Proceedings of the 2006 Telecom Paris Conference on the Economics of Information and Communication Technologies

ENST Paris, France — June 23-24, 2006

Editor : Patrick Waelbroeck (ENST)

Organized by :
Department of Economics and
Social Sciences of Telecom Paris

ENST 2006 S 003

École Nationale Supérieure des Télécommunications

Contents

Music Variety and Retail Concentration, *M-L. Allain and P. Waelbroeck* 1

Information Congestion, *S. P. Anderson and A. de Palma* 25

Piracy and Competition, *P. Belleflamme and P. M. Picard* 58

The Economics of Digital Assemblers, *E. Brousseau and T. Pénard* 95

Information Exchange and Competition in Communications Networks, *C. Cambini and T. M. Valletti* 112

Monopoly Practices and Competitive Behavior in the French Satellite Pay-TV Market, *M. Cincera and A. G. Noury* 158

Commercial or Open Source Software? Winner-Take-All Competition, Partial Adoption and Efficiency, *E. Darmon, T. Letexier and D. Torre* 184

Bundling in Vertically Differentiated Communication Markets, *T. Diallo* 209

Antitrust Policy, the Theory of Market Leaders and the New Economy: Comments, *D. Encaoua* 244

Antitrust Policy, the Theory of Market Leaders and the New Economy, *F. Etro* 251

Usage and Diffusion of Cellular Telephony: 1998-2004, *M. Grajek and T. Kretschmer* 286

Simulation of Merger in Mobile Telephony, *L. Grzybowski and P. Pereira* 316

The Virtual Location of E-Tailers: Evidence from a B2C E-Commerce Market, *J. Haering* 340

Multi-Sided Markets: Competing with Network Externalities and Price Discrimination, *B. Jullien* 363

Do Consumers Buy Less of a Taxed Good? *H. J. Kind, M. Koethenbuerger, G. Schjelderup* 405

The Determinants of e-business Strategies: Empirical Evidence from Luxemburg Data, *L. Martin* 425

Reexamining the Effects of the Most-Favored-Nation (MFN) Provision in Input Prices on R&D Incentives: The Effect of the MFN on Access Charge Agreements, *J. Y. Kim and J. Nahm* 472

Access Regulation and the Adoption of VoIP, *P. W. J. de Bijl and M. Peitz* 498

The Competitiveness of Upstream Markets with Vertically Integrated Firms, *M. Bourreau, J. Hombert, J. Pouyet and N. Schutz* 538

Diffusion of Electronic Business Technologies, K. Sülze 567

Are Digital Rights Valuable? Theory and Evidence from eBook Pricing, G. *Oestreicher-Singer and A. Sundararajan* 586

Two-sided Markets with intra-group Competition, *P. Belleflamme and E. Toulemonde* 625

Network Externalities and Interconnection Incentives, *M. A. Vanberg* 658

The Long-Run Impact of ICT, *F. Venturini* 697

Piracy on the Silver Screen, *R. Rob and J. Waldfogel* 739

Economic and Technical Drivers of Technology, Choice: Browsers, *T. F. Bresnahan and P-L. Yin* 771

Music Variety and Retail Concentration

Marie-Laure Allain*and Patrick Waelbroeck†

first draft-january 2006

Abstract

In this paper, we examine the impact of horizontal and vertical market structure on product variety. We consider a market for horizontally differentiated products in which the cost of launching a new product is fixed and spread between the manufacturing and the retail industries. While this framework can be applied to a large number of industries we focus on music variety. We show that a vertically intergrated firm offers a wider variety of products than a chain of monopolies. If the cost of launching a new product is equally shared among the vertical structure or mostly supported by upstream firms, retail competition partially restores the incentives to innovate of the vertical structure. Yet when the cost of launching a new product is mostly supported by the retail sector, downstream competition leads to even more innovation than vertical integration.

1 Introduction

Over the recent years, sales of CDs have decreased by as much as 20 percent in some countries. According to the International Federation of the Phonographic Industry's statistics, world sales of recorded music (audio and video) for the year 2003 fell by 7.6 percent in value, marking a fourth consecutive year of decline. The global music market has thus declined by US 6.2 billion since 1999, a fall of 16.3 percent in constant dollar terms. The reasons for this fall are widely discussed among the actors of the

*CNRS, Laboratoire d'Econométrie de l'Ecole Polytechnique (1 rue Descartes 75 005 Paris, France, E-Mail : marie-laure.allain@shs.polytechnique.fr) and London Business School.

†ENST, Rue Barrault, Paris.

industry. One of the most common culprits pointed at by the record companies is internet or file-sharing piracy. Using peer-to-peer network, internet users can freely and anonymously exchange copyrighted files without authorization of the copyright owners. This has lead to legal measures against peer-to-peer network developers and users as well as increased technological protection of digital music using Digital Rights Management technologies. Whether or not such measures will restore CD sales to their 1999 level is questionable.

Another trend that is pointed at by some industry observers but overlooked by most analysts is the decrease in the number of new releases, and the possible role of the retail structure in this decrease in music variety. The Record Industry Association of America has strangely stopped producing these numbers after 2000. The number of new releases reached an all-time high in 1999 with 38900 new releases according to Soundscan to drop to 31734 new releases in 2001. The Canadian Recording Industry Association continues to produce these numbers: the number of new releases decreased from 6728 in 1998 to 5619 in 2003. Similarly, in Australia the four major record labels (Sony BMG, EMI, Universal and Warner) released 2906 albums in 2004 compared to 4480 in 2003, a 35 percent drop in 1 year, continuing the downward trend.[1] This recent decrease in music variety offered to consumers should be paralleled to two factors: the increasing concentration of the retail segment with the emergence of large superstore selling to the masses and their increasing share in total music sales to consumers.

Secondly, we observe in several OECD countries that an increasing share of music sales occurs in large department stores or supermarket chains. In the US, share of music sold in department stores and hardware and Audio/Video stores has increased from 26 percent in 1999 to 39 percent in 2003.[2] Walmart accounted for 13.5 of music sales in 1998. Its share has increased to 34.8 percent in 2003.[3] A similar trend is observed in the UK where share of music sales in supermarkets and big chains has increased from 29 percent in 1999 to 36 percent in 2003. In France, supermarket chains sell 38.5 percent of the CD in 2003, compared to 34.4 percent ten years before, and 37.7 percent are sold by large disc store chains (compared to 29 percent in 1993) owned by large retail groups.[4]

[1] See "CD sales fall disguises a lack of choices" in The Register, September 15, 2005

[2] See IFPI (2004)

[3] See "Big stores make exclusive deals to bring in music buyers" in the New York Times, December 29, 2003

[4] In 2003, these large disc store chains sales were splitted as follows : 56 percent for FNAC (part of the Pinault-Printemps-La Redoute retail group), 20 percent Virgin, 11 percent Starter, and 10

Some analysts of the music industry have associated these two factors with the recent turmoil in sales of pre-recorded music, citing high turnover of titles due to competition for shelf spaces of other entertainment products such as DVDs, discount prices, availability of only most popular titles on shelves, reducing variety to consumers. For instance, according to the music and video director of Carrefour, the music variety offered on the shelves was reduced by 30 percent over a couple of years, with the offer mainly concentrated on new releases and hit titles.[5]

This phenomenon is not limited to the music sector. The concentration of the retail industry is nowadays an important policy issue in most developed countries. Since the seventies, the emergence of new store formats and the development of large and increasingly international retail chains, through diversification and external growth, have considerably modified the retail landscape. The increasing concentration of the retail industry has resulted in an oligopolistic structure in most European countries: the 5 main retail chains control about 65% of the food sales in the UK, 80% in France, 65% in Germany, 56% in Spain and up to 98.5% in Norway. Large mergers among retail groups have occurred in the nineties. The 2000 merger between Carrefour and Promodès has given birth to the second largest worldwide retail group with sales above 70 billions euro. In 1999, the American giant Wal Mart acquired the British supermarket chain ASDA, but the same year, the European Commission set restrictive conditions to the merger of the German retailers REWE and Meinl. The Commission even prohibited merger of the Finnish retail groups Kesko and Tuko in 1996. The retail industry may be compared to the bottleneck of an hourglass, controlling the links between numerous manufacturers of consumer goods and their consumers. This trend towards increasing retail concentration leads to increased buying power from the retailers, and thus induces a shift in the balance of power between retailers and their suppliers, which has generated many conflicts and may even induce structural changes in vertically related industries.

Public authorities have debated over the last years issues related to the bargaining power between producers and retailers in order to assess the economic consequences of increasing retail concentration that go beyond the music industry (see, for instance, the green book on vertical restraints published by the European commission in 1997 or the official report by the British Office of Fair Trading of 1999). Broadly speaking, arguments against large retailers relate to the fact that a marked disequilibrium in the

percent for the specialised stores lauched by the supermarket group Leclerc (source : SNEP).

[5]See Le Monde , January, 22, 2005

bargaining power between suppliers and retailers can be detrimental to the survival of small producers and especially to the variety of products available to the consumers. Both retailers and suppliers consider the breadth of the product line as a crucial point in the bargaining process. On the one hand, producers implement innovation strategies that segment the market in order to discriminate between different types of consumers. They thus tend to extend their product lines. On the other hand, retailers fear brand proliferation as it increases costs associated with inventory control and involves increasing marketing and promotion expenses. As a consequence, retailers often impose contracts that limit the number of products available on the shelves. Hence the economic incentives to produce and distribute a new variety differ according to the side of the vertical relation. While producers expect their new products to increasing demand by building new niches, retailers fear market segmentation that increases distribution costs (Marvel and Peck, 2000). Upstream and downstream firms' goals can thus diverge and harsh commercial bargaining talks reflect these divergences.

In order to understand the economic forces behind the conflict, we analyze the incentives to supply variety in a vertically related industry between innovating producers, retailers and consumers. We determine the effect of the vertical relation on social welfare by studying the incentives to increase product variety according to the competitiveness of the downstream sector. We focus on two main points: the degree of novelty, and the fixed distribution costs of the new product related to inventory control and promotion. It is well known that the costs of launching a new musical product are fixed, and are mainly related to recording, producing and pressing the CD on the one hand and to expensive promotion and marketing campaigns (see Peitz and Waelbroeck, 2005, section 2.1). We study a vertical relation between a monopolist producer and one or two potential retailers, in a framework related to the market structure of the music industry. The music label develops a new album that is costly to produce and that generates fixed distribution and promotion costs for the retailers. We model the upstream industry as a monopoly for several reasons. First, the music industry is very concentrated as four labels control more than 80% of the global market, which at an aggregate level is a lot more concentrated than the retail industry. Furthermore, the price of pre-recorded music is rather uniform across music labels. Finally, the music labels have been accused of price coordination several times. Recently, the music labels settled a law suit on this matter out of court in 2002 in the US. We analyze how the repartition of these costs between the producer and retailers influence the incentives to increase producrt variety.

While we focus on the music industry, the main issue is related to the increase in variety in vertical structure. Some empirical studies on different markets lead to similar conclusions. For instance, a recent survey of German food producers (Weiss and Wittkopp, 2005) highlights a negative relationship between the bargaining power of large retailers and the introduction of new products. However, this effect is reduced by the market power of producers. From an empirical perspective, the difficulty to correctly measure the cost of launching a new product and the strategic dimension of the issue certainly explain why there are few studies on this topic. From a theoretical perspective, such studies are also very rare: while there is a huge literature on the economic analysis of vertical relationships on the one hand, and on innovation and incentives to innovate on the other hand, there has been little work on the incentives to innovate in a vertically separated industry. Inderst and Shaffer (2003) analyze the effect of a horizontal merger between non-competing retailers (assuming for instance that they are operating on two distinct geographical areas) on the variety offered to consumers. They show that after the merger and in order to improve their bargaining power with the producers, retailers might have to remove some products from the shelves: making their product lines more uniform would enable them to get better bargaining terms with their suppliers. In this case, an increase in the bargaining of the retailers leads to a decrease in product variety. However, rather than looking at product line simplification, we address the question of launching new products when it incurs specific costs; furthermore, we consider retailers who compete on the downstream market, but who do not bilaterally negotiate with their suppliers but interact on an upstream market. A related literature deals with quality improvement in vertical structures. For instance, Economides (1999) shows that vertical separation might lead to lower quality on the market. We do not consider a single-product market but we endogenize the decision of increasing product range.

Our article contributes to the literature on three points. First, we show that a vertically integrated structure better internalizes the fixed costs of production than a chain of separated monopolies and offer a larger variety of products. This first result rests on a classical reduction of the inefficiency related to the double margin that limits the introduction of new products. Next, we show that a situation in which retailers are competing gives more incentives to extend the product line than a chain of monopolies, mainly by reducing vertical inefficiencies. Finally, we show that competition between retailers can surprisingly lead to more product variety than a vertically integrated structure when the cost of launching the new product is mainly supported by the upstream firm. This result stems from the fact that the producer might strategically reduce competition between retailers by charging them retail prices that force them to specialize. In this case, one of the retailers specializes in the new product, while the other only sell the old product. By reducing competition in such a way, the producer can reduce the costs associated with the new product, where a vertically integrated structure would not have increased variety. In addition, we show that this increase in product variety is welfare-enhancing.

The remainder of the article is organized as follows. First, we study the decision of a producer who sells his products through a vertically separated chain of monopolies in Section 2. Next, we analyze in section 3 a situation where the same producer faces two retailers who compete for his products. The last section concludes.

2 The model

We consider a vertical relation between a producer P (the music label) who sells her products (albums) through a retailer D. We assume that the producer is unable to set up shop and sell independently. She initially produces product A at a constant marginal cost k_A that we normalise to zero without loss of generality. The producer can invest a fixed cost Ω to extent its product line and then produces also a substitute product B with a constant marginal cost $k_B = k_A = 0$. Products A and B are horizontally differentiated: we consider here product novelty as a mean to increase product variety and not to improve product quality. We consider a linear inverse demand function translating consumers' taste for variety, with P_I the price of one unit of product I, and q_I the quantity of product I on the market ($\{I, J\} = \{A, B\}$):

$$P_I(qI, q_J) = 1 - q_I - cq_J$$

Parameter c, that we assume to be in [0,1[, measures the substitution between the two goods. The retailer distributes the product A without cost. However if he decides to introduce the new product on the final market, he has to incur a fixed cost of distribution and inventory control, noted F. The retailer also faces a constant marginal cost of distribution independent of the type of product that is distributed, which we normalise to zero. The exogenous parameters of the models are: c, Ω, F.

We compare the producer's incentives to invest the fixed cost of increasing product variety Ω in two different cases: when the producer and the retailer are vertically integrated and when they are separated. ‘

2.1 The benchmark case: vertically integrated monopoly

We first consider as a benchmark a situation in which the producer and the retailer are vertically integrated. The integrated unit only extend its product line and sells both products if it is profitable to do so, regardless of the number of products distributed.

If the vertically integrated structure does not introduce a new variety, it only sells A in quantity $q_A = \frac{1}{2}$, and with profit $\Pi_A^{VI} = \frac{1}{4}$. As introducing the new product involves a fixed cost, the structure that increases product variety has to choose whether to continue selling product A in addition to product B or not (selling only B is a dominated strategy, as it would lead to a maximum profit of $\Pi_B^{VI} = \Pi_A^{VI} - \Omega - F$). If on the contrary the vertically integrated structure chooses to distribute both products, the profit-maximizing quantities are then $q_A = q_B = \frac{1}{2(1+c)}$, and the profit $\Pi_{A+B}^{IV} = \frac{1}{2(1+c)} - F - \Omega$.

The vertically integrated monopoly thus increases variety and sells both products if and only if it leads to higher profits than through selling only the old product, which is equivalent to the following condition:

$$F + \Omega \leq \frac{1-c}{4(1+c)}$$

Thus the new product is profitable to market as long as the total fixed cost of producing and retailing the new good are not above a threshold level that decreases with the substitutability between products: as c tends to 1 and products become more substitutable, the firm is less likely to introduce the new product as its profits become smaller.

In a vertically integrated structure, this classical cannibalisation effect is driving product line decisions.

2.2 Product variety in a vertically separated chain

We now study how the investment decisions of a vertically separated industry depend on how the total fixed cost is shared between the producer and the retailer. When the producer and the retailer are vertically separated, the product line decisions are taken sequentially. Formally, the producer and the retailers play the following game: in the first stage, the producer decides whether to increase her product line and accordingly spends the fixed cost Ω. Then she sets the two wholesale prices w_A and w_B , each in[6] $[0,1]$. In the second stage, the retailer decides which products to sell to the consumers (and whether or not to pay the fixed cost F) and which quantities q_A and q_B to order. The last stage is consumption in the downstream market. We are looking for the subgame perfect equilibria of this game. Vertical separation, inducing a double margin externality, modifies the incentives to increase product variety for the producer.

2.2.1 Downstream listing and pricing strategy

In the second stage, the retailer chooses his listing strategy given the wholesale prices charged by the producer. If, on the one hand, he decides to only distribute the old product , he orders the optimal quantity $q_A = \frac{1-w_A}{2}$ and gets a profit $\Pi^D = \frac{(1-w_A)^2}{4}$ while the producer gets $\Pi^P = \frac{w_A(1-w_A)}{2}$ possibly less the fixed cost Ω. If, on the other hand, the retailer decides to distribute product B only, provided that the upstream firm has introduced the new product, he has nevertheless to pay the fixed cost F. He then maximizes his profit by ordering the quantity $q_B = \frac{1-w_B}{2}$ and makes a profit of $\Pi^D_B = \frac{(1-w_B)^2}{4} - F$. Finally, if the retailer chooses to distribute both products, he orders quantities $q_I = Max\{0, \frac{1-w_I-c(1-w_J)}{2(1-c^2)}\}$ (with $\{I,J\} = \{A,B\}$).

Given wholesale prices w_A and w_B, the retailer determines his listing strategy by comparing his profits with or without the new product. Regardless of w_A and w_B, the retailer always prefers to distribute both goods instead of only product B: the strategy of selling the new product only is dominated by the strategy of selling both products. In addition, if $w_B \geq 1 - c + cw_A$, the retailer would make losses if he sold both products, in which case he prefers to save on the fixed cost F and sell product A only. In the other cases, the optimal listing strategy depends on the fixed cost F. Finally, the retailer distributes the new product only if the fixed distribution cost F is smaller than a threshold level that decreases with wholesale price w_B:

$$F \leq \frac{(1-w_A)^2 + (1-w_B)^2 - 2c(1-w_A)(1-w_B)}{4(1-c^2)} - \frac{(1-w_A)^2}{4} \tag{1}$$

The following figure illustrates the listing choice of the retailer in the (w_B, F) plane for a given value of w_A.

[6] Any wholesale price above 1 would lead to a zero demand, and would thus not be rational, so that we can make this assumption without loss of generality.

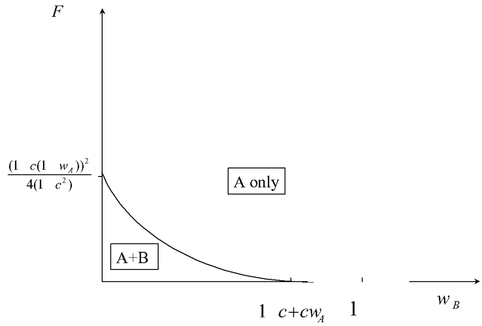


Figure 1: listing strategies of the retailer

2.2.2 Upstream strategy

In the first stage, the producer decides whether to produce the new product and sets the wholesale prices anticipating the outcome of the second stage. If she only produces the old product, she sets a wholesale price of $w_A = \frac{1}{2}$ that correponds to a maximal profit of $\Pi_A^P = \frac{1}{8}$. If on the contrary she innovates, she has to make sure that the retailer will list the new product as she would make at most $\Pi_B^P = \frac{1}{8} - \Omega$ otherwise. She then sets the two wholesale prices in order to maximize her profit under the constraint (1), which guarantees that the retailer will list both products. The only interior solution is $w_A^* = w_B^* = \frac{1}{2}$ as long as $F \leq \frac{1-c}{16(1+c)}$. For higher values of the fixed cost of distributing the new product, the producer has to adopt a limit-pricing strategy that induces the retailer to sell both products. The corner solution is to set a price $\widetilde{w}_A = \frac{1}{2}$ for the old product and $\widetilde{w}_B = 1 - \frac{c}{2} - \sqrt{4F(1-c^2)}$ for the new one. Finally, the comparison of the expected profits in each case determines the optimal strategy of the producer in the first stage (see appendix 1 for the details). Figure 2 compares the resulting equilibria with the corresponding solution under the vertically integrated structure. The necessary condition under which a chain of monopolies increases variety is more binding than the corresponding condition for a vertically integrated structure.

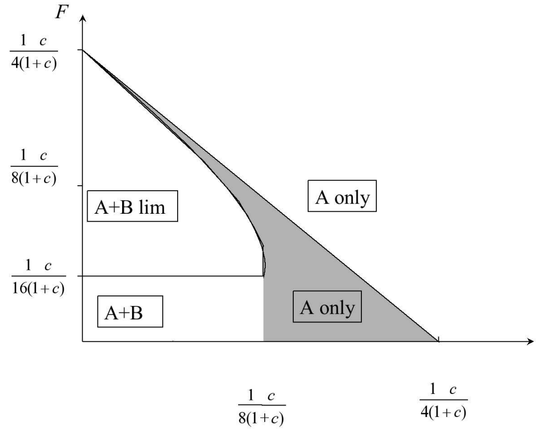


Figure 2

.

The grey area indicates the values of fixed costs for which a separated chain of monopolies would not innovate even though an integrated firm would.

Proposition 1 *Vertical separation in a chain of monopolies can reduce product variety.*

In other words, an integrated structure has better incentives to distribute a new product than a separated structure. This results from the double margin externality: the standard issue of coordination in a non-integrated vertical relation generates a new form of inefficiency by reducing the profitability of the new product.

Notice that in this simple case, a two-part tariff associated to a tying contract, or a two-part tariff with a fixed fee independent of the range of products sold by the retailer, would be sufficient to restore the incentives: when the new product increases total profits, the upstream firm can set wholesale prices equal to the marginal costs (here, zero) and get the whole profit[7] through the fixed fee. However, if the two goods have to be priced seperately, with two-part tariffs w_A, F_A and w_B, F_B, the upstream

[7] This point relies on the assumption that the producer has all bargaining power, and is only to enable a comparison with the classical principal-agent literature on double margin. Of course, this assumption would be unrealistic in most industries, including the music sector.

producer is not able to get all the profit anymore because she has to leave a rent to the retailer in order to have him selling the two goods. The producer has to give the retailer an incentive to list both products rather than only one of them, which requires the following incentives constraints to be satisfied, where Π_D^I is the variable part of the retailer's profit (excluding the payment of the fixed costs) : $\Pi_D^{A+B} - F_A - F_B \geq \Pi_D^A - F_A$ and $\Pi_D^{A+B} - F_A - F_B \geq \Pi_D^B - F_B$, which implies that $2\Pi_D^{A+B} - \Pi_D^A - \Pi_D^B \geq F_A + F_B$. Yet $\Pi_D^{A+B} < \Pi_D^A + \Pi_D^B$ because the products are substitutes and thus $F_A + F_B < \Pi_D^{A+B}$: finally, even if the producer can delegate the optimal choices to the retailers by setting variable prices to her marginal costs of production, she cannot get the whole profit of the vertical structure through the fixed part $F_A + F_B$.

Furthermore, it is interesting to observe that the incentives to distribute the new product are more sensitive to the fixed cost of production Ω than to the fixed cost distribution F. Indeed, when the latter is high, the producer can adapt its wholesale price by setting a limit price that leads the retailer to distribute both products. On the contrary, when the fixed cost of production is high, the retailer can not commit to share the cost spent by its vertical supplier who unilaterally decides not to introduce the new variety.

3 Competition between retailers

We have seen in the first part that vertical separation of the activities of production and distribution can reduce product variety. However, it is well known that downstream competition reduces double margin problems: we adress here the question of how competition between retailers can affect product variety, when variety brings about fixed costs at both levels. We thus analyze the effect of imperfect competition between two retailers on the incentives of an upstream firm to introduce a new variety. We analyze the following situation: two retailers D_1 and D_2 sell producer P 's production to the consumers. The 3-stage game is as follows. In the first stage the producer decides whether to innovate or not and sets the wholesale prices. In the second stage, the retailers simultaneously decide whether to invest the fixed cost to be able to sell the new product. This cost is sunk. In a third stage, as the outcome of the investment decisions are made public, the retailers simultaneously order the quantities of the two goods they are going to put on their shelves, and the prices on the final market are determined by the consumer inverse demand. Retail competition is thus à la Cournot. The fixed cost F is sunk and represents a commitment[8] of the retailers on their listing choices: if a retailer does not pay F in the second stage, he will not be able to sell the new product in stage 3. We solve this game for subgame perfect equilibria.

3.1 Downstream quantity competition

In this section we determine the equilibrium outcome of downstream competition, given wholesale prices (w_A, w_B) and the investment decisions of the second stage. We assume that wholesale prices are smaller than 1, a necessary condition for products to be profitable to market. At the third stage of the game, retailers are already commited to their listing strategies, and there are three different subgames to analyze (plus the symmetric ones): either both retailers have invested the sunk cost F, or one only, or none of them.

3.1.1 No retailer has invested

In this first subsection, only one good is distributed: A. Downstream competition is thus a simple monoproduct Cournot game. There exists a unique equilibrium where the two retailers sell the same quantity of the old product A: $q_A^1 = q_A^2 = \frac{1-w_A}{3}$. Both retailers make $\Pi_A^D = \frac{(1-w_A)^2}{9}$.

3.1.2 Both retailers have paid the fixed cost

In this configuration, each retailer chooses two quantities (possibly setting them to zero). Solving the Cournot game leads to the following strategies according to the values of the wholesale prices (technical details are given in the Appendix). If the wholesale price of good B is too high, only good A is distributed. On the contrary, for small values of w_B, only the new good is distributed. Finally, there exists an equilibrium in which both goods coexist on the shelves for intermediary values of w_B. In addition, the set of values of w_B for which both products are distributed shrinks with c, the degree of substitutability of the two products: the lower bound on w_B below which the retailers only distribute B increases with c, while the upper bound

[8] In a previous version of this article, we solved the game without this commitment effect of the sunk cost, considering that stages 2 and 3 were simultaneous. This led to more equilibria: for a given configuration of retail costs, several equilibria existed. However, our results were qualitatively similar.

above which the retailers only sell the old good decreases with c. Indeed, for high values of c, products are highly substitutable and compete for shelf space, in which case the retailers prefer to only distribute the most profitable good. We also show that the same set of values of w_B shrinks with w_A. However, now, both the upper and the lower bound of the interval shift to the right as w_A increases. This shift translates the fact that the profitability of A decreases with w_A regardless of whether product B is also distributed or not. We should also point out that in this subgame, none of the asymmetric market configurations arises at equilibrium, although they were *a priori* possible.

3.1.3 Asymmetric configuration: one retailer only has paid the sunk cost

In this subgame, one of the retailers can only sell product A. We refer to this retailer as retailer 2. The other retailer chooses his listing strategy. We completely solve the downstream Cournot subgame in the Appendix. There are 4 configurations to analyze according to the values of w_B. Only good A is distributed if the wholesale price of B is too high, and this threshold is identical to the one found in the previous subsection. For values of w_B slightly below this threshold, both goods are distributed by the retailer who has invested the fixed cost of distributing the new product. For even smaller values of w_B, this retailer only distributes good B while his competitor is constrained to sell only good A. Finally, for very small values of w_B, the retailer who did not spend the fixed cost to distribute the new good must exit the market, leaving his competitor in a monopoly situation in the market for good B. Notice that in that case, retailer 1 still leaves good A out of the shelves to avoid cannibalization[9] of sales of good B. As in the previous subsection, the set of values of w_B for which the new product is distributed shifts to the right as w_A increases and the size of the interval decreases as parameter c increases.

It is interesting to observe that asymetric equilibria with downstream specialization are due to the commitment value of the sunk cost F. Indeed, in a Cournot game without this commitment effect, the retailers do not have incentives to specialize: a retailer who did not pay the fixed cost could always deviate from the equilibrium strategy by reducing the quantities of A on the shelves and by offering a small but positive quantity of B. When the fixed cost F is sunk, the retailer who has paid it

[9] Listing product A and selling a zero quantity of this good does not raise a credibility issue as there is no listing cost associated with the old product, so there is no commitment at the listing stage.

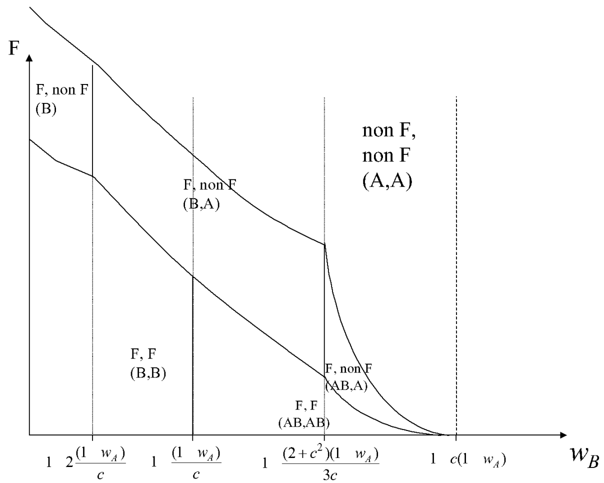
knows that, at the last stage of the game, his competitor can not sell good B. Under this assumption, for small values of w_B, distributing good A will only cannibalize sales from good B and this retailer preferes to leave his competitor in a monopoly position on the market for good A, while enjoying a monopoly position on the market for good B.

We can now analyze the investment decisions of the retailers at the second stage of the game.

3.2 Investment decisions

This stage of the game is only played if the producer has developed the new product. Retailers have to choose whether to invest the fixed cost or not in order to distribute the new product. They take wholesale prices w_A and w_B as given and anticipate downstream market outcomes.

There are five market configurations in this subgame. In the symmetric equilibria each retailer only sells the new good, or only the old one, or both. In the first asymmetric configuration, each retailer specializes in only one good. In the second asymmetric case, one retailer only sells the old good and his competitor sells both goods. The following figure summarizes these configurations, which are detailed in appendix 2.



For given values of the wholesale prices, equilibria in which good B is sold disappear as the fixed cost of marketing the new product increases. Moreover, the higher the value of w_B the lower the profits generated by sales of B. These results confirm the intuition that for low values of the wholesale prices and of the fixed costs, both retailers invest to distribute good B, while for high values of F and w_B, the total cost of distributing the new product is too high and both retailers symmetrically choose to stick to the old product. The commitment value of the sunk cost F has an interesting implication: for intermediate values of F, retailers adopt a "specialization" strategy that is characterized by the fact that only one retailer invests in the distribution of the new good (possibly together with product A) and the other retailer only distributes the old product.

3.3 Product line decision

In the first stage of the game, the producer decides whether to introduce the new variety and determines the wholesale prices. She anticipates the strategies of the retailers in stage 2 and sets her product lines and the wholesale prices in order to maximize her profits. The subgame perfect equilibrium outcomes are detailed in Appendix A4 where we also compare profits of the different players with those obtained when we studied the chain of monopolies. The main results are summarized in proposition 2.

Proposition 2 *A producer facing a competitive downstream market increases product variety more often than if she faced a single retailer.*

Proof : see appendix A4. ∎

More precisely, when the fixed costs of introducing the new variety are such that the chain of monopolies extend its product line, a producer who faces a competitive downstream market also increases product variety. However, there are parameter configurations in which the chain of monopolies does not introduce the new product whereas the downstream competition leads to product variety. This situation occurs when one of the fixed costs is large and the other is small (see Figure 4). In the first area, the total cost of introducing the new variety is mainly supported by the producer . When the downstream fixed cost is relatively small ($F \leq \frac{1-c}{36(1+c)}$ and $\Omega \in \left[\frac{1-c}{8(1+c)}, \frac{1-c}{6(1+c)}\right]$), a chain of monopolies does not introduce the new product, while downstream competition allows the producer to charge wholesale prices that are below the unconstrained optimum $(w_A = w_B = 1/2)$. Downstream competition increases the quantities of both goods sold by the producer who can then bear a larger fixed cost of product introduction than when she faces a single retailer. For larger values of F, the producer reduces the wholesale price of the new product to give incentives to the retailer to distribute it. This limit-price strategy is profitable as long as the the fixed cost is not too large and as retailers keep distributing the new product, i.e. until $F = \frac{1-c}{16(1+c)}$. In the area being discussed, the competitive downstream market leads to more product variety mainly because competition reduces the double margin externalities, which makes the new product more profitable to introduce for the producer.

For intermediate values of the fixed costs, downstream competition does not lead to more product introduction than the chain of monopolies: the area in which the new product is marketed is the same under the two structures. Indeed, competition between retailers reduce profits and make it harder to support the fixed costs. As a matter of fact, in this area, only one retailer distributes the new product, while both retailers keep distributing the old product. The quantity of good B sold under this configuration is the same as in the monopoly case; the profits generated by sales of good B are also identical. It would be too costly for the producer to charge wholesale prices that give more incentives to the retailers to distribute the new product, as the producer also faces a fixed cost of product introduction. Thus, the producer facing a competitive downstream market has the same incentives to increase product variety as when she only faces a single retailer.

On the contrary, as the fixed distribution cost increases even more and the share of the total cost supported by the producer shrinks, downstream competition leads to a new area where the competitive structure introduce more variety than the chain of monopolies. This area only exists when products A and B are rather close substitutes (for $c \geq 1/2$). In this case, for F in the interval $[\frac{1-c^2}{36c^2}, \overline{F}]$ where $\overline{F} \geq \frac{1-c}{4(1+c)}$, product B is distributed when Ω is relatively small. This area is larger under donwnstream competition than with the chain of monopolies. Indeed, the fixed cost of the producer being small, she can afford a lower wholesale price w_B that leads one of the retailers to distribute the new good. This product line extension increases total demand. In this area, the commitment value of the fixed F analyzed in the second stage of the game leads the the retailers to specialize: each retailer sells only one of the goods and has a monopoly position on its market. This market configuration leads to a paradoxical outcome. When fixed cost F is large, retailers specialize in the distribution of only one good, which increases the profitability of the new product but limit competition between retailers. It is worth stressing that the strategy of specialization is only feasible when both products are relatively close substitutes, i.e. when products are competing for shelf space. This implies that the retailer who chooses to distribute the new product gives up the old product to avoid cannibalization.

To summarize, downstream competition increases product variety through two mechanisms: a classical mechanism related to a reduction in the vertical externality and a strategic mechanism related to the specialization of the retailers, which is conditioned by the commitment value of the fixed cost of marketing the new product. We can now compare the incentives to increase product variety when retailers are competing to the incentives of a vertically integrated structure.

Proposition 3 *If the two goods are poor substitutes ($c \leq 1/2$), a producer selling its products through a competitive downstream sector introduces less variety than a vertically integrated monopoly;*
If the two goods are close substitutes ($c \geq 1/2$), a competitive retail industry innovates less than a vertically integrated monopoly except when the share of the total fixed cost of introducing the new variety supported by the producer is small ($F >> \Omega$).

Proof: see appendix A.5. ■

We illustrate Proposition 2 and 3 in the following figure in the (Ω, F) plane (for $c \geq 1/2$).

Even if downstream firms are competing, the vertical externality related to the double margin remains and lowers the incentives of the producer to extend its product line. This effect dominates when the distribution cost (F) is low. In this case, the vertically separated structure innovates less than a vertically integrated monopoly. However, an opposite vertical effect appears when the new product is less profitable to market (i.e. when F is large compared to Ω) and is a close substitute to the old product (i.e. c is large). Now, the upstream firms softens downstream competition by setting wholesale prices so as to enforce an asymmetric retail market in which one firm distributes the old product and another firm distributes the new product. Hence, specialized firms do not directly compete for the new product. This market environment can sustain more product variety than a vertically integrated firm: a vertically integrated firm would avoid introducing the new product, as it would cannibalize the sales of the old one.

Proposition 3 has several implications. First, from an empirical perspective, the strategy of the upstream firm of relaxing the competitive pressure in the downstream market is observed for a new product that is costly to distribute and that strongly substitutes to the older product. In this case, even if retailers specialize, downstream competition is relatively strong (at the second stage of the game).

Secondly, competition authorities do not generally frown upon vertical mergers because of the vertical externality. In our model, a vertical integration can have a negative effect on investment strategies if the new product is costly to market but relatively cheap to produce (a new and relatively unknown band for instance), whereas a vertically integrated structure innovates more when the innovation is costly to produce but not too costly to market (the new album of a superstar for instance).

Finally, total surplus (net of the fixed costs) is defined as $W(Q_A, Q_B; c) = Q_A + Q_B - \frac{1}{2}(Q_A^2 + Q_B^2) - cQ_AQ_B$. It is easy to show that the total welfare at equilibrium, $W^*(c) = W(Q_A^*(c), Q_B^*(c))$, is decreasing in c for $0 < c < 1$ in each product configuration. Moreover, for each equilibirum listing strategies, total surplus is higher under vertical integration followed by downstream competition and then monopoly. In the case in which the downstream retail sectors distribute the new product but the vertically integrated structure does not, total surplus is higher when there product variety is highest: competition increases social surplus by increasing the variety offered to consumers when the cost of launching the new product is mainly supported by the upstream firm and when the new product is a close substitute to the old one.

4 Conclusion

We have explained how retail concentration could reduce music variety that involves fixed costs of production and distribution. We have highlighted several mechanisms -both horizontal and vertical- behind this influence. First of all, the profit-cutting effect of double marginalization reduces the incentives to launch a new product. In a chain of monopolies, vertical integration increases the scope and the variety of products distributed to consumers with heterogeneous tastes. Thus, vertical separation of the production and the distribution activities may generate conflicts of interest between the vertically related firms, which translates into a shorter product line and leads to too few products distributed to the consumers. To restore the vertical efficiency, sophisticated contracts including full-line forcing clauses would be necessary. Downstream competition may however soften vertical inefficiencies. When we analyze a more complex framework with a producer releasing a new product and two competing retailers, the effect of competition on the incentives to increase product variety depends on the degree of substitution between the old and the new products and also from the repartition of the fixed costs between upstream and downstream firms. If manufacturing and retail activities are vertically separated, then downstream competition leads to more variety than does retail concentration. In addition, vertical separation with downstream competition may lead to more or less innovation than vertical integration, depending again on the allocation of fixed costs and on the degree of product substitution. When the retail costs are less than the manufacturing costs of launching the new product, retail competition, by reducing downstream profits, lessens the retailer's ability to invest in the fixed cost, and thus hinders the development of the new product. In that case, a vertically integrated firm would launch the new product more often than an upstream monopoly facing two competing retailers. On the contrary, when the new product is more costly to sell than to manufacture, a vertically separated structure with downstream competition may innovate more than a vertically integrated monopolist because retailers are ready to sell the new product even with high costs in order to segment the downstream market. In terms of policy implications, our model stresses the necessity to preserve competition at the retail level, especially in order to promote cultural diversity.

