and J. Tirole (1994a) "Access pricing and competition", European Economic Review, 38, 1673-1710.

Laffont, J. and J. Tirole (1994b) "Creating competition through interconnection: theory and practice", paper presented at the conference on "New developments in access pricing for network utilities", London, May 24, 1995.

Lewis, T. and D. Sappington (1988) "Regulating a monopolist with unknown demand", American Economic Review, 78, 986-998.

Littlechild, S. (1983) REGULATION OF BRITISH TELECOMMUNICATIONS' PROFITABILITY, Department of Industry, London.

Lockwood, B. (1995) "Multi-firm regulation without lump-sum taxes", Journal of Public Economics, 56, 31-53.

Loeb, M. and W. Magat (1979) "A decentralized method of utility regulation", Journal of Law and Economics, 22, 399-404.

Mankiw, G. and M. Whinston (1986) "Free entry and social inefficiency", Rand Journal of Economics, 17, 48-58.

Maskin, E. and J. Riley (1984) "Monopoly with incomplete information", Rand Journal of Economics, 15, 171-208.

Maskin, E. and J. Riley (1990) "Optimal multi-unit auctions" in Hahn, F. THE ECONOMICS OF MISSING MARKETS AND INFORMATION, Oxford University Press, Oxford, 312-33.

McAfee, R.; Preston, and J. McMillan (1994) "Auctions and bidding", Journal of Economic Literature, 25, 699-738.

McGuire, T and M. Riordan (1995) "Incomplete information and optimal market structure: public purchases from private providers", Journal of Public Economics, .

McMillan, J. (1994) "Selling spectrum rights", Journal of Economic Perspectives, 8, 145-162.

Meyer, M. and J. Vickers (1995) "Performance comparisons and dynamic incentives", CEPR, Discussion Paper n. 1107.

Mirrlees, J. (1971) "An exploration in the theory of optimal income taxation", Review of Economic Studies, 38, 175-208.

Mookherjee, D. (1984) "Optimal incentive schemes with many agents", Review of Economic Studies, 51, 433-466.

Mussa, M. and S. Rosen (1978) "Monopoly and product quality", Journal of Economic Theory, 18, 301-317.

Myerson, R. (1981) "Optimal auction design", Mathematics of Operations Research, 6, 58-73.

Myerson, R. (1983) "Mechanism design by an informed principal", Econometrica, 47, 61-73.

Nalebuff, B. and J. Stiglitz (1983) "Prizes and incentives: towards a general theory of compensation and competition", Bell Journal of Economics, 14, 21-43.

Oftel (1993) INTERCONNECTION AND ACCOUNTING SEPARATION, Oftel, London.

Oftel (1995) ANNUAL REPORT, London, HMSO.

Oren, S.; Smith, S. and R. Wilson (1983) "Competitive non-linear tariffs", Journal of Economic Theory, 29, 49-71.

Phlips, L. (1983) THE ECONOMICS OF PRICE DISCRIMINATION, Cambridge University Press, Cambridge.

Pint, E. (1991) "Nationalization vs. regulation of monopolies: the effect of ownership on efficiency", Journal of Public Economics, 44, 131-164.

Riordan, M. and D. Sappington (1987) "Awarding monopoly franchises", American Economic Review, 77, 375-87.

Rob, R. (1986) "The design of procurement contracts", American Economic Review, 76, 378-89.

Sappington, D. (1980) "Strategic firm behavior under a dynamic regulator adjustment process", Bell Journal of Economics, 11, 360-72.

Sappington, D. (1991) "Incentives in principal-agent relationships", Journal of Economic Perspectives, 5, 45-66.

Sappington, D. and D. Sibley (1992) "Strategic nonlinear pricing under price-cap regulation", Rand Journal of Economics, 23, 1-19.

Schwartz, M. (1989) "Investments in oligopoly: welfare effects and tests for predation", Oxford Economic Papers, 41, 698-719.

Sherman, R. and M. Visscher (1982) "Rate of return regulation and two-part tariffs" Quarterly Journal of Economics, 97, 26-41.

Shleifer, A. (1985) "A theory of yardstick competition", Rand Journal of Economics, 16, 319-27.

Sibley, D. (1989) "Asymmetric information, incentives and price-cap regulation", Rand Journal of Economics, 20, 392-404.

Spence, A. (1981) "The learning curve and competition", Bell Journal of Economics, 12, 49-70.

Spence, M. (1980) "Multi-product quantity-dependent prices and profitability constraints", Review of Economic Studies, 47, 821-841.

Tirole, J. (1986) "Hierarchies and bureaucracies: on the role of collusion in organizations", Journal of Law, Economics and Organization, 2, 181-214.

Tirole, J. (1988) THE THEORY OF INDUSTRIAL ORGANIZATION, MIT Press, Cambridge.

Tye, M. (1984) "Some subtle pricing issue in railroad regulation: Comment", International Journal of Transport Economics, 11, 2-3, 207-16.

Varian, H. (1989) "Price discrimination" in Schmalensee, R. and R. Willig (eds.) HANDBOOK OF INDUSTRIAL ORGANIZATION, North Holland, Amsterdam.

Vagliasindi, M. (1994) NON-LINEAR PRICING, REGULATION AND COMPETITION, M.Phil Thesis, Oxford.

Vagliasindi, M. and M. Waterson (1995a) "Access and cream skimming in network industries. A note on agency problems and competitive issues in the theory of regulation", Warwick Economic Discussion Paper, 9508.

Vagliasindi, M. and M. Waterson (1995b) "New insights on the interactions between regulation and competition in vertically related markets", Warwick Economic Research Paper, University of Warwick, 438.

Vickers, J. (1995) "Competition and regulation in vertically related markets", Review of Economic Studies, 62, 1-17.

Vickers, J. and G. Yarrow (1988) PRIVATIZATION, MIT Press, Cambridge.

Vickrey, W. (1961) "Counterspeculation, auctions and competitive sealed tenders", Journal of Finance, 16, 8-37.

Vogelsang, I. (1990) PUBLIC ENTERPRISE IN MONOPOLISTIC AND OLIGOPOLISTIC INDUSTRIES, Chur, Harwood Academic Publisher.

Vogelsang and Finsinger (1979) "A regulatory adjustment process for optimal pricing by multiproduct monopoly firms", Bell Journal of Economics, 10, 157-71.

Waterson, M. (1984) ECONOMIC THEORY OF THE INDUSTRY, Cambridge University Press, Cambridge.

Waterson, M. (1988) REGULATION OF THE FIRM AND NATURAL MONOPOLY, Basil Blackwell, Oxford.

Waterson, M. (1992) "A comparative methods of regulating public utilities", Metroeconomica, 43, 205-26.

Waterson, M. (1994) "The future for utility regulation", in Corry, D.; Souter, D. and M. Waterson (eds.), REGULATING OUR UTILITIES, Institute for Public Policy Research, 104-30.

Willig, R. (1978) "Pareto superior nonlinear outlay schedules", Bell Journal of Economics, 9, 56-9.

Willig, R. (1979) "The theory of network access pricing" in Trebing, H. (ed) ISSUES IN PUBLIC UTILITY REGULATION, Michigan State University, East Lansing.

Wilson, R. (1993) NONLINEAR PRICING, Oxford University Press, New York.