5 References

Allain, ML et C. Chambolle (2003) "Economie de la distribution", Repères, La Découverte.

Boone, J. (2001), "Intensity of competition and the incentive to innovate", *International Journal of Industrial Organization*, 19, 5, 705-726.

Brocas, I. (2003), "Vertical Integration and Incentives to Innovate", *International Journal of Industrial Organization,* 21 (4),457-488.

Bonanno, G. and J. Vickers (1988) "Vertical separation", *The Journal of Industrial Economics,* Vol. 36 N°3, 257-265.

Deloitte and Touche (1990) "Managing the Process of Introducing and Deleting Products in the Grocery and Drug Industry, Washington DC, Grocery Manufacturers of America.

Desiraju (2001) "New product introduction, slotting allowances and retailer discretion", *Journal of Retailing*, 77 335-358.

Dobson Consulting (1999) "Buyer Power and its Impact on Competition in the Food Retail Distribution Sector of the European Union", report for the European Commission, Brussels.

Economides, N. (1999) "Quality choice and vertical integration", *International Journal of Industrial Organization* 17, 903-914.

Inderst, R. and G. Shaffer (2003) "Retail Mergers, Buyer Power and Product Variety", forthcoming *Economic Journal.*

International Federation of the Phonographic Industry (2004) "The recording industry in numbers"

Larivière, M. and V. Padmanabhan (1997) "Slotting allowances and new product introductions", *Marketing Science*, 16 (2), 112-128.

Marvel, H. P. and J. Peck (2000) "Vertical Control, Retail Inventories and Product Variety", Working Paper n.00-09 , Ohio State University.

Motta, M. (2004) "Competition Policy : Theory and Practice", Cambridge University Press.

Peitz, M. and P. Waelbroeck (2005) "An Economist's Guide to Digital Music", *CESIfo Economic Studies*, 51, 363-432.

Rey, P. (1997) "Impact des accords verticaux entre producteurs et distributeurs", *Revue Française d'Economie*, Vol. 12, N. 2, 3-55.

Rey, P. and J. Tirole (1986) "The Logic of Vertical Restraints", *American Economic Review*, Vol. 76 N°5, pp. 921-939.

Spengler, J. (1950) "Vertical Integration and Antitrust Policy", *Journal of Political Economy*, 58, 347-352.

Tirole, J. (1988) "The Theory of Industrial Organization", MIT Press.

Weiss, C. and A. Wittkopp (2005) "Retailer concentration and product innovation in food manufacturing", *European Review of Agricultural Economics*, 32-2, 219-244.

A Appendix

A.1 Chain of monopolies

Retailer's strategy If only product A is available, the retailer orders $q_A = \frac{1-w_A}{2}$, gets profit $\Pi^D = \frac{(1-w_A)^2}{4}$,and the producer's profit is $\Pi^P = \frac{w_A(1-w_A)}{2}$.

If the retailer lists both products, his maximum profit is:

$\Pi_{A+B} = \frac{(1-w_A)^2+(1-w_B)^2-2c(1-w_A)(1-w_B)}{4(1-c^2)} - F$ and is attained for the following quantities $(\{I,J\} = \{A,B\})$:

$$q_I^{A+B} = Max\{0, \frac{1-w_I-c(1-w_J)}{2(1-c^2)}\}$$

Producer's strategy The comparison of her anticipated profits gives the producer's optimal strategy in the first stage:

-if $F \leq \frac{1-c}{16(1+c)}$ and $\Omega \leq \frac{1-c}{8(1+c)}$, she innovates, sets the optimal wholesale prices $w_A^* = w_B^* = \frac{1}{2}$ and gets the interior optimal profit $\Pi_{A+B}^P = \frac{1}{4(1+c)} - \Omega$.

-if $F \geq \frac{1-c}{16(1+c)}$ and $\Omega \leq \sqrt{F(\frac{1-c}{1+c})} - 2F$, she innovates, sets the optimal wholesale price $w_A^* = \frac{1}{2}$ and the limit-price $\widetilde{w}_B = 1 - \frac{c}{2} - \sqrt{4F(1-c^2)}$, and gets profit $\widetilde{\Pi}_{A+B}^P = \frac{1-16F}{8} + \frac{\sqrt{F(1-c^2)}}{1+c} - \Omega$.

-if $\Omega \geq \frac{1-c}{8(1+c)}$ or $F \geq \frac{1-c}{16(1+c)}$ and $\Omega \geq \sqrt{F(\frac{1-c}{1+c})} - 2F$, she does not innovate, sets $w_A = \frac{1}{2}$, and gets profit $\Pi_A^P = \frac{1}{8}$.

A.2 Downstream competition : third stage of the game

If both retailers have paid the fixed cost F, downstream Cournot equilibrium are as follows, for given values of the wholesale prices :

- if $w_B \geq 1 - c(1 - w_A)$, only A is sold.
- if $1 - c(1 - w_A) \geq w_B \geq 1 - (1 - w_A)/c$, each retailer sells both goods.
- if $1 - (1 - w_A)/c \geq w_B$, only B is sold.

If only one retailer, say 1, has paid the fixed cost F, downstream Cournot equilibrium are as follows :

- if $w_B \geq 1 - c(1 - w_A)$, only A is sold by both retailers.
- if $1 - c(1 - w_A) \geq w_B \geq 1 - \frac{(1-w_A)(2+c^2)}{3c}$, retailer 1 sells both goods in quantities $q_A^1 = \frac{(2+c^2)(1-w_A)-3c(1-w_B)}{6(1-c^2)}$, $q_B^1 = \frac{1-w_B-c(1-w_A)}{2(1-c^2)}$ and his competitor sells only good A in quantity $q_A^2 = \frac{1-w_A}{3}$.
- if $1 - \frac{(1-w_A)(2+c^2)}{3c} \geq w_B \geq 1 - \frac{2(1-w_A)}{c}$, the retailers specialise in a narrower range of products: retailer 1 sells only B and his competitor only A .
- if $1 - 2(1 - w_A)/c \geq w_B$, there exists a unique equilibrium where the retailer who did not invest the fixed cost exits the market (or sells a zero quantity of good A) while the other one enjoys a monopoly situation over the two goods, but chooses not to sell good A in order to avoid canibalisation of his sales of good B. Then he chooses to sell the monopoly quantity of the new product: $q_{B,\varnothing}^{1M} = \frac{1-w_B}{2}$.

A.3 Retailers' investment strategies

In the second stage, in the subgame where the producer has innovated, and given the wholesale prices w_A and w_B :

1) if $w_B \geq 1 - c(1-w_A)$:

both retailers decline to invest in the fixed cost, and in the following stage A will be the only product available.

2) If $1 - \frac{(1-w_A)(2+c^2)}{3c} \leq w_B \leq 1 - c(1-w_A)$,

if $F \leq \frac{(c(1-w_A)-(1-w_B))^2}{9(1-c^2)}$, both retailers pay F and sell both goods;

if $\frac{(c(1-w_A)-(1-w_B))^2}{9(1-c^2)} \leq F \leq \frac{(c(1-w_A)-(1-w_B))^2}{4(1-c^2)}$, only one retailer invests F to sell both goods, and his competitor sells only A;

if $F \geq \frac{(c(1-w_A)-(1-w_B))^2}{4(1-c^2)}$, both retailers give up the selling of the new product: none pays F, and both sell A.

3) If $1 - \frac{1-w_A}{c} \leq w_B \leq 1 - \frac{(1-w_A)(2+c^2)}{3c}$,

if $F \leq \frac{(1-w_A)^2-2c(1-w_A)(1-w_B)+(1-w_B)^2}{9(1-c^2)} - \frac{(2(1-w_A)-c(1-w_B))^2}{(4-c^2)^2}$, both retailers pay F and sell both products;

if $\frac{(1-w_A)^2-2c(1-w_A)(1-w_B)+(1-w_B)^2}{9(1-c^2)} - \frac{(2(1-w_A)-c(1-w_B))^2}{(4-c^2)^2} \leq F \leq \frac{(c(1-w_A)-2(1-w_B))^2}{(4-c^2)^2} - \frac{(1-w_A)^2}{9}$, only one retailer invests F to sell only B, and his competitor sells only A;

if $F \geq \frac{(c(1-w_A)-2(1-w_B))^2}{(4-c^2)^2} - \frac{(1-w_A)^2}{9}$, no retailer pays F, both sell only product A.

4) If $1 - \frac{2-2w_A}{c} \leq w_B \leq 1 - \frac{1-w_A}{c}$ (this zone may exist only if $w_A \geq 1-c$),

if $F \leq \frac{(1-w_B)^2}{9} - \frac{(2(1-w_A)-c(1-w_B))^2}{(4-c^2)^2}$, both retailers pay F and sell only B;

if $\frac{(1-w_B)^2}{9} - \frac{(2(1-w_A)-c(1-w_B))^2}{(4-c^2)^2} \leq F \leq \frac{(c(1-w_A)-2(1-w_B))^2}{(4-c^2)^2} - \frac{(1-w_A)^2}{9}$, only one retailer pays F to sell only B, his competitor sells only A;

if $F \geq \frac{(c(1-w_A)-2(1-w_B))^2}{(4-c^2)^2} - \frac{(1-w_A)^2}{9}$, no retailer pays F, both sell only A.

5) if $w_B \leq 1 - \frac{2-2w_A}{c}$ (this zone may exist only if $w_A \geq 1 - \frac{c}{2}$),

if $F \geq \frac{(1-w_B)^2}{4} - \frac{(1-w_A)^2}{9}$, no retailer pays F, both sell only A;

if $\frac{(1-w_B)^2}{9} \leq F \leq \frac{(1-w_B)^2}{4} - \frac{(1-w_A)^2}{9}$, only one retailer pays F, his competitor exits the market. The monopolist retailer sells only product B;

if $F \leq \frac{(1-w_B)^2}{9}$ both retailers pay F and sell only B.

A.4 Upstream choice: proof of proposition 2

In the first stage, the producer innovates if the profit she gets by selling the new product is higher than $\Pi^P_{AA} = \frac{1}{6}$, the profit she gets with product A only. If she innovates, her profit depends on the quantities sold by the retailers in stage 3. We summarize here the producer's optimal choices in equilibrium.

(i) If $F \leq \frac{1-c}{36(1+c)}$, she innovates if and only if $\Omega \leq \frac{1-c}{6(1+c)}$, and both retailers sell both goods in the interior optimum. For such values of F, the chain of monopolies would innovate only if $\Omega \leq \frac{1-c}{8(1+c)}$: downstream competition leads here to more innovation than a chain of monopolies would offer.

(ii) If $\frac{1-c}{36(1+c)} \leq F \leq \frac{1-c}{16(1+c)}$, the producer has to use a limit-pricing strategy in order to induce the two retailers to sell both goods each in equilibrium. The producer innovates if and only if $\Omega \leq 2\sqrt{\frac{F(1-c)}{1+c}} - 6F$, with $2\sqrt{\frac{F(1-c)}{1+c}} - 6F \geq \frac{1-c}{8(1+c)}$ for $F \in [\frac{1-c}{144(1+c)}, \frac{1-c}{16(1+c)}]$. For such values of F, the chain of monopolies would innovate only if $\Omega \leq \frac{1-c}{8(1+c)}$: downstream competition leads here again to more innovation than a chain of monopolies would offer.

(iii) If $F \geq \frac{1-c}{16(1+c)}$, the producer sets the wholesale prices in order to induce one of the retailers to list the new product, the other retailer selling only the old one. In that case, if $c \leq 1/2$, the producer chooses a limit-pricing strategy, denoted $\widetilde{AB, A}$, such that one only of the two retailers invests F and sells both goods, the other selling only good A. This strategy brings about more profit than no innovation for $\Omega \leq \sqrt{\frac{F(1-c)}{1+c}} - 2F$, which corresponds exactly to the fronteer of innovation in the chain of monopolies case.

On the contrary if $c \geq 1/2$, this strategy is no more possible for $F \geq \frac{1-c^2}{36c^2}$, and the best the producer can do is then to set prices inducing the retailers to specialize, one of them paying F to sell only the new product B, and the other selling only A without investing. This strategy always dominates[10] the absence of innovation for fixed costs such that the chain of monopolies would innovate, and even in a wider zone defined by $\Omega \leq \Omega^{A,B}$ with $\Omega^{A,B} = -2F - \frac{7-c^2}{72} + \frac{c^2}{2}F + (1-\frac{c}{2})\sqrt{F + \frac{1}{36}}$, so $\Omega^{A,B} \geq \sqrt{\frac{F(1-c)}{1+c}} - 2F$. In other words, for such values of F and c, downstream competition leads here to more innovation than a chain of monopolies would offer.

A.5 Proof of proposition 3

If $c \geq 1/2$, proposition 2 showed that for $F \geq \frac{1-c^2}{36c^2}$, the strategy to develop the new product and set prices inducing the retailers to specialize dominates the strategy without innovation for $\Omega \leq \Omega^{A,B} = -2F - \frac{7-c^2}{72} + \frac{c^2}{2}F + (1-\frac{c}{2})\sqrt{F + \frac{1}{36}}$. In the plan (Ω, F), this fronteer intersects the F axe in $\overline{F}^{A,B} \geq \frac{1-c}{4(1+c)}$ for any $c \geq 1/2$. Thus the zone in which retailers' specialization allows the development of the new product is wider than the zone in which the vertically integrated monopoly would innovate for such values of c.

[10]Notice that this particular pricing strategy is not necessary the optimal one, but it is enough to show that the optimal stategy will lead to innovation in this zone.

Information Congestion

Simon P. Anderson and André de Palma*

June 2003; this version May 2006
COMMENTS WELCOME

Abstract

Advertising must compete for the attention of prospective consumers, and attention is a scarce resource. "Junk" mail, "spam" e-mail, telemarketing 'phone calls, and advertising messages generally have the common feature that both sides of the market (advertisers and prospective customers) need to apply effort to generate a transaction. Message recipients supply attention according to average expected benefits of messages received. Senders are motivated by expected profits. The equilibrium number of messages transmitted depends on the profit of the marginal firm along with the communication probability induced from the receiver attention decision. In equilibrium, the wrong messages or too many messages may be sent. A higher cost of message transmission may improve message quality, which raises the number of messages examined: the optimal message rate reflects this. Product class competition also dissipates rents. An opt-out option (e.g. the Federal Do-Not-Call list) beats an outright ban, but too many individuals opt out. With several media, higher price ones are less congested in equilibrium, and attract fewer advertisers, but with more profitable products. When profitability is correlated with consumer benefits, "the medium is the message" for consumers to pay more attention to higher priced media. There are insufficient incentives to gather market research information on receivers since senders do not internalize the benefits of reducting congestion or improving consumer response from more tailored messages.

JEL Classification: D11, D60, L13.

Keywords: information overload, congestion, advertising, common property resource, overfishing, lottery, two-sided markets, junk mail, email, telemarketing, Do Not Call List, message pricing, the Medium is the Message, market research.

*Simon P. Anderson: Department of Economics, University of Virginia, PO Box 400182, Charlottesville VA 22904-4128, USA, sa9w@virginia.edu. André de Palma: Senior Member of the Institut Universitaire de France, ThEMA, Université de Cergy-Pontoise, 33 Bd. du Port, 95011, Cergy Cedex, France. andre.depalma@eco.u-cergy.fr. The first author gratefully acknowledges research funding from the NSF under grant SES-0137001 and from the Bankard Fund at the University of Virginia. We thank Bruno Jullien and Joshua Gans for comments, seminar participants at Cincinatti, Iowa State, Universite de Montreal, and the European Institute in Florence, and conference participants at EARIE 2004 (Berlin), the 2004 Chinese IO Society meeting (Beijing), and the 2005 Internet conference (Toulouse).

1 Introduction

One in every six pieces of mail sent in the world is US "junk" mail. This adds up to a tree's worth of paper per US household. The lowest rate charged by the US post office for bulk mailing is 8.8 cents per item, way below the current rate of 37 cents for first class mail.[1] Nations are still looking for an appropriate way to deal with "spam" (junk email).[2] A spammer can send 650,000 messages in an hour, at virtually no cost: spam filters cause people to lose possibly important messages, or even valid commercial offers that they might have taken up had they not been lost in a swamp of other propositions. Telemarketing (junk telephone calling) in the US has seen a dramatic decline since the recent advent of the FTC-sponsored Do-Not-Call list. More generally, people see thousands of advertising messages a day. An average of one message is remembered. Advertising technology (for unsolicited advertising) is therefore characterized by a huge degree of congestion in the competition for a consumer's attention. We address the economics of such unsolicited advertising.[3]

The model builds on two strands of research. The first is the model of parking in Anderson and de Palma (2004), which treats parking as a common property problem. A novel feature of the analysis is to treat multiple locations prone to congestion; locations are ranked by intrinsic quality (closeness to the CBD). The common property problem is greatest at locations of highest quality. The analogue is that advertisers must compete for the consumer's attention by sending them messages. The quality aspect is that consumers differ by how responsive they are to advertisements. The second research antecedent is work by Anderson and Coate (2005) on advertising-financed media, which contributes to the nascent literature on "two-sided" markets with bilateral network externalities. The message transmission (advertising) side of the model is drawn from here.

We consider two groups of economic agents, senders and receivers of messages. For concreteness here, think of them as firms and consumers. Firms need to communicate their wares. They do so by sending messages (bulk mail, etc.) to prospective consumers on the other side. Firms (message senders) differ by the profitability conditional on reaching potential consumers (message

[1] This lowest rate applies to non-profit organizations. The rate is up to 7 cents higher for private firms. For USPS rate information, see http://pe.usps.gov/text/dmm/R600.htm#Xgc81619

[2] The term spam comes from an early anecdote in the annals of computer geekdom. Someone sent his friends a message which contained just the word "spam" (after the Monty Python Flying Circus song) repeated hundreds of times: htpp://www.templetons.com/brad/spamterm.html describes the "origin of the term spam to mean net abuse". Spam can be around 55% of email, or even rise to 80%: see http://www.eprivacygroup.com/pdfs/SpamByTheNumbers.pdf

See also: http://cobb.com/spam/numbers.html

[3] Of course, other advertising is unsolicited: think of billboards and radio/television for examples. Classified ads, and ads in specialist magazines, may be more sought after. TV ads finance the programs they run on, and so are bundled commodities. The distinguishing feature we consider is the crowding of attention. This feature applies to billboards and TV ads too.

receivers). Sending messages is costly. To communicate successfully with a consumer, the firm must first send the consumer messages and then she must examine a message (open the particular mail containing the advertisement, for example). Both sides of the market need to exert effort to arrive at transactions. Clearly, firms with higher profitability have a larger benefit from sending ads. We can then find a critical profitability level such that all superior sender types send messages to receivers, while all other potential senders do not.

In equilibrium the cost of sending a message should equal the expected benefit for the marginal sender type. This process leads to three types of outcome: recipients receive more messages than they examine, they examine all messages received, or they receive no messages (although they would examine everything they got if they got any!)[4]

We also deal with the consumer choice of how many messages to examine (envelopes to open, say). This issue is an important feature of the junk message phenomenon because it explicitly treats the joint input nature of information production. Specifically, both receiver inputs and sender inputs are needed in order for a sale to be made. The sender must send the message at some cost and the receiver must examine it (open the envelope and process the information therein), also at some cost. The interesting economics of this market interaction are that the number of messages sent depends on expected profit of the marginal sender, which in turn depends on the number of messages read by the receiver. However, the number of messages the receiver examines depends on the benefit that the receiver expects, which in turn depends on the composition in terms of average quality of the messages sent. In this regard, we shall see below that a higher cost of sending messages will increase the number of messages examined because the expected quality rises.

Consider the effects of raising the costs of sending messages. If the recipients are examining all messages received, they will receive fewer messages. There is a social loss on this account due to a reduction in socially beneficial transactions. However, if receivers do not examine all messages, they will examine more messages since the average quality of the messages they receive increases.[5] This is a source of welfare improvement.

Congestion in transportation represents about 1% of GNP in European nations. Various policies mitigate the social loss. Yet the major flows nowadays are in information rather than physical goods. It has been recognized for decades that excess information is costly (Toffler's, 1970, Information Overload, for example.) But there is curiously little work on the economics of information overload.[6] A striking exception is Van Zandt (2004), who considers a model similar to ours but with a different emphasis: he is interested in targeted recipients of messages. In both our frameworks, each sender can send a message

[4] A similar situation occurs for referee reports: some only respond to some requests, others do all they are asked, and some would do them but are never asked.

[5] A similar finding (though in the opposite direction) arises in Engers and Gans (1998): paying referees is not as effective as might be thought since referees may be more likely to refuse knowing that other referees are induced by payments.

[6] The term Information Overload was coined by Toffler (1970).

to a given receiver. We treat all receivers as identical where he treats them as having different worth to different senders, and therein lies the efficiency benefit in his model from a tax on messages. In Van Zandt (2004), receivers examine a fixed number of messages so that there is a common property problem in accessing recipients. In our set-up, the receiver must also exert effort to examine messages: there is a cost and an associated benefit that depends on the average expected quality of messages. Van Zandt shows that a small tax on sending messages may be Pareto improving for all firms. This is because such a tax will cause marginal firms to refrain from sending messages to those consumers unlikely to be much interested. However, those firms will gain from becoming more prominent with those consumers from whom they expect larger profits. This matching aspect does not arise in our set-up.

Other authors have looked at some aspects of the overall problem. Hui and Png (2005) survey theoretical and empirical research on privacy, and note that information may impart direct externalities on market participants (for example, in the sending of unsolicited messages, which is our focus). The problem of spam email has attracted attention because of its prominence in the information age. Ayres and Nalebuff (2003) suggested that senders ought to pay for consumer access through email stamps. Van Alstyne et al. (2004) propose a system whereby the sender must post a bond that can only be recouped if the receiver likes the message content.

The outline of the paper is as follows. The next section gives the background in terms of the behavior of the agents on the two sides of the potential transaction. Section 3 derives the building blocks of the equilibrium analysis, namely the sender transmission function and the receiver examination function. Section 4 puts these together to derive equilibrium solution. Section 5 then provides welfare analysis and finds conditions under which the optimum can be decentralized through choice of the transmission price. Section 6 allows for intrinsic nuisance costs and looks at the possibility of receivers opting out completely (for example, the federal Do Not Call list in the US), and the pros and cons of outright bans. Section 7 addresses the solution chosen by a monopoly information gatekeeper and relates the current work to the analysis of two-sided markets. Section 8 discusses the effects of allowing for competition between senders in the product market. Section 9 underscores the role and benefits of multiple media, Section 10 considers the insufficiency of equilibrium targeting, and Section 11 concludes.

2 Congestible information

In order for a (mutually beneficial) transaction to be consummated, information must be transmitted by a sender, and the receiver must both process it and react positively (by purchasing an advertised good, say, or joining a club). In this case, a successful transaction occurs. The size of the conditional surpluses to each party (following a positively processed message) depends only on the sender type. We suppose that there is a single receiver and a continuum of senders. Initially, we assume that each sender can send at most one message to the receiver.

The basic idea of the model is the following. Senders transmit costly messages that may result in positive profits if the receiver is aware of the opportunity communicated in the message. That is, the receiver needs to examine the message and wish to act on it. Senders have different expected profits from communicating with the receiver. For example, senders may have different mark-ups, and/or the probability that the receiver is interested in buying a product may differ across products. The receiver chooses how many messages to examine, bearing in mind the average expected surplus from a message. Congestion arises when the receiver chooses not to examine all the messages received. Even though (as argued below) congestion is not optimal at the first best solution, it may arise in equilibrium because a sender may transmit a message that is not examined with certainty if it has a high enough expected profitability (even if it is disregarded with some positive probability). The expected receiver surplus from any particular message can reflect the probability of buying (being interested in an advertised product), as well as different conditional surpluses from buying, as noted above.

Expected surpluses are assumed independent across messages. This means that there is no "business stealing," so that messages are only in competition with each other insofar as they compete for a receiver's attention in examining them. Thus, the senders decide whether or not to send a message, and the receiver decides how many of the messages received to examine. In the junk mail context, households decide how many letters to examine. For telemarketing calls, they decide how often to answer the telephone. For spam e-mail, recipients determine how much mail to read (or perhaps, how tight to set the spam filter setting).

2.1 Information senders

There is a mass of senders of size M. We assume that at most one message can be sent by each sender to the receiver.[7] Let $\pi(\theta)$ denote the conditional expected sender benefit from sender type θ's match with the receiver. We order the sender types by $\theta \in [0,1]$ such that higher θ implies higher conditional expected sender benefits. The distribution of sender types is denoted by $F(\theta)$, with associated density $f(.) > 0$.[8]

Assumption 1: $\pi(\theta) > 0$ is continuous and strictly increasing in θ.

We assume the profit function is strictly increasing in θ in order to avoid dealing with ties: nothing really rides on this except for simplicity of presentation.[9] Assumption 1 implies that we seek allocations at which all senders above

[7] In Anderson and de Palma (2006) we address multiple messages sent to the same receiver. In terms of the notation below, all senders will transmit only one message if $\pi(1) < \pi(\theta_1)/[1-\gamma/\pi(\theta_1)]$ where θ_1 solves $\pi(\theta_1)\phi/M[1-\theta_1]$, with $\pi(\theta_1) > \gamma$.

[8] We could, without loss of generality, set $f(\theta) = 1$ and have the $\pi(\theta)$ function pick up relative clustering of types (alternatively, we could set $\pi(\theta)$ linear and have $f(\theta)$ do the work). For the present, it is useful for exposition to retain both $f(\theta)$ and $\pi(\theta)$.

[9] For what follows we could equally well suppose that $\pi(0) = 0$ to reflect the idea that

a critical threshold are transmitting. The threshold is denoted θ^*. In such cases, the number of messages sent is the same as the number of active senders, and is denoted $n(\theta^*) = M[1 - F(\theta^*)]$.

We assume that the unit cost to a sender for transmitting a message is constant and denoted by γ.

Assumption 2: $\pi(0) < \gamma < \pi(1)$.

We make this assumption to avoid a couple of rather trivial cases. If $\gamma \geq \pi(1)$, no sender will transmit messages, while if $\gamma < \pi(0)$ all senders will transmit if the receiver examines a large enough number of messages (so each message has a high enough probability of being examined). This latter part of Assumption 2 will also imply that there are always some senders "waiting in the wings" in equilibrium: not all transmission options are exhausted.

The total benefits to senders from transmitting messages are the total expected profits weighted by the probability each message is examined, minus the cost of sending the messages.

2.2 Information receivers

We assume that receivers are identical and normalize their mass to one. Let $s(\theta)$ denote the conditional receiver expected benefit (or surplus), conditional on the receiver examining the message. Thus, $s(\theta)$ allows for the probability that the receiver wishes to act upon the message (buy the advertised good, say) or not, and that sender types may entail different probabilities that the receiver is interested. Different θ types may also entail different surpluses when the receiver is interested. Any product pricing information is included in the message.

Assumption 3: $s(\theta) \geq 0$ is continuous in θ and $s(\theta) > 0$ almost everywhere.

Clearly the expected surplus should be non-negative since the receiver always has the option of ignoring the information received. In the sequel, we shall mainly consider the function $s(\theta)$ to be monotone.

Suppose that it costs the receiver $C(\phi)$ to examine ϕ messages. For the time being, we assume that messages received but not processed by the receiver incur no cost (free disposal). The examination cost function is assumed to have the following properties.

Assumption 4: $C(.)$ is a twice continuously differentiable, non-negative, increasing, and strictly convex function, with $C'(0) < s^{\max}$.

If the receiver chooses not to examine all the messages received, then each is assumed to be equally likely to be examined. That is, if n messages are received, from which ϕ are examined, with $\phi \leq n$, then the probability that the receiver examines any one is $\phi/n \leq 1$. Examination is a necessary precursor to a consummated message (a sale). Clearly, if $C'(0) \geq s^{\max}$, the market is closed off.

The receiver's surplus from examining ϕ messages is the expected benefit over ϕ examinations, given the expected surplus of the messages sent, minus

there are many potential opportunities running the gamut from positive through negative profitability. We can censor those with zero or negative profits.

the examination cost, $C(\phi)$. The total benefits from market interaction due to message transmission is the sum of the receiver and sender benefits.

3 Sender and receiver behavior

Senders take the number of messages, ϕ, examined by the receiver as given. In sub-section 3.1 we look at the mass of messages sent and the associated sender types as a function of ϕ. As we show, this yields a threshold value, θ^*, such that all higher θ's transmit, with an associated mass of messages sent. The number of messages sent, as a function of ϕ, is the *sender transmission function* (or transmission function, for short), which is denoted $N(\phi)$. Since only the highest profit messages are sent in equilibrium, $N(\phi)$ is a sufficient statistic for describing the threshold message, θ^*.

Similarly, on the other side of the market, the receiver takes the mass and composition of the messages sent as given. In sub-section 3.2 we determine the examination condition (choice of ϕ) as a function of the number of messages sent, n. It is also understood that, since the highest profit firms send messages, the average expected benefit from a message depends on n. This yields the *receiver examination function* (or reception function, for short), which is denoted $\Phi(n)$.

3.1 Information senders

The sender transmission function, $N(\phi)$, describes the number of messages sent as a function of the receiver examination level, ϕ. Senders maximize profits for a given receiver examination level. Assumption 1 (that the ranking of π is strictly increasing) leads to a preliminary characterization of a sender transmission function. The proof is in the Appendix.

Lemma 1 *For any given ϕ, all sender types θ above a threshold θ^* send messages.*

Lemma 1 indicates that the most profitable senders are active, so the messages sent are those in $[\theta^*, 1]$. Higher profit senders crowd out lower profit ones. The threshold θ^* induces a mass of messages $n = M[1 - F(\theta^*)]$, so there is an inverse and monotonic relation between θ^* and n since we assumed $F(\theta)$ is strictly increasing.

Lemma 2 *There is a one-to-one relation between the mass of messages sent, n and the threshold, $\theta^* = F^{-1}\left(1 - \frac{n}{M}\right)$, and $\frac{d\theta^*}{dn} = \frac{-1}{Mf(\theta^*)} < 0$.*

Proof: This follows directly from Lemma 1, since $n = M[1 - F(\theta^*)]$, so $\frac{dn}{d\theta^*} = -Mf(\theta^*)$. Q.E.D.

In the sequel, for notational convenience we shall not explicitly denote the functional dependence of θ^* on n. However, this relation should be recalled whenever a θ^* is encountered.[10]

[10] For example, $\pi(\theta^*) = \pi\left(F^{-1}\left(1 - \frac{n}{M}\right)\right)$, etc.

The sender transmission function $N(\phi)$ maps values of ϕ into numbers of senders transmitting, n. Any value of ϕ is interpreted as a number of messages that will be examined if available. Thus if the value of ϕ exceeds (or equals) the actual number of messages sent, n, the probability is one that a message sent is examined. Conversely, for $\phi < n$, the probability that a sent message is examined is $\phi/n < 1$.

The function $N(\phi)$ is implicitly defined by the condition that the least profitable sender (type θ^*) earns zero profit (by Assumption 2). Hence, $N(\phi) = M(1 - F(\theta^*))$ where θ^* is implicitly defined by $\min\left\{\frac{\phi}{n}, 1\right\}\pi(\theta^*) = \gamma$. Note that all messages sent (and received) are examined (i.e. $N(\phi) = \phi$) when $\gamma = \pi(\theta^*)$ or, equivalently, when the threshold value satisfies $\theta^{\min} = \pi^{-1}(\gamma)$ or $n^{\max} = M\left[1 - F\left(\theta^{\min}\right)\right] \equiv \phi^c$. Thus $\theta^{\min}$ is the lowest possible threshold for θ^* and is attained when so many messages are examined that each one sent is examined. Then the profit on the marginal message just equals the transmission cost, γ.

For $\phi \leq \phi^c$, the sender transmission function is given implicitly by

$$N(\phi) = \frac{\phi}{\gamma}\pi(\theta^*), \tag{1}$$

or, in inverse form, $\phi = \gamma\frac{n}{\pi(\theta^*)}$. The transmission function is sketched in Figure 1, in (ϕ, n) space.

INSERT FIGURE 1. The Sender Transmission Function $N(\phi)$.

The properties of the sender transmission function, $N(\phi)$, are proved in the Appendix.

Proposition 1 *The* ***sender transmission function*** *is given by*

$$N(\phi) = \min\left\{\frac{\phi}{n}, 1\right\}\pi(\theta^*) = \gamma.$$

It is continuous and increasing over $\phi \in [0, \phi^c]$, where $\phi^c = M\left[1 - F\left(\theta^{\min}\right)\right] = n^{\max}$ and $\theta^{\min} = \pi^{-1}(\gamma)$. The inverse of $N(\phi)$ is $\phi = \gamma\frac{n}{\pi(\theta^*)}$. For $\phi \in (0, \phi^c)$, $N(\phi) > \phi$ and $N(\phi) = \phi$ for $\phi \in \{0, \phi^c\}$. For $\phi \geq \phi^c$, $N(\phi) = n^{\max}$.

In summary, the sender transmission function is increasing from $(0, 0)$ to $(n^c, \phi^{\max})$. Receivers do not examine all messages received except at the end points where either no messages or else the maximum number of messages n^c are sent. The sender transmission relationship $N(\phi)$ is thus a number of messages sent that increases with the number of messages examined. In that sense it is like a supply relation, although one that depends on the quantity demanded. Equivalently, there is a critical sender type, θ^*, that is decreasing in the number of messages examined by the receiver. There is no consistent solution where receivers examine more messages than senders want to transmit.

Corollary 1 *The ratio ϕ/n is increasing along the Sender Transmission Function.*

This is a fundamental property of the function. As ϕ rises, the only way that more (marginal) sender types can be attracted is that the examination probability must rise. Note that this property implies the one given in the Proposition that the STF slopes up.

The important economic problem on this side of the market is the common property problem. The receiver's attention is "over-fished" by the sender. As we just showed, for only two levels of ϕ do the senders respond by sending exactly the number of messages that the receiver examines - and one of these is the zero level! At all other feasible levels, the senders are creating information congestion by sending more messages than the receiver examines. We now need to determine receiver behavior.

3.2 Information receivers

The receiver faces the decision of how many messages to examine given the profile of messages received. Lemma 1 implies that only the highest θ-types transmit, so if the critical θ value is θ^*, the number of messages sent is $n = M[1 - F(\theta^*)]$. Since n is inversely related to θ^*, we shall analyze the receiver examination function, which is denoted $\Phi(n)$, by determining the receiver's actions (choice of how many messages to examine, ϕ) as a function of θ^*. Each time an expression involves θ^*, the reader should remember that this is a decreasing function of n. The receiver's calculus depends on the benefits she expects to get, and the cost of examination. Given she cannot tell a priori which messages contain which offers, she examines them at random.[11] If all senders with $\theta \geq \theta^*$ are transmitting, the average expected surplus per message scrutinized by the receiver is[12]:

$$s_{av}(\theta^*) = \int_{\theta^*}^{1} s(\theta) \frac{f(\theta)}{1 - F(\theta^*)} d\theta. \tag{2}$$

The integral term is the expected surplus per message examined from the total set of messages received, where $f(\theta)/(1 - F(\theta^*))$ is the density of message types conditional on the set of messages sent by senders in $[\theta^*, 1]$. If the receiver examines $\phi \leq n$ messages, her expected benefits are then

$$S(\phi; n) = \min\{\phi, n\} s_{av}(\theta^*),$$

which is the number of messages examined times average expected surplus.

The cost of examining ϕ messages is given by $C(\phi)$, where $C(.)$ is increasing, strictly convex, and twice continuously differentiable (by A4). The receiver's problem is then

$$\max_{\phi \in [0,n]} \{S(\phi; n) - C(\phi)\}.$$

[11] We allow for different media below. The fact that the sender has paid a higher price provides information before the letter is opened. For example, there may be two types of letter (with different postage costs, like registered and bulk mail).

[12] Here $s(\theta)$ is the expected surplus from a message from sender θ.

The solution to this problem yields the receiver examination function, $\Phi(n)$. If $s_{av}(\theta^*) \leq C'(0)$, no messages are examined, so the receiver's solution is to set $\Phi(n) = 0$. If some but not all messages are examined, $0 < \phi < n$, the first-order condition is

$$s_{av}(\theta^*) = C'(\phi),$$

where the LHS is simply the average surplus over the messages received. This is an important principle in the problem: average benefit equals marginal cost on the sender side if the sender is at an interior solution of not examining all messages received. Otherwise (if $s_{av}(\theta^*) > C'(n)$) the receiver is supply constrained in the sense that she would examine more messages if she received them, and $\Phi(n) = n$. To summarize:

Proposition 2 *The **receiver examination function** $\Phi(n)$ is given by*

$$\Phi(n) = \min\left\{n, C'^{-1}(s_{av}(\theta^*))\right\}, \tag{3}$$

where $\theta^* = F^{-1}\left(1 - \frac{n}{M}\right)$. It is a strictly increasing function of n if $s_{av}(\theta^*)$ is strictly decreasing in θ^*.

Consider first the simple case when $s(\theta) = \bar{s}$ (so $s(\theta)$ is constant at level $\bar{s}$). The resulting receiver examination function is given in Figure 2.

INSERT FIGURE 2. Receiver Examination Function $\Phi(n)$ when $s(\theta) = \bar{s}$.

The optimal examination decision is simply described as a value for ϕ equal to $C'^{-1}(\bar{s})$ as long as this level is below n, and n otherwise (see (3) with $s_{av}(\theta^*) = \bar{s}$. Hence the receiver's response is to examine a constant number of messages: when θ is so high that this number is no longer attainable, the receiver examines as many messages as possible, which is the number sent. The response curve therefore follows the constraint $\phi = n$ for high enough θ (i.e., for low enough $n \leq C'^{-1}(\bar{s})$). The case of non-constant s is described in more detail in the equilibrium section below.

4 Equilibrium configurations

Equilibrium is a market clearing (or consistency) condition that the agents on each side rationally and correctly anticipate the actions of the agents on the other side of the market. This is simply the intersection of the sender transmission function and the receiver examination function. Thus, an equilibrium will be described by a pair (ϕ^e, n^e) such that $N(\phi^e) = n^e$ and $\Phi(n^e) = \phi^e$.

4.1 Constant receiver benefits

We first determine the full equilibrium to the model when $s(\theta)$ is constant. As well as the trivial equilibrium at which no messages are sent and none are examined, there is one other equilibrium. Depending on parameter values, either all messages are examined, or else only a subset of those sent are. Three cases are illustrated in Figure 3.

INSERT FIGURE 3. Equilibrium with constant receiver surplus, $\bar{s}$.

The first case, with Sender Transmission Function STF1 and high γ, involves all messages examined, and receivers would be willing to examine more if more were sent. However, the senders wish to send no more even though they are examined with probability one.

In the second case, with Sender Transmission Function STF2 and intermediate γ, senders examine all messages, but would not examine more if more were sent. Again, the senders wish to send no more even though they are examined with probability one.[13] In the third case, with Sender Transmission Function STF3 and low γ, senders examine fewer messages than the number sent. Here, the senders would wish to send more messages if more were examined, and there is message congestion at equilibrium. Note that an increase in γ has no effect on the receiver examination function. The sender transmission function shifts left though. Therefore, $\gamma_1 > \gamma_2 > \gamma_3$ in Figure 3. This discussion proves the following result.

Proposition 3 *Suppose that receiver surplus is constant, $s(\theta) = \bar{s}$. Let $\gamma_2 = \pi\left(F^{-1}\left(1 - \frac{C'^{-1}(\bar{s})}{M}\right)\right)$. Then for $\gamma > \gamma_2$, all messages sent are examined. For $\gamma < \gamma_2$, only a fraction of the messages sent are examined.*

As we discuss further below, if there is over-fishing ($\gamma < \gamma_2$), then this will be diminished by raising γ: there will be less rent dissipation by senders, and better θ types will transmit. There is a clear welfare gain here to the "better" senders from eliminating the worse rivals, meaning that the higher profit senders are more likely to get attention. Overall sender benefits may rise even if they are not compensated with the extra rate revenues because the senders with the lowest benefits are foreclosed, rendering the remainder more prominent.[14] The receiver though is unaffected because the examination decision is unchanged.

Thus a higher γ can raise welfare starting from a congested regime. This does not mean that the first best optimum can be reached by raising γ. Indeed,

[13] This case arises when ϕ_2^c and $\theta_2^{\min}$ satisfy $C'(\phi_2^c) = \bar{s}$, $\pi(\theta_2^{\min}) = \gamma_2$, and $\phi_2^c = M(1 - F(\theta_2^{\min})) = n_2^{\max}$. Hence γ_2 is defined by $\gamma_2 = \pi\left(F^{-1}\left(1 - \frac{C'^{-1}(\bar{s})}{M}\right)\right)$.

[14] Van Zandt (2004) allows targeting of different consumer types. Then a rate increase may benefit all senders because low profit opportunities are crowded out. This raises the profits of the remaining senders (which now have better prospects for being examined). All sellers' profits may rise if different senders have high profits with different receivers.

as we show in Section 5, the first best optimum is unattainable if the status quo has congestion. The reason is that the optimum has no congestion. If (higher) pricing were used to price out congestion, then the solution (say the "elbow" point in Figure 3, at $n_2^{\max}$) would be sub-optimally low transmission because the receiver examination decision does not account for sender profits. On the other hand, if the status quo has no congestion, then it may be possible to attain the optimum, by reducing the transmission rate and so inducing more messages to be sent (which will be examined as long as the receiver is not sated, that is, as long as the solution lies on the $\phi = n$ locus).[15]

As we shall see below, when sender surplus rises with θ, there is an additional social benefit on the receiver side from raising γ. This benefit stems from improving the quality of messages received.

4.2 Increasing receiver benefits

Consider now the case where $s(\theta)$ is increasing with sender type θ so that profits and consumer surpluses are positively correlated. This positive correlation is more likely when receivers have a high probability of not liking offers proposed in messages: profits and receiver surpluses are only earned when receivers are interested in taking up the offers. In this case, the average benefit increases with θ^*, and, differentiating (2), $\frac{ds_{av}(\theta^*)}{d\theta^*} = \left(\frac{f(\theta^*)}{1-F(\theta^*)}\right)[s_{av}(\theta^*) - s(\theta^*)] > 0$. An interior solution to the receiver's problem is given by $s_{av}(\theta^*) = C'(\phi)$. Consider the Receiver Examination Function, starting with high n. As n falls, θ^* rises, and so ϕ rises. However, as the number of messages sent falls sufficiently, the constraint $\phi = n\ (= M[1 - F(\theta^*)])$ is reached. Thereafter, a lower n (higher θ^*) leads to a smaller ϕ. Thus (still using reading from right to left), the receiver's choice relation traces out an increasing curve until the constraint $\phi = n$ is attained, and then it follows a declining path with $\phi = n$ (see Figure 4).

INSERT FIGURE 4. Equilibrium with increasing surplus, $s'(\theta) > 0$.

The interesting feature of this case is the beneficial effects of a higher γ when there is congestion. Then, as noted above, the number of messages examined rises because a higher γ crowds out senders with lower θ. This raises the average surplus from examination, causing the receiver to examine more messages.[16]

Proposition 4 *Suppose that receiver surplus, $s(\theta)$, is increasing. Let γ^u be defined by $\pi(\theta^u) = \gamma^u$, $\phi^u = M(1 - F(\theta^u))$, and $s_{av}(\theta^u) = C'(\phi^u)$. Then for $\gamma > \gamma^u$, all messages sent are examined. For $\gamma < \gamma^u$, only a fraction of the messages sent are examined. In the latter case, a small rise in γ leads to higher receiver welfare by improving the quality of the messages as perceived by both the sender and the receiver: fewer messages are sent and more are examined.*

[15] A similar point applies to the analysis of the next sub-section.

[16] It is possible for aggregate sender surplus to go up too.

This Proposition highlights the possibility of an extra social benefit ensuing a transmission price rise. The price rise causes higher quality messages to be sent, which leads to higher examination rates. The higher examination rate somewhat curtails the reduction in transmission, but not as much as to overturn the initial reduced transmission.

4.3 Decreasing receiver benefits

This case leads to an upward-sloping receiver examination function. It may join the constraint, $\phi \leq n$ and leave it, then join it again, etc. Such a upward-sloping relation implies that multiple equilibria are possible. Figure 5 illustrates such a case with four equilibria at the intersections of the REF and STF.[17]

INSERT FIGURE 5. Equilibria with decreasing surplus, $s'(\theta) < 0$.

The stable equilibria among these are the two for which the reception function cuts the transmission function from above. The logic is akin to that for duopoly reaction functions. In particular, the zero-equilibrium in the Figure is unstable, and the one at $(\phi^c, n^{\max})$ is stable.[18] More generally, if the reception function meets the transmission function from above at the no examination/no send equilibrium, then it is unstable.

Multiple equilibria entail different levels of sender transmission. For example, a low level of transmission is sustained as an equilibrium because receiver rationally anticipate a low average surplus from the highest profit sender types. The receiver then examines few messages, inducing few senders to transmit. A higher level of transmission can also be sustained as an equilibrium because the receiver examines more messages in rational anticipation of high numbers sent and therefore high average surplus from the messages. Senders respond to high examination with high transmission. Now consider the receiver welfare at alternative equilibria. Clearly, higher sender numbers are better for the receiver because the average expected surplus (s_{av}) is higher. For the senders, the equilibrium in the Figure at $n^{\max}$ is best because all messages sent are examined. Another example is illustrated in Figure 6.

INSERT FIGURE 6. Equilibria with decreasing surplus, $s'(\theta) < 0$, second example.

[17] The Figure is drawn with a positive vertical intercept for the REF. This arises if $C'(0) < s(1)$, meaning that if only the top profit sender were present, the receiver is interested. If instead $C'(0) > s(1)$, the REF will hug the horizontal axis until a sufficient number of senders transmit that the average surplus is high enough to make it worthwhile to start examining messages.

[18] This logic also implies that the no examination/no send equilibrium for the constant and increasing sender benefits cases of the two preceding sub-sections is unstable.

There are again four equilibria illustrated. Consider the (stable) equilibrium that involves the highest level of examination and messages sent, which is the one furthest to the right. For a small rise in the transmission price, γ, the STF moves left and this equilibrium moves down the REF. A higher transmission price causes fewer messages to be sent, and this in turn leads to lower expected receiver surplus, causing even fewer messages to be sent. There is thus a vicious circle for the receiver. Indeed, for high enough γ, the top two equilibria disappear. This means that the market can collapse down to a much lower level of transmission (and examination). Indeed, the decreasing receiver benefits case may also entail that the only equilibrium is the one with no messages sent at all. This case may arise if the receiver's examination function lies everywhere below the senders' transmission function.

We summarize the discussion with the following Proposition.

Proposition 5 *Suppose that receiver surplus, $s(\theta)$, is decreasing. Then, multiple equilibria are possible. These equilibria can be Pareto ranked: equilibria with higher levels of messages transmitted are Pareto superior for both senders and the receiver. Selecting the equilibrium with the highest transmission level, an increase in γ causes both examination and transmission to fall. This fall may be drastic.*

Proof The possibility of multiple equilibria is illustrated in Figure 6. Equilibria with higher levels of messages imply the equilibrium ϕ/n is higher, by Corollary 1. Hence active senders are better off at such equilibria since they have a better chance of examination. Receivers are better off with more messages since average quality is higher (because $s(\theta)$ is decreasing). The effects of a higher γ are described just prior to the Proposition. Q.E.D.

Choosing among the multiple equilibria would seem straightforward: the equilibrium with the highest level of transmission is indicated by dint of it being Pareto superior. Note that the coordination problem is essentially on the side of the senders insofar as the receiver does examine sequentially in practice – and would therefore discover the average quality of messages. Nonetheless. this makes no difference (to the equilibria) in that each sender is too small to influence the average quality.

We give two examples below of extreme market failure. The first example shows that the market may be closed down when it ought optimally to be functioning. The second example shows that, with two types of firm, the wrong type may enter in equilibrium.

EXAMPLE 1. Suppose that $s(\theta) = 1$ for $\theta \in [0, 1/2]$, and $s(\theta) = 0$ for $\theta \in (1/2, 1]$. Let too $F(\theta) = \theta$. Then, $s_{av}(\theta^*) = \frac{\frac{1}{2}-\theta^*}{1-\theta^*}$ for $\theta^* \in [0, 1/2]$, and $s_{av}(\theta^*) = 0$ for $\theta^* \in (1/2, 1]$. Suppose that $C(\phi) = \frac{\phi^2}{2}$. Then the receiver's examination function, as a function of θ^*, satisfies $\phi^r(\theta^*) = 1 - \frac{1}{2[1-\theta^*]}$ for $\theta^* \in [0, 1/2]$, and $\phi^r(\theta^*) = 0$ for $\theta^* \in (1/2, 1]$. Note that the first part of the curve is a decreasing and convex function. Suppose too that $\pi(\theta) = k\theta$, with

$k > \gamma$, $M = 1$. Then the senders' transmission function is given by inverting the relation $k\theta\frac{\phi}{1-\theta} = \gamma$, as long as $\phi \leq 1 - \theta$, i.e., $\theta \geq \gamma/k$. Hence, the sender transmission function is $\phi^s(\theta^*) = \frac{\gamma[1-\theta]}{k\theta}$ for $\theta \geq \gamma/k$. If k is large enough, the only intersection with the receiver's examination function is at $\phi = 0$, and so the only equilibrium is the one where the market unravels.[19] However, for low k (and γ not too large) the social optimum involves the lowest set of products.

Example 2. This example illustrates a major potential problem with unsolicited advertising. Suppose that there were two different classes of product. Let π_i denote the sender benefit of each product in class i, $i = 1, 2$, and similarly let s_i denote the receiver benefit. We assume that $\pi_1 > \pi_2$ and $s_1 < s_2$, so that the first class has higher sender benefit and lower receiver benefit than the second class. There is a large enough number of independent products in each class. (This is similar to the main model except now products are clustered on two points.) In equilibrium, only the high profit senders survive: any low profit sender is driven out of the market since a high profit sender has a bigger incentive to send a message (advertise). Put another way, if the high-profit senders are earning zero expected profits from sending messages, then the low-profit ones cannot enter the market given that the consumer chooses at random which messages to examine. However, the optimum arrangement will have only the low profit senders sending messages if the sum $b = \pi + s$ is higher. In equilibrium, the low-profit senders are chased from the market by the others, even though the social surplus associated with them is higher. This is reminiscent of Gresham's law - bad junk mail crowds out good. Note that the receiver would examine more messages if she got more of the low-profit messages, but she does not rationally expect to get it. An extreme form of this phenomenon is when the surplus on the highest profit messages is zero. Then the high-profit ones crowd out all others, and the market unravels completely because no receiver finds it worthwhile to examine any messages. This is the "lemons" problem of e-mail - some people have closed their accounts because of the preponderance of junk.

Proposition 6 *Suppose that there are two product classes, with many products in each class. Benefits satisfy $\pi_1 > \pi_2 > \gamma$ and $s_1 < s_2$. If $C'(0) \geq s_1$, the only equilibrium has no messages sent. If $C'(0) < s_1$, and $\pi_1 + s_1 < \pi_2 + s_2$, the only equilibrium at which a positive number of messages is sent involves only messages of type 1, but the optimum entails only messages of type 2 being sent.*

More generally, the fundamental property is that senders transmit messages based on their profitability and receivers examine messages based on average expected surplus. We then expect the greatest problem in the market system

[19] For higher k, there are two other equilibria where the receiver and senders' functions intersect. For example, if $k/\gamma = 20$, then there are two solutions: $\theta^* = 0.22$ and $\theta^* = 0.42$. From the argument in the text, the former solution describes the stable equilibrium; the no examination/no send equilibrium is also stable. Welfare is lowest at the no examination/no send equilibrium.

is that high profit but low surplus deals survive. The bigger the negative correlation, the bigger the problem of the market failure: even fewer messages are examined but more and “worse” ones are trying to get through.

5 Decentralizable optimum

We now consider conditions under which the first-best optimum can be attained by judiciously choosing γ. To derive the first best allocation, let $b(\theta) = s(\theta) + \pi(\theta)$ denote the social benefit of a message that is examined.[20] Clearly, the optimum allocation entails all messages received being examined and only the messages associated to the highest benefit, b, should be sent. If the set of optimal messages sent is not convex, or even if does not comprise all messages above some critical θ (for example, if b is decreasing, it involves all messages below a threshold, θ^l) then it is not possible to replicate the first-best optimum with a single price instrument, γ. The reason is that the market equilibrium necessarily involves all senders above some critical profit level (equal to γ) transmitting messages. The market therefore cannot deliver other patterns.

Suppose then that the optimum does entail all messages being sent above some critical level of θ, θ^h. A sufficient condition for this is that b is increasing, which we assume in what follows. Then the critical θ^h is given as follows. A marginal message costs $\bar{\gamma}$ to send and delivers social benefit $b\left(\theta^h\right)$ when examined. The incremental examination cost is $C'\left(n^h\right)$ where we have defined $n^h = M\left[1 - F\left(\theta^h\right)\right]$. Thus the optimality condition (marginal social benefit equals marginal social cost) is

$$s(\theta^h) + \pi(\theta^h) = \bar{\gamma} + C'(n^h). \tag{4}$$

It is useful to compare this condition with those describing equilibrium. The latter has two separate equations for the senders and the receivers: that the *examination-probability-weighted* profit equal the transmission cost, and that the *average* surplus equal the marginal examination cost.

Consider now the conditions under which the optimum is decentralizable by choice of γ. We need to check whether two equilibrium conditions can hold at the optimal allocation. First, the senders must be induced to transmit the top n^h messages. Given that all messages are to be read at the optimum, the choice of $\gamma = \gamma^o$ such that $\pi(\theta^h) = \gamma^o$ will ensure that all messages that are optimally sent will be sent in equilibrium. However, full decentralization also requires that the receivers voluntarily examine all the messages sent. That is, it must be true that $C'\left(n^h\right) \leq s_{av}(\theta^h)$. To check whether the examination condition holds, it is useful to consider a benchmark case when $s(\theta) = \bar{s}$. In this case, we can write (4) as

$$\bar{\gamma} + C'\left(n^h\right) = \gamma^o + \bar{s}. \tag{5}$$

[20] Let $b^{\max} = \max_{\theta}[b(\theta)]$. If $\bar{\gamma} + C'(0) \geq b^{\max}$, no messages are sent. If $\bar{\gamma} + C'(0) < b^{\max}$, all messages with $b(\theta) \geq y^o$ are sent and examined, where y^o solves $\bar{\gamma} + C'(\Psi(y^o)) = y^o$ and $\Psi(y)$ is the measure of senders who generate a social benefit of at least y.

Suppose that $\gamma^o \leq \bar{\gamma}$. Then necessarily $C'(n^h) \leq \bar{s}$, as required for the receiver to examine all n^h messages received. This implies that the optimum is decentralizable. On the contrary, if $\gamma^o > \bar{\gamma}$, it is impossible to decentralize the optimum. In that case, the marginal examination cost necessarily exceeds the average (expected) sender surplus.

Proposition 7 *Assume that $s(\theta) = \bar{s}$, and let θ^h solve (5). Then, the first-best optimum can be decentralized (by setting $\gamma^o = \pi(\theta^h)$) if and only if $\pi(\theta^h) \leq \bar{\gamma}$.*

The interesting point about the Proposition is that it involves subsidizing message transmission! Notice that we only allow for action on the transmission price: we do not simultaneously allow consumers to be subsidized (say) for examining more messages.[21] It would seem hard in practice to do so since the opportunities for misrepresentation would be great. The receiver could simply claim to have carefully examined the messages.

The intuition for the result is as follows. At the optimum, all messages ought to be examined. This means that if it can be decentralized, then the receivers want to examine at least as many messages as they actually do receive. However, if the transmission price were at or above marginal transmission cost, then a price cut would have a positive effect on both sides of the market by inducing more senders to transmit, and consumers would necessarily examine those messages and be better off, a contradiction.

The above analysis effectively focuses on the case of no information congestion as the one in which decentralization is feasible. The more interesting case is when there is congestion. Then, if there were congestion at a message transmission rate of $\bar{\gamma}$, dropping the message transmission rate (by subsidizing) will only raise congestion. Therefore, the optimum will not be decentralizable. Instead, as pointed out in Section 4, the optimal (second best) policy is a higher rate in the absence of direct subsidies to receivers for examining.

6 Pas de publicité s.v.p. and the Federal Do-Not-Call List

Some Belgians and Frenchmen have a little sign on their letterboxes saying they do not want advertising flyers. In the US, the Do Not Call List is a

[21] Consider Figure 3, and suppose that $\bar{\gamma}$ is low enough that the relevant sender curve is STF (which leads to a congested equilibrium). The first best optimum allocation must have all messages examined, and so must be a point on the 45-degree line $\phi = n$. We now argue that it must be at a point between $n_2^{\max}$ and $n_3^{\max}$. Indeed, restricting attention to $\phi = n$, the derivative of social welfare with respect to θ is (see (4)): $-s(\theta) - \pi(\theta) + \bar{\gamma} + C'(n^h)$ (the welfare function is clearly concave under the assumption that b and $C''(.)$ are increasing). At the θ value corresponding to $n_2^{\max}$, $C'(n_2^{\max}) = s(\theta) = \bar{s}$. However, above the STF, $\pi(\theta_2) > \bar{\gamma}$ so that welfare is decreasing in θ (equivalently, increasing in n) at this point. On the other hand, at $n_3^{\max}$, $C'(n_3^{\max}) > s(\theta) = \bar{s}$ (since the receiver is unwilling to examine so many messages). Since $n_3^{\max}$ is on the STF, $\pi(\theta_3) = \bar{\gamma}$ so that welfare is increasing in θ (decreasing in n) at this point.

successful initiative orchestrated by the Federal Trade Commission that allows people choose not to receive calls from telemarketers.

If there is a nuisance cost to receiving the ads, then it makes sense to the individual to decline all of them.[22] However, she then foregoes some socially beneficial prospects. Our next aim is to introduce this receiver option into the model, so that the receiver may refuse to accept the messages.

Let S^* denote the equilibrium value of expected receiver surplus from examining messages, and let C^* denote the corresponding examination cost. Assume that receiving each message has a constant annoyance cost, ω. Note that the surplus and examination cost are independent of the nuisance cost, which is sunk if receivers intrinsically dislike receiving messages (telemarketing calls especially). Then the private benefit while receiving messages is $S^* - C^* - n\omega$.With the option of electing to receive no messages, this benefit is to be compared to the zero benefit the receiver gets by opting out.[23]

Suppose that ω is distributed in the population (which has unit mass) with support $[\underline{\omega}, \bar{\omega}]$ and density $g(\omega)$ with distribution $G(\omega)$, so that different individuals face different annoyance costs. An equilibrium with some, but not all, individuals opting out is then a critical value, $\hat{\omega} \in (\underline{\omega}, \bar{\omega})$ with

$$\hat{\omega} = \frac{S^* - C^*}{n}.$$

The social benefit associated with a receiver accepting messages is

$$S^* - C^* - n\omega + \Pi^* - n\bar{\gamma}.$$

The two extra terms added here are the senders' gross benefits and the transmission costs (assumed priced at social cost). A blanket ban is preferred to allowing the nuisance if

$$B^A \equiv S^* - C^* - n \int_{\underline{\omega}}^{\bar{\omega}} \omega g(\omega)\, d\omega + \Pi^* - n\bar{\gamma} < 0,$$

so that the the social perspective may not prescribe a blanket ban. Allowing individuals to opt out is better than a total ban if

$$B^{oo} \equiv [S^* - C^* + \Pi^* - n\bar{\gamma}]\, G(\hat{\omega}) - n \int_{\underline{\omega}}^{\hat{\omega}} \omega g(\omega)\, d\omega > 0,$$

where $G(\hat{\omega})$ is the fraction of receivers opting in. This is necessarily positive since $S^* - C^* - n\omega$ must be positive for all those opting in (by revealed preference)

[22] In the sequel we suppose that the receiver does not consider the option (in the status quo of no restrictions, and therefore free access) of closing down the message medium entirely. For example, she could disconnect her telephone to block out telemarketers' calls, or she could close her e-mail account to stop spam. Allowing for such options enriches the analysis that follows.

[23] The zero here reflects the surplus from the message transmission: the basic message medium may have a positive benefit (telephone service or mail delivery) that is netted out on both sides of any comparison.

and also $\Pi^* > n\bar{\gamma}$, which also holds by revealed preference because senders only transmit when they get non-negative net benefits.

Conversely, allowing opt-out from a status quo of no opt-out improves welfare if $B^{oo} > B^A$, or

$$[S^* - C^* + \Pi^* - n\bar{\gamma}]\,[1 - G(\hat{\omega})] - n\int_{\hat{\omega}}^{\bar{\omega}} \omega g(\omega)\, d\omega < 0.$$

This condition may or may not hold. For example, if $\hat{\omega}$ is close to $\bar{\omega}$ then nearly all individuals opt out. If $\Pi^* - n\bar{\gamma}$ is large, allowing opt-out is socially disadvantageous. Indeed, the reason is because "too many" individuals opt out. To see this, notice that the optimal opt-out cut-off, ω^o, is lower than the privately chosen one. This is because the individual does not account for the profit of senders at the margin. In summary, in a comparison of regimes we have:

Proposition 8 *The opt-out regime is socially preferable to banning messages entirely, but may be worse or better than allowing free access.*

The first part follows because opting improves welfare by letting in information when both receiver and senders essentially agree; the second part depends simply on whether surplus is greater with or without messages, given that the receiver side achieves a negative total benefit at the status quo of receiving all that senders wish to transmit.

The Do-Not-Call opt-out option may also change the quality of messages received through a volume effect, once we aggregate over different consumers. For example, suppose that a particular message type is sent by a sender that produces with increasing returns to scale. Insofar as its equilibrium price decreases with volume, then a reduction of sales through less information getting through to consumers who opt out of receiving messages may lead to a higher price for the remaining consumers. On the other hand, a decreasing returns to scale technology may have the opposite impact. It may also be (even with constant returns to scale) that the consumers who exclude themselves have a more inelastic demand (for example) and so when they opt out the price falls. Note that when the price falls, the remaining consumers expect higher surplus and so end up examining more messages.

7 A Monopoly Gatekeeper

The open access market organization described above may arise spontaneously in the marketplace. Another market system has an intermediary control the volume of messages transmitted through a conduit. Then, the price of transmitting messages may be determined by pure profit maximization concerns. For example, there could be a profit maximizing Broadcast Company, telephone company, Internet Service Provider, or Post Office. Such an entity is the intermediary (or "platform") in a two-sided market with negative externalities, in the parlance of the recent literature. The present context emphasizes the common property problem of an open access system. Since access can be priced by a platform which will account for the common property problem, pricing may have a positive effect of reducing congestion. But the intermediary is also interested in the volume of messages since it takes its mark-up on the total number of messages sent. A priori, it is unclear whether it will encourage mailings or just concentrate on the high willingness-to-pay senders. We show below that the intermediary will fully price out the common property problem.

First assume that the receiver examines at most $\bar{\phi}$ messages. This case arises when $s(\theta)$ is constant, so that $C'(\bar{\phi}) = \bar{s}$. Let $\pi(\theta) = \bar{\pi}\theta$ (without loss of generality[24]) and suppose that the density of advertiser types is $f(\theta)$ with distribution $F(\theta)$, and let $\frac{f}{1-F}$ be an increasing function (this is the monotone likelihood ratio property; equivalently, the log-concavity of $1-F$).

Suppose first that there were congestion, so that the price, p, that a critical sender type $\hat{\theta}$ is willing to pay to be included in the mailing is $\bar{\phi}\bar{\pi}\hat{\theta}/n = p$, where $n = M\left[1 - F\left(\hat{\theta}\right)\right]$. If there is congestion, then some mail is unexamined, so that $n > \bar{\phi}$. Let the constant marginal cost of delivering a message be $\bar{\gamma}$.[25] The gatekeeper's profit is then

$$M\left[1 - F\left(\hat{\theta}\right)\right](p - \bar{\gamma}) = \left(\bar{\phi}\bar{\pi}\hat{\theta} - M\left[1 - F\left(\hat{\theta}\right)\right]\bar{\gamma}\right),$$

which is increasing in $\hat{\theta}$. This means that a monopoly intermediary would price out any congestion.[26] Equivalently, if there is congestion, the monopoly is in the inelastic part of its demand curve and can therefore raise profits by pricing higher. Thus, a monopolist will not tolerate congestion (assuming indeed that all consumers are the same); hence, $\bar{\pi}\hat{\theta} = p$. Then the gatekeeper faces a simple monopoly problem of maximizing $(p - \bar{\gamma}) M (1 - F(p))$, which has a straightforward solution. It either prices so high that the message-examination constraint is slack, or else holds with equality.[27]

---

[24] We standardize $\pi(\theta)$ and look at a distribution of types since this is familiar from monopoly pricing analysis. A later footnote gives the parallel analysis for the function $\pi(\theta)$ with $f(\theta)$ normalized.

[25] This is a gross over-simplification. The cost structure is an important determinant of the Post Office's pricing policy. Costs, and associated tariffs, are lower per unit in bulk mailings that group similar destinations and when the sender uses a bar-code for addresses. Price discrimination over different classes of mail is another important dimension to the Post Office's behavior. Non-profit organizations also benefit from lower tariffs.

[26] If each message had a 50-50 chance of being read, a (risk-neutral) sender would pay twice as much to be read for sure. This means that the monopoly could ration message delivery by half and keep revenues the same. It would save on costs and raise profits.

[27] We can rephrase this analysis using the equivalent representation of the model where we take $f(\theta)$ uniform and $\pi(\theta)$ is the profit function. Under congestion, the intermediary's profit function, $(M\left(1-\hat{\theta}\right)(p - \bar{\gamma}) = \bar{\phi}\pi\left(\hat{\theta}\right) - \bar{\gamma}M\left(1-\hat{\theta}\right))$, is increasing in $\hat{\theta}$ (since $\pi'(\theta) > 0$). Without congestion, the intermediary's problem is

$$\max_{\{\hat{\theta}\}}\left[\pi\left(\hat{\theta}\right) - \bar{\gamma}\right] M\left(1-\hat{\theta}\right).$$

In the case of endogenous examination with $s'(\theta) > 0$, the receiver either examines all messages sent or else a number that decreases in the number sent (see Figure 4). Again the monopoly prices out congestion. This now has the extra benefit that higher message price improves the receive's selection. Consequently, the receivers examines more (and better) messages.

Consider now the socially optimal choice of intermediary price, first with $s(\theta) = \bar{s}$. A higher γ moves the STF left (see Figure 1) and leads to a socially better selection of messages (the higher profit ones), and the intermediary's profit rises if the initial position was congested. Hence, welfare rises by pricing out congestion. This is a fortiori true when $s(\theta) > 0$ since receivers also benefit from a better selection induced from better senders (and they respond with an improved examination rate). Indeed, as long as the free-access equilibrium has congestion, the optimum price is at the kink (for beyond that point, higher p eliminates senders with profit above γ, since they were willing to transmit at price γ without congestion).

Although neither a monopoly nor the optimum involve congestion, the monopoly does not necessarily implement the optimum arrangement since it tends to price too high. However, both monopolist and optimum may price at the kink in the REF, i.e., where congestion just ceases.

Clearly, there are cases where either limited access or open access would be preferred in a binary comparison. The monopoly platform likely restricts access too much. However, the common-property solution has too much access when it is congested (which is the interesting case).[28]

8 Firm Competition

The models above have assumed that each ad is sent by a firm producing a different new good. However, most junk mail is for credit cards, much spam concerns Viagra or mortgages, and many telemarketers call about time-sharing. This leads us to consider the effects of competition when there are many firms that offer the products for sale.

Suppose that all junk mail is from credit card companies and all credit cards are perfectly homogeneous except for possibly their price. Assume that consumers open messages randomly but there is a cost of each additional message

---

The first order condition to this problem yields $\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} = \frac{1}{(1-\hat{\theta})}$. The LHS is decreasing under the assumption that the function $\pi\left(\hat{\theta}\right)$ is log-concave (while the RHS is increasing), so there is a unique solution.

[28] The two-sided market literature usually considers access pricing for *both* sides of the market. In the current setting, this might mean charging receivers for access too: an access price would extract all receiver surplus. One insight from the two-sided market literature is that the platform may not want to charge for access (even if it could): getting "on board" with sufficient gusto the side that is more desirable to the other side may enable the platform to charge more for access. Thus it could be that the optimal price to the monopoly intermediary could still be zero (or even a subsidy, just like "free" entertainment on the television or radio could be seen as a subsidy to entice prospective customers to advertisers).

examined. The equilibrium has consumers rationally anticipate the prices announced in the mail, and firms maximize profits rationally anticipating other firms' prices and consumer mail-opening. Then, as per the Diamond (1971) paradox, the only equilibrium is that all firms set the monopoly price.[29]

Now though, firms will enter to dissipate all rents. Raising the message transmission price now necessarily raises welfare by decreasing the amount of rent dissipation of the monopoly profit. With a higher rate, fewer messages are sent to vie for the fixed profit.

This suggests that the junk message problem may be a double common property resource problem when there is competition within product classes. First there is over-fishing for a consumer's attention and second there is over-fishing in any product class (business stealing). The case for high postage rates on junk mail (email) is the strongest when most consumers do not open all of their mail and there are high rents so that there are many competing products within any class. What comes to mind is credit card ads through the regular mail and Viagra through email.

There is though a caveat to this conclusion. When there are multiple senders within a product class and multiple products, the logic of the Diamond paradox no longer holds because a receiver may get a second (or further) price quote while searching for other product class offers. This breaks the monopoly price equilibrium because consumers may then have several price quotes before choosing (as in Burdett and Judd, 1983, although this paper considers a single product class). The consequent pricing and transmission/examination equilibrium are left for future investigation.

9 The medium is the message[30]

When there are multiple media for reaching the consumer, the medium through which the message arrives may indicate useful information to the receiver, since different media cost the sender differently, and suffer different congestion levels in equilibrium. The fact that the sender has paid a higher price provides information before the message is examined. For example, two types of letters may arrive, with different postage costs, like registered and bulk mail. In equilibrium, we would expect different media to be used by different firm types. Clearly spam email appears to be used by the lowest expected gross profit options, while telemarketing calls or notifications by FedEx presumably have higher returns

[29] This analysis supposes that all credit cards are homogeneous. Introducing product heterogeneity tempers the extreme results of the Diamond Paradox. Consumers will typically open several envelopes to find a suitable product. This brings firms into competition and brings equilibrium prices down (the original set-up has no competition because only one letter is opened). The rent dissipation problem is muted because consumers typically choose the best of several offers. However, as shown in Anderson and Renault (1999), the number of firms is excessive, implying that an increase in the postage rate is optimal.

[30] The aphorism is due to Marshall McLuhan (1964). McLuhan and Fiore (1967), in "The Medium is the Massage" (which title apparently began as a typo that tickled the authors) gives a precis of this and related work.

than bulk mail.

We introduce alternative media with the minimal departure from our base model. We retain the assumption that each sender transmits a single message, and now must choose which medium to use (or not to transmit at all).[31] For simplicity, each medium entails the same examination cost function. We also suppose that examination costs on any medium are independent of the levels of examination on other media. For comparisons of different media such as telemarketing calls, email messages, bulk postal mail, television commercials and exposure to billboards on the evening commute, this is perhaps not too egregious an assumption.

Let there be $j = 1, ..., J$ media. The different media entail different transmission costs, and for clarity we ignore ties by assuming

$$\gamma_1 < \gamma_2 < ... < \gamma_J.$$

In equilibrium, therefore, lower priced media will be more congested because senders allocate themselves across media and a lower access price must be offset by a higher congestion level – otherwise senders would join any medium (platform) with both low access price and low congestion. Thus we expect email spam to have a lower equilibrium response rate (i.e., lower ϕ/n) than telemarketing.

This self-selection result implies that higher priced media will be used by senders with higher expected profits from reaching receivers. That is, the expected profit differential between two media is higher for a high θ type (and in favor of the more expensive medium) than for the low θ type. To see this property, note that the net benefit of a sender of type θ using medium j is

$$\pi(\theta_j)\frac{\phi_j}{n_j} - \gamma_j,$$

where ϕ_j is the equilibrium examination rate on medium j and n_j is the mass of senders using that medium. Here $\frac{\phi_j}{n_j}$ portrays the "quality" to the sender of the medium, and, just as in models of vertical product differentiation, higher quality is most appreciated (and therefore bought) by those with highest willingness to pay for it. Mathematically, the net benefit expression rises fastest across options available in different media for the highest θ_j.[32]

[31] As before, if transmission costs are high enough relative to profit differences across senders, this will be an equilibrium in an extended context where multiple messages are possible.

[32] The equilibrium relation is simply illustrated for sender types by graphing net benefit for each option on the vertical axis as a function of type on the horizontal. For given access prices and congestion rates, this yields a series of rays with increasing slopes and lower intercepts (each intercept is at $-\gamma_j$) as j rises. The senders choose the options with the highest net benefit, yielding the upper envelope of the rays as the equilibrium benefit levels as a function of type. Then, the horizontal distance between the intersections of the rays gives the mass of senders using each medium. In equilibrium, this needs to be consistent with the congestion levels that determine the rays' slopes, and the examination levels themselves must be consistent with the receiver's choice of examination.

The indifferent sender type, $\hat{\theta}_j$ that determines the cross-over point between medium $j-1$ and j is

$$\pi\left(\hat{\theta}_j\right)\frac{\phi_{j-1}}{n_{j-1}}-\gamma_{j-1}=\pi\left(\hat{\theta}_j\right)\frac{\phi_j}{n_j}-\gamma_j$$

which may be rearranged as

$$\pi\left(\hat{\theta}_j\right)=\frac{\gamma_j-\gamma_{j-1}}{\frac{\phi_j}{n_j}-\frac{\phi_{j-1}}{n_{j-1}}}$$

and these $\hat{\theta}_j$ are increasing in j. Hence more expensive media attract more profitable senders: email is used by senders with low expected profits while telemarketing attracts those with high expected profits.

It is also the case that the more expensive media are those that are more attractive to the receiver as along as consumer surplus and sender profit are positively correlated (i.e., $s'(\theta)>0$). This is because the most expensive media then carry the most profitable messages and therefore the most desirable ones. In response, the receiver therefore examines more messages. Nevertheless, this does not necessarily mean more will be sent through such media since they are also the least congested. To see this, consider the special case $s(\theta)=\bar{s}$ so that all messages carry the same expected surplus. Then the receiver will choose to exert the same amount of examination effort across all media (since the expected benefit is the same) so that all media carry the same level of examination (the solution to $\phi=C'^{-1}(\bar{s})$). Then, since congestion must be lowest at the top, the transmission rates across media satisfy

$$n_1>n_2>...>n_J.$$

Thus the cheapest media are the ones that are the most used and the most expensive ones the least used. In this model, it is also true that only the most expensive media can be uncongested: if this were true for any cheaper one, senders would move there, contradicting the equilibrium allocation.

One social benefit of multiple media is to enable a better triage of messages because they are sorted into more homogenous groups. As long as receiver surplus is not negatively correlated with sender profit, receivers will pay more attention to messages in the highest-cost media, which will reinforce the incentives for the high profit types to use them. We now show how this sorting mechanism works with an example of two channels. The simplest case has constant surplus per message, $\bar{s}$, and so the same number of messages ($\phi=C'^{-1}(\bar{s})$) is examined on each medium.

The higher priced medium, 2, will attract the top θ senders, in number $n_2=M(1-\theta_2)$, while medium 1 will attract $n_1=M(\theta_2-\theta_1)$ senders. Assume that the media have no social cost, and we want simply to show that differential pricing is better than equal pricing. The surplus function is then measured by aggregate profit. This is

$$W=\frac{M\int_{\theta_1}^{\theta_2}\pi(\theta)\,d\theta}{n_1}\phi+\frac{M\int_{\theta_2}^{1}\pi(\theta)\,d\theta}{n_2}\phi,$$

where the first term is the average gross profit to the sellers in the low-θ group and the second term that in the high group. Suppose we take the number of messages as fixed and then vary the composition across media. This entails varying θ_2, with a surplus effect of

$$\frac{1}{M\phi}\frac{dW}{d\theta_2}=\frac{\pi\left(\theta_2\right)}{n_1}-\frac{\pi\left(\theta_2\right)}{n_2}-\frac{M\int_{\theta_1}^{\theta_2}\pi\left(\theta\right)d\theta}{n_1^2}+\frac{M\int_{\theta_2}^{1}\pi\left(\theta\right)d\theta}{n_2^2}.$$

At the limit as γ_2 approaches γ_1 from above, n_2 approaches n_1 and so the support of the high profit senders has the same size as that of the low profit ones ($\theta_2-\theta_1=1-\theta_2$). The welfare derivative above then reduces to the last two terms, i.e.,

$$\frac{M}{n^2}\left[-\int_{\theta_1}^{\theta_2}\pi\left(\theta\right)d\theta+\int_{\theta_2}^{1}\pi\left(\theta\right)d\theta\right],$$

and this is necessarily positive because $\pi\left(\theta\right)$ is increasing. This means that starting from an allocation induced by equal transmission costs, surplus (profits) must rise by moving these costs apart in such a manner that the total volume of messages remains the same. The reason is that the better quality messages will now get more prominence - even though the examination rate is the same in their group - because the group size falls and relieves the congestion in that group by dropping out the lowest quality messages into the other group.

10 The Inadequacy of Market Research

While van Zandt (2004) emphasizes different receiver types, we have so far treated the same receiver type. Van Zandt's model is one of targeted transmission, so a sender will not send to a receiver who has low likelihood of reacting to its message. We now consider market research that allows senders to determine which receivers (households) are more likely to be interested in their products. A sender who does not discriminate between households imposes a greater externality by increasing congestion on all other senders and to all households. This implies there is likely insufficient investment in market research (coupled with our overall finding of excessive message transmission).

For simplicity of exposition, let there be two classes, the m-agents and the f-agents, with equal mass of each class. Profitability for senders differs across the two types. We assume that receiver types are diametrically opposed in terms of their profitability – what is a high profit product if it reaches one consumer type could be a very low profit type if it reaches another (think of the classic Hotelling linear city model). For the m-agents, profit is given by $\pi_m\left(\theta\right)$ which is an increasing function on $[0,1]$ with $\pi_m\left(0\right)=0$. Correspondingly, $\pi_f\left(\theta\right)$ is a decreasing function, and we impose symmetry by assuming that $\pi_f\left(\theta\right)=\pi_m\left(1-\theta\right)$. This means that the profits earned from the m-types are higher for senders with higher θ whereas the f-types yield higher profits for products/senders with lower θ.[33] The middle sender types, who have similar profits from either consumer type, have little reason to screen while the extreme types want to screen (each wanting to pick up a different consumer type). This means the extreme θ senders have more incentive to do market research to determine agent type since they gain more from excluding the unprofitable class. Conversely, the middling θ's gain little from knowing agent types: they want to sell to both since both have similar valuations.

The expected profit to a sender of type θ who does is not invest in market research is

$$\frac{\pi_f(\theta) + \pi_m(\theta)}{2}\frac{\phi}{n} - \gamma \tag{6}$$

where we have normalized the *total* receiver population to unity, and used symmetry to let n and ϕ denote the common mass of messages sent to each agent type and the common examination rate (these endogenous variables are determined below).

Now suppose that the sender can pay κ to successfully identify the receivers. Let κ be low enough that some senders do use the discriminatory strategy. The total profit from discriminating for a high θ-type (who wants to find the m-agents if $\pi_f(\theta) < \gamma$ since the f-agents do not cover the transmission cost) is

$$\left[\pi_m(\theta)\frac{\phi}{n} - \gamma\right]\frac{1}{2} - \kappa, \tag{7}$$

where the term in square brackets is the expected profit per m-agent. Hence, the marginal sender type is defined by $\tilde{\theta}$, obtained by equating (7) with (6). This means that

$$\frac{\pi_f(\tilde{\theta}) + \pi_m(\tilde{\theta})}{2}\frac{\phi}{n} - \gamma = \left[\pi_m(\tilde{\theta})\frac{\phi}{n} - \gamma\right]\frac{1}{2} - \kappa, \tag{8}$$

where we note that the LHS exceeds the RHS for $\theta < \tilde{\theta}$, so that lower θ types necessarily prefer not to discriminate. Solving out from (8),

$$\pi_f(\tilde{\theta}) = \frac{n}{\phi}\left[\gamma - 2\kappa\right]. \tag{9}$$

For starters, suppose that ϕ is fixed (for example, if all senders' products are worth $\bar{s}$ to the receiver). Then the only other endogenous variable to determine

[33] As one example, demand for one type could be a multiple of the demand for the other. As long as marginal cost is constant, the price the sender would want to charge each receiver type is the same, and also the same as when the sender does not distinguish the type (the price discrimination motive for determining types is discussed later). Then the differences in the $\pi(.)$ functions fully reflect differences in quantity demanded: if the quantity demanded by an m-agent for sender $\hat{\theta}$'s product were twice that of an f-agent, we would have $\pi_m(\hat{\theta}) = 2\pi_f(\hat{\theta})$.

An alternative underpinning is to assume both m-agents and f-agents have the same demands conditional on being interested at all in the product. Then, different interest probabilities give rise to different values for $\pi(.)$: if the m-agents were statistically twice as likely to be interested in sender $\hat{\theta}$'s product than f-agents, we would again have $\pi_m(\hat{\theta}) = 2\pi_f(\hat{\theta})$.

is n. The set of senders to an f-type is all those below $\tilde{\theta}$; equivalently, the set of senders to an m-type is all those above $1-\tilde{\theta}$. In both cases, we have

$$n = M\tilde{\theta}. \tag{10}$$

In conjunction with (9), this clearly yields a unique interior solution.

Now consider the welfare properties of the solution. Continue to treat ϕ as fixed, and the receiver surplus as the same across missives. Then the total surplus is the same as total profit, up to a constant. Total profit is

$$\begin{aligned}\Pi\left(\tilde{\theta},\phi,n\right) &= 2\int_0^{1-\tilde{\theta}}\left\{\left[\pi_f(\theta)\frac{\phi}{n}-\gamma\right]\frac{1}{2}-\kappa\right\}d\theta \\ &\quad +2\int_{1-\tilde{\theta}}^{1/2}\left\{\frac{\pi_f(\theta)+\pi_m(\theta)}{2}\frac{\phi}{n}-\gamma\right\}d\theta\end{aligned}$$

where we have used symmetry. In $\Pi\left(\tilde{\theta},\phi,n\right)$, n is given by (10), but we leave is as a separate argument since it represents the negative externality across senders: $\Pi\left(\tilde{\theta},\phi,n\right)$ is clearly decreasing in n in the expression above. The effects on $\Pi\left(\tilde{\theta},\phi,n\right)$ of raising $\tilde{\theta}$ is

$$\frac{d\Pi\left(\tilde{\theta},\phi,n\right)}{d\tilde{\theta}} = \left\{\gamma+2\kappa+\pi_m\left(1-\tilde{\theta}\right)\frac{\phi}{n}-2\gamma\right\}+\frac{\partial\Pi\left(\tilde{\theta},\phi,n\right)}{\partial n}M.$$

When evaluated at the equilibrium as given by (9), the term in curly brackets is zero, so there remains the pure externality effect, $\frac{\partial\Pi(\tilde{\theta},\phi,n)}{\partial n}M<0$. This indicates that there is too little investment in screening since senders disregard the beneficial effects of withdrawing their messages whence they have lowest social value.

We now add in the receiver examination condition into the mix. Assume that $s_m(\theta)$ is increasing (so that profits and sender surplus are aligned) and let $s_f(\theta)=s_m(1-\theta)$ (symmetry). Now, the examination level is determined by the equality of marginal search cost with average surplus, i.e.,

$$C'(\phi)=\frac{\int_0^{\tilde{\theta}}s_f(\theta)\,d\theta}{n},$$

where again $n=M\tilde{\theta}$ by (10).

Now consider the social optimand:

$$W\left(\tilde{\theta},\phi\right)=\Pi\left(\tilde{\theta},\phi,n\right)+\frac{\int_0^{\tilde{\theta}}s_f(\theta)\,d\theta}{n}\phi-C(\phi)$$

The derivative with respect to ϕ reveals a negative externality that remains at the equilibrium solution:

$$\frac{dW}{d\phi}=\frac{\partial\Pi\left(\tilde{\theta},\phi,n\right)}{\partial\phi}>0,$$

which arises since a greater examination rate by receivers would yield a higher profit to senders. Second,

$$\frac{dW}{d\tilde{\theta}} = \frac{\partial \Pi \left(\tilde{\theta}, \phi, n\right)}{\partial n} M + \frac{1}{n} \left[s_f \left(\tilde{\theta}\right) - s_{av}\right] < 0,$$

where the first term is the negative one identified earlier, and the second is also negative since the marginal surplus is lower than the average one. This second effect reflects the fact that discrimination raises average quality (to the receiver) of messages received and so induces more examination, but the marginal sender does not internalize this effect.

The model sketched above has the motivation for targeting as saving on transmission costs. There are other variants of the model with different reasons for targeting, that may give different partitions of sender types across behavior. One such motivation is price discrimination. A sender may wish to target different types with different offers – as, famously, does American Express with its card feature offers. Then, the senders that gain most by differential offers may be the ones engaging in more market research. More broadly, the topic here applies not just to targeted mailing and buying of lists of names of appropriate demographics (or past positive response rate to similar ads, for example), but to the question of mass market appeal vs. niche market appeal. Mass market appeal means a blanket strategy, like indiscriminate mailing, advertising in a general interest magazine or on network television. Niche market products are likely to gain most from careful vetting of households for mailings, are more likely to advertise in specialized magazines or programs, etc.

It remains that the incentives to invest in targeting are likely to be socially insufficient. In concurrence with our general theme of congestion, an untargeted message creates a greater externality on other senders. In our simple version with ϕ fixed, untargeted messages are more in number since they hit all households, and so cause greater congestion. When we further allow for recipients to respond to expected quality of messages, the problem is exacerbated because untargeted messages are likely to be of lower expected benefit to the receiver. Hence these two externalities together suggest that not enough resources will be allocated to careful screening of appropriateness of receiver type. One factor that might mitigate this tendency for excess blanketing is that the demand for information is likely to be efficiently addressed to intermediaries, i.e., since there are significant returns to scale in gathering market research data that can be spread among many users, one would expect specialized firms to set up and deliver information at low cost. This could greatly reduce the potential inefficiencies. A careful modeling of pricing of providers of such information is needed here.

11 Conclusions

This paper has studied the economics of communication through unsolicited advertising. Receiver (consumer) attention is a scarce resource, but may be considered as "common property" by senders (advertisers), and so be over-utilized. Consumers also need to exert effort to process and absorb the content of unsolicited advertising. In determining how much effort to exert, they consider only the average benefit from the advertising sent. This market set-up is unlike standard markets in which the marginal agents on both sides determine the volume of transactions. The important policy implications from understanding these markets have to do with optimally restricting access to the consumers, for example through pricing or direct regulation.

In particular, junk email has gotten so out of hand that many people automatically trash many messages, or quarantine them through a spam filter. But this may prevent legitimate firms getting their message across. Bill Gates has suggested an email tax might help: "At perhaps a penny or less per item, e-mail postage wouldn't significantly dent the pocketbooks of people who send only a few messages a day. Not so for spammers who mail millions at a time." Indeed, some proposals (for example, by Ayres and Nalebuff, 2003) allow recipients to set their own rates: a college student might accept e-mail with a one-cent stamp; a busy chief executive might demand a dollar. One drawback is that this solution gives all the power to the recipient, who neglects advertiser benefits.

With junk mail, too, people do not pay much attention because the average message is of too poor quality. Raising the postage rate on bulk mail may have the advantageous effect of improving the allocation of resources, and through two sources. People recognize only the better offers will be sent, and therefore pay more attention, so getting better mail. This surprising possibility may raise more revenue for the Postal Service because firms are prepared to pay more to mail – more messages will be opened if they cost more to send (because sending then signals it must be a worthwhile offer).

One regulatory solution actually enacted in the case of telemarketing is the DNC list. Despite widespread success, one problem with this is that it does not account for advertiser surplus. The more general problem is that of unsolicited ads. The research here stresses the intricate nature of the market equilibrium: consumer expected surplus is determined by the average message quality and determines the attention to devote. The average message quality is determined by the number of messages absorbed and the expected profit of the marginal firm.

Different media differ by transmission cost. The fact that the sender has paid a higher price provides content information before the message is examined, giving credence to the maxim "The Medium is the Message." For example, two types of letters may arrive, with different postage costs, like registered and bulk mail. In equilibrium, different media are used by different firm types. Clearly spam email appears to be used by the lowest expected gross profit options, while telemarketing calls of notifications by FedEx presumably have higher returns than bulk mail. Thus the medium through which the message arrives may indicate useful information to the receiver, since different media cost the sender differently, and are subject to different congestion levels in equilibrium.

The current analysis is restrictive in several respects. We assumed a single receiver. Nevertheless, the main part of the analysis goes through with multiple types of receiver if the senders know who they are. The equilibrium and optimum then applies to each group separately. Different groups will then be in different circumstances as regards whether or not they examine all their messages. Some groups (with greater proclivity to examine, for example) will receive more messages, as will those with greater attractiveness for senders. The latter are likely to be the most congested: overfishing is greatest for the biggest fish. However, if the message transmission rate must be universally applied (as opposed to individually tailored) then an optimum analysis must balance out the marginal benefits across individual groups. For example, a marginal rise in the transmission rate will improve the welfare of some receivers who are currently in a congested state, while simultaneously reducing the welfare of others who are constrained by the number of messages received. For the latter, a rise in the transmission rate will make them worse off because they will get fewer messages when they already desired more.

Engaging market research firms to determine the consumer demographics most likely to buy one's product is costly, but benefits the firm by being able to target and adjust its advertising message. Along similar lines, senders also choose which media to employ for which consumer types, and different sectors tend to use different media. We show here that targeting is likely to be insufficient in equilibrium. Targeting also raises the issue of how the information about consumers is collected and the concomitant worries about consumer privacy. Hui and Png (2005) give a valuable survey on the Economics of Privacy; Hann, Hui, Lee, and Png (2005) introduce consumer avoidance (and marketers' consequent efforts to reach them) into an equilibrium marketing analysis. Targeting is a key ingredient in van Zandt's (2004) model of sender equilibrium, and integrating his analysis with the additional dimensions of consumer effort exertion and multiple sender messages may be a useful extension.

References

[1] Anderson, Simon P. and Coate, Stephen C. (2005) Market Provision of Broadcasting: A Welfare Analysis. *Review of Economic Studies*, 72, 947-972.

[2] Anderson, Simon P. and de Palma, Andre (2006) Shouting to be Heard in Competing for Attention: Sending Multiple Ads to Get the Message Through. Mimeo, University of Virginia.

[3] Anderson, Simon P. and de Palma, Andre (2004) The Economics of Pricing Parking. *Journal of Urban Economics*, 55, 1-20.

[4] Anderson, Simon P., de Palma, Andre, and Thisse, Jacques-François (1992): *Discrete Choice Theory of Product Differentiation*, M.I.T. Press, Cambridge, Mass.

[5] Anderson, Simon P. and Renault, Regis (1999): Pricing, Product Diversity and Search Costs: a Bertrand-Chamberlin-Diamond Model. *RAND Journal of Economics*, 30, 719-735.

[6] Ayres, Ian and Nalebuff, Barry (2003) Want to call me? Pay me! *Wall Street Journal,* October 8th, 2003.

[7] Bagwell, Kyle (2006) The Economic Analysis of Advertising, in *Handbook of Industrial Organization*, III, forthcoming, Elsevier Science, North Holland.

[8] Burdett, Kenneth and judd, Kenneth L. (1983) Equilibrium Distributions of Sales and Advertising Prices. *Review of Economic Studies*, 44, 465-491.

[9] Butters, Gerard R. (1977): Equilibrium Price Dispersion. *Econometrica*, 51, 955-969.

[10] Caplin Andrew, and Nalebuff, Barry (1991): Aggregation and Imperfect Competition: On the Existence of Equilibrium. *Econometrica*, 59, 25-59.

[11] Diamond, Peter A. (1971): A Model of Price Adjustment. *Journal of Economic Theory*, 3, 156-168.

[12] Dixit, Avinash K. and Norman, Victor D. (1978): Advertising and Welfare. *Bell Journal of Economics.* 9, 1-17.

[13] Engers, Maxim P. and Joshua Gans (1998): Why Referees are Not Paid (Enough), *American Economic Review*, 88, 1341-1349.

[14] Hann, Il-Horn, Hui, Kai-Lung, Lee, Sang-Yong T., and Png, Ivan P. L. (2005) Consumer Privacy and Marketing Avoidance. Working Paper, Department of Information Systems, National University of Singapore.

[15] Hui, Kai-Lung, and Png, Ivan P. L. (2005) The Economics of Privacy, in *Handbook of Information Systems and Economics*, Ed. Terry Hendershot, forthcoming, Elsevier Science, North Holland.

[16] McFadden, Daniel. (2001): Economic Choices. *American Economic Review*, 93, 3, 351-378.

[17] McLuhan, Marshall (1964) *Understanding Media.*

[18] McLuhan, Marshall and Quentin Fiore (1967): *The Medium is the Massage*, Random House, 2000 reprint by Gingko.

[19] Shapiro Carl. (1980). Comment on Dixit and Norman. *RAND Journal of Economics.*

[20] SHIRMAN, D. R. (1996) When E-mail becomes Junk Mail: the Welfare Implications of the Advancement of Communications Technology. *Review of Industrial Organization,* 11, 35-48.

[21] TOFFLER, ALVIN (1970). *Future Shock.* Random House.

[22] VAN ALSTYNE, MARSHALL, THEDE LODE, AND RICK WASH (2004): Information asymmetry and thwarting spam, *Center for eBusiness, Research Brief III(3),* 1-4, MIT Sloan School.

[23] VAN ZANDT, TIMOTHY (2004): Information Overload in a Network of Targeted Communication, *RAND Journal of Economics*, 35(3), 542-560.

[24] WILLMORE, C. (1999) A Penny for your Thoughts: E-mail and the Under-Valuation of Expert Time. *Working Paper,* University of Warwick.

12 Appendices

Proof of Lemma 1.

Proof. Suppose instead that the support of the messages sent is not the interval from a critical θ to 1. Indeed, suppose there exists some positive measure of senders with $\theta \in [\theta_3, \theta_4]$ that do not send messages, and another positive measure of senders with $\theta \in [\theta_1, \theta_2]$ that do send messages, with $\theta_2 \leq \theta_3$. Let the mass of messages sent at the purported sender transmission relation be n. Then, for all $\theta \in (\theta_1, \theta_2)$, it must be that $\frac{\phi}{n}\pi(\theta) > \gamma$. Since $\pi(\theta)$ is strictly higher for all $\theta \in [\theta_3, \theta_4]$ than for any $\theta \in (\theta_1, \theta_2)$, there exists some positive measure, μ, of senders from the interval $[\theta_3, \theta_4]$ that can transmit messages and earn profits $\frac{\phi}{n+\mu}\pi(\theta) > \gamma$. This contradicts the starting position that excludes senders in $[\theta_3, \theta_4]$. ■

Proof of Proposition 1

It remains to prove that $N(\phi)$ is increasing over the domain $\phi \in [0, \phi^c]$, that $N(\phi) > \phi$ for $n \in (0, \phi^c)$, and that $N(\phi) = \phi$ for $\phi \in \{0, \phi^c\}$. First note that $N(\phi)$ is given by (1) for $\phi \in [0, \phi^c]$. The first property then follows simply from writing (1) in inverse form as $\phi = \gamma\frac{n}{\pi(\theta^*)}$ and then noting the RHS is increasing in n (the denominator is decreasing since $\pi(.)$ is increasing in θ^* and θ^* is decreasing in n). To prove the second, note that for $\phi \in (0, \phi^c)$,

$$\gamma = \frac{\phi}{n}\pi(\theta^*) > \frac{\phi}{n}\pi\left(\pi^{-1}(\gamma)\right) = \frac{\phi}{n}\gamma,$$

where we recall that $\pi^{-1}(\gamma) = \theta^{\min}$. The third property follows from setting the solution on the LHS of (1) as $N(\phi) = \phi$. One solution is $\phi = 0$; the other has $\gamma = \pi(\theta^*)$, which defines $\theta^{\min}$ (and hence $n^{\max}$ and ϕ^c). Q.E.D.

Proof of Proposition 6.

If $C'(0) \geq s_1$, then if any messages were examined, type 1 senders would crowd out type 2 senders. Suppose not, and that some type 2 senders were transmitting. They would need to earn non-negative profits. But then further type 1 senders would transmit, up till the margin where profits for the type 1 senders were zero. This would mean that the type 2 senders would make losses, and therefore not transmit. Given then that only type 1 senders could transmit at any putative equilibrium with positive examination, the receiver would not wish to examine any messages since the marginal cost exceeds the marginal benefit from type 2 messages even at zero examination.

If $C'(0) < s_1$, the above argument again implies that only type 1 senders will transmit messages, and in this case there exists an equilibrium with positive examination. However, if $\pi_1 + s_1 < \pi_2 + s_2$, the socially optimal messages transmitted are of type 2 since they imbue higher social surplus. Q.E.D.

Piracy and competition*

Paul Belleflamme† Pierre M. Picard‡

September 30, 2005

Abstract

The effects of (private, small-scale) piracy on the pricing behavior of producers of information goods are studied within a unified model of vertical differentiation. Although information goods are assumed to be perfectly differentiated, demands are interdependent because the copying technology exhibits increasing returns to scale. We characterize the Bertrand-Nash equilibria in a duopoly. Comparing equilibrium prices to the prices set by a multiproduct monopolist, we show that competition drives prices up and may lead to price dispersion. Competition reduces total surplus in the short run but provides higher incentives to create in the long run.

JEL Classification Numbers: L13, L82, L86, K11, O34.

Keywords: Information goods, piracy, copyright, pricing.

*We are grateful to Rabah Amir, Nicolas Boccard and Xavier Wauthy for helpful discussions on the paper. We also thank seminar participants at SERCI (Boston, 2003), ESEM (Stockholm, 2003), ECARES (2005), CESIfo (2005) and EEA (Amsterdam, 2005) for their comments.

†Corresponding author: CORE and IAG, Université catholique de Louvain, 34 Voie du Roman Pays, B-1348 Louvain la Neuve, Belgium (tel: +32 10 47 82 91, fax: +32 10 47 43 01, e-mail: belleflamme@core.ucl.ac.be).

‡University of Manchester, School of Economics Studies, U.K. and CORE, Université catholique de Louvain, Belgium.

1 Introduction

Over the last decade, the fast penetration of the Internet and the increased digitization of information goods like music, movies and software have turned the issue of piracy into a topic of intense debate. Not surprisingly, economists have recently shown a renewed interest in information goods piracy.[1] Recent contributions revive the literature on the economics of copying and copyright, which was initiated some twenty years ago.[2] While early contributions focus on the effects of photocopying and examined how publishers can indirectly appropriate some revenues from illegitimate users (Novos and Waldman, 1984, Liebowitz, 1985, Johnson, 1985, and Besen and Kirby, 1989), later papers concentrate on the intellectual property (IP) protection and discuss the trade-off between the ex ante benefit of preserving the intellectual creation incentives and the ex post cost of the consequent restrain on the use of those goods (Landes and Posner, 1989, Besen and Raskind, 1991, and more recently Bae and Choi, 2004).[3]

Yet, the literature on the economics of copying abstracts away the strategic interaction among producers of information goods. Indeed, all the above mentioned contributions focus on the study of monopolies, with the exception of Johnson (1985) who considers price-taking producers of information goods. However, this perspective sharply contrasts with the reality of information good industries where production decisions are concentrated in the hands of a small number of major players. The oligopolistic nature of those industries leaves economic observers unsatisfied with the monopoly or price-taking representation of production under the threat of copying. In particular, such a representation is unable to relate to some important characteristics of information good industries

[1]See the excellent survey by Peitz and Waelbroeck (2003) and the references therein.

[2]For a recent survey (and extension) of this literature, see Watt (2000).

[3]From an *ex ante* point of view, IP protection preserves the incentive to create information goods, which are inherently public goods. On the other hand, IP rights encompass various potential inefficiencies from an *ex post* point of view. The protection grants de facto monopoly rights, which generates the standard deadweight losses; also, by inhibiting imitation, IP rights might limit the creators' ability to borrow from, or build upon, earlier works, and thereby increase the cost of producing new ideas.

such as the recent consolidation phases[4] and the existence of price dispersion.[5] The present paper aims to address those issues.

The aim of the present paper is to study the strategic interactions among the producers of information goods in the presence of piracy. In particular, we want to analyze how oligopolistic producers set the prices of their information goods when users are able to purchase copying devices and copy the goods. We also want to investigate the ex ante and ex post benefits of oligopolistic competition under this threat of copying. We finally want to relate those results to those obtained under the monopoly assumption (Bae and Choi, 2004) or under the perfect competition assumption (Johnson, 1985).

Our modeling strategy is the following. As usually done in the literature, we consider information goods with independent content.[6] All other things being equal, the demands for the (two, for simplicity) goods would be independent. However, we allow consumers to make lower-quality copies by using a technology that exhibits increasing returns to scale. The copying technology appears thus as a common substitute for the information goods. The main effect of such common substitute is to make the demands for the goods interdependent over some range of prices. In particular, the goods become complementary when their prices are similar enough and they remain independent otherwise. More precisely, the demand function for a particular information good typically exhibits three segments and two kinks. Each segment corresponds to a different category of consumers. We call these categories "buyers", "copiers", and "switchers". For the first two categories, the demand for a good does not depend on the price of the other good: indeed, whatever the price of the other good, "buyers" buy this other good

[4]In the music industry, the 50-50 joint venture of BMG and Sony was cleared in 2004 both by the European Commission and by the FTC. In the software industry, in September 2005, Oracle Corporation acquired Siebel Systems just one day after eBay acquired Skype. In the movie industry, Sony absorbed MGM and United Artists studios in 2005.

[5]For instance, Brynjolfsson and Smith (2000) observe prices for a matched set of 20 books and 20 CDs sold through conventional and Internet outlets, and report average price differences ranging from 25% for books to 33% for CDs. For an updated measure of price dispersion over a wider range of products (including computer software), see the ongoing research project of Michael Baye, John Morgan and Patrick Scholten on http://www.nash-equilibrium.com/.

[6]Think, for instance, of software applications for games and word processor, or CD recordings of classic and pop music.

and "copiers" copy this other good. In contrast, for "switchers", the best use of one good depends on the best use of the other good: if they purchase (copy) one good, they also purchase (copy) the other one. Therefore, the demand for one good depends on both prices.

Using this framework, we contrast the behavior of a multiproduct monopolist with the behavior of two Bertrand duopolists. We also perform a welfare analysis, both from a static and from a dynamic perspective. Our main results are the following.

As far as the multiproduct monopoly is concerned, we show that the firm may set different prices for its goods. Actually, the multiproduct monopolist may follow two strategies. The first strategy is to set close prices and target the demand by switchers. The second strategy is to set a high price and target the buyers in one market while setting a low price and targetting copiers in the other market. It turns out that the former strategy dominates the latter. The firm prefers to target switchers for whom only the sum of the two prices matters. Therefore, prices are nor unique neither symmetric. The monopolist might well set different prices for the two goods although consumers value these goods exactly in the same way. Yet, the two prices cannot be too distant so as to avoid that some consumers become buyers or copiers. This result can be interpreted as a first explanation for a form of limited price dispersion, confirming the empirical evidence.

In a duopoly, the interaction between firms leads to interesting properties. A firm's best-response to the price set by the competitor can depict up to four different attitudes. Because the nature of the marginal buyer suddenly changes as the competitor's price rises, the best response function shows discontinuities and equilibria in pure strategies cannot be guaranteed. Intuitively, the inexistence of equilibria stems from the firms' free-riding behavior with respect to the threat of piracy. If all firms take this threat seriously and quote low prices to accommodate consumers, then they set too low a price and there exists an opportunity for any individual firm to raise its price while keeping a sufficiently large demand and making a larger profit. Technically, increasing returns to scale in the copying technology introduce non-convexities in the profit functions and undermine the existence of market equilibrium.

Consequently, we have to distinguish between two regions of parameters. The first region corresponds to a sufficiently large cost of copying. In this region, there exists a symmetric equilibrium in pure strategies. At this equilibrium, both firms target switchers and quote identical prices. The interesting feature of this equilibrium is that the duopolists set a higher price than the (average) price of a multiproduct monopoly. This is so because the two goods are perfect complements over the segment of demand corresponding to the switchers. We observe thus a manifestation of the so-called 'Cournot effect' (Cournot, 1838). That is, the multiproduct monopolist has an incentive to decrease prices further than the duopolists do because it realizes that decreasing the price for one good increases demand for the other good by making copying less attractive. Although Johnson (1985) briefly mentions this effect, he does not analyze it in detail. As will be seen in the sequel, the existence of complementarities for some price ranges is important in understanding why the industry may not reach a single price equilibrium.

In the second region of parameters, where the fixed cost of copying is sufficiently low, an equilibrium in pure strategies fails to exist. Yet, we show that a symmetric equilibrium in mixed strategies exists. Each firm quotes two prices with positive probabilities. As price realizations may be different, we have an explanation for equilibrium price dispersion. Noteworthy is the fact that this explanation does not rely, as often proposed in the literature, on asymmetric information and search frictions (see for instance Varian, 1980 and Baye and Morgan, 2001). It is also interesting to note that the expected price in the mixed-strategy equilibrium, though smaller than the price that would prevail in the pure-strategy equilibrium, remains above the average price set by a multiproduct monopolist. The Cournot effect eases but does not disappear.

Finally, we perform a welfare analysis. Considering first *ex post* efficiency, we stress that industry concentration is welfare improving in the present context. Because of the Cournot effect, the multiproduct monopoly leads to larger consumer and producer surpluses than the duopoly. A merger is thus beneficial under the threat of copying because it eliminates the negative externality resulting from the lack of coordination about how firms should deter copying. This is a novel point that has not yet been considered by anti-trust agencies. We also assess the effects of policy measures aiming at strengthening IP protection; we show that increasing the cost of copying and decreasing the quality of copies do not have the same qualitative effects. Considering next *ex ante* efficiency, we compare our framework with an economy where only a single information good is available. This exercise allows us to measure the (gross) incentives to create a new information good. Whether those incentives are larger for an entrant or for an incumbent firm is not clear a priori. Indeed, the entrant's incentives are reduced by the free-riding effect observed in a duopoly, whereas the incumbent's incentives are reduced by a cannibalization effect (copying becomes more attractive as the number of goods increases). Yet, we can conclude in our framework that incentives to create are always higher for an entrant, i.e., if the ex post economy is organized as a duopoly. Therefore, although industry concentration improves welfare from a static perspective, it reduces welfare from a dynamic perspective. In other words, ex post competition can be seen as a necessary evil that enhances ex ante incentives to create. This conclusion turns on its head the traditional argument underlying IP protection, which considers ex post monopoly–and not competition–as the necessary evil.

To sum up, our main message is the following. The interactions between producers of information goods under the threat of piracy dramatically alter the equilibrium outcome compared to the outcome obtained under a one-good monopoly setting. Equilibrium prices in pure strategies may not exist and, if they do, they may be higher than those in the one-good monopoly case. Inferences about dynamics and welfare implications are not obvious anymore in oligopolistic industries. For instance, industry concentration may enhance static efficiency while being detrimental to dynamic efficiency.

The rest of the paper is organized as follows. In Section 2, we lay out the model and we derive the demand schedule for a particular original. In Section 3, we characterize the two-good monopoly case. In Section 4, we present the two-good duopoly case. In Section 5, we perform a welfare analysis. We conclude and propose an agenda for future research in the last section.

2 Demand for originals

There is a continuum of potential users who can consume at most two information goods. These information goods are assumed to be perfectly (horizontally) differentiated and equally valued by the users. In particular, users are characterized by their valuation, θ, for any information good. We assume that θ is uniformly distributed on the interval $[\underline{\theta}, \overline{\theta}]$, with $\underline{\theta} > 0$.

Each information good is imperfectly protected and thus "piratable". As a result, users can obtain each information good in two different ways: they can either *buy* the copyrighted product (an "original") or make a *copy* of the product. It is reasonable to assume that all users see the copy as a lower-quality alternative to the original.[7] Therefore, in the spirit of Mussa and Rosen (1978), we posit some vertical (quality) differentiation between the two variants of any information good: letting s_o and s_c denote, respectively, the quality of an original and a copy, we assume that $0 < s_c < s_o$.[8]

As for the relative cost of originals and copies, we let p_i denote the price of original i ($i = 1, 2$) and we assume that users have access to a copying technology with increasing returns to scale. To keep things simple, we assume that to be able to copy, consumers must incur a fixed cost $K > 0$. Finally, for the sake of the exposition, we further assume that all users prefer copying a single original over not using any information good:

A1: $\underline{\theta} s_c - K \geq 0$.

Assumption A1 is likely to be satisfied in industries like music, software and video because, in the digital era, (i) the quality of copies (s_c) is high, (ii) the fixed cost of copying (K) is low, and (iii) users have a high valuation (θ) for information goods.[9] This assumption greatly simplifies the exposition while it

[7] This assumption is common (see, e.g., Gayer and Shy, 2003) and may be justified in several ways. In the case of analog reproduction, copies represent poor substitutes to originals and are rather costly to distribute. Although this is no longer true for digital reproduction, originals might still provide users with a higher level of services, insofar as that they are bundled with valuable complementary products which can hardly be obtained otherwise.

[8] Similar models are used by Koboldt (1995) to consider commercial copying and by Yoon (2002) and Bae and Choi (2004) to analyze the market for a single information good.

[9] Another way to justify this assumption is to say that the fraction of users we choose not to

retains the main properties of the model. Indeed, it follows that all users will always consume both goods, either by purchasing the original or by copying it.

The demand function for good i is therefore derived as follows. One can write that a user with type θ buys original i iff

$$\theta s_o - p_i + \max\{\theta s_o - p_j; \theta s_c - K\} \geq \max\{\theta s_o - p_j + \theta s_c - K; 2\theta s_c - K\}. \quad (1)$$

This inequality compares user θ's value of purchasing original i, and either purchasing or copying good j, to the best option available given that he/she does not purchase original i, namely, copying good i whereas either buying or copying good j.

In the extreme case where $K = 0$ (copies are free), expression (1) rewrites as

$$\begin{aligned} \theta s_o - p_i + \max\{\theta s_o - p_j; \theta s_c\} &\geq \theta s_c + \max\{\theta s_o - p_j; \theta s_c\} \\ &\Longleftrightarrow \theta (s_o - s_c) \geq p_i. \end{aligned}$$

In that case, the demand for good i only depends on p_i: the two goods are independent.

However, for $K > 0$, increasing returns to scale in copying make the demands interdependent. Inequality (1) can take three different forms, each form corresponding to a specific category of users.

First, for high valuation users such that $\theta (s_o - s_c) \geq p_j$, expression (1) rewrites as

$$\theta s_o - p_i + \theta s_o - p_j \geq \theta s_o - p_j + \theta s_c - K \Longleftrightarrow \theta \geq \frac{p_i - K}{s_o - s_c}.$$

Because these users purchase the other original whether they purchase good i or copy it, we call them *'buyers'*. The maximum price they are willing to pay for original i is equal to

$$p_i^b (\theta) = \theta (s_o - s_c) + K.$$

consider (i.e., those for whom $\theta s_c - K < 0$) is constantly narrowing, as copying devices become widely and cheaply available and as the "moral barrier" to illegal copying is increasingly fading. The widespread use of copied music and software in less developed countries corroborates this assumption.

That is, they are willing to pay up to the extra value that an original brings on top of a copy, augmented by the cost of the copying technology (which they save once they decide to buy i instead of copying it).

Second, for intermediate valuation users such that $p_j - K \leq \theta(s_o - s_c) \leq p_j$, expression (1) rewrites as

$$\theta s_o - p_i + \theta s_o - p_j \geq 2\theta s_c - K \iff \theta \geq \frac{p_i + p_j - K}{2(s_o - s_c)}.$$

For these users, the best use of one good depends on the best use of the other good: if they purchase good i, they also purchase good j; if they copy good i, they also copy good j. We therefore call them *'switchers'*. How much are switchers willing to pay for good i? Going from two copies to two originals, they earn twice the extra value of an original compared to a copy, and they trade the cost of the copying technology for the price of the other original. So, their maximum price is given by:

$$p_i^s(\theta, p_j) = 2\theta(s_o - s_c) + K - p_j.$$

Finally, for low valuation users such that $\theta(s_o - s_c) \leq p_j - K$, expression (1) rewrites as

$$\theta s_o - p_i + \theta s_c - K \geq 2\theta s_c - K \iff \theta \geq \frac{p_i}{s_o - s_c}.$$

Because these users copy good j no matter what they decide about good i, we call them *'copiers'*. What they are willing to pay for good i is just the extra value of an original compared to a copy (for they have already sunk the cost of the copying technology). Their maximum price is thus equal to:

$$p_i^c(\theta) = \theta(s_o - s_c).$$

The three price functions are depicted in Figure 1. We observe that depending on the price of good j, the price function for good i can have up to two kinks. The price function is (increasing and) concave in θ in the neighborhood of $p_j/(s_o - s_c)$ (which separates switchers from buyers), and (increasing and) convex in θ in the neighborhood of $(p_j - K)/(s_o - s_c)$ (which separates copiers from switchers).

[Insert Figure 1 about here]

As a result, the demand function for good i has three segments and two kinks,

$$D_i(p_i, p_j) = \frac{1}{\overline{\theta} - \underline{\theta}} \times \begin{cases} \left(\overline{\theta} - \frac{p_i - K}{s}\right) & \text{if } \; p_j + K \leq p_i \text{ (buyers)} \\ \left(\overline{\theta} - \frac{p_i + p_j - K}{2s}\right) & \text{if } \; p_j - K \leq p_i < p_j + K \text{ (switchers)} \\ \left(\overline{\theta} - \frac{p_i}{s}\right) & \text{if } \; p_i < p_j - K \text{ (copiers)} \end{cases} \tag{2}$$

where, to ease the exposition, we denote by s the quality difference between originals and copies:

$$s \equiv s_o - s_c.$$

Two remarks are in order. First, as long as $K > 0$, there exists a range of prices for which goods are *complements*. Indeed, for $p_j - K \leq p_i < p_j + K$, an increase in the price of one good decreases the demand for the other. When p_j is close to p_i, the marginal users choose to copy according to the value of the bundle, $p_i + p_j$. For other prices, the goods remain independent: when p_j is relatively smaller than p_i, the marginal users are buyers; when p_j is relatively larger than p_i, the marginal users are copiers.

Second, as long as $K > 0$, the demand function has the same concave and convex kinks as the price function depicted above. These kinks result from the fact that the marginal consumer alters his/her behavior about copying at some pivot prices p_i. Similar kinks would appear for other distributions of consumers' types θ than the uniform distribution assumed here.

The case of price competition over perfect complements is well known since Cournot (1838). Yet, the present model differs from Cournot's case as complementarity only takes place over a limited range of prices. A similar property appears in Gabszewicz *et al.* (2001) who consider imperfect complements (i.e., the joint consumption of two products provides an extra utility but products can still be consumed individually). In our setting, however, complementarity is not built in consumers' preferences but stems, indirectly, from the existence of a common substitute.

We now analyze the pricing decisions in the monopoly and duopoly cases. To make this analysis relevant we assume that in any possible demand regime, decision makers never find it optimal to cover the whole market. Under the present demand system, this means that high valuation consumers are willing to pay significantly more for the goods than low valuation consumers. A sufficient condition is given by the following assumption:[10]

A2: $\overline{\theta}s > \underline{\theta}\,(s + s_o)$.

3 Multiproduct monopoly

Since Cournot (1838), the analysis of a multiproduct monopoly selling perfect complements is well known: the firm sets a unique price reflecting only the value of the bundle. Yet, in the present paper, the case of multiproduct monopoly deserves some attention. Indeed, as goods are complement only over a limited range of prices, it is not sure whether the firm's optimal prices will lie in that range. Indeed, when prices are close, the firm faces only switchers. By contrast, when prices differ by more than the fixed cost K, the firm faces two groups of consumers: copiers and buyers. The firm may follow two strategies. On the one hand it may set close prices and target the demand by switchers. On the other hand, the firm may set a high price and target the buyers in one market while it sets a low price in the other market and targets the copiers, the prices being sufficiently distant to avoid some consumers to become switchers. Doing so, the firm is able to collect high revenues on the buyers who have a high willingness to pay. It nevertheless turns out that the latter strategy is not optimal.

Formally, the monopoly chooses prices p_1 and p_2 so as to maximize profits:

$$\max_{p_1,p_2} \pi_m = p_1 D_1\left(p_1, p_2\right) + p_2 D_2\left(p_1, p_2\right).$$

where demands are given by (2) and where the firm is assumed to have zero production cost.

Proposition 1 *The multiproduct monopolist sets any price* (p_1, p_2) *such that* $(p_1 + p_2)/2 = p_m \equiv \overline{\theta}s/2 + K/4$ *and* $p_2 - K < p_1 \leq p_2$.

Proof. See Appendix 7.1. ■

[10]Together, Assumptions A1 and A2 give a fair representation of the reality as they guarantee that there are always some consumers who pirate at least one good.

We prove in the appendix that the strategy where the monopolist targets buyers and copiers on separate markets is dominated because the optimal prices on each market are not distant enough (i.e., they do not satisfy $|p_i - p_j| > K$).[11] As a result, the monopolist sells the two goods at prices such that marginal buyers are switchers. The difference between the two prices is limited upward to avoid that marginal users become copiers or buyers.

It is important to note that prices are nor unique neither symmetric. The monopolist might well set different prices for the two goods although consumers value these goods exactly in the same way. It is because the copying technology offers a common substitute for the goods and because the monopolist prefers to target switchers that only the sum of the two prices matters for the consumers, as well as for the monopolist. Our setting slightly contrasts Cournot's (1838) discussion about perfect complements because products are here not genuine complements. Henceforth, the monopolist is constrained to set its prices in some price range. For instance, the firm has no possibility to set a zero price for one good and to collect its revenue on the other as it would be possible for perfect complements. Consumers are likely to take the free good and copy the other good.

Several additional comments are in order. First, Assumption A1 implies that the monopolist sets an average price p_m which is smaller than the price it would set for each good if there were no threat of copy, namely $\overline{\theta}s_o/2$; profits are also smaller. Second, it is easily seen that p_m decreases as copies become a closer substitute for originals (i.e., if K decreases and/or s_c increases). Finally, in the extreme case where $K = 0$, the demands for the two goods are independent and the monopolists sets the same price for both: $p_1 = p_2 = \overline{\theta}s/2$.

We now examine the case of the duopoly.

4 Duopoly

Under a duopoly, each information good $i \in \{1, 2\}$ is produced and sold by a separate firm. In the limiting case where $K = 0$, demands are independent and the producers act as local monopolists; they set the same prices as the

[11] It is possible to show that this result holds for any distribution of consumer types.

multiproduct monopolist would do in this situation: $p_1 = p_2 = \overline{\theta}s/2$. However, for $K > 0$, the interdependence between the demand functions is a source of strategic interaction. To analyze this interaction, we proceed in two steps: first, we derive firm i's best reponse and then we compute the Bertrand-Nash price equilibria.

4.1 Best response function

Best response functions are derived from the demand functions (2). Because the demand functions are piece-wise linear and include a convex kink, firms' best response functions are expected to be discontinuous. In fact, the point of discontinuity will take place when marginal users shift from being switchers to copiers. We now characterize the portion of the best response of firm i below and above the discontinuity.

Targeting buyers or switchers? The optimal price on the buyers of good i is equal to

$$p_i^{b*} \equiv \arg\max_{p_i} p_i \left(\overline{\theta} - \frac{p_i - K}{s}\right) = \frac{1}{2}\left(\overline{\theta}s + K\right).$$

Firm i's best response is to set $p_i = p_i^{b*}$ as long as the competitor's price does not to entice the marginal consumer to become a switcher. Using (2), this is so as long as

$$p_j \leq p_i^{b*} - K \iff p_j \leq p^f \equiv \frac{1}{2}\left(\overline{\theta}s - K\right).$$

Note that $p^f > 0$ under Assumptions A1 and A2.

For $p_j > p^f$, some low valuation users are enticed to switch to copying. Firm i can either accommodate these switching users by lowering its price, or it can avoid them and concentrate on higher valuation users by increasing its price. On the one hand, when p_j is low enough, firm i sets a 'limit price' to 'deter' switchers. By (2), it sets a price equal to

$$p_i^D(p_j) = p_j + K$$

(or just a small amount below this price) and achieves a corresponding profit of $\pi_i^D(p_j)$. This price is an increasing function of p_j. Since more users tend to switch to the copying technology when the competitor raises its price p_j, firm i must raise its price p_i to avoid the switchers. Hence, there exists a range of prices such that prices are *strategic complements.*

When p_j gets larger, firm i has no other choice but to accommodate switchers. It sets a price equal to $p_i^{s*}(p_j)$ where

$$p_i^{s*}(p_j) \equiv \arg\max_{p_i} p_i \left(\overline{\theta} - \frac{p_i + p_j - K}{2s}\right) = \overline{\theta}s + \frac{K - p_j}{2},$$

and achieves a corresponding profit of $\pi_i^{s*}(p_j)$. The price $p_i^{s*}(p_j)$ is a decreasing function of the competitor's price; prices are then *strategic substitutes* in this range of prices.

The transition between deterrence and accommodation of switchers takes place at the price p^d such that deterence and accommodation of switchers yield the same profit and thus the same price: $p_i^D(p^d) = p_i^{s*}(p^d)$, or equivalently

$$p^d = \frac{1}{3}\left(2\overline{\theta}s - K\right) > p^f.$$

Targeting copiers? Because of the convex kink in the demand function, the shift from switchers to copiers has to be analyzed by comparing profit levels. The optimal price and profit on copiers are equal to:

$$p_i^{c*} \equiv \arg\max_{p_i} p_i \left(\overline{\theta} - \frac{p_i}{s}\right) = \frac{1}{2}\overline{\theta}s \text{ and } \pi_i^{c*} = \frac{1}{4}\overline{\theta}^2 s.$$

We readily get that

$$\pi_i^{s*}(p_j) > \pi_i^{c*} \iff p_j < p^e \equiv \left(2 - \sqrt{2}\right)\overline{\theta}s + K.$$

The regime including accommodation of switchers is part of the best response function as long as

$$p^d < p^e \iff K > \frac{3\sqrt{2} - 4}{4}\overline{\theta}s. \tag{3}$$

In this case there exits a downard jump at $p_j = p^e$.

Otherwise, accommodation of switchers is not part of the best response function and the latter has a downward jump from deterrence of switchers to accommodation of copiers for another price $p_j = p^{e\prime}$, where $\pi_i^D(p^{e\prime}) = \pi_i^{c*}$, which is equivalent to

$$p^{e\prime} = \frac{1}{2}\left(\overline{\theta}s - K\right) + \frac{1}{2}\sqrt{K\left(2\overline{\theta}s + K\right)}.$$

Summarizing our results, we have that, under condition (3), the best response function is given by

$$p_i^*\left(p_j\right) = \begin{cases} p_i^{b*} & \text{if } p_j \leq p^f, \\ p_i^D\left(p_j\right) & \text{if } p^f \leq p_j \leq p^d, \\ p_i^{s*}\left(p_j\right) & \text{if } p^d \leq p_j \leq p^e, \\ p_i^{c*} & \text{if } p_j > p^e. \end{cases}$$

Otherwise it is given by

$$p_i^*\left(p_j\right) = \begin{cases} p_i^{b*} & \text{if } p_j \leq p^f, \\ p_i^D\left(p_j\right) & \text{if } p^f \leq p_j \leq p^{e\prime}, \\ p_i^{c*} & \text{if } p_j > p^{e\prime}. \end{cases}$$

Figure 2 displays these functions (in black for firms 1 and in grey for firm 2) for 'high' and 'low' fixed cost of copying (resp. in the left- and right-hand panel).

[Insert Figure 2 about here]

4.2 Existence of equilibria in pure strategies

Because of discontinuities in the best response functions, equilibria in pure strategies might fail to exist. Intuitively, the possible inexistence of equilibria stems from firms' free-riding behavior with respect to the threat of copying. If both firms take this threat seriously and quote low prices to accommodate copiers, then there exists an opportunity for either firm to raise its price while keeping a sufficiently large demand and making a larger profit. This situation is shown in the right-hand panel of Figure 2 where best response functions do no intersect. By contrast, the left-hand panel shows the situation where firms reach an equilibrium as their best response function intersect at a symmetric equilibrium. More formally, we can state the following proposition.

Proposition 2 *There exists a unique symmetric Nash equilibrium in which both firms focus on switchers and set the price* $p_S \equiv \frac{1}{3}\left(2\overline{\theta}s + K\right)$ *if and only if*

$$K > \widehat{K} \equiv \frac{3\sqrt{2}-4}{2}\overline{\theta}s. \tag{4}$$

Otherwise, there is no Nash equilibrium in pure strategies.

Proof. See Appendix 7.2. ■

Proposition 2 tells us that the only possible Nash equilibrium in pure strategies is symmetric and is such that both firms target switchers. This result could not have been guessed at the outset. Moreover, we also see that the market fails to reach an equilibrium for small fixed costs of copying because the price p_S and the profit associated to this strategy decrease with K. For a low enough value of K, profits under accommodation of copiers become more attractive and firms tend to cut their price to p_i^{c*}. As a result, the absence of duopoly equilibria for low fixed costs of the copying technology casts some doubts about traditional analyses of the threat of copying in one-good monopoly settings.

4.3 Pure-strategy equilibrium: Cournot effect

We first focus on the situation in which the market reaches a symmetric equilibrium. One observes that

$$p_S = \frac{1}{3}\left(2\overline{\theta}s + K\right) > p_m = \frac{1}{4}\left(2\overline{\theta}s + K\right).$$

Corollary 4.1 *The price set by duopolists at the pure-strategy equilibrium is* higher *than the average price set by the multiproduct monopolist.*

At the symmetric equilibrium in pure strategies, the duopolists focus on switchers. In such a case, the copying technology constitutes a common substitute for their original good. The presence of this common substitute turns goods i and j (which are a priori perfectly horizontally differentiated) into complementary goods. As a result, the so-called 'Cournot effect' (Cournot, 1838) applies. That is, the multiproduct monopolist has an incentive to decrease prices further than the duopolists do because it realizes that decreasing the price for one good increases demand for the other good by making copying less attractive.

The externality that each firm imposes on the other can be quite important. Indeed, if the quality of copies is sufficiently low, the duopolists end up setting *prices larger than the price they would set under no threat of copying.* To see this, note that a user $\theta \in [\underline{\theta}, \overline{\theta}]$ who purchases the original good i gets a utility of $\theta s_o - p_i$. If copying is not an option, the demand for this good is simply equal to $D_i(p_i) = \left(\overline{\theta} - p_i/s_o\right) / \left(\overline{\theta} - \underline{\theta}\right)$ and is independent of the demand for the other good. Firms in duopoly and monopoly thus set the same price, the monopoly price: $p_M = \frac{1}{2}\overline{\theta}s_o$. Comparing it to p_S, we observe that the price set by duopolists at the pure-strategy equilibrium is *higher* than the monopoly price under no threat of copying if and only if $K \geq \frac{1}{2}\overline{\theta}\,(4s_c - s_o)$, which is clearly satisfied if $s_c < s_o/4$.

The last two findings qualify the argument that the threat of piracy forces firms to lower their prices and that the usage of the copyrighted product increases with piracy. Instead, our model gives some evidence to the common claim of copyright holders, who assert that piracy reduces their sales. These results also cast some doubt about the social benefits of stronger competition in information good markets that are subject to potential piracy. The last two findings indeed suggest that a more concentrated industry is better equipped to provide surplus both to legal consumers and to producers. We consider the latter issue in more detail in Section 5.

4.4 Mixed-strategy equilibria: price dispersion

When $K < \widehat{K}$, there exist no equilibrium in pure strategies. Nevertheless, by Glicksberg (1952), there exists a mixed-strategy Nash equilibrium because profits are continuous. Therefore, firms randomize prices at equilibrium when the cost of the copying technology is low enough. As a result, price dispersion can be observed. In this section, we first characterize a simple and intuitive class of mixed-strategy equilibria; we then discuss the impact of the copying technology on price dispersion.

As in Boccard and Wauthy (1997, 2003), the piecewise linearity of the demand function allows us to show that firms do not use continuous densities. This allows us to focus on mixed strategy equilibria in which firms play two prices with the same probability distributions. The following proposition shows that such equilibria exist provided that fixed costs are not too small.

Proposition 3 *When $\widehat{K} > K > \tilde{K} \equiv 0.0274\overline{\theta}s$, there exists an equilibrium where firms randomize between the prices*

$$p_a = \frac{2\overline{\theta}s + xK}{4-x} \quad \textit{and} \quad p_b = \frac{2\overline{\theta}s + (x+1)\,K}{x+3} > p_a$$

with probabilities x and $1-x$. The probability x is equal to zero when K is equal to $\widehat{K}$, it increases when K decreases below $\widehat{K}$ and it is equal to $x = 0.3603$ when K tends to $\tilde{K}$. Prices are such that $p_b > p_a + K$.

Proof. See Appendix 7.3. ■

Unfortunately the probability x has no explicit expression. Numerical simulations show that for any admissible set of parameters, the probability x monotonically increases when K falls from $\widehat{K}$ to $\tilde{K}$. When $K < \tilde{K}$, symmetric mixed strategies with two prices are not equilibria; firms have to randomize over a larger number of prices. The characterization of such equilibria goes beyond the scope of this paper.

Because $p_b > p_a + K$, the mixed-strategy equilibrium yields ex post realizations that include the three regimes with a positive probability. Each firm faces switchers when its price realization is equal to the other firm's realization; when price realizations are different, a firm faces buyers when it quotes the highest price and copiers when it quotes the lowest price. Hence, by playing mixed strategies, firms are able to avoid the negative Cournot effect of competing on a segment of demand where goods are perfect complements. Indeed, with probability $x(1-x)$, they set the prices (p_b, p_a) and the firm setting p_b collects revenues on high-valuation consumers (i.e., the "buyers").

One can check that prices can be ranked as in the following corollary.

Corollary 4.2 *(i) Prices are ranked as follows: $p_i^{c*} < p_a < p_a + K < p_b < p_S$. (ii) The expected price is larger than the average monopoly price: $xp_a + (1-x)p_b > p_m$.*

Figure 3 illustrates these results. The intuition behind part (i) goes as follows: firm i has no incentive to set prices below p_i^{c*} because, if it does, it gets a positive marginal revenue irrespective of firm j's mixed strategy; similarly, it gets a *negative* marginal revenue whenever it sets a price *above* p_S. This result demonstrates that the Cournot effect is less acute than in the pure-strategy equilibrium. However, part (ii) shows that the Cournot effect does not disappear completely: the expected price at the mixed-strategy equilibrium is still larger than the average price set by a multiproduct monopolist. In other words, firms would like to coordinate on a limit-price to eliminate piracy but are unable to commit to do so as they prefer to free-ride on the rival's effort. Therefore, they choose random prices as a way to share the burden of deterring piracy. Yet, this joint effort remains insufficient.

[Insert Figure 3 about here]

The absence of a pure-strategy equilibrium in prices is often presented as an explanation for *equilibrium price dispersion*. In consumer search models (see, e.g., Varian, 1980 and Baye and Morgan, 2001), equilibrium price dispersion arises because firms are tempted to lower price to attract informed consumers, but realize that if they do so, they forego rents from uninformed consumers; therefore, each firm intentionally randomizes prices to reduce the ability of rival firms to undercut its own price.

Our model generates similar price dispersion results and shares a similar intuition. Yet, it does not rely on search frictions and information asymmetries because our consumers are perfectly informed about prices and, therefore, do not need to search. Here, firms also face the tension between lowering price to attract low-valuation users (the copiers) and increasing price to extract rents from high-valuation consumers (the buyers). The difference with search models is that, in our setting, the existence of various classes of consumers results, endogenously, from the presence of the copying technology. In fact, price dispersion is even a prerequisite for two of these classes to exist at equilibrium; indeed, all consumers are switchers when prices do not differ by more than K.

A number of empirical studies document significant and persistent price dispersion on markets for information goods.[12] The previous argument relates such price dispersion to piracy. It is interesting to study the impact of the copying technology on price dispersion. On the one hand, supposing that firms quote two prices, it can readily be shown that the price dispersion $p_b - p_a$ decreases when K falls. On the other hand, when firms quote more than two prices, we can also show that the range $[p_i^{c*}, p_S]$, within which price atoms must lie, also shrinks as K decreases. Hence, the model predicts that price dispersion is likely to be lower

[12]See, e.g., Bailey (1998), Brynjolfsson and Smith (2000), and Iyer and Pazgal (2003).

for information goods which are more exposed to piracy.[13]

5 Welfare analysis

As indicated in Section 1, the economics of IP protection discusses the trade-off between *ex ante* and *ex post* efficiency considerations: to remedy the long-run underproduction problem that might arise from insufficient incentives to create, the law grants exclusive rights to creators, which entail a short-run underutilization problem.

Our simple framework allows us to shed some new light on this policy debate. First, in a short-run perspective, we can perform comparative statics exercises to assess the effects of stronger IP protection; we can also compare the welfare performances of two market structures, namely a multiproduct monopoly versus a duopoly. Second, in a long-run perspective, we can measure incentives to create and compare again the relative merits of monopoly and duopoly. As we now explain, policy implications are not clear cut. Indeed, *from a static perspective, the multiproduct monopoly enhances welfare with respect to the duopoly, but from a dynamic perspective, the duopoly provides higher incentives to create than the monopoly.*

5.1 Ex post efficiency considerations

In many discussions, the protection of IP rights calls for an increase in the cost of piracy (Novos and Waldman (1984), Yoon (2002), etc). In this model, this would call for two policy measures: first, one can increase the fixed cost of the copying technology K by, e.g., applying a tax on the reproduction devices; second, one can take actions to decrease the value of a copy s_c. Therefore, we assess the effects of a change in K and in s_c for the multiproduct monopoly and for the duopoly. As for the duopoly, we restrict for now our attention to the case where K is sufficently large so that an equilibrium in pure strategies exists. We consider the mixed-strategy equilibrium case below using numerical simulations.

[13]Recall, nevertheless, that no dispersion is observed for $K > \hat{K}$ (as a unique pure-strategy equilibrium exists).

When a pure-strategy equilibrium exists in the duopoly, we can easily analyze the effects of a marginal strengthening of IP rights. Let $\theta_S \equiv \left(4\overline{\theta}s - K\right) / (6s)$ be the type of the switching user at the equilibrium prices p_S. This is the lowest type amongst the consumers who purchase an original good. Users with type $\theta \in [\theta_S, \overline{\theta}]$ purchase an original whereas, by Assumption A1, users with type $\theta \in [\underline{\theta}, \theta_S)$ make use of copies (we call them 'pirates'). Similarly, in the multiproduct monopoly, let $\theta_m \equiv (2p_m - K) / (2s) = \left(2\overline{\theta}s - K\right) / (4s)$ denote the marginal user at price p_m.

We do not need to perform separate comparative statics analyses for the monopoly and for the duopoly as the equilibrium price and demand vary in the same direction in the two cases. We observe that the two policy measures have different effects. On the one hand, a rise in the copying cost ($dK > 0$) implies an upward parallel shift of the demand for original goods. It then increases equilibrium prices ($dp_S, dp_m > 0$) and increases the set of consumers buying original goods ($d\theta_S, d\theta_m < 0$). The number of pirates falls. On the other hand, a deterioration of the value of copies ($ds_c < 0$) implies a rotation of the demand, thereby reducing its elasticity. The deterioration of the value of copies then leads to an increase in price ($dp_S, dp_m > 0$) and to a reduction of the set of consumers buying original goods ($d\theta_S, d\theta_m > 0$). The number of pirates increases.

In Appendix 7.4, we formally establish the following results, which generalize those obtained by Bae and Choi (2004) for a single-product monopoly.[14] First, *the two ways of strengthening IP protection increase (optimal or equilibrium) profits.* This result is trivial for an increase in K as both the price and the quantity demanded increase. As for a decrease in s_c, price and quantity move in opposite directions, meaning that the result is a priori ambiguous. However, under Assumptions A1 and A2, it can be shown that the positive effect on price outweighs the negative effect on demand.

Second, *the two ways of strengthening IP protection decrease consumer surplus.* To see this, we observe that both categories of users suffer from a stronger protection: first and obviously, pirates are negatively affected by the decrease

[14] What Bae and Choi (2004) call the "copying regime" corresponds to the regime faced by our multiproduct monopolist and by our duopolists around the symmetric pure-strategy equilibrium.

in the attractiveness of copies (either through a larger cost or through a lower quality); second, buyers of originals are charged a higher price (as $dK > 0$ and $ds_c < 0$ both imply $dp_S > 0$ and $dp_m > 0$).

Finally, we evaluate the effect on social surplus by adding the two previous effects. The overall effect is negative if the additional costs imposed on all pirates is smaller than the additional revenue from having an additional legal consumer. This is so when the policy measure consists in deteriorating the copies ($ds_c < 0$): *strenghtening IP protection by making copies less valuable decreases social surplus.* On the other hand, when the policy measure consists in an increase in the fixed cost of copying ($dK > 0$), the sign of the welfare change is ambiguous. We show in the appendix that *strenghtening IP protection by making copies more costly decreases social surplus if K is not too large; it increases social surplus otherwise.*

To close our static welfare analysis, we repeat the conclusion we drew in the previous section: the multiproduct monopolist always sets lower prices than duopolists do; as the monopolist also achieves higher profits, *social surplus is undoubtedly higher under a multiproduct monopoly than under a duopoly.*

Finally, we provide results from numerical simulations pointing that some of the previous results might not hold in the region of parameters where symmetric mixed strategies with two prices exist. For the sake of the exposition, we concentrate on the impact of K. Figure 4 depicts an example that shows one noticeable difference with the symmetric equilibrium regime: as K increases, expected duopoly profits ($E\pi$) *decrease* in some range. The explanation is the following: as K increases, the prices p_a and p_b move in opposite directions and therefore, although the expected price increases, expected profits decrease.

[Insert Figure 4 about here]

5.2 Ex ante efficiency considerations

Comparing our framework with an economy where only a single information good is available allows us to measure the (gross) incentive to create a new information good. In the previous model, if only one information good is available instead of two, it is easy to see that under Assumption A1, the producer of this good only faces 'buyers'. The demand function is thus given by $D_i(p_i) =$

$\left(\overline{\theta} - (p_i - K)/s\right) / \left(\overline{\theta} - \underline{\theta}\right)$, and the optimal price and profit are given by $p_i^{b*} = \left(\overline{\theta}s + K\right)/2$ and $\pi_i^{b*} = \left(\overline{\theta}s + K\right)^2 / (4s)$.

There are two cases to consider when going from one to two goods. A first possibility is that the new good is created by an incumbent firm that already produces the extant good; the ex post economy is then organised as a multiproduct monopoly. From Proposition 1, we know that the multiproduct monopolist's optimal (average) price and profit are given by $p_m = \left(2\overline{\theta}s + K\right)/3$ and $\pi_m = \left(2\overline{\theta}s + K\right)^2 / (8s)$. One readily observes that, under Assumptions A1 and A2, we have that $\pi_m > \pi_i^{b*}$, meaning that an incumbent firm has a gross incentive to introduce a second good. Still, although goods are genuinely independent, the profit per good decreases when the number of goods rises: $\pi_m < 2\pi_i^{b*}$. Indeed, *the monopolist jeopardizes the sales of the first good when it introduces the second good.* Copying becomes more attractive when the number of goods is larger and the firm is compelled to reduce the average price of originals.

Alternatively, the new good could be created by an entrant firm, turning the ex post economy into a duopoly. Supposing for now that the condition of Proposition 2 is met, prices and profits at the pure-strategy equilibrium are given by $p_S \equiv \left(2\overline{\theta}s + K\right)/3$ and $\pi_S = \left(2\overline{\theta}s + K\right)^2 / (18s)$.

Let us now compare the two scenarios. Comparing prices, it is easily checked that Assumptions A1 and A2 imply the following ranking: $p_m < p_i^{b*} < p_S$. Therefore, *the average price decreases when the new good is introduced by an incumbent firm, whereas it increases when the new good is introduced by an entrant.* This is another illustration of the negative externality independent producers impose on each other, and on consumers, in the presence of copying.

Next, comparing profits, we can gauge the (gross) incentive to create in the two settings. We say that an entrant has higher incentives to introduce a second good than an incumbent if the following condition is met: $\pi_S > \pi_m - \pi_i^{b*}$. Under no threat of piracy, goods are independent and incentives for the incumbent and the entrant are exactly equal. However, under piracy, the question is whether the free-riding problem between duopolists harms the entrant to a greater or to a lesser extent than the cannibalization effect hurts the incumbent. In Appendix 7.4, we show that *when a pure-strategy equilibrium exists in the duopoly game, the duopoly always yields larger incentives to create than the multiproduct monopoly.*

Formally, we show that $K > \hat{K}$ implies that $\pi_S > \pi_m - \pi_i^{b*}$: the cannibalization effect is stronger than the free-riding effect.

As illustrated in Figure 4, the previous conclusion may still apply in the region of parameters where symmetric mixed-strategy equilibria with two prices exist.

6 Conclusion

In this paper, we qualify the traditional results and insights about the impact of piracy obtained in a one-good monopoly setting. When there exist more than one information good, increasing returns to scale in the copying technology create an interdependence between the demands for information goods, which would be independent otherwise. We first show that a multiproduct monopoly may set different prices for its goods. We then show that two-product duopolies are subject to free-riding behaviors with respect to the threat of piracy. If the two firms take this threat seriously by quoting low prices, then there exists an opportunity for a firm to take advantage of this situation and to raise its price. This can lead to the absence of an equilibrium in pure strategies if the fixed cost of copying is low enough. In this case, firms may randomize between several prices. To the best of our knowledge, this is the first contribution showing that price dispersion in information good industries can be generated by the presence of piracy. When the fixed cost of copying is not too small, the market can yield a symmetric equilibrium with prices that are larger than the (average) price of the multiproduct monopoly. Furthermore, those prices can even become larger than the price of a monopoly which faces no threat of piracy. The externality that firms impose on each other can therefore be quite important and it can drastically reduce the demand for legal copies.

In sum, the interactions between producers of information goods under the threat of piracy dramatically alter the equilibrium outcome compared to the outcome obtained under a one-good monopoly setting. Taking those interactions into account also yields surprising welfare implications: concentration appears as welfare-enhancing from a static perspective but welfare-detrimental from a dynamic perspective.

The present model suggests several avenues of future research. First, the current study is limited to the production of two perfectly differentiated information goods. It would be worthwhile to explore the pricing decisions and welfare aspects under piracy threat in a setting with more numerous and less differentiated varieties. Second, by assuming exogenous production and pricing of the copying technology, the current model sets aside the strategic issue of integration between the creators (or distributors) of information goods and the sellers of copying devices. It seems natural to investigate about the competition and welfare implications of such integration processes.

7 Appendices

7.1 Proof of Proposition 1

Suppose w.l.o.g. that $p_1 \leq p_2$. Then, the monopolist gets the following profits according to whether its two prices significantly differ or not:[15]

$$\begin{aligned} \text{either } \max_{p_1,p_2} \pi_m^{(1)} &= p_1\left(\overline{\theta} - \frac{p_1}{s}\right) + p_2\left(\overline{\theta} - \frac{p_2 - K}{s}\right) \text{ s.t. } p_1 \leq p_2 - K \\ \text{or } \max_{p_1,p_2} \pi_m^{(2)} &= (p_1 + p_2)\left(\overline{\theta} - \frac{p_1 + p_2 - K}{2s}\right) \text{ s.t. } p_1 \geq p_2 - K. \end{aligned}$$

The unconstrained solution to the first problem is $p_1 = \overline{\theta}s/2$ and $p_2 = \overline{\theta}s/2+K/2$. This solution does not meet the constraint because $p_1 > p_2 - K$. The solution of the first problem is thus the corner solution with $p_1 = p_2 - K$ and profit equal to

$$\pi_m^{(1)} = \frac{\overline{\theta}}{2}\left(\overline{\theta}s + K\right).$$

The second problem is equivalent to

$$\max_p \pi_m = 2p\left(\overline{\theta} - \frac{2p - K}{2s}\right)$$

where $p \equiv p_1 + p_2$. Optimal price and profit are easily found as

$$p_m = \frac{\overline{\theta}}{2}s + \frac{K}{4} \text{ and } \pi_m^{(2)} = \frac{\left(2\overline{\theta}s + K\right)^2}{8s}.$$

[15]Profits are actually multiplied by the constant $\left(\overline{\theta} - \underline{\theta}\right)$, which we forget from now on as it does not affect optimal decisions.

Noting that $\pi_m^{(2)} - \pi_m^{(1)} = K^2/(8s) > 0$ completes the proof. The second problem includes an infinity of prices (p_1, p_2) such that $(p_1 + p_2)/2 = p_m$ subject to the contraint set in this second problem: $p_1 \geq p_2 - K$.

One can check that this solution is interior under Assumptions A1 and A2.

7.2 Proof of Proposition 2

We first show that the optimal prices p_i^{b*}, $p_i^{s*}(p_j)$ and p_i^{c*} are interior solutions and do not lead to full market coverage under Assumptions A1 and A2. Indeed, the price p_i^{b*} is an interior solution iff $p_i^{b*} > p_i^b(\underline{\theta}) \iff K < (\overline{\theta} - 2\underline{\theta})\, s$, which is true under A1 and A2. The price $p_i^{s*}(p_j)$ with $p_j \geq 0$ is an interior solution iff $p_i^s(\underline{\theta}, p_j) < p_i^{s*}(p_j) \iff K < p_j + 2(\overline{\theta} - 2\underline{\theta})(s_o - s_c)$, which is less stringent than the previous condition. Finally the price p_i^{c*} is an interior solution iff $p_i^{c*} > p_i^c(\underline{\theta}) \iff \overline{\theta} - 2\underline{\theta} > 0$, which follows from A2.

Each best response function $p_i^*(\cdot)$ and $p_j^*(\cdot)$ can have four segments. Removing symmetric configurations, we need to check the existence of a pure strategy equilibria for the 10 following configurations. For some configurations we will need to distinguish equilibrium conditions in which the switcher's branch $p_j^{s*}(\cdot)$ exists (i.e. $p^d \leq p^e$ or condition (3)) or in which it does not (i.e. $p^d > p^e$ or the reverse of condition (3)).

1. The configuration (p_i^{b*}, p_j^{b*}) cannot be an equilibrium because $p_j^{b*} > p^f$ and thus the best response of i cannot be equal to $p_i^{b*} : p_i^*(p_j^{b*}) \neq p_i^{b*}$.

2. The configuration $(p_i^D(\cdot), p_j^D(\cdot))$ cannot be an equilibrium because the system $p_i = p_i^D(p_j)$ and $p_j = p_j^D(p_i)$ has no solution.

3. The configuration $(p_i^{s*}(\cdot), p_j^{s*}(\cdot))$ is an equilibrium if and only if $K > \widehat{K} \equiv \frac{3\sqrt{2}-4}{2}\overline{\theta}s$. Indeed, solving the system $p_i = p_i^{s*}(p_j)$ and $p_j = p_j^{s*}(p_i)$, we find $p_i = p_j = p_S \equiv \frac{1}{3}(2\overline{\theta}s + K)$. It is a best response for both firms to set $p_i = p_i^{s*}(p_j)$ if and only if $p^d \leq p_S \leq p^e$. The first inequality is clearly met, whereas the second is met provided that $p_S < p^e \iff K > \frac{3\sqrt{2}-4}{2}\overline{\theta}s$.This last condition is compatible with $p^d \leq p^e$.

4. The configuration (p_i^{c*}, p_j^{c*}) cannot be an equilibrium because one can check that $p_j^{c*} < p^e$ and $p_j^{c*} < p^{e\prime}$. Hence, $p_i^*(p_j^{c*}) \neq p_i^{c*}$.

5. The configuration $\left(p_i^{b*}, p_j^D(\cdot)\right)$ cannot be an equilibrium because $p_j^D\left(p_i^{b*}\right) = p_i^{b*} + K > p^f$ and thus $p_i^*\left[p_j^D\left(p_i^{b*}\right)\right] \neq p_i^{b*}$.

6. Similarly, the configuration $\left(p_i^{b*}, p_j^{s*}(\cdot)\right)$ cannot be an equilibrium because $p_j^{s*}\left(p_i^{b*}\right) = \frac{3}{4}\overline{\theta}s + \frac{K}{4} > p^f$, and hence $p_i^*\left[p_j^{s*}\left(p_i^{b*}\right)\right] \neq p_i^{b*}$.

7. The configuration $\left(p_i^{b*}, p_j^{c*}\right)$ cannot be an equilibrium because when $p^d \leq p^e$, one can easily check that $p_i^{b*} < p^e$ so that $p_j^*\left(p_i^{b*}\right) \neq p_j^{c*}$. Also, when $p^d > p^e$, we get $p_i^{b*} < p^{e\prime}$ iff $K < \frac{2}{3}\overline{\theta}s$ which follows from Assumptions A1 and A2. So, $p_j^*\left(p_i^{b*}\right) \neq p_j^{c*}$ when $p^e < p^d$.

8. The configuration $\left(p_i^D(\cdot), p_j^{s*}(\cdot)\right)$ cannot be an equilibrium because solving for $p_i = p_i^D(p_j)$ and $p_j = p_j^{s*}(p_i)$, we get $p_j = \widetilde{p}_j \equiv \frac{2}{3}\overline{\theta}s > p^d$, meaning that $p_i^*(\widetilde{p}_j) \neq p_i^D(\widetilde{p}_j)$.

9. The configuration $\left(p_i^D(\cdot), p_j^{c*}\right)$ cannot be an equilibrium because when $p^d \leq p^e$, we have $p_i^D\left(p_j^{c*}\right) = p_j^{c*} + K = \frac{1}{2}\overline{\theta}s + K < p^e$. When $p^d > p^e$, we have $p_i^D\left(p_j^{c*}\right) = \frac{1}{2}\overline{\theta}s + K < p^{e\prime}$ iff $K < \frac{1}{4}\overline{\theta}s$, which is always true. Therefore $p_j^*\left[p_i^D\left(p_j^{c*}\right)\right] \neq p_j^{c*}$.

10. The configuration $\left(p_i^{c*}, p_j^{s*}(\cdot)\right)$ cannot be an equilibrium because, for this to be an equilibrium, we should have (a) $p_j^{s*}\left(p_i^{c*}\right) = \frac{3}{4}\overline{\theta}s + \frac{K}{2} \geq p^e \iff K \leq 0.328\overline{\theta}s$, and (b) $p_i^{c*} \geq p^d \iff K \geq 0.5\overline{\theta}s$, which is incompatible with (a).

7.3 Proof of Proposition 3

Firm i's profit is equal to $\pi(p_i, p_j)$ where

$$\pi(p_i, p_j) = \begin{cases} \pi^b(p_i) = p_i\left(\overline{\theta} - \frac{p_i - K}{s}\right) & \text{if } p_j \in [0, p_i - K) \\ \pi^s(p_i, p_j) = p_i\left(\overline{\theta} - \frac{p_i + p_j - K}{2s}\right) & \text{if } p_j \in [p_i - K, p_i + K), \\ \pi^c(p_i) = p_i\left(\overline{\theta} - \frac{p_i}{s}\right) & \text{if } p_j \in [p_i + K, \infty) \end{cases}$$

Each section of the profit function is concave in p_i.

We consider mixed-strategy equilibria that include two price atoms p_{ai} and p_{bi} played by player i with probabilities x_i and $1 - x_i$. Firm i's expected profit is equal to

$$\begin{aligned}\Pi_i &= x_i x_j \pi(p_{ai}, p_{aj}) + x_i(1-x_j)\pi(p_{ai}, p_{bj}) \\ &\quad + (1-x_i)x_j\pi(p_{bi}, p_{aj}) + (1-x_i)(1-x_j)\pi(p_{bi}, p_{bj}).\end{aligned}$$

We look for a symmetric mixed-strategy equilibrium. This exists if we can find probabilities and prices such that $(x, p_a, p_b) = (x_i, p_{ai}, p_{bi})$, $i = 1, 2$ and

$$(x_i, p_{ai}, p_{bi}) = \arg\max_{(x_i, p_{ai}, p_{bi})} \Pi_i \text{ s.t. } (x_j, p_{aj}, p_{bj}) = (x, p_a, p_b),\ i = 1, 2.$$

First, suppose that the symmetric equilibrium is such that $p_b \leq p_a + K$. For (x_i, p_{ai}, p_{bi}), $i \in \{1, 2\}$ close enough to (x, p_a, p_b), firm i's payoff is given by the function

$$\begin{aligned}\Pi_i &= x_i x_j \pi^s(p_{ai}, p_{aj}) + x_i(1-x_j)\pi^s(p_{ai}, p_{bj}) \\ &\quad + (1-x_i)x_j\pi^s(p_{bi}, p_{aj}) + (1-x_i)(1-x_j)\pi^s(p_{bi}, p_{bj}).\end{aligned}$$

Because π^s is strictly concave, it is easy to show that there is a unique symmetric equilibrium with $p_S = p_{ai} = p_{bi}$, $i \in \{1, 2\}$. This is the equilibrium with pure strategies found in Proposition 2.

Second, suppose that the symmetric equilibrium is such that $p_b > p_a + K$. For (x_i, p_{ai}, p_{bi}), $i \in \{1, 2\}$ close enough to (x, p_a, p_b), firm i's expected payoff writes as

$$\begin{aligned}\Pi_i &= x_i x_j \pi^s(p_{ai}, p_{aj}) + x_i(1-x_j)\pi^c(p_{ai}) \\ &\quad + (1-x_i)x_j\pi^b(p_{bi}) + (1-x_i)(1-x_j)\pi^s(p_{bi}, p_{bj}).\end{aligned}$$

There are three first order conditions for equilibrium:

$$\begin{aligned}\frac{\partial \Pi_i}{\partial p_{ai}} &= 0 \iff x_i = 0 \quad \text{or} \quad x_j \pi^s_{p_i}(p_{ai}, p_{aj}) + (1-x_j)\pi^c_{p_i}(p_{ai}) = 0 \\ \frac{\partial \Pi_i}{\partial p_{bi}} &= 0 \iff x_i = 1 \quad \text{or} \quad x_j \pi^b_{p_i}(p_{bi}) + (1-x_j)\pi^s_{p_i}(p_{bi}, p_{bj}) = 0 \\ \frac{\partial \Pi_i}{\partial x_i} &= 0 \iff \left\{ \begin{array}{c} x_j \pi^s(p_{ai}, p_{aj}) + (1-x_j)\pi^c(p_{ai}) \\ -x_j\pi^b(p_{bi}) - (1-x_j)\pi^s(p_{bi}, p_{bj}) \end{array} \right\} = 0 \qquad (5)\end{aligned}$$

where the subscript p_i denotes a partial differentiation w.r.t. to p_i. These three conditions guarantee a maximum because $\frac{\partial^2 \Pi_1}{\partial p_{a1}^2} < 0$, $\frac{\partial^2 \Pi_1}{\partial p_{b1}^2} < 0$ and the Hessian determinant is zero:

$$|H| = \begin{vmatrix} \frac{\partial^2 \Pi_i}{\partial p_{ai}^2} & 0 & 0 \\ 0 & \frac{\partial^2 \Pi_i}{\partial p_{bi}^2} & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0$$

We now determine the symmetric equilibrium by setting $(x, p_a, p_b) = (x_i, p_{ai}, p_{bi})$, $i = 1, 2$ and $x_i = x \in (0, 1)$. We successively get

$$p_a = \frac{2\overline{\theta}s + xK}{4 - x}, \quad p_b = \frac{2\overline{\theta}s + (x+1)K}{x+3}$$

and

$$x = \frac{\pi^s(p_b, p_b) - \pi^c(p_a)}{\pi^s(p_a, p_a) + \pi^s(p_b, p_b) - \pi^c(p_a) - \pi^b(p_b)}.$$

One can check that $p_b > p_a + K$ which is consistent with the condition $p_{bi} > p_{ai} + K$, $i \in \{1, 2\}$.

Equilibrium profits can be computed as the following functions of x:

$$\begin{aligned}
\pi^c(p_a) &= \left((2-x)\overline{\theta}s - xK\right)\frac{\left(xK + 2\overline{\theta}s\right)}{(x-4)^2 s}, \\
\pi^s(p_a, p_a) &= \frac{1}{2}\left(2(2-x)\overline{\theta}s + (4-3x)K\right)\frac{\left(xK + 2\overline{\theta}s\right)}{(x-4)^2 s}, \\
\pi^s(p_b, p_b) &= \frac{1}{2}\left((x+1)2\overline{\theta}s + (1-x)K\right)\frac{\left(2\overline{\theta}s + (x+1)K\right)}{(x+3)^2 s}, \\
\pi^b(p_b) &= \left(\overline{\theta}s(x+1) + 2K\right)\frac{\left(2\overline{\theta}s + (x+1)K\right)}{(x+3)^2 s}.
\end{aligned}$$

After some substitutions, x solves :

$$x = \frac{\begin{bmatrix} \left(K^2 + 4\overline{\theta}sK\right)x^4 + \left(20K^2 - 2\overline{\theta}sK + 8\overline{\theta}^2 s^2\right)x^3 \\ + \left(3K^2 + 38\overline{\theta}sK - 12\overline{\theta}^2 s^2\right)x^2 + \left(-8K^2 + 20\overline{\theta}^2 s^2\right)x \\ +16K^2 - 8\overline{\theta}^2 s^2 + 64\overline{\theta}sK \end{bmatrix}}{2K(x+3)(4-x)\left(x^2K + 2\overline{\theta}sx - \overline{\theta}s - 2K\right)}. \quad (6)$$

Hence x is the solution of a polynomial with degree 5. There is at least one real solution. We have found no analytical solution.

Two cases can readily be studied. On the one hand, when $K = \widehat{K} \equiv \frac{3\sqrt{2}-4}{2}s\overline{\theta}$, expression (6) implies that $x = 0$. Furthermore, one can check that

$\left[\frac{\partial^2 \Pi_i}{\partial x_i \partial x_j}\right]_{x_j=0} < 0$ and that $\left[\frac{\partial^2 \Pi_i}{\partial x_i \partial K}\right]_{x_j=0} < 0$ so that $\left[\frac{\partial x_j}{\partial K}\right]_{x_j=0} = \left[\frac{\partial x}{\partial K}\right]_{x=0} < 0$. A smaller K increases the probability x above zero. Hence, mixed strategy equilibria occur for $K < \widehat{K}$ and pure strategy equilibria occur otherwise. Furthermore, when $x = 0$ and $K = \widehat{K}$, we have that $\Pi_i = \pi_i^{c*}$ and that $d\Pi_i/dK = \partial \Pi_i/\partial K + (dx/dK)(\partial \Pi_i/\partial x) = (-10 + \sqrt{2})\overline{\theta} < 0$. Therefore, expected profits under symmetric mixed strategy increase above π_i^{c*} as K decreases below $\widehat{K}$. For K smaller and close enough to $\widehat{K}$, symmetric mixed strategy dominate the price strategy p_i^{c*}.

On the other hand, when $K \to 0$, we have that $\pi_1^s(p_a, p_a) = \pi_1^s(p_b, p_b) = \pi_1^c(p_a) = \pi_1^b(p_b)$. So the RHS of expression (6) is indefinite. To solve this problem, we approximate expression (6) by dropping terms in K of order larger than one and we get

$$K\left(x^4 - x^3 - x^2 + 3x + 8\right) + \frac{1}{2}\overline{\theta}s\,(2x-1)\left(x^2 - x + 2\right) = 0$$

which yields the unique solution $x = 1/2$ when $K \to 0$. Applying this result, we get that the expected profits is equal to $\pi_i^{c*} - \overline{\theta}s/196$. Therefore, the symmetric strategy is dominated by the strategy p_i^{c*} when $K \to 0$.

The previous argument suggests that the two-price symmetric mixed strategy is a maximum as long as the firms set prices such that $p_{ai}+K < p_{bi}$, but that it can be dominated by the one-price strategy p_i^{c*} for small enough K. To check when the symmetric mixed strategy is a global maximum, we fix firm j's strategy as $(x_j, p_{aj}, p_{bj}) = (x, p_a, p_b)$ where $x > 0$, and we verify whether firm i can profitably deviate by fixing any other pair of prices. To this purpose, we sketch firm i's expected profit as a function of price p_i:

$$\Pi_i(p_i) = x\pi(p_i, p_a) + (1-x)\,\pi(p_i, p_b)\,.$$

Recall that $\pi(p_i, p_a)$ and $\pi(p_i, p_b)$ are combinations of three quadratic and concave functions. Then, it is readily shown that $\Pi_i(p_i)$ is a piece-wise quadratic and concave function. Consider p_i increasing from zero. One can check that the first section of $\Pi_i(p_i)$ is either increasing if $p_i^{c*} \geq p_a - K$, or bell-shaped with a maximum at p_i^{c*} if $p_i^{c*} < p_a - K$; the second section is bell-shaped with maximum at $p_i = p_a$; the third section is monotonically increasing; the fourth section is bell-shaped with maximum at $p_i = p_b$ and the last section is monotonically decreasing.

If $p_i^{c*} \geq p_a - K$, $p_i = p_a, p_b$ are the only candidates for a global maximum and the above solution is the unique mixed-strategy equilibrium with two price atoms.

On the other hand, if $p_i^{c*} < p_a - K$, there are three candidates for a global maximum: $p_i = p_a, p_b$ *and* $p_i = p_i^{c*}$. We need thus to check if $p_i = p_i^{c*}$ can yield a higher profit. If firm i plays $p_i = p_a$ or $p_i = p_b$, it achieves an expected profit equal to:

$$\Pi_i^* \equiv x\pi^s\left(p_a, p_a\right) + (1-x)\,\pi^c\left(p_a\right) = (2-x)\,\frac{\left(2s\overline{\theta} + Kx\right)^2}{2s\left(4-x\right)^2}.$$

Otherwise, if firm i plays $p_i = p_i^{c*}$, it achieves a profit equal to π_i^{c*}. Some computations show that $\Pi_i^* \geq \pi_i^{c*}$ is equivalent to

$$2x\left(2-x\right)K^2 + 8\overline{\theta}s\left(2-x\right)K - xs^2\overline{\theta}^2 \geq 0$$

Hence, $K \geq K_1(x) \equiv \overline{\theta}s\left(2\left(4-2x\right) + \left(4-x\right)\sqrt{4-2x}\right)/\left[2x\left(x-2\right)\right]$. Note that the probability x depends on K. To get the probability x that makes this inequality binding and that is simultaneously compatible with a mixed strategy equilibrium, we insert $K_1(x)$ in expression (5), we evaluate at the symmetric mixed strategy equilibrium to get

$$32 - 4(4-x)\sqrt{4-2x}\left(1+x^2\right) + x\left(48 + x\left(-1 + x\left(-8 + x\left(-1+2x\right)\right)\right)\right) = 0$$

This equation has a unique solution in the interval $x \in (0,1)$ which is equal to $\tilde{x} = 0.3603$. The associated level of fixed cost of copying is equal to $\tilde{K} \equiv 0.02739\overline{\theta}s < \widehat{K}$.

Hence, the above solution is not an equilibrium if $K < \tilde{K}$ and if $p_i^{c*} < p_a - K$. The latter condition rewrites as

$$K < \frac{x}{4\left(2-x\right)}\overline{\theta}s.$$

Using equation (6), this is equivalent to $K < \tilde{K}' \equiv 0.04233\overline{\theta}s$, which is implied by $K < \tilde{K}$.

We conclude that the solution $(x_i, p_{ai}, p_{bi}) = (x, p_a, p_b)$, $i = 1, 2$, is a mixed-strategy equilibrium with two atoms provided that $K > \tilde{K}$.

7.4 Welfare properties

As indicated in Section 5, we can can apply the same analysis to the multiproduct monopoly and to the duopoly with a symmetric pure-strategy equilibrium. We note p_k and θ_k respectively for the optimal or equilibrium price and for the marginal user, with $k = m, S$.

Effects on profit. An increase in K induces variations in profits through effects on price and demand. Changes in the surplus of the two producers can be written as $dPS = 2\left(\overline{\theta} - \theta_k\right) dp_k - 2p_k d\theta_k$. This expression is clearly positive for $dK > 0$ as it implies $dp_k > 0$ and $d\theta_k < 0$. The sign for $ds_c < 0$ is a priori ambiguous. It is readily verified that $dPS/ds_c < 0$ iff $K < 2\overline{\theta}s$, which follows from the combination of Assumptions A1 and A2. Therefore, profits increase when copies are damaged or made more expensive.

Effects on consumer surplus. The consumers' surplus obtained from the use of both information goods includes four effects: the negative effect on illegal copiers because of the increase in the copying cost K; the negative effect of the deterioration of copies on the copying users; the negative effect of larger prices on legal consumers; and finally the effect on the switching users who move from copying to purchasing an original. At the price p_k, the latter effect on marginal users is nil since they are indifferent between copying and purchasing the orginals. Hence, $dCS = -\left(\theta_k - \underline{\theta}\right) dK + \left(\theta_k - \underline{\theta}\right)^2 ds_c - 2\left(\overline{\theta} - \theta_k\right) dp_k + 0 * d\theta_k$, which is negative because all terms are non positive $(dK > 0, ds_c < 0, dp_k > 0)$. Consumers are negatively affected by both policy measures.

Effects on social surplus. The change in social surplus writes as $dW = dPS + dCS = -\left(\theta_k - \underline{\theta}\right) dK - \left(\theta_k - \underline{\theta}\right)^2 ds_c - 2p_k d\theta_S$. This expression is negative if the additional costs imposed on all copiers is smaller than the additional revenue from having an additional legal consumer. This is the case for $ds_c < 0$. For $dK > 0$, the sign of the welfare change is ambiguous.

- In the monopoly case, one can show that $dW < 0$ iff $K < K_m \equiv \frac{2}{3}s\left(\overline{\theta} - 4\underline{\theta}\right)$. The following numerical example shows that both cases can occur under our assumptions. Take $s_o = 7$, $s_c = 4$, $\overline{\theta} = 500$ and $\underline{\theta} = 100$. One checks that these values satisfy Assumption A2. To meet Assumption A1, we need $K < \underline{\theta}s_c = 400$. We check that $K_m = 200 < \underline{\theta}s_c$.

- In the duopoly case, one can show that $dW < 0$ iff $K < K_S \equiv \frac{2}{5}s\left(4\overline{\theta} - 9\underline{\theta}\right)$. The following numerical example shows that both cases can occur under our assumptions. Take $s_o = 4$, $s_c = 1$, $\overline{\theta} = 244$ and $\underline{\theta} = 100$. One checks that these values satisfy Assumption A2. To meet the condition for a pure-strategy equilibrium and Assumption A1, we need $K \in \left(\hat{K}, \underline{\theta}s_c\right) = (88.8, 100)$. We check that $K_S = 91.2$ falls in the latter interval.

Higher incentives to create in duopoly. It is easy to check that $\pi_S > \pi_m - \pi_i^{b*}$ is equivalent to $K > \left(3\sqrt{10} - 8\right)\overline{\theta}s/13 \simeq 0.114\overline{\theta}s$, which is implied by $K > \hat{K} \simeq 0.121\overline{\theta}s$.

References

[1] Bae, S.H. and Choi, J.P. (2004). A model of piracy. Mimeo. Michigan State University.

[2] Bailey, J.P. (1998). Electronic commerce: prices and consumer issues for three products: books, compact discs, and software. *OECD/GD*, 98(4).

[3] Baye, M. and Morgan, J. (2001). Information gatekeepers on the Internet and the competitiveness of homogeneous product markets. *American Economic Review* **91**: 454-474.

[4] Besen, S.M. and Kirby, S.N. (1989). Private copying, appropriability, and optimal copying royalties. *Journal of Law and Economics* **32**: 255-280.

[5] Besen, S.M. and Raskind, L.J. (1991). An introduction to the law and economics of intellectual property. *Journal of Economic Perspectives* **5**: 3-27.

[6] Boccard, N. and Wauthy, X. (1997). The Hotelling model with capacity precommitment. CORE Discussion paper 9783, Université catholique de Louvain, Belgium.

[7] Boccard, N. and Wauthy, X. (2003). Optimal quotas, price competition and products' attributes. *The Japanese Economic Review* **54**: 395-408.

[8] Brynjolfsson, E. and Smith, M.D. (2000). Frictionless commerce?: A comparison of Internet and conventional retailers. *Management Science* **46**: 563-585.

[9] Cournot, A. (1838), *Recherches sur les principes mathematiques de la theorie des richesses.* Paris: Hachette.

[10] Gabszewicz, J., Sonnac, N. and Wauthy, X. (2001). On price competition with complementary products. *Economics Letters* **70**: 431-437.

[11] Gayer, A. and Shy, O. (2003). Copyright protection and hardware taxation. *Information Economics and Policy* **15**: 467-483.

[12] Glicksberg, I. (1952). A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points, *Proceedings of the American Mathematical Society* **3**: 170–174.

[13] Iyer, G. and Pazgal, A. (2003). Internet shopping agents: virtual co-location and competition. *Marketing Science* **22**: 85-106.

[14] Johnson, W.R. (1985). The economics of copying. *Journal of Political Economy* **93**: 158-174.

[15] Koboldt, C. (1995). Intellectual property and optimal copyright protection. *Journal of Cultural Economics* **19**: 131-155.

[16] Landes, W.M. and Posner, R.A. (1989). An economic analysis of copyright law. *Journal of Legal Studies* **38**: 325-363.

[17] Liebowitz, S.J. (1985). Copying and indirect appropriability: Photocopying of journals. *Journal of Political Economy* **93**: 945-957.

[18] Mussa, M., and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory* **18**: 301-317.

[19] Novos, I.E. and Waldman, M. (1984). The effects of increased copyright protection: An analytical approach. *Journal of Political Economy* **92**: 236-246.

[20] Peitz, M. and Waelbroeck, P. (2003). Piracy of digital products: A critical review of the economics literature. CESifo Working Paper No. 1071.

[21] Varian, H.R. (1980). A model of sales. *American Economic Review* **70**: 651-659.

[22] Watt, R. (2000). *Copyright and Economic Theory. Friends or Foes?* Edward Elgar Publishing, Cheltenham (UK) and Northampton (Mass.).

[23] Yoon, K. (2002). The optimal level of copyright protection. *Information Economics and Policy* **14**: 327-348.

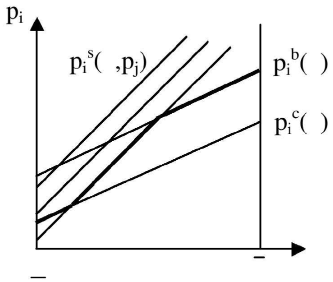


Figure 1: Prices as functions of θ

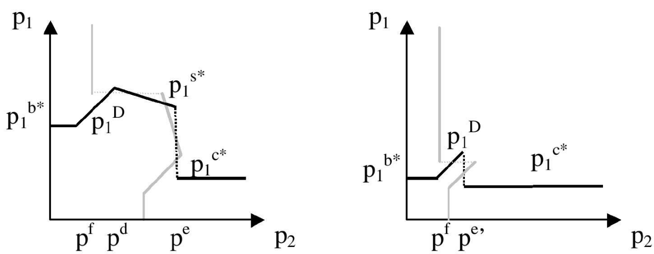


Figure 2: Best Response Functions

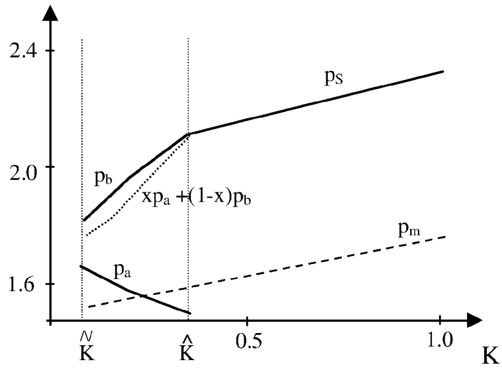


Figure 3: Prices as a function of K

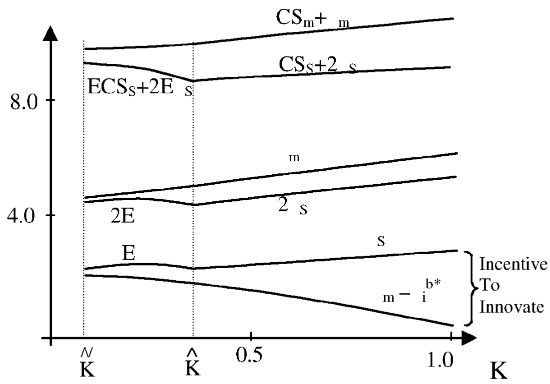


Figure 4: Profits, Consumer Surpluses and Incentive to Innovate as a function of K

THE ECONOMICS OF DIGITAL ASSEMBLING

Eric Brousseau*

EconomiX, University of Paris X & Institut Universitaire de France

Thierry Pénard**

Crem, University of Rennes 1, Marsouin

* Université de Paris X, EconomiX, Bâtiment K; 200 avenue de la république; F-92001 Nanterre Cedex; France
Tel: + 33 1 40 97 59 22; Fax: + 33 1 40 97 59 07; eric@brousseau.info

** CREM, Faculté de Sciences Economiques, 7 place Hoche, 35 065 Rennes Cedex; France
Tel +33 (0)2 23 23 35 20; Fax: (33) (0)2.23.23.35.09; thierry.penard@univ-rennes1.fr

CURRENT DRAFT: APRIL 2006

***Abstract*:**

The development of information and communication technologies has largely renewed and stimulated intermediation. New forms of intermediation and business models have emerged on the Internet (shopbots, referrals, portals ...). This paper focuses on a particular form of intermediation, digital assembling that consists in gathering information and digital contents and arranging them in packages to provide customers with modular goods or services that fit to their needs and preferences. We propose a theoretical framework aimed at understanding digital assemblers' strategies. It enables us to highlight the tradeoff faced by providers of components that can either use the services of the intermediary or bypass it and market directly their components. Our main findings are the existence of dual distribution equilibria where the core components are packaged and sold by the assembler and the peripheral component are directly offered by the producers.

***Keywords**: assembling, intermediation, digital markets, dual distribution*

1-INTRODUCTION

The massive diffusion of information and communication technologies (ICTs) has been deeply transforming the economics of intermediation. ICTs have been providing agents with extended capabilities to coordinate and manage transactions. These technologies have also induced new opportunity and needs for intermediation. Indeed, companies like E-Bay, Amazon, Google, Yahoo, Autobytel, have rapidly understood the potentiality of intermediation business on the Internet and have built their success stories on their ability to organize exchanges between providers and customers of informational or tangible goods.

In general, intermediaries can perform various tasks. By managing information about customers' preferences and suppliers' proposals, they contribute to the matching of buyer and sellers' plans (Spulber, 2004). By performing logistic operations (such as holding inventories or moving goods), they adapt these plans when they do not spontaneously match. By dealing with information asymmetries between the two sides of the markets, they help to solve problems of adverse selection, moral hazard, and defaults of enforcement (Biglaiser, 1993). By ensuring the liquidity of exchanges when necessary, they allow market clearing. Obviously intermediaries can choose to perform only one of these tasks or to provide all of them, leading to a wide range of possible intermediation business models. Currently, there is an increasing tendency for intermediaries to conduct a part of their activities on digital networks. In this perspective, a digital intermediary corresponds to an intermediary whose activity is exclusively on the Internet or deals with informational goods.

Our paper focuses on a particular form of digital intermediation: assembling that consists in gathering information and services and arranging them in packages to provide customers with modular goods that fit to their needs and preferences. An assembler is not a simple broker that matches existing supply of goods, whose characteristics are already qualitatively fitting to the users needs. The assembler re-organizes a supply of functions and operates a matching on quality in addition to a matching on the terms of exchange. Assemblers range from the infomediaries that market information packages such as TV channels or Internet Portals, to the developer of an operating system (like Microsoft), and include on-line retailers (since they assemble portfolios of consumer goods). Hence, a digital assembler can be compared to the owner of a technology platform that provides the required infrastructure and standards to perform the assembling process of digital components (Church and Gandal, 2004; Evans et al. 2004, Economides and Katsamakas, 2004). In this paper, we propose a theoretical model aimed at understanding digital assemblers' strategies and the underlying business models.

The paper is organized as follows. Section 2 proposes a brief survey on digital intermediation. Section 3 attempts to characterize digital assembling. Section 4 presents the theoretical framework for analyzing the strategies of assembling and establishes several propositions on the market equilibrium configurations. Section 5 concludes.

SECTION 2 LITERATURE ON DIGITAL INTERMEDIATION

Digital and informational goods are made of complementary components that must be assembled together in order to address complex needs. One of the roles played by intermediaries is precisely to assemble components or to bundle them in packages that correspond to consumer needs. In our view they do not match a demand with a supply that *ex ante* fits consumers qualitative preferences, they qualitatively adapt the supply to the demand.

A typical assembler is the software company Microsoft that buys many software components and combines them with its built-in components to provide users with a ready-to-use package. A portal like Yahoo is another example of assembler, that provides information services to the Internet user by arranging access to various information services provided by third part. Many media companies are also assemblers in the sense that they buy entertainment or news to assemble them in packages adapted to various customers' categories (corresponding to channels in TV broadcasting). There are also assemblers outside the digital world. Automakers, for instance, are typically assemblers of various components made by independent producers specialized on various technologies and functions.

The activity of digital assembling has to be contrasted with infomediary's activities. In the literature on the Internet markets, infomediaries include referrals, shopbots, matchmakers or gatekeepers. These players are specific in the sense that they are essentially information brokers aimed at facilitating the matching between suppliers and consumers and reducing search costs (i.e. improving market transparency). They do not directly participate in the transaction and are in that sense very different of what characterize commercial intermediaries for Spulber [1996] since they don't hold any property rights on the exchanged products. This later characteristic is essential according to Spulber because it allows intermediaries to rearrange the supply, to make it compatible with the customers' preferences.

Among infomediaries, business models can be rather heterogeneous. Baye and Morgan (2001, 2002) study the business strategy of Internet gatekeepers. Such intermediaries enable buyers to compare prices and other characteristics like delivery conditions, of a product. The gatekeeper only advertises the offers of retailers that subscribe to its service (by paying advertising fees). Baye and Morgan seek to determine the fees that the gatekeeper should fix for retailers, but also for customers (subscription fees). Baye and Morgan find that the gatekeeper should set low subscription fees to attract most of the consumers and set high advertising fees to limit firms' participation and thus to maintain price dispersion on the market. The latter condition is important since the added value of a gatekeeper arises from the persistence of price dispersion. Whereas full participation by firms would lead to a unique price, a partial participation maintains price dispersion, thus enabling the gatekeeper to make profit. According to the authors, the firms that advertise on the gatekeeper's site set lower prices than the firms that do not advertise on the site, but get larger market shares1.

---

[1] See also Saliba (2003) for an extension of this model to heterogeneous markets. For an empirical approach of gatekeepers, see Smith and Brynjolfson (2001) and Smith (2002) that analyze the impact of a gatekeeper on behaviors of sellers and buyers of books. Similarly, Baye, Morgan and Scholten (2001,

The business models of referrals have also received a large attention (Chen et al. 2002, Ghose et al. 2002, Arnold and Pénard 2006). The role of a referral is to indicate to potential buyers one or a few selected suppliers that could fit with buyers' needs. Chen et al. (2002) find that the profits of the referral infomediary can strongly decline when its reach becomes too large. The intuition is that as more and more buyers use the referral service, competition between sellers increases and drives their profits to zero. Their second finding is that the infomediary is better when it offers exclusive contracts to a single seller (geographical exclusivity)2.

The business model of matchmakers (Jullien and Caillaud, 2003) is quite similar to the gatekeeper's and referrals. However, the economics of matchmaking relies strongly on network externality (direct and indirect), and the role of a matchmaker consists in attracting as many providers and consumers as possible on its site to stimulate network effects. Therefore, it is optimal for a matchmaker to target full participation of customers and firms, whereas it is in the interest of the gatekeeper to target partial participation of firms. Matching has to be linked to the growing literature on two-sided markets, on which a service provider faces two interdependent demands due to indirect network externalities (e.g. between credit card holders and affiliated retailers). Intermediaries can then implement discriminating pricing strategies to ensure the maximization of their surplus by playing on cross-subsidization (e.g. Rochet and Tirole [2001], Parker and Van Alstyne [2000])

An assembler sharply differs from a matchmaker or a referral because it modifies and rearranges the supply. Moreover the rationale for an assembler is not price dispersion and information asymmetry, but modules dispersion (in variety and quality) and the heterogeneity of preferences. Few theoretical models on the economics of assembling exist. An interesting approach of assembling is proposed by Economides and Katsamakas (2004). They consider a technology platform (an assembler) that is essential for the providers of contents to access the end-user. Moreover, the assembler sells a direct service to the end-users. Economides et al. seek to characterize the pricing strategies of the assembler depending on the number of modules that must be assembled and the degree of vertical integration of the assembler (the number of modules that it produces itself). They find that the platform access fee (paid by the providers of modules to access the end-users) can sometimes be below marginal cost (cross-subsidization). Our paper proposes another theoretical framework of assembling that enables us to examine the conditions under which an assembler can emerge on a digital market and to understand which components of the digital system the assembler is more likely to control and resell.

2002) study the pricing strategies of retailers of electronic and digital products (camera, laptops, printers, scanners…) on the gatekeeper Shopper.com.

[2] It is the case of Autobytel.com, a referral in the automobile sector. This Website offers consumers detailed information about cars and enables the buyers to obtain a price from the geographically closest automobile dealer with which Autobytel has an exclusive agreement (in exchange of a subscription fee paid by the dealer). See Scott Morton F., Zettelmeyer F., Risso J. (2001).

SECTION 3 THE ECONOMICS OF DIGITAL ASSEMBLING

Understanding the economics of assembling requires to build a framework based on the three following assumptions.

Assumption 1: Digital goods and services are of a modular nature. *Basic components — corresponding to "functionalities" — are assembled to produce services that meet consumers' needs. Consumable services that have a value of use are therefore made of "packages" of basic functions.*

This fits with the Lancaster [1979] vision of consumption and goods. Consumers are seeking for a collection of attributes. In the case of digital goods, basic attributes are the contents or the functionalities that form "composite goods". In the line of Economides and Salop [1992] and Economides [1996], one can analyze how different industry configurations result in various patterns of competition among the suppliers involved in the production of these functions. Indeed, the way the assembling of this package is made is a major factor of singularization of the different models. Moreover, the competition between "provision schemes" plays at two levels. First, a single need can be met by "packages" of different functions (just as transportation need can be met by different technological systems: road, rail, air, corresponding to various packages of services). This leads to a competition among packages that are close substitutes or not. Second, a given package can be assembled though different (organizational) processes. That leads to a competition among organizational models.

Assumption 2: Digital activities are characterized by three basic operations:

- *The* ***production*** *of functionalities. Functionalities result from the providing either of tangible goods or information services, or a mix of them. It corresponds either to an informational content or to an informational process.*
- *The* ***assembling*** *of functionalities. Each module does not produce any value of uses by itself. They are multipurpose, and have to be combined with each other to result in a service that fits to particular needs of a specific category of users.*
- *The* ***consumption*** *of services by which a user transforms a service in value of use (either because he is satisfied — in the case of final users — or because he can use the service to generate new service or new knowledge — in the case of producers).*

This second assumption mostly draws from the first one. Digital goods are composite goods made of multipurpose components in the sense that they can be implemented in different goods[3], but that do not produce any value of use *per se*. Only the combination of these functions leads to some useful services. Software is a good example. A spreadsheet or a word processor are made of basic components that manage the screen, the hard-disk, the printing resources, and that operate basic — e.g. sorting, calculating — or more sophisticated operations on information — e.g. checking grammar or turning tables into graphical representations —. Any software is made of basic components that are useless when considered alone, and that are recombined differently in various softwares. The same applies for many information services or databases. They are produced by combining data or information flows (usually coming from heterogeneous

[3] This characteristic is obviously reinforced by the public nature of information goods and the tending to zero reproduction costs of digitized information.

sources in contrasted formats), with processes to organize, stock, retrieve and distribute them.

In practice an assembler can rely on contrasted economics tools to buy and integrate the various components of its service. It can, as Microsoft does, acquire the (start-up) firms that develop and market the functionality. It can, as many portals do, simply buy a license of exploitation to an information provider.

Assumption 3: Users (or consumers) are not neutral in the process of value creation, *since they can themselves assemble the functionalities, and since their use of the digital goods generate information that is itself generating value of use.*

First, with digital technologies most components are compatible among each others (as long as interfaces are appropriately managed). This gives the consumers the opportunity to credibly become their own assembler. There are many examples of that in the digital industries. Most PC owners, for instance, assemble themselves a set of hardware and software resources to customize their information system. The users of open-source software are also most of the time their own assemblers. This capability organizes therefore a virtual competition between consumers and specialized assemblers (i.e. between various modes of assembling McKelvey [2001]). Understanding the nature of the interactions among these alternative ways of producing digital goods is the objective of this paper.

Secondly, on a market for information or for knowledge, the supply and the demand sides are difficult to disentangle since any agent is both information (or knowledge) user and producer. Actively or unintendedly, the user of a service generates information, feedback, and sometimes knowledge, which can be of value, either for the other users (who could benefit of information on the quality of the service or on its optimal condition of use) or for the agents involved in its production (who could better target or design their products or services). To the limit the "user" can be an intermediate producer whose information output results from the information he consumed. This is typically the case for the developers in open source software communities who "invent around" existing inventions, for many digital artists that rearrange digital contents to produce new works of authorships, or for many consultants that reorganize and synthesize existing information.

To sump up, digital markets are characterized by the fact that there is no complete separation between the demand and the supply sides. The users are both final users, in the sense that they consume the services provided by the suppliers, and suppliers of intermediate goods and services. They can provide the assemblers with information enabling the later to target more efficiently the users, and can also decide to substitute to them by assembling the final services for themselves.

Finally, assembling generates costs that correspond to an addition of transaction costs and technical costs. Transaction costs are generated when the assembler deals with the producers of functions and with the final customers. Technical costs result from the operations necessary to actually make compatible and interoperable the various components bought on the market for functionalities. Indeed, these components do not spontaneously interact together. Assembling tasks and associated costs are performed and born by the final user when an assembler does not intervene in the process. When there are several users, the presence of an assembler enables to avoid the performance of many redundant operations since the assembler transact once with each of the functionalities providers — whatever the number of times he resell it to users — and design once and for all the interfaces among them. Assembling costs being to a large extent fixed and reproduction costs being nearly equal to zero, assemblers generate efficiency gains. However, they have to be remunerated for the service they provide, and might eventually capture rents. This generate a tradeoff both for the users and the producer of functionalities, who will balance efficiency gains with the cost of the intermediation service, and with the redistribution due to the presence of a third part that can extract a part of the surplus.

In addition, if users have different preferences, they face another trade-off. On the one hand, an assembler seeking to maximize economies of scale might provide the users with a standardized package — that correspond to the functionalities required by the majority of customers, but that will not meet some specific needs. They will benefit from low cost of assembling, but will have to bear maladptation costs (of the package to their needs). In addition, the assembler of standardized package could benefit from a market power that could enable it to capture rents. On the other hand, the assembler can seek to customize assembling. In that case, it and the user loose part of efficiency gains in assembling. Moreover the user can be discriminated on a customized basis. Variables like preference heterogeneity, the concentration of the provision of assembling service, the ability to vertically or horizontally differentiate packages… are likely to impact on this complex trade-off. In depth investigation are necessary to better understand the properties of business models based on the presence of not of an assembler.

4- A THEORETICAL MODEL OF DIGITAL ASSEMBLING

41- THE BASIC FRAMEWORK

We consider a digital service composed of a continuum of technologically separable modules. Each module is produced by a single provider and needs to be assembled with the other modules to form a system before being used or consumed. But, a consumer can decide to buy either all the modules or only a subset of these modules.

The assembling task can be directly performed by the end users or by an assembler. The cost of assembling a module with the rest of the modules is equal to c_a for the assembler. When a user buys a package of modules through the assembler, these modules are ready to use and he incurs no additional assembling cost. If the module comes directly from the provider (bypassing the intermediary), then he has to bear assembling costs.

There is a continuum of end users with a total mass of 1, who are heterogeneous in their assembling cost: some are more skilled than the others to assemble the modules they have directly bought on the market. Assembling costs are uniformly distributed in the interval $[c_a, c_h]$. The users are always less skilled to assemble digital modules than the assembler, because the intermediary is specialised in this task and can benefit from economies of scale or learning by doing.

Each digital module is produced by a distinct provider at a fixed cost F, and the cost for reproducing them is close to zero. Although modules are produced in identical cost conditions, they are heterogeneous in their value. All the consumers are identical in their willingness-to-pay a module, but their reservation value varies from a module to another, depending on the utility of the component in the system. Some modules are at the core of the system and correspond to essential functionalities. Others are at the periphery of the system and represent ancillary functions. The reservation value of a module is defined by αq, where α is uniformly distributed between 0 and 1. The core modules are characterised by $\alpha=1$ and the peripheral modules by a α close to 0. As the modules characterized by α below $\frac{c_a}{q}$ are not viable on the market (they cannot be profitably assembled by the assembler or by the most skilled user), we only consider modules belonging to the interval $[\frac{c_a}{q},1]$.

The sequence of decision in our model is as follows:

- Stage 1: each provider chooses either to contract with an assembler (and to cede its module in counterpart of fixed fee) or to market directly its module at a unit price *p*.
- Stage 2: the assembler assembles its modules together under the form of packages. Then, the assembler and the independent providers set the retail prices for the single modules and packages of modules. When a consumer chooses to buy modules from the independent providers, he has to assemble them and to integrate them to the package bought to the assembler before using them.

The aim of this model is to understand how the presence of an assembler can affect the relationships between the providers of modules and the end-users. When does a provider prefer directly selling its digital modules to the end users instead of using the service of the assembler? Another issue is to know whether dual distribution with the coexistence of direct selling and intermediated selling can be a market equilibrium?

4.2 THE BENCHMARK: NO INTERMEDIARY

In the absence of assembler, each provider must directly sell its module. If the provider of the module of value α sets a price p for its module, then the surplus of the consumer wuth an assembling cost c is equal to $\alpha q - p - c$. The covered market corresponds to all the users having an assembling cost c such that $c<\alpha q - p$.

As the assembling cost are uniformly distributed on $[c_a, c_h]$, then the demand is equal to

$$D(p)=\text{Min}\left\{\frac{\alpha q - p - c_a}{c_h - c_a},1\right\}$$

It means that by setting a price below $\alpha q - c_h$, the provider is guaranteed to sell its module to any end user and to cover all the market (even the less skilled users will gain to buy and assemble the module).

As we consider digital modules, there is no variable cost of production and distribution (only a fixed cost) and the gross profit of the provider is then defined by

$$\pi(p) = p\min\left\{\frac{\alpha q - p - c_a}{c_h - c_a},1\right\}$$

The optimal pricing is given by

$$p = \frac{\alpha q - c_a}{2} \qquad \text{if } \alpha < \frac{2c_h - c_a}{q} \text{ (partial coverage)}$$

and $$p = \alpha q - c_h \qquad \text{if } \alpha \geq \frac{2c_h - c_a}{q} \text{ (full coverage)}$$

The price and the market coverage of a module increase with the willingness-to pay of end users. Moreover, the price tends to decreases with user assembling costs (through the increase of the lower and upper bounds of assembling costs), because the provider tends to absorb a part of the additional cost incurred by the end users, in order to sustain demand. Finally, if the quality of the core module is not enough valued by user (if $q < 2c_h - c_a$), then all the providers (even the core providers) obtain a partial coverage of the market.

The profit of the provider of the α module is given by

$$\pi(\alpha) = \frac{(\alpha q - c_a)?}{2(c_h - c_a)} \qquad \text{if } \alpha < \frac{2c_h - c_a}{q} \text{ (partial coverage)} \quad (1)$$

and $$\pi(\alpha) = \alpha q - c_h \qquad \text{if } \alpha \geq \frac{2c_h - c_a}{q} \text{ (full coverage)}$$

The profit of the provider increases with the value of the module (i.e. with the degree of centrality of the module) and decreases with the assembling costs of end-users.

4.2 A MARKET WITH AN ASSEMBLER

We now consider the existence of an assembler that can acquire some modules and package them. The end user does not incur any cost of assembling, when buying from an assembler. So by setting a uniform price equal to αq, the assembler can extract all the users' willingness-to-pay for the module of quality α,. obtaining a gross profit of αq-c_a. This pricing policy can be compared to a uniform delivered price in a competition *à la Hotelling* (Thisse and Vives, 1988; Aguirre and Martin, 2001). In our model, the consumers are heterogeneous in their assembling costs instead of being heterogeneous in their location and thus in their transportation costs in a spatial model. In the absence of assembler, each provider proposes a Free on board price or equivalently a mill price, the consumers having to incur assembling costs that can be assimilated to delivering costs. The presence of an assembler enables to discriminate among consumers. By announcing a uniform assembled price, the assembler subsidizes the high assembling cost consumers and can facilitate their access to the market, by alleviating the full price of the assembled modules.

Thus, the assembler contributes to increase social welfare, thanks to savings on the cost of assembling and a better coverage of the market. When $\alpha \geq \frac{2c_h - c_a}{q}$, the welfare increases by $\int_{c_a}^{c_h} \frac{c - c_a}{c_h - c_a} dc = \frac{c_h - c_a}{2}$ (it corresponds to the amount of cost savings, since the coverage is complete in this case). When $\alpha < \frac{2c_h - c_a}{q}$, the welfare is also improved through the better coverage of the market.

Now it is important to specify how the assembler acquires the different modules from the providers. At least, the assembler needs to propose a bid (or a price) slightly above the expected profit of the provider when marketing directly its modules. If it decides to bypass the assembler, then it can always secure a positive profit defined in (1). Let suppose that the assembler acquires the module at the reservation profit of the provider[4]. Then, the assembler profit generated by the module of value α is given by:

$$\pi_a(\alpha) = \alpha q - c_a - \frac{(\alpha q - c_a)^2}{2(c_h - c_a)} = (\alpha q - c_a)\left(1 - \frac{(\alpha q - c_a)}{2(c_h - c_a)}\right) > 0 \quad \text{if } \alpha < \frac{2c_h - c_a}{q}$$

$$\pi_a(\alpha) = (\alpha q - c_a) - (\alpha q - c_h) = c_h - c_a > 0 \qquad \text{otherwise}$$

The assembler's profit is always positive whatever the value of the module and increases with its assembling cost advantage or efficiency and with the heterogeneity of end users (measured by $(c_h - c_a)$).

$(c_h - c_a)$ also measures the cost advantage of an assembler or equivalently the difficulty for a end user to bypass the assembler. This cost differential is likely to increase if the assembler owns a dominant proprietary standard or platform (like Microsoft).

The profit of the assembler is independent of α if $\alpha \geq \frac{2c_h - c_a}{q}$. But as long as $\alpha < \frac{2c_h - c_a}{q}$, there is a non monotonous relationship between the value of the modules and the expected profit of the assembler. The profit first increases with the reservation value and then decreases.

Proposition 1: The assembler has more incentives to acquire and package intermediary modules than peripheral or core modules.

Proof: The derivative of profit with respect to the module's value is

$$\frac{\partial \pi_a(\alpha)}{\partial \alpha} = q\left(1 - \frac{(\alpha q - c_a)}{(c_h - c_a)}\right)$$

[4] Obviously, the assembler can bargain the price of the module with its provider and converge to a more equitable sharing of the value created by intermediation.

Then $\frac{\partial \pi_a(\alpha)}{\partial \alpha} > 0$ if $\alpha < \frac{c_h}{q}$ and $\frac{\partial \pi_a(\alpha)}{\partial \alpha} < 0$ if $\alpha > \frac{c_h}{q}$

As $\frac{c_h}{q} < \frac{2c_h - c_a}{q}$, then the expected profit of the assembler reaches a maximum for the functionality $\alpha = \frac{c_h}{q}$.

The highest expected profit is not obtained on the peripheral modules (because their value is limited) or on core modules (because the assembler has to concede a higher price to the providers), but on intermediary modules. The rationale of the assembler lies in its capacity to extract a higher rent on these medium values modules.

Here, we have implicitly supposed that each module assembled by the assembler has its value unchanged. Hence, the value of a package is given by the sum of the values of the modules inside this package. It means that an assembler has no constraint on the modules that he wants to integrate in its package. He can assemble low value, medium value and high value modules, even with discontinuity in the set of modules. Thus, as the assembler can always extract some surplus from every module, the highest profit is obtained when he controls all the modules. But, there are other situations where the assembling can degrade the quality of the modules. What happens if the value of a package is not the addition of the individual values of the modules?

4.3 INTERMEDIATED ASSEMBLING AS A SOURCE OF VALUE LOSS

Now, we consider that the assembler tends to damage the value of the modules when putting them together. Indeed, the assembler cannot personalize or customize the assembling and only offers a rather standardized package that is less valued than a package self-assembled by each end user. A trade-off exists between the lower price of a ready-to-use package and the enhanced value of a self-assembled package.

Let suppose that the value of all the modules contained in a ready-to-use package is aligned on the weakest value module. It refers to the idea of the weakest link and is quite relevant to analyze the assembling of digital system, where the quality and security of a system is strongly dependent of all the modules. If a module is more vulnerable than the others, it can affect the security and the quality of all the modules that composed the package. In this context, an end user could be better to assemble these kinds of modules, because he can customize the system and assemble only the modules essential for his own usage and needs.

Let consider an assembler who controls the core modules. Then, does he have incentives to integrate additional modules that will degrade the value of the core modules contained in the package?

To address this question, we have to distinguish two cases.

In the first case, the assembler acquires and packages all the modules between α and 1, with $\alpha \geq \frac{2c_h - c_a}{q}$ (meaning that if the provider decided alternatively to sell them, it will cover all the consumers). Then the profit of the assembler is given by

$$\pi_a(\alpha) = \int_\alpha^1 \left(\alpha q - c_a - (\theta q - c_h)\right) d\theta$$

Proposition 2: If $q > 3c_h - 2c_a$, it is optimal for the assembler to only acquire a finite number of modules. The assembler offers a package containing all the modules having a value $\alpha \in [\hat{\alpha}, 1]$, with $\hat{\alpha} = 1 - \left(\frac{c_h - c_a}{q}\right)$ and the remaining modules are directly sold by the providers and assembled by the end-users.

Proof : $\frac{\partial \pi_a(\alpha)}{\partial \alpha} = (1-\alpha)q - (\alpha q - c_a) + (\alpha q - c_h) = (1-\alpha)q - (c_h - c_a)$.

$\frac{\partial \pi_a(\alpha)}{\partial \alpha} < 0$ if $\alpha > 1 - \left(\frac{c_h - c_a}{q}\right)$. The assembler has incentives to acquire additional modules as long as the reservation value of consumers for these modules is still high. As $\frac{\partial ?\pi_a(\alpha)}{(\partial \alpha)?} < 0$, the module $\hat{\alpha} = 1 - \left(\frac{c_h - c_a}{q}\right)$ maximizes the profit of the assembler.

Since by assumption $\frac{2c_h - c_a}{q} < \alpha < 1$, we need to have $\hat{\alpha} = 1 - \frac{c_h - c_a}{q} > \frac{2c_h - c_a}{q}$ for establishing the existence of this interior solution. This condition of existence is equivalent to $q > 3c_h - 2c_a$.

Proposition 2 implies that the assembler strongly prefers to be specialized in the assembling of the core modules and some intermediary modules and let the providers market the peripheral modules when the willingness-to-pay for the core modules is sufficiently high.

Now let consider the situation where $q < 3c_h - 2c_a$. Then $\frac{\partial \pi_a(\alpha)}{\partial \alpha} < 0$ for $\alpha = \left(\frac{2c_h - c_a}{q}\right)$, meaning that the assembler has still incentives to add other modules in its package. When these additional modules are directly sold, their price tends to exclude a part of the consumers, those facing the highest assembling costs). The profit of the assembler is thus defined by:

$$\pi_a(\alpha) = \int_{\alpha}^{\frac{2c_h - c_a}{q}} \left(\alpha q - c_a - \frac{(\theta q - c_a)^2}{2(c_h - c_a)} \right) d\theta + \int_{\frac{2c_h - c_a}{q}}^{1} \left(\alpha q - c_a - (\theta q - c_h) \right) d\theta$$

if $3c_h - 2c_a > q > 2c_h - c_a$

and by $\pi_a(\alpha) = \int_{\alpha}^{1} \left(\alpha q - c_a - \frac{(\theta q - c_a)^2}{2(c_h - c_a)} \right) d\theta$ if $q < 2c_h - c_a$

In both cases, the derivative of profits with respect to α is given by

$$\frac{\partial \pi_a(\alpha)}{\partial \alpha} = (1-\alpha)q - \alpha q + c_a + \frac{(\alpha q - c_a)^2}{2(c_h - c_a)} \tag{2}$$

The additional gain to add a module corresponds to the uniform price αq and the additional costs are equal to the cost of assembling this additional module (c_a) plus the cost linked to the degradation of the value of the existing modules $(1-\alpha)q$ and the price for acquiring this module $\frac{(\alpha q - c_a)^2}{2(c_h - c_a)}$.

Proposition 3: For any q such that $(2c_h - c_a) - \sqrt{\Delta} < q < 3c_h - 2c_a$, with $\Delta = 2(c_h - c_a)(2c_h - c_a - q)$, it is optimal for the assembler to offer a package containing a finite number of modules characterized by a value $\alpha \in [\tilde{\alpha}, 1]$, with $\tilde{\alpha} = \frac{(2c_h - c_a) - \sqrt{\Delta}}{q}$ and the remaining modules are directly sold by the providers and assembled by the end-users.

Proof : (2) can be rearranged as follows:

$$\frac{\partial \pi_a(\alpha)}{\partial \alpha} = \frac{\alpha^2 q^2 - 2\alpha q(2c_h - c_a) + c_a(2c_h - c_a) + 2q(c_h - c_a)}{2(c_h - c_a)}$$

Let $\Delta = 2(c_h - c_a)(2c_h - c_a - q)$ be the determinant of the polynomial equation in α, then we have a root that maximises the profit of the assembler $\tilde{\alpha} = \frac{(2c_h - c_a) - \sqrt{\Delta}}{q}$. As $(2c_h - c_a) > \sqrt{\Delta}$, then $\tilde{\alpha} > 0$. However, this is an interior solution if $\tilde{\alpha} < 1$ (otherwise at α=1, $\frac{\partial \pi_a(\alpha)}{\partial \alpha} > 0$ and the assembler would have no incentives to acquire additional modules). $\tilde{\alpha} < 1$ if $q > (2c_h - c_a) - \sqrt{\Delta}$. As $q < 3c_h - 2c_a$ by assumption and $3c_h - 2c_a > (2c_h - c_a)$, the existence of an interior solution is satisfied when $(2c_h - c_a) - \sqrt{\Delta} < q < 3c_h - 2c_a$.

When the value of the core modules (q) decreases, then the assembler is more tempted to acquire and assemble all the modules, because the degradation of quality is less severe and more than counterbalanced by the economies on assembling tasks. Therefore, the quality of the core module must reach a certain level ($q > (2c_h - c_a) - \sqrt{\Delta}$) to enable the emergence of a dual distribution on this market (a direct channel for the peripheral modules and the intermediary channel for the core modules).

However the repartition of modules between these two channels depends on the competitive advantage of the assembler in its technology of assembling and on the willingness-to-pay for the core module. Indeed, the number of modules packaged by the assembler increases in q and decreases in the assembling cost differential between the assembler and the end-users ($c_h - c_a$).

5. CONCLUSION

The theoretical framework developed in the paper gives some fruitful insights on digital business models. First, the presence of an assembler does not prevent independent providers from selling directly their modules. Thus, alternative channels of distribution can perfectly coexist on a digital market, especially if intermediated assembling induces a loss of value for the modules incorporated in the package.

Our model presents several limits. For example, it does not consider the possibility of versioning strategy for the assembler (Varian and Shapiro, 1998). If an assembler controls a set of modules, it can design a high quality package with all the modules sold at a high price and a low quality package with only a subset of modules but at a lower price. What impact could this strategy of versioning have on the market structure? Should it benefit to the assemblers and reinforce their role of intermediation between providers and end-users? For the assembler, there is a trade-off between an intensive strategy of versioning that implies higher packaging cost and a strategy of bundling (a single package including all the modules) that can exclude customers with a low willingness-to-pay. Versioning enables to discriminate between end users and to increase sales.

For the providers, as versioning increases the expected rent, the price of their modules should be higher. This can reinforce the role of intermediaries on digital markets. By selling directly their modules, the providers cannot exert any versioning and do not enhance the value of their modules as efficiently as the assemblers do. Therefore, the role of intermediaries should be strengthened on the markets where versioning is easy to implement like in digital markets. Our future researches will attempt to integrate versioning in our theoretical framework of digital assembling.

REFERENCES:

Acquisti A., Varian H. R., [2001], "Conditioning Prices on Purchase History", Working Paper, University of California, Berkeley

Arnold M., Pénard T. [2006] "Bargaining and Posted Prices: What does it change?" Working Paper.

Baye M.R., Morgan J., [2001] "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets", *American Economic Review*, 91(3), 454-474.

Baye M.R., Morgan J., [2002] "Information Gatekeepers and Price Discrimination on the Internet", *Economics Letters*.

Baye, M. R., Morgan, J., Scholten, P., [2001] "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site", Working Paper.

Baye, M. R., Morgan, J., Scholten, P., [2002] "Persistet Price Dispersion in Online Markets", Working Paper.

Bessy C. & E. Brousseau [1998], "Technology Licensing Contract: Features and Diversity", *International Review of Law and Economics*, 18, 451-489.

Biglaiser G. [1993] "Middlemen as experts", *Rand Journal of Economics*, Vol.24, N°2, Summer, pp.212-223.

Brousseau E. [2003], "Property Rights in the Digital Space", in Enrico Colombatto [ed], *Companion to Economics of Property Rights*, Edward Elgar, forthcoming

Brousseau, E. [2002]. "The governance of transaction by commercial intermediaries: An analysis of the re-engineering of intermediation by electronic commerce", *International Journal of the Economics of Business* 8(3), 2002.

Brynjolfsson E., Smith M.D. [2001], "Consumer Decision-Making at an Internet Shopbot", *The Journal of Industrial Economics*, XLIX(4), pp. 541-558.

Caillaud B., Jullien B., [2003] "Chicken & Egg: Competition Among Intermediation Service Provider", *RAND Journal of Economics*, 34(2), Summer 2003.

Chen Y., Iyer G., Padmanabhan V. (2002) "Referral Infomediaries", Marketing Science, Vol. 21, N°4, Fall, pp. 412-434.

Church J., N. Gandal [2004] "Platform Competition in Telecommunications", in *The Handbook of Telecommunications, Vol.2*, M. Cave et al. (Eds.), forthcoming.

Donsimoni, M.P. [1985] "Stable Heterogeneous Cartels ", *International Journal of Industrial Organization,* 3, 451-467.

Economides N, [1996], "The Economics of Networks", *International Journal of Industrial Organization*, 14(2), March

Economides N, Salop S. C., [1992], "Competition and Integration Among Complements, and Network Market Structure", *Journal of Industrial Economics*, XL(1), March, pp. 105-12.

Economides N., E. Katsamakas [2004] “Two-sided competition of proprietary vs. Open source technology platforms and the implications for the software industry”, WP, Stern School of Business, New York University.

Elkin-Koren N., E. Salzberger [2000], ‘The economic analysis of cyberspace: challenges posed by cyberspace to legal theory and legal rules”, *International Review of Law and Economics*, 19, 553-582

Evans D.S., A. Hagiu, R. Schmalensee [2004] “A Survey of the Economic Role if Software Platforms in Computer-Based Industries”, Cesifo Working Paper N° 1314.

Ghose A., Mukhopadhyay T., Rajan U. (2002) “Strategic Benefits of Referral Services”, Review of Marketng Science Working Papers, Vol.2, Issue 2.

Iñaki A., Martin, A. M. [2001] "On the strategic choice of spatial price policy: the role of the pricing game rules," Economics Bulletin, Economics Bulletin, vol. 12, pages 1-7.

Lancaster K., [1979], *Variety Equity and Efficiency*, Basil Blackwell, Oxford

McKelvey M., [2001] “The Economic Dynamics of Software: Three Competing Business Models Exemplified through Microsoft, Netscape and Linux, in Innovation in the Telecommunication Industry”, *Economics of Innovation and New Technology*, Special Issue, ed by E. Böhlin, E. Brousseau, S. Hulten, 10(2-3)

Parker J. G., Van Alstyne, M. W., [2000] “Information Complements, Substitutes, and Strategic Product Design”, Working paper, Tulane University - A.B. Freeman School of Business, November

Rochet J.-C., Tirole J., [2001] “Platform Competition in Two-Sided Markets”, Working Paper IDEI, University of Toulouse, November 26, 2001

Saliba [2003] “Pricing Behavior and Equilibrium Price Dispersion in online Markets”, PhD University of Delaware.

Scott Morton F., Zettelmeyer F., Risso J. [2001] “Internet Car Retailing “, *The Journal of Industrial Economics*, Vol. XLIX, N°4, pp.501-519.

Shapiro C., Varian H. [1998] *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press.

Smith, M.D. [2001] “The Law of one Price: The Impact of IT-enabled Markets on Consumer Search and Retailer Pricing “,, http://e-commerce.mit.edu.

Spulber, D.F., [1996] “Market Microstructure and Intermediation”*, Journal of Economic Perspectives*, , 10[3], pp. 135–52.

Thisse, J.-F., Vives, X., [1988] "On the Strategic Choice of Spatial Price Policy," American Economic Review, American Economic Association, vol. 78(1), pages 122-37.

Varian, H., [2000] “Market Structure in the Network Age”, in Brynjolfsson, E., Kahin, B., *Understanding the Digital Economy*, MIT Press.

Information Exchange and Competition in Communications Networks[§]

Carlo Cambini (Politecnico di Torino)[*]

and

Tommaso M. Valletti (Imperial College London and CEPR)[**]

This version: November 2005

Abstract

We develop a model of information exchange between calling parties. We characterize the equilibrium when two interconnected networks compete for such users by charging both for outgoing and incoming calls. We show that networks have reduced incentives to use off-net price discrimination to induce a connectivity breakdown when calls originated and received are complements in the information exchange. This breakdown disappears if operators are allowed to negotiate reciprocal access charges. We also study the relationship between sending and receiving retail charges as a function of the level of access charges. We identify circumstances where private negotiations over access charges induce first-best retail prices. In other words, endogenously chosen and unregulated access charges can internalize externalities between callers belonging to competing networks.

JEL Classification Codes: L41, L96.

Keywords: Interconnection, Access charges, Reception charges, Bill-and-keep, Information exchange.

---

§ We thank seminar participants and discussants in Helsinki, Florence (EUI), Keele, Munich, Pavia, Rome, Porto, San Diego and Toulouse for useful comments and discussions.

* Politecnico di Torino, DSPEA, C.so Duca degli Abruzzi, 24, 10139 Turin, Italy, e-mail carlo.cambini@polito.it

** Corresponding author. Imperial College London, Tanaka Business School, South Kensington campus, London SW7 2AZ, UK, e-mail t.valletti@ic.ac.uk, http://www.imperial.ac.uk/people/t.valletti

1. Introduction

People use electronic communications in order to exchange information. This exchange may take very different forms. Take, as an example, a situation with two individuals, A and B. A and B are academics trying to work on a joint paper but live in different places, thus they rely on telephone conversations and e-mail exchanges to conduct their work. One day, A and B are struggling to solve an equation. A finds the solution and calls B: B is obviously happy, but then does not need to call A back since the problem has been solved. Another day, A has an embryo of an idea and calls B. B needs some time to ponder over this rudimental idea and then calls A back to tell what she thinks about it: in this second case B is still happy of having received a call from A since it may lead to an improvement of their paper, but now the original call has initiated a chain reaction that has led B to return the call.

These simple examples illustrate a more general feature, namely that communications services are not consumed in isolation and involve interdependencies between callers. They also exemplify a phenomenon which seems particularly relevant for users, where an exchange of information may create the need for further exchanges of information, leading to a stimulation of calls. From a theoretical standpoint, two things are worth noticing. Firstly, receivers of an electronic communication enjoy a positive benefit (a *call externality*). Secondly, each individual has a demand for calls that depends not only on the price she has to pay, but also on the amount of calls received from the other party. Thus there is interdependency among incoming and outgoing messages and the demand functions depend on the prices charged to *both* parties. This paper studies a model of network competition when demand functions depend on the way information is exchanged between users.

Existing models of competition between interconnected networks typically assume that consumers derive utility only from making calls. This literature,[1] initiated by the seminal works of Armstrong (1998) and Laffont et al. (1998), asks if free negotiations over inter-carrier (wholesale) access charges can lead operators to choose (retail) fees that are detrimental to social welfare, for instance because they may be used to sustain collusive linear prices in the retail market. Another source of concern arises when access fees provide a commitment mechanism to limit competition over investment in infrastructure (Valletti and Cambini, 2005). If operators are of differing size they may not be able to reach a deal over the level of

[1] See the surveys of Armstrong (2002) and Vogelsang (2003).

reciprocal access charges (Carter and Wright, 2003). These works ignore the role played by call externalities which, as we argued above, are pervasive in electronic communications.

Given the normative intent of this literature, the absence of call externalities is potentially worrying since they have a strong impact on the (efficiency properties of the) competitive process. Recent literature has started to look into the problem of how to set retail charges in an environment with call externalities (Kim and Lim, 2001; DeGraba, 2003; Jeon et al., 2004; Berger, 2004; Hermalin and Katz, 2005). Jeon et al. (2004; JLT hereafter) and Hermalin and Katz (2005; HK) are the papers closest to ours. They both determine competitive retail and reception charges in a setting where outgoing and incoming calls are perfectly separable. JLT show that, in the presence of a reception charge, the call price goes down as it reflects a pecuniary externality: when a network lowers its price, the subscribers to the *rival* network will have to pay more as they receive more calls. Hence an increase in the rival's reception charge makes it more desirable for a network to expand its output. JLT also show that, in the presence of network-based discrimination both on final prices and on receiving charges, "connectivity breakdowns" arise. In order to discourage subscribers from connecting to the rival, a network has an incentive to charge extremely high off-net call prices or off-net receiver prices. These results are worrying from a policy perspective and call for intervention (e.g., regulation of reception charges). HK also find that carriers have a tendency to reduce off-net traffic and that, under rather general conditions, access charges cannot be used to achieve full efficiency, even if set by a benevolent regulator.

All these works, however, ignore the interdependencies among users' calls and assume that the sets of calls/messages sent and received are independent of each other.[2] This basic assumption contrasts with empirical findings on telecommunications demand that show how incoming and outgoing calls are strongly interdependent. The starting point of our approach derives from this empirical evidence, started by the work of Larson *et al.* (1990) on point-to-point demand. This framework disaggregates traffic into two distinct elements: autonomous traffic – independent of the traffic level between the networks – and induced traffic, which is influenced by the traffic volume. Induced traffic strictly depends on how information between two customers is exchanged. Taylor (2004) reviews the empirical literature from the early 90's that applied point-to-point models to communications between city

[2] An exception is represented by Hermalin and Katz (2004) who analyze two-way communications. Their focus is on the possible strategic game between the communicating parties as to who will be the sender and who the receiver. The concern of Hermalin and Katz (2004) is on pricing under uncertainty about the parties' values of an exchange while they do not analyze multiple networks and the problem of interconnection.

pairs in Canada (inter-provincial traffic) and in the US (inter-LATA traffic), and between countries (Canada/US among others). This literature *rejects* separability between outgoing and incoming traffic, and finds strong and statistically relevant reverse-calling phenomena. Calls tend to generate calls in return.[3]

In this paper we take at face value these empirical findings that have been neglected so far by the theoretical literature. Call externalities have been either ignored or modeled in a simple fashion by assuming perfect separability between incoming and outgoing calls. We start by studying the process of information exchange which is the primitive that leads parties to enjoy utility from calls. Utility is not generated by calls *per se*, but rather calls are inputs to the production of information. Calls made and received could be substitutes (if the information received does not require the receiver to call back) or complements (if information becomes "productive" only by responding to queries initiated by one the parties). We embed the fundamental process of information exchange between two parties calling each other into the IO framework of network competition.

We propose a general formulation of the information exchange process between senders and receivers to study the role of inter-network access charges. Firstly, we show that, when access charges are taken as given, the risk of a "connectivity breakdown" is much diluted when induced traffic is positive (i.e., calls made and received are complements in the information exchange). In addition, we consider if mild forms of intervention might allow the industry to self-regulate. We show that the breakdown is completely eliminated if operators can choose reciprocal access charges (neither JLT nor HK address the case of access charges endogenously chosen by operators). Intuitively, although reciprocal deals give firms an instrument to tacitly collude, there is no need to destroy off-net traffic. On the contrary, operators prefer to choose low access charges so that off-net prices are also low: in this way the intensity of competition for the market is reduced as customers find a "big" network less appealing than a "small" one. While concerns for a breakdown are weakened, we still show how there exists some (limited) room for intervention to improve the outcome of private negotiations, for instance by using suitable reception charges.

Secondly, we describe the interaction between access charges and retail prices, both for outgoing and incoming calls, when all of them are determined endogenously. We show that operators set positive reception charges *only* when access charges are sufficiently low. Thus we provide an explanation for the common practice in telecommunications to charge

[3] "A call in one direction stimulates something like one-half to two-thirds of a call in return" (Taylor, p. 129).

only for outgoing calls and not for incoming calls when access (termination) charges are high enough, although receiver charges could also be set in principle.[4] We also show the result that, when induced traffic is positive, there exist conditions such that unregulated operators sets reciprocal wholesale access charges that induce efficient retail pricing structures. This finding is quite striking and in stark contrast with both JLT and HK. It also clarifies the crucial role played by induced traffic: if incoming and outgoing calls are perfectly separable instead, we also find it impossible to achieve the first-best, as in the rest of the literature.

From the positive point of view our paper is about competitive pricing in the presence of externalities. Our analysis is also relevant, from the normative point of view, to assess the impact of recent regulatory interventions over inter-network charges. The discussion on the regulation of fixed-to-mobile termination charges in Europe, where CPP is in place, includes the adoption of charges based on Long Run Incremental Costs (LRIC), and the introduction of RPP. We establish a link between outgoing and incoming charges as a function of the access charge and we argue that RPP might emerge endogenously only if access charges are set sufficiently low. Thus we clarify some debate in the existing literature on the topic, where it is sometimes considered that CPP or RPP are mutually exclusive pricing policies (Crandall and Sidak, 2004; Littlechild, 2005). The determination of mobile-to-mobile termination is typically less intrusive and it includes reciprocal arrangements, and in some instances "bill-and-keep" deals. "Bill-and-keep" has been advocated as a way of sharing efficiently the value created by a call when both callers and receivers benefit from it (DeGraba, 2002). We give conditions under which this system is indeed efficient and we study if it can ever emerge from private negotiations.

The remainder of the paper is organized as follows. Section 2 introduces a model of information exchange between calling parties. Section 3 solves a model of network competition with uniform pricing. Section 4 analyzes the case with network-based price discrimination. Since price discrimination does not exhibit the result of profit neutrality with respect to access charges, we also ask what level would be chosen by negotiating operators as opposed to a social planner. In Sections 3 and 4 reception charges are fixed or set by a regulator, while Section 5 considers the case of market-determined reception charges. Section 6 concludes.

[4] The practice of charging customers both for originating and receiving calls is usually referred to as the Receiver Pays Principle (RPP) which is applied in the mobile markets of countries such as Canada, China, Hong Kong, and the U.S.. Positive reception charges make sense only if the receiver cares about being called, otherwise no one would ever answer a call. The practice of charging only for sending calls is known as the Calling Party Pays (CPP).

2. The information exchange process and call externalities

Imagine there are two users. One user subscribes to network i and the other user to network j and they communicate with each other. For the moment assume the networks are different, hence i can also denote the first user and j the second user. Each user is assumed to derive utility from the *amount of information created from originating and receiving calls* in a typical point-to-point connection. Information for user i, denoted I^{ij}, is produced according to the following production function: $I^{ij} = f^i(Q^{ij}, Q^{ji})$, where Q^{ij} is the quantity of outgoing calls user i originates to contact user j and Q^{ji} is the quantity of incoming calls that user i receives from user j. The production function $f(\cdot, \cdot)$ is one of the important primitives in our model and it can describe different cases when calls are substitutes or complements in the information creation process. $\partial I^{ij} / \partial Q^{ij} = f_1^i$ and $\partial I^{ij} / \partial Q^{ji} = f_2^i$ denote respectively how information changes at the margin with calls sent and received by i when communicating with j, and they are both assumed to be positive. We use superscripts ij to denote the exchange of information between i (the originator) and j (the receiver). This notation can also accommodate the case when both users are connected to the same network, in which case $j = i$ and $I^{ii} = f^i(Q^{ii}, Q^{ii})$.

Denote with p_i (respectively p_j) the price charged by network i (j) for calls originated by user i (j). Users have identical preferences over consumption of information and $U_i(I^{ij})$ denotes a standard (concave) utility function of user i when communicating with user j. We assume for the moment that users cannot be charged for receiving calls. Users i and j determine simultaneously the quantity of calls to consume by maximizing the following problems:

$$\text{User } i\text{: } \begin{cases} \max\limits_{Q^{ij}} U_i(I^{ij}) - p_i Q^{ij} \\ s.t. \quad I^{ij} = f^i(Q^{ij}, Q^{ji}) \end{cases} \qquad \text{User } j\text{: } \begin{cases} \max\limits_{Q^{ji}} U_j(I^{ji}) - p_j Q^{ji} \\ s.t. \quad I^{ji} = f^j(Q^{ji}, Q^{ij}) \end{cases}$$

The first-order conditions are respectively $U'_i f_1^i = p_i$ and $U'_j f_1^j = p_j$, which determine the equilibrium level of calls originated by user i (directed to j) and user j (directed to i): $Q^{ij} = Q^{ij}(p_i, p_j)$ and $Q^{ji} = Q^{ji}(p_j, p_i)$. Thus the quantity originated by a user is affected both by the price she pays and by the price other people pay.[5] Let $Q_1^{ij} = \partial Q^{ij}(p_i, p_j) / \partial p_i$ denote the "direct" price effect on quantities and $Q_2^{ij} = \partial Q^{ij}(p_i, p_j) / \partial p_j$ the "induced" price effect.

[5] We assume that both callers send a positive amount of calls. We discuss below the role played by this assumption.

Their sign is found by totally differentiating the first-order conditions. Denoting with $A^{ij} = U''_i (f_1^i)^2 + U'_i f_{11}^i < 0$ and $B^{ij} = U''_i f_1^i f_2^i + U'_i f_{12}^i$, it results:

$$\begin{bmatrix} A^{ij} & B^{ij} \\ B^{ji} & A^{ji} \end{bmatrix} \begin{bmatrix} dQ^{ij} \\ dQ^{ji} \end{bmatrix} = \begin{bmatrix} dp_i \\ dp_j \end{bmatrix}$$

In a symmetric equilibrium with $p_i = p_j$, $A^{ij} = A^{ji} = A$, $B^{ij} = B^{ji} = B$ and:

$$(1) \qquad \begin{cases} Q_1^{ij} = A/(A^2 - B^2) \\ Q_2^{ij} = -B/(A^2 - B^2) \end{cases}$$

If both users are connected to the same network, i.e., calls are "on-net", then $p_i = p_j$. For "on-net" calls the total effect of a price change is:

$$dQ^{ii}/dp_i = Q_1^{ii} + Q_2^{ii} = 1/(A+B).$$

In summary, a point-to-point exchange of information between two parties leads them to call each other, generating demand functions, $Q^{ij}(p_i, p_j)$ and $Q^{ji}(p_j, p_i)$. These demand functions are determined as a Nash equilibrium of the game played by the two users, where each user controls only one input of the information exchange. Demand functions react to the prices charged to both parties. The direct effect has a standard sign: $sign[\partial Q^{ij}/\partial p_i] = sign[A]$ $= sign[U''(f_1)^2 + U'f_{11}] < 0$. There is also an indirect effect that induces additional traffic: $sign[\partial Q^{ij}/\partial p_j] = sign[-B] = sign[-(U''f_1 f_2 + U'f_{12})]$. The sign of B depends on the properties of the information production function. It turns out to be zero only when utility is perfectly separable between the benefit derived from outgoing and incoming calls, as it is assumed by JLT and HK. As anticipated by the Introduction, separability does not seem to hold in the data. Calls tend to generate calls in return, thus sign[B] > 0. This would imply that incoming and outgoing calls are complements in the information exchange ($f_{12} > 0$).[6]

[6] The sign of the cross-price effects could also be thought in terms of the slope of the "best-response function" of each user, thus paralleling the cases of strategic complements and substitutes that are common to I.O. readers. In the problem of information exchange, however, users are not competing against each other in the usual sense.

2.1 A discussion of the information exchange

The exchange of messages is clearly a dynamic phenomenon. Our formulation should be seen as a reduced-form static representation of a more complex dynamic interaction between senders and receivers that we do not model explicitly. Thus one should not interpret the static Nash formulation described above as a game where the two end users must decide at a single, specific point in time how much they are going to call each other. To continue with our co-author example from the Introduction, we are not assuming that A, at the beginning of the collaboration with B, must decide how many messages he will send to B over the entire course of their collaboration, while B makes a simultaneous decision of the messages he will send to A.[7] However, it is still true that when A calls B because, at a given moment, he has information relevant to both, this would benefit B and might trigger further messages from B to A in a later period. It is this sequence of calls that the model of information exchange captures, not the one-way calling at each point in time. The cumulate amount of calls $Q^{ij}(p_i, p_j)$ from i to j is affected by the price paid by both parties, and is a simple way to capture, in a reduced-form, the interaction between users, which represents a significant departure from the existing literature that assumes independence between messages sent and received.

Having clarified the interpretation to the Nash representation, we also emphasize that the Nash solution described in Section 2 presupposes that there is an *interior* solution where both parties make some calls, i.e., both $Q^{ij} > 0$ and $Q^{ji} > 0$. It is easy to find light restrictions on functional forms that ensure that the Nash equilibrium game between users has an interior solution with a positive quantity of calls for both users and degenerate cases do not arise.[8]

These restrictions are reasonable to describe the information exchange. To see why, let us discuss what our approach rules out. We rule out corner solutions that would arise, for instance, if the amount of information sent and received were perfect substitutes, e.g., $I^{ij} = f(Q^{ij} + Q^{ji})$.[9] This functional form would be problematic for two reasons. Firstly, it is not clear why, any time the amounts of calls sent and received differ, the sender and the receiver must benefit in equal terms, as the functional form would imply. On the contrary, the amount of calls sent and/or received is expected to confer different benefits to the sender and to the

[7] To date, no one of the authors of this paper has had a privilege of working with such an organized co-author!

[8] These conditions are sufficient to ensure an interior solution: (i) $U_i(.)$ is differentiable and concave, (ii) $U_i(f^i(0,0)) = 0$, (iii) $U_i(f^i(0, Q_{ji}) > 0$ and there exists a value $\overline{Q}_{ij}$ such that $U_i(f^i(\overline{Q}_{ij}, Q_{ji}) - p_i\overline{Q}_{ij} > U_i(f^i(0, Q_{ji})$, (iv) $\lim_{Q_{ij}\to\infty} U_i(f^i(Q_{ij}, Q_{ji})) - p_i Q_{ij} < 0$, (v) $U(f^i(Q_{ij}, Q_{ji})) \neq U(f^j(Q_{ji}, Q_{ij}))$, when $Q_{ij} \neq Q_{ji}$.

[9] This functional form violates condition (v) of footnote 8.

receiver. Secondly, this functional form that we rule out would imply that, when the two parties are symmetric, and p_i differs from p_j by a very small amount, only one party (the one that faces the cheaper price) would call, while the other party would never make a call, *for the entire duration of the information exchange*. This is not realistic in our view. Even in the extreme case where the parties are able to achieve a good level of co-ordination, so that the party that faces the cheaper price would make most calls, we can easily think of some point in time when the party that faces the more expensive prices would still have something to say and initiate the call, notwithstanding the price difference. This is precisely what our approach tries to capture and justifies its reliance on the manipulation of the first-order conditions of interior solutions.

Information exchange is a metaphor for representing the complex interaction between callers. It is easy to accommodate this framework also in the presence of reception charges. If network i (j) charges its customers a price r_i (r_j) per call for receiving calls, then the Nash equilibrium described in Section 2 that generates $Q^{ij}(p_i, p_j)$ and $Q^{ji}(p_j, p_i)$ is unchanged as long as customers accept every call, which indeed would happen if $U'_i f_2^i \geq r_i$ and $U'_j f_2^j \geq r_j$. We call this situation where reception charges do not "bind", either because they are zero or "low" enough, a situation of "sender sovereignty".

When reception charges do bind, the same analysis can be applied. In particular, if both reception charges are "high" enough compared to sending charges, then each customer effectively controls the amount of calls *received*. The first-order conditions in this case are respectively $U'_i f_2^i = r_i$ and $U'_j f_2^j = r_j$, which determine the volume of calls at the Nash equilibrium: $Q^{ij}(r_j, r_i)$ and $Q^{ji}(r_i, r_j)$. This situation of "receiver sovereignty" arises when sending charges do not bind: $U'_i f_1^i \geq p_i$ and $U'_j f_1^j \geq p_j$.

Finally, a last possibility arises when the binding prices (both for sending and receiving calls) are set by one network alone. This case is denoted as a situation of "network sovereignty". For instance, if network i is sovereign, then user i chooses both inputs in the information exchange. The first-order conditions in this case are $U'_i f_1^i = p_i$ and $U'_i f_2^i = r_i$, which determine the volume of calls: $Q^{ij}(p_i, r_i)$ and $Q^{ji}(r_i, p_i)$. Notice that this is a simple optimization problem conducted by user i, since user j is passive. This situation of "network i sovereignty" arises when user j accepts all calls received ($U'_j f_2^j \geq r_j$) and cannot send additional calls since they are truncated by i ($U'_j f_1^j \geq p_j$). The four situations of sender/receiver/network *i*/network *j* sovereignty exhaust all possible cases for the determination of the volume of calls. Our analysis in Sections 3 and 4 concentrates on sender sovereignty since it is the more relevant case from an empirical standpoint. We do so by considering reception charges that are regulated and "low" enough, so that they do not bind. We do examine also the other situations when we analyze market-determined reception charges in Section 5.

3. The model of network competition

Demand for calls between a pair of users is generated according to the model described in the previous Section and summarized by the price effects of eq. (1). The model of competition follows Armstrong (1998), Laffont et al. (1998) and its extension by JLT. There are two networks, differentiated à la Hotelling. A unit mass of consumers is uniformly located on the segment [0,1] while the network operators are located at the two extremities. We denote by 1 (respectively 2) the firm located at the origin (respectively at the end) of the line. Customers pay a three-part tariff, $T_i(q) = F_i + p_i q_o + r_i q_r$, $i = 1, 2$, where the fixed fee F_i can be interpreted as a subscriber line charge, p_i as the unit price for originating q_o calls, r_i as the unit price that network i's consumer pays for receiving q_r calls.[10]

When a consumer located at l buys from firm i located at l_i, her utility is given by $y + v_0 - |l - l_i|/(2\sigma) + w_i$, where y is the income, v_0 is a fixed surplus component from subscribing (sufficiently large such that all customers always choose to subscribe to a network) and w_i is the indirect utility derived from making and receiving calls that we describe below. The parameter σ represents an index of substitutability between the networks.[11]

The consumer indifferent between the two networks determines the market share of the two firms. In particular firm i's share is α_i where:

$$(2) \qquad \alpha_i = \alpha(w_1, w_2) = 1/2 + \sigma(w_i - w_j)$$

where:

$$(3) \qquad w_i = \alpha_i v(p_i, p_i) + \alpha_j v(p_i, p_j) - r_i \left(\alpha_i Q^{ii}(p_i, p_i) + \alpha_j Q^{ji}(p_j, p_i)\right) - F_i$$

[10] The case with network-based price discrimination is considered in Section 4.

[11] HK concentrate most of their analysis on the case of undifferentiated networks ($\sigma \to \infty$) without fixed costs, which leads to possible tipping effects where all users subscribe to a single network.

(4) $$v(p_i, p_i) = U_i(f^i(Q^{ii}(p_i, p_i), Q^{ii}(p_i, p_i)) - p_i Q^{ii}(p_i, p_i)$$

(5) $$v(p_i, p_j) = U_i(f^i(Q^{ij}(p_i, p_j), Q^{ji}(p_j, p_i)) - p_i Q^{ij}(p_i, p_j).$$

This system of equations describes a situation where a customer has many point-to-point information exchanges. A customer belonging to network i has α_i exchanges "on net" and α_j exchanges "off net". Each on-net exchange generates a quantity of calls Q^{ii} originated by the user connected to network i and destined to a user connected to the same network. This quantity depends only on p_i, and $v(p_i, p_i)$ is the corresponding (indirect) utility derived from each point-to-point exchange on net. Similarly, $v(p_i, p_j)$ is the (indirect) utility derived from each off-net exchange (this depends not only on p_i but also on p_j since the reverse traffic is originated by the rival network).

Both networks have full coverage. Serving a customer involves a fixed cost f of connection and billing. The marginal cost is c per call at the originating end and t at the terminating end. The total marginal cost is $c + t$. Networks pay each other a reciprocal two-way access charge, denoted by a, for terminating each others calls. Thus, network i's profit is given by:

(6) $$\begin{aligned}\pi_i &= \alpha_i \{\alpha_i (p_i - c - t + r_i) Q^{ii}(p_i, p_i) + \alpha_j (p_i - c - a) Q^{ij}(p_i, p_j) + F_i - f\} + \\ &+ \alpha_i \alpha_j (a - t + r_i) Q^{ji}(p_j, p_i)\end{aligned}$$

The first term in (6) is the profit deriving from originating calls terminated on the same network ($\alpha_i^2 (p_i - c - t + r_i) Q^{ii}$); the second term is profit from originating calls destined to the rival network ($\alpha_i \alpha_j (p_i - c - a) Q^{ij}$); the last term is the profit from termination and reception of incoming calls originated by rival customers ($\alpha_i \alpha_j (a - t + r_i) Q^{ji}$).

We consider the following timing of the game. First interconnection (a) and reception (r) terms are set in stage I. Then operators compete in two-part prices (F and p) in stage II. In particular, we assume that reception charges are always set "low" enough. In this way, we can concentrate on the case of "sender sovereignty" only. This analysis obviously includes the case where reception charges do not exist ($r = 0$) as is the case in many countries that adopt CPP. It also deals with the case of reception charges regulated in stage I by a benevolent social planner that makes sure that they do not "bind" in stage II.

3.1. Equilibrium with uniform pricing under "sender sovereignty"

In the Appendix we obtain the following expression for calling prices at equilibrium:

$$(7) \qquad p_i^* = c + t + \alpha_j (a - t)(\Phi - \Gamma) - \alpha_i r_j \Phi + (\alpha_i U_j' f_2^j \Phi - U_i' f_2^i \Theta)$$

where:

- $\Phi = \dfrac{\partial Q^{ij} / \partial p_i}{\partial Q_o^i / \partial p_i} = \dfrac{Q_1^{ij}}{\alpha_i (Q_1^{ii} + Q_2^{ii}) + \alpha_j Q_1^{ij}}$, where $Q_o^i = \alpha_i Q^{ii} + \alpha_j Q^{ij}$ denotes the total quantity of calls originated by network i and destined both on-net and off-net. Φ represents the ratio between the marginal change in outgoing off-net calls and the average marginal change in all outgoing calls (both off- and on-net, including the return effect on the latter), caused by an increase of price p_i; Φ is a summary of the "outgoing return phenomenon".
- $\Gamma = \dfrac{\partial Q^{ji} / \partial p_i}{\partial Q_o^i / \partial p_i}$ is the ratio between the marginal change in incoming off-net calls and the average marginal change in all outgoing calls, caused by an increase of price p_i. Γ is a summary of the "incoming return phenomenon".
- $\Theta = \dfrac{\partial Q_r^i / \partial p_i}{\partial Q_o^i / \partial p_i}$, with $Q_r^i = \alpha_i Q^{ii} + \alpha_j Q^{ji}$ denoting the total quantity of calls received by users connected to network i. Θ represents the ratio between the average marginal change in all incoming calls and the average marginal change in all outgoing calls, caused by an increase of price p_i. Θ is a summary of "incoming-to-outgoing" calls.

Eq. (7) can be interpreted as follows:

- $c + t + \alpha_j (a - t)(\Phi - \Gamma)$ represents the network's direct perceived marginal cost. The possible termination mark-up for off-net calls is multiplied by the difference between the off-net outgoing and incoming ratios, reflecting the two-way effect of call propagation;[12]

[12] Empirically, one would expect $\Phi \geq \Gamma$. Also notice that the sign of the termination mark-up does not matter if $\Phi = \Gamma$, i.e., when outgoing and incoming calls are perfect complements.

- $-\alpha_i r_j \Phi$ is the pecuniary externality due to the presence of the reception charge r_j. This is multiplied only by the factor Φ, the outgoing return ratio from *i* to *j*, since this is the only direction that matters for the pecuniary externality imposed on *j*'s customers.
- $\alpha_i U'_j f_2^j \Phi - U'_i f_2^i \Theta$ represents the difference between the marginal utility of user *j* from receiving information through calls originated by users *i* (the marginal utility per point-to-point exchange is multiplied by the off-net outgoing ratio Φ and by *i*'s share) and the marginal utility of receiving calls by user *i* (user *i* gets information both on- and off-net, thus the marginal utility per exchange is multiplied by the incoming-to-outgoing ratio Θ). Prices thus depend on the difference between the positive externalities generated by the networks, as in a typical two-sided market setting (Rochet and Tirole, 2003 and 2004).

Eq. (7) generalizes previous results obtained in the literature. When the utility function is perfectly separable between outgoing and incoming calls, as it is in JLT, it results $Q_1^{ii} = Q_1^{ij}$ and $Q_2^{ij} = Q_2^{ii} = 0$, so that the incoming ratio is Γ = 0 and the outgoing ratio is Φ = 1, while the incoming-to-outgoing ratio simplifies to Θ = α_i. Then eq. (7) simplifies to $p_i^* = c + t + \alpha_j (a-t) - \alpha_i r_j + \alpha_i \left(U'_j f_2^j - U'_i f_2^i \right)$, that corresponds to (and partly amends) the result of JLT.[13] If users do not receive utility from being called (f_2 = 0) and are not charged for reception (r = 0), then (7) becomes simply $p_i^* = c + t + \alpha_j (a-t)$, as in Laffont et al. (1998). Under our more general specification we obtain the following:[14]

Proposition 1. *Under "sender sovereignty", the symmetric equilibrium with uniform multi-part pricing is characterized by the following conditions:*

a) the call price is given by:

(8) $$p^* = c + t + \frac{2}{2+x}\left[\frac{a-t}{2}(1-x) - \frac{r}{2} - xU'f_2 \right]$$

where x is a "propagation" factor given by: $x = \dfrac{\Gamma}{\Phi} = \dfrac{\partial Q^{ji} / \partial p_i}{\partial Q^{ij} / \partial p_i}$;

[13] Proposition 1 of JLT (p. 95) misses the last term in brackets, which is zero only in a symmetric equilibrium. This term represents the difference in users' marginal utility for receiving calls. It shows that if the marginal utility from receiving calls is higher for a rival user than for a own user, then, ceteris paribus, a network would *increase* its own price in order to reduce its outgoing calls and reduce the benefits of users of the rival network.

[14] All the proofs are relegated to the technical Appendix.

b) the fixed charge is given by: $F^* = f + 1/2\sigma - (p^* + r - c - t)Q(p^*)$*;*

c) each network's profit is equal to $1/4\sigma$*.*

The result of profit neutrality (part c)) is common in symmetric models of network competition with multi-part uniform pricing. The more interesting feature of Proposition 1 lies in the expression for the call price given by eq. (8) where we have introduced a new element, the propagation factor x. This denotes the ratio between the change of off-net (return) calls and the direct change of on-net calls when the own price changes. In other words the propagation factor measures the relative impact of the reverse traffic phenomenon.

Using eq. (1) and the notation from Section 2, the expression for the propagation factor in a symmetric equilibrium is:

$$x = -B/A = -\left(U'' f_1 f_2 + U' f_{12}\right)/[U''\left(f_1\right)^2 + U' f_{11}].$$

Since A is negative, the sign of x is the same as the sign of B. As we argued in the Introduction, the empirical evidence available to us suggests that incoming and outgoing calls are complements in the information production function, hence x should take positive values, although there is nothing in theory to prevent substitution effects, leading to negative values of x. It is also realistic to imagine that $x < 1$, while $x = 1$ corresponds to the case of perfect complements. This factor reduces to zero when the separability condition on utility holds, as in JLT, in which case eq. (8) reduces to $p^* = c + t + (a - t - r)/2$.

The impact of the propagation factor on the operators' pricing is two-fold. On the one hand, a higher x tends to decrease the price in order to induce the "call generates calls" phenomenon which is enjoyed by customers: this is the last term of the square bracket in eq. (8). On the other hand, call propagation also impacts on call prices via access imbalances and reception charges. The "perceived" off-net cost $(a - t)/2$ is pushed down the higher is the propagation factor x as every call off-net (for which $a - t$ is paid) induces a fraction of return calls (for which $a - t$ is received).

3.2 Welfare analysis

We now compare the equilibrium call charges with the social optimum. This is found by maximizing total welfare, i.e., the utility of the calling and the receiving parties minus the total cost, leading to $U'(f_1 + f_2) = c + t$. The optimal outcome reflects the externality and, under sender sovereignty (i.e., $U'f_2 \geq r$), could be implemented by setting the following call charge:

$$p = c + t - U'f_2 .$$

The optimal call charge should be set below the marginal cost, accounting for the positive externality from receiving calls. By equating eq. (8) to the optimal charge, one derives the following condition that would ensure efficiency:

(9) $$(a - t)(1 - x) - r = -(2 - x)U'f_2 .$$

Since there are two prices to be set (*a* and *r*), several cases can arise:

- Imagine there is no reception charge ($r = 0$), as is the case under a CPP system, then efficiency can be ensured only with below-cost access charges. In this case in fact from eq. (9) one gets $a = t - (2 - x)U'f_2 /(1 - x) < t$. To understand this result, recall that efficiency dictates that the call price is set below cost: without a reception charge the only way to decrease the retail price is to cut the "perceived" marginal cost of a network, thus $a < t$. If the benefit from receiving calls is high, this might imply a "bill-and-keep" ($a = 0$).
- Imagine termination is set at the LRIC level, $a = t$, then efficiency is increased with a positive reception charge; however the first-best is not compatible with "sender sovereignty" if $x < 1$. In this case in fact from eq. (9) one gets $r = (2 - x)U'f_2$. The level of the reception charge decreases with the level of the propagation factor but violates the constraint $U'f_2 \geq r$. Only the case of perfect complements ($x = 1$) makes LRIC regulation compatible with efficiency and sender sovereignty. This last result is more general: if $x = 1$ efficiency can always be ensured for *any* access charge if $U'f_2 = r$.

- A corollary of the previous case is that efficiency cannot be ensured under sender sovereignty if $a > t$ since from eq. (9) $r = (2-x)U'f_2 + (a-t)(1-x) > U'f_2$; thus no reception charge can induce to "call more" since the receivers would hang up.
- Imagine the reception charge is set at the network termination cost, $r = t - a$, then efficiency can be reached only if access is priced below marginal cost. In this case in fact from eq. (9) one gets $a - t = -U'f_2 < 0$. Notice that these conditions imply $U'f_2 = r$ and $a = t - r$, thus they just satisfy the constraint of "sender sovereignty". When the externality is high, this might imply a zero termination charge ($a = 0$) and a reception charge equal to the marginal cost of termination ($r = t$).

4. Network-based discrimination

Under uniform pricing operators are indifferent with respect to the level of the reciprocal access charge. This makes the model not very interesting to understand if the industry could "self-regulate" with a minimal degree of intervention, e.g., by simply imposing a reciprocity clause. To address this question we now consider the case of network-based discrimination, both on final prices and on receiving charges. In other words, customers are offered five-part tariffs of the form: $T_i(q) = F_i + p_i q_o + \hat{p}_i \hat{q}_o + r_i q_r + \hat{r}_i \hat{q}_r$, $i = 1, 2$, where $\hat{r}_i$ is the unit price for receiving calls originated on network $j \neq i$ and $\hat{p}_i$ is the unit price for off-net calls.

Firm i's share is still given by eq. (2) where now:

$$w_i = \alpha_i v(p_i, p_i) + \alpha_j v(\hat{p}_i, \hat{p}_j) - r_i \alpha_i Q^{ii}(p_i, p_i) - \hat{r}_i \alpha_j Q^{ji}(\hat{p}_j, \hat{p}_i) - F_i$$

is the net surplus for customers connected to network i. Notice that with the notation above we have again implicitly assumed that reception charges are "low" enough so that they do not bind in determining call volumes that are determined instead by senders alone. Network i's profit becomes:

$$\begin{aligned} \pi_i &= \alpha_i \{\alpha_i (p_i - c - t) Q^{ii}(p_i, p_i) + \alpha_j (\hat{p}_i - c - t) Q^{ij}(\hat{p}_i, \hat{p}_j) + F_i - f\} + \\ &+ \alpha_i \{\alpha_i r_i Q^{ii}(p_i, p_i) + \alpha_j \hat{r}_i Q^{ji}(\hat{p}_j, \hat{p}_i)\} + \alpha_i \alpha_j (a - t)[Q^{ji}(\hat{p}_j, \hat{p}_i) - Q^{ij}(\hat{p}_i, \hat{p}_j)] \end{aligned} \tag{10}$$

The timing of the game is as before. First interconnection (a) and reception (r and $\hat{r}$) terms are set in stage I. Then operators compete in multi-part prices (F, p and $\hat{p}$) in stage II.

4.1. Equilibrium in stage II: price competition

Operators can internalize everything on-net, resulting in an efficient amount of on-net calls. On the contrary, there are uninternalized externalities off-net and thus inefficiency can arise with off-net communications. In the Appendix we prove the following:

Proposition 2. *When a symmetric equilibrium with sender sovereignty and termination-based discrimination exists, this equilibrium is characterized by:*

a) the on-net unit price is at the socially optimal value: $p^* = c + t - U'f_2$;

b) the off-net unit price is given by the following expression:

(11) $\hat{p}^* = c + a - (a-t)x - \hat{r} + U'f_2(1-x)$;

c) the fixed charge is given by the following expression:

(12) $F^* = f - (p^* - c - t)Q(p^*, p^*) + 1/2\sigma - \hat{r}Q(\hat{p}^*, \hat{p}^*) - \left(v(p^*, p^*) - v(\hat{p}^*, \hat{p}^*)\right)$;

d) a "connectivity breakdown" is less (more) likely to arise when the propagation factor x is positive (negative).

JLT find that when the utility function is separable and the utility from incoming calls is a fraction β of the utility from outgoing calls, there is always a "connectivity breakdown" if $\beta \geq 1$ via infinitely high off-net prices. To compare our results with JLT, denote with $\beta = f_2 / f_1$ so that $U'f_2 = \beta\hat{p}$. From eq. (11) the equilibrium off-net price is re-written as:

(11bis) $\hat{p}^* = \dfrac{c + a - (a-t)x - \hat{r}}{1 - \beta(1-x)}$,

which takes finite values for $0 \leq \beta < \frac{1}{1-x}$. Even if β tends to 1 we do not have connectivity breakdown if the propagation factor is positive. In our framework, the breakdown appears (and is even more pronounced) when calls are substitutes ($x < 0$) in the production of information since it could occur also for values of $\beta < 1$, while the phenomenon of network breakdown is less likely when calls are complements ($x > 0$). Notice that when $x = 1$, i.e., when a change of price induces the same change in outgoing and incoming volumes, then the off-net price simplifies to $c + t - \hat{r}$. The access charge a does not matter at all since for any outgoing call (for which a has to be paid) there will also be an incoming return call (for which a will be received).

A positive propagation factor helps ensure that the off-net exchange of information is not cut via a breakdown induced by high retail call charges when senders determine volumes.[15] A mirror result arises in the case of "receiver sovereignty" when calling prices do not bind. In this case, connectivity breakdown would occur in JLT if $\beta < 1$, via infinitely high off-net reception prices. In our model, we do not have a breakdown if the propagation factor is positive since reception charges take finite values for $\beta > 1 - x$.[16]

All in all, the magnitude of the connectivity problem is reduced compared to JLT if calls originated and received are complements in the information production function. To understand this, recall that the breakdown typically happens because one network (say the receiving) has the opposite interest to the other network (say the originating). However, once calls are complements to generate information, then the receiving network will need to allow termination of calls if it wants also its own originating calls to become "productive".

4.2 Stage I and the role of access charges

We now ask what level of access charges would be chosen in the first stage by negotiating firms as opposed to a social planner, to see if and how private and social interests diverge. In case of a discrepancy, in stage I a benevolent regulator may also set a reception charge.

The stage-II profit in a symmetric equilibrium is given by:

$$\pi^* = \tfrac{1}{4}\left(p^* + r - c - t\right)Q(p^*, p^*) + \tfrac{1}{4}\left(\hat{p}^* + \hat{r} - c - t\right)Q(\hat{p}^*, \hat{p}^*) + \tfrac{1}{2}\left(F^* - f\right),$$

where the expressions for the equilibrium off-net prices and fixed fees are given respectively by eq. (11bis) and (12): these are the only prices affected by access charges and by off-net reception charges. We imagine that in the first stage network operators are free to negotiate reciprocal access prices. We also suppose that a low-enough $\hat{r}$ is set exogenously and may be

[15] Since we have assumed so far that the volume of traffic is determined by callers, it must be $\beta\hat{p} \geq \hat{r}$, which from eq. (11bis) can be re-written as $\hat{r} \leq \beta\left[a(1-x) + c + tx\right]/(1 + \beta x)$.

[16] A formal proof is available from the authors.

controlled only by the regulator. Given the symmetry, negotiations over access charges are equivalent to maximizing each operator's profit. The reciprocal access charge is set such that

$$(13)\qquad \frac{\partial \pi^*}{\partial a} = \frac{1}{4}\left[\frac{\partial Q(\hat{p}^*,\hat{p}^*)}{\partial a}\left(\hat{p}^* - c - t + \hat{r}\right) + Q(\hat{p}^*,\hat{p}^*)\frac{\partial \hat{p}^*}{\partial a}\right] + \frac{1}{2}\frac{\partial F^*}{\partial a} \leq 0 .$$

The effect of a on the volume of off-net calls is:

$$(14)\qquad \frac{\partial Q(\hat{p}^*,\hat{p}^*)}{\partial a} = \left[Q_1(\hat{p}^*,\hat{p}^*) + Q_2(\hat{p}^*,\hat{p}^*)\right]\frac{\partial \hat{p}^*}{\partial a} = (1+x)Q_1(\hat{p}^*,\hat{p}^*)\frac{\partial \hat{p}^*}{\partial a} .$$

The effect of a on the off-net price $\hat{p}^* = \dfrac{c + a - (a-t)x - \hat{r}}{1-\beta(1-x)}$ is quite intricate in general. We cannot suppose that the ratio β of marginal utilities does not change with a. To simplify the problem, we assume from now that the propagation effect does not substantially change with a change in prices. In other words, if calls "propagate" at certain vector of prices – for instance every call sent induces 0.5 calls in return – a change of prices would not change the information exchange and calls would still "propagate" in the same manner with x = 50%. This assumption is restrictive but is arguably realistic for small price changes.[17] Under this simplifying assumption, we obtain the following:

Proposition 3. *If an interior solution exists, network operators jointly agree on the following reciprocal access charge:*

$$(15)\qquad a^* = t + \frac{1}{(1-x)(1+2\beta)}\left[-\beta(3-x)(c+t) + \hat{r}(2+\beta+\beta x) + \frac{Q(\hat{p}^*,\hat{p}^*)}{Q_1(\hat{p}^*,\hat{p}^*)}\frac{1-\beta(1-x)}{1+x}\right].$$

In order to understand eq. (15), consider the standard case with no reception charge ($\hat{r}=0$) and no benefit from receiving calls (β = 0). Then eq. (15) simplifies to $a^* = t + Q/Q_1 < t$. This is the finding of Gans and King (2001): firms "collude" by setting be-

[17] This assumption is strictly true with utilities functions such as $U^i = (Q^{ij} + \rho Q^{ji})^{\alpha}$ with $\alpha < 1$ and $0 < \rho < 1$. In this case the propagation factor is constant and equal to $x = -\rho < 0$. This particular specification also exhibits β = constant. An example with a constant and positive propagation factor is $U^i = Q^{ij}(1+\rho Q^{ji}) - Q^{ij^2}/2$, with $x = \rho > 0$. In this case the ratio β is not constant.

low-cost access charges, in the limit by adopting a "bill-and-keep" system. This is because the off-net calling charge becomes cheaper than the on-net charge, and customers then prefer to belong to the "smaller" network so that they make many cheap off-net calls. Operators are less aggressive, since it pays to be small. In equilibrium, this results in higher profits. This mechanism carries forward to our more general framework. The first term in the square bracket of eq. (15) is negative when customers care about receiving calls: part of this benefit is internalized by the operators, and they tend to push off-net calling prices down via lower access charges. The second term in the bracket reflects the impact of the reception charge. As operators try to avoid competition by pushing off-net prices down, and as the reception charge already achieves this via the pecuniary externality, then there is no need to push a too low when this effect is ensured by the reception charge. The last term is the effect of Gans and King, corrected by the call externality and by the propagation factor.

Eq. (15) is only a candidate equilibrium charge since, as we discuss in the Appendix, restrictions are needed to ensure that profit is concave in a. Also, this solution holds so long as the reception charge does not bind, i.e., $\hat{r} \leq \beta\hat{p}$. When it exists and is positive, an interesting implication is that connectivity breakdown would *not* happen even when β is very high: when a is endogenized and takes the value of eq. (15), then $\hat{p}^*$ from (11bis) becomes:

$$(16) \qquad \hat{p}^* = \frac{c+t+\hat{r}}{1+2\beta} + \frac{Q}{Q_1(1+2\beta)(1+x)},$$

that does not go to infinity for *any* value of the propagation factor and of the utility from receiving calls. This contrasts with Proposition 2, where we found that, under some parameter configuration, a connectivity breakdown would happen. The difference is that now operators can adjust the access charge as well: as the utility from receiving calls increases or as the propagation factor goes up, the access charge is pushed down so that the off-net price always takes finite values. In fact, it is easy to see that, if the reception charge is zero, then from eq. (16) the off-net price is strictly lower than the on-net price: $\hat{p}^* < p^* = (c+t)/(1+\beta)$.

This result has two immediate implications. Firstly, connectivity breakdown concerns are totally eliminated if operators are allowed to negotiate over access charges. Secondly, in the absence of reception charges, negotiations would not necessarily deliver efficient outcomes, since they would result in off-net prices that are "too" low. The latter point leads to our next policy question. How does the privately chosen access charge compare to the socially optimal one? The access charge that maximizes social welfare is the one that makes the privately optimal off-net price in eq. (11bis) equal to the socially efficient price:

$$\hat{p}^* = p^* \Rightarrow \frac{c+a-(a-t)x-\hat{r}}{1-\beta(1-x)} = \frac{c+t}{1+\beta}.$$

Thus, the socially optimal access charge is:

$$(17) \qquad a^W = t - \frac{(c+t)\beta(2-x)}{(1-x)(1+\beta)} + \frac{\hat{r}}{1-x}.$$

Note that $a^W = t$, the so-called LRIC rule, is typically not socially optimal in the absence of reception charges, as we already discussed in Section 3.2 under uniform pricing. If $\hat{r} = 0$, then $a^W < t$ is socially efficient and so is a "bill-and-keep" system when β is sufficiently high. The same arguments apply here. Evaluating the difference between eq. (15) and eq. (17), we get:

$$(18) \qquad a^* - a^W = \frac{1-\beta(1-x)}{(1+\beta)(1+2\beta)(1-x)}\left[-\beta(c+t) + \hat{r}(1+\beta) + \frac{Q}{Q_1}\frac{1+\beta}{1+x}\right].$$

Eq. (18) is informative about the difference between private and social optimal access charges. In the special case when $\beta = 1/(1-x)$ there is no wedge between the two charges: the industry would self-regulate with no need of intervention. However, this result does not hold in general. In the absence of reception charges, negotiations over reciprocal access charges would set them too low from a social point of view, as both the first and the third term of the RHS of (18) are negative. Still, this may not be a problem if the socially-optimal charge from eq. (17) is already equal to zero: operators could not go below this level, thus they would agree on an efficient "bill-and-keep" system without having to mandate it.

If, on the other hand, both a^* and a^W take positive values, then it possible for the regulator to use a regulated reception charge to achieve efficiency. This level of the reception charge is determined by equating the LHS of (18) to zero, finding:

(19) $$\hat{r}^W = -\frac{Q(\hat{p}^*, \hat{p}^*)}{Q_1(\hat{p}^*, \hat{p}^*)} \frac{1}{1+x} + \frac{\beta}{1+\beta}(c+t).$$

Our results on welfare are summarized in the following proposition:

Proposition 4. *In stage I of the game:*

a) unregulated negotiations over reciprocal access can eliminate the "connectivity breakdown" problem for off-net calling charges;

b) an efficient "bill-and-keep" system can be chosen by negotiating operators;

c) when negotiations lead to a positive access charge, the regulator can use positive reception charges to get closer to efficient choices, although he never achieves the first-best.

In summary, we have found two encouraging results, and a mildly encouraging one. We have shown that free negotiations over the reciprocal access charge would ensure that connectivity breakdown problems disappear. This is the first encouraging result. The intuition can be obtained once again by recalling that operators are better off by trying not to compete too hard against each other: only by allowing *cheap* off-net calls they can ensure that this is achieved. Their mutual interest goes *opposite* to a breakdown that is equivalent to infinitely high off-net charges. We have also seen that, under some circumstances, both a regulator and private firms would choose a "bill-and-keep" system, in which case efficient outcomes are reached with minimal intervention (only a reciprocity clause is needed). This is the second encouraging result. When negotiations are not efficient, it is still possible for a regulator to improve efficiency with an appropriate reception charge, although a lot of information on costs and demand is required on the regulator's side to achieve this outcome: this is our mildly encouraging result. In fact, it is impossible even for a perfectly informed regulator to achieve the first best via the regulation of reception charges: they should be set so high that they would violate sender sovereignty.

5. Market-determined reception charges

In the previous section we discussed the role that can be played by reception charges when set by a benevolent regulator. If they were instead negotiated by operators in stage I, it is clear that operators would have an interest divergent from the regulator's. The intuition that we gave for negotiated access charges would carry forward to negotiated reception charges as well. Operators would try to obtain cheap off-net calls to reduce the intensity of competition. When off-net charges are already the lowest possible (i.e., "bill-and-keep" or $a = 0$), then the off-net price from (11bis) is $\hat{p}^* = [c + tx - \hat{r}]/[1 - \beta(1 - x)]$: operators could start introducing *positive* reception charges as this allows them to push off-net retail prices even lower, via the pecuniary externality. This typically results in reception charges that are "too high" from the social point of view. Conversely, if the freely negotiated access charge is positive (but "too low" from a social point of view), then eq. (16) holds and the off-net prices increases with the reception charge. The regulator and the negotiating operators have conflicting interests: the former would like to introduce a positive reception charge as the overall result is to push up the price, but the latter have exactly the opposite interest and they would set the reception charge to zero, the lowest possible level.

It is doubtful however that operators would be allowed to negotiate over reciprocal reception charges as these are retail prices, contrary to access charges that are wholesale prices. An antitrust authority would almost surely forbid this kind of retail contracts given their collusive nature. In this last section we ask what type of retail reception charges would be chosen individually by competing operators.

In order to answer this question we must amend our model since, as explained in Section 2, the amount of incoming and outgoing traffic is determined by two "binding" prices only. For instance, under "sender sovereignty" only outgoing call charges matter as long as reception charges are low enough, but the exact level of reception charges is indeterminate. Similarly, under "receiver sovereignty", only reception charges matter as long as outgoing charges are low enough. Since we want to determine simultaneously the four call prices (outgoing and reception charges for both operators) we follow JLT by introducing some noise when receiving calls. This captures a random element, for instance the recipient of a call may not be in the ideal situation for answering the call (e.g., because driving or engaged in other activities). Its technical role is that both the sender and the receiver now have a positive probability of cutting a communication.

We suppose that the utility from an exchange of information is as in Section 2 plus an additive term with noise from receiving calls. For consumer i the gross utility is thus $U_i(I^{ij}) + \varepsilon_i Q^{ji}$ where ε_i has a distribution function $G(\cdot)$, with wide enough support $[\underline{\varepsilon}, \overline{\varepsilon}]$, zero mean, and strictly positive density $g(\cdot)$ over the support. ε_i is i.i.d. for each pair of customers. Consider an off-net information exchange between i and j:

- If $\varepsilon_i \geq \hat{\varepsilon}_i = \hat{r}_i - U'_i f_2^i$ and $\varepsilon_j \geq \hat{\varepsilon}_j = \hat{r}_j - U'_j f_2^j$, then there is a "sender sovereignty" regime that generates calls $Q^{ij} = Q^{ij}(\hat{p}_i, \hat{p}_j)$ and $Q^{ji} = Q^{ji}(\hat{p}_j, \hat{p}_i)$;
- If $\varepsilon_i < \hat{\varepsilon}_i$ and $\varepsilon_j \geq \hat{\varepsilon}_j$, then there is a "network i sovereignty" regime that generates calls $Q^{ij} = Q^{ij}(\hat{p}_i, \hat{r}_i - \varepsilon_i)$ and $Q^{ji} = Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{p}_i)$;
- If $\varepsilon_i \geq \hat{\varepsilon}_i$ and $\varepsilon_j < \hat{\varepsilon}_j$, then there is a "network j sovereignty" regime that generates calls $Q^{ij} = Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{p}_j)$ and $Q^{ji} = Q^{ji}(\hat{p}_j, \hat{r}_j - \varepsilon_j)$;
- If $\varepsilon_i < \hat{\varepsilon}_i$ and $\varepsilon_j < \hat{\varepsilon}_j$, then there is a "receiver sovereignty" regime that generates calls $Q^{ij} = Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i)$ and $Q^{ji} = Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{r}_j - \varepsilon_j)$.

The volume of calls between i and j can be written as:

$$\begin{aligned}
D^{ij} &= D^{ij}\left(\hat{p}_i, \hat{p}_j, \hat{r}_i, \hat{r}_j\right) = Q^{ij}(\hat{p}_i, \hat{p}_j)(1 - G(\hat{\varepsilon}_i))(1 - G(\hat{\varepsilon}_j)) + \\
&(1 - G(\hat{\varepsilon}_j))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} Q^{ij}(\hat{p}_i, \hat{r}_i - \varepsilon_i) g(\varepsilon_i) d\varepsilon_i + (1 - G(\hat{\varepsilon}_i))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{p}_j) g(\varepsilon_j) d\varepsilon_j \\
&+ \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i) g(\varepsilon_i) g(\varepsilon_j) d\varepsilon_i d\varepsilon_j.
\end{aligned}$$

The first term of the RHS is the demand under sender sovereignty, the second term is the demand when network i is sovereign and customer j accepts incoming calls, the third term is the demand when network j is sovereign and customer i accepts the calls, finally the fourth term is the demand under receiver sovereignty. The indirect utility for a customer belonging to network i from off-net exchanges with network j is $\alpha_j\left(\hat{v}^{ij} - \hat{r}_i D^{ji}\right)$ where:

$$\begin{aligned}
\hat{v}^{ij} &= \hat{v}\left(\hat{p}_i, \hat{p}_j, \hat{r}_i, \hat{r}_j\right) = U_i(f^i(Q^{ij}(\hat{p}_i, \hat{p}_j), Q^{ji}(\hat{p}_j, \hat{p}_i)))(1 - G(\hat{\varepsilon}_i))(1 - G(\hat{\varepsilon}_j)) \\
&+ (1 - G(\hat{\varepsilon}_j))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} U_i(f^i(Q^{ij}(\hat{p}_i, \hat{r}_i - \varepsilon_i), Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{p}_i))) g(\varepsilon_i) d\varepsilon_i \\
&+ (1 - G(\hat{\varepsilon}_i))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} U_i(f^i(Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{p}_j), Q^{ji}(\hat{p}_j, \hat{r}_j - \varepsilon_j))) g(\varepsilon_j) d\varepsilon_j \\
&+ \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} U_i(f^i(Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i), Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{r}_j - \varepsilon_j))) g(\varepsilon_i) g(\varepsilon_j) d\varepsilon_i d\varepsilon_j - \hat{p}_i D^{ij}.
\end{aligned}$$

which is used to determine market shares as in eq. (2). The profit of network i is:

(20) $$\pi_i = \alpha_i\left[\alpha_i\left(p_i + r_i - c - t\right)D^{ii} + \alpha_j\left(\hat{p}_i - c - a\right)D^{ij} + F_i - f + \alpha_j(\hat{r}_i + a - t)D^{ji}\right].$$

We now illustrate the equilibrium in stage II, when operators compete in five-part prices. As before, on-net prices are efficient as everything is internalized and their expressions are not reported. The following proposition characterizes off-net prices as noise converges to zero:

Proposition 5. *As the noise vanishes,*

- *the only symmetric candidate equilibrium without connectivity breakdown when senders are sovereign most of the time is given by:*

(21) $$\begin{cases} \hat{p}^* = \dfrac{c + a + (a - t)\left[1 - x(1 + x)\right]}{1 - \beta + \beta x(1 + x)} \\ \hat{r}^* = t - a + x^2 \dfrac{a - t + \beta(c + a)}{1 - \beta + \beta x(1 + x)} \end{cases} \quad \text{iff} \quad \frac{t - \beta c}{1 + \beta} < a \le t + x^2 \frac{\beta(c + t)}{(1 - x)(1 + x - \beta)} \quad \text{and}$$

$\beta < \dfrac{1}{1 - x - x^2}$;

(22) $$\begin{cases} \hat{p}^* = \dfrac{c + a - (a - t)x}{1 - \beta(1 - x)} \\ \hat{r}^* = 0 \end{cases} \quad \text{iff } a > t + \frac{x^2 \beta(c + t)}{(1 - x)(1 + x - \beta)} \text{ and } \beta < \frac{1}{1 - x}.$$

- *The only symmetric candidate equilibrium without connectivity breakdown when receivers are sovereign most of the time is given by:*

(23) $$\begin{cases} \hat{p}^* = a + c + x^2 \dfrac{a - t + \beta(c + a)}{\beta - 1 + x(1 + x)} \\ \hat{r}^* = \dfrac{\beta\left[-2a - c + t + (a + c)x(1 + x)\right]}{\beta - 1 + x(1 + x)} \end{cases} \quad \text{iff} \quad -c + \frac{x^2(c + t)}{\beta - 1 + \beta x} \le a < \frac{t - \beta c}{1 + \beta} \quad \text{and}$$

$\beta > 1 - x - x^2$;

(24) $$\begin{cases} \hat{p}^* = 0 \\ \hat{r}^* = \dfrac{\beta\left[t - a(1 - x) + cx\right]}{\beta - 1 + x} \end{cases} \text{iff } a \le -c + \frac{x^2(c + t)}{\beta - 1 + \beta x} \text{and } \beta > 1 - x.$$

- *There is a symmetric candidate equilibrium without connectivity breakdown and efficient prices only in the presence of a positive propagation factor:*

$$(25)\quad \begin{cases} \hat{p}^* = p^* = \dfrac{t+c}{1+\beta} \\ \hat{r}^* = \beta\hat{p}^* = \dfrac{\beta(t+c)}{1+\beta} \end{cases} \quad \textit{iff } a = \frac{t-\beta c}{1+\beta} \textit{ and } 1-x < \beta < 1/(1-x).$$

We discuss eq. (21) and (22) as they relate to the case of sender sovereignty most of the time, which is the case people are familiar with in practical terms, as the number and length of calls is predominantly set by the caller and not the receiver for price reasons. The range of validity of the two equations is also plotted in Figure 1, where we have put the marginal utility from receiving calls on the horizontal axis and the level of the access charge on the vertical axis. Connectivity breakdown occurs in the bottom left corner (the downward-sloping curve describes the limiting condition for having sender sovereignty most of the time: $a > (t - \beta c)/(1+\beta)$; in this region reception charges are set so high that a connectivity breakdown arises) and to the right of the diagram (the vertical line is $1/(1 - x - x^2)$; in this region sending charges are set very high). In the other regions there is a candidate for a symmetric equilibrium with senders determining volumes most of the time.

Eq. (21) says that when termination charges are set above cost ($a > t$ = LRIC), then it is possible that competing operators should set *negative* reception fees. By paying consumers to receive calls, operators directly promote increased termination revenue. Indeed we have found several practical examples of such behavior in mobile telephony. However, negative reception charges may raise several problems. Firstly, telecom operators might lack the ability to set such charges because of the technical complexity associated to keeping track of receiving calls – this difficulty was present in the past. Secondly, negative prices may give rise to opportunistic behavior. For instance, some time ago in Italy mobile operators did set negative reception fees for a while only to find that people were calling their mobile phones from office lines so as to obtain these rebates. The scheme had to be withdrawn.[18] Thus in Proposition 5 and Figure 1 we also impose a non-negativity constraint on reception charges. This generates the upward-sloping curve in Figure 1. In the region above this curve, reception charges are zero and off-net call charges are given by eq. (22).

[18] http://www.cellularitalia.com/contrattigsm.html

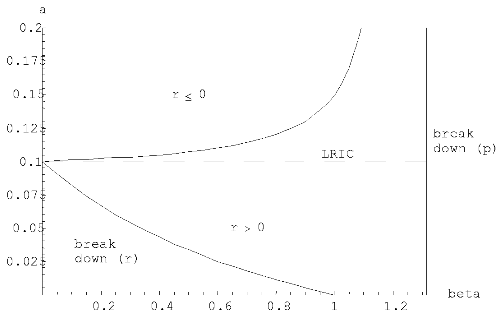


Figure 1 – Market-determined reception charges under "sender sovereignty"
[Parameter values: $x = 0.2$, $t = 0.1$, $c = 0.1$]

If, on the other hand, negative reception charges do not represent a problem and could be implemented, then eq. (21) is also valid in the region above the upward sloping curve. From eq. (21), it can also be found that the sum of outgoing and incoming off-net charges is:

$$\hat{p}^* + \hat{r}^* = c + t + \frac{(1-x)\left[a(1+\beta) - t + \beta c\right]}{1 - \beta + \beta x(1+x)} > c + t\,.$$

Thus, in the domain of validity of eq. (21) ($a > (t - \beta c)/(1+\beta)$) there is always an inefficiency result: too few calls are sent and received as the benefits of the receiver are not accounted for and the access charge cannot be set low enough to revert this tendency. The inefficiency disappears only when $a = (t - \beta c)/(1+\beta)$, in which case eq. (25) is valid. This is the limiting value of the access charge that indicates a shift between receiver and sender sovereignty. Efficiency emerges as firms are moving along the line $\beta\hat{p} = \hat{r}$ and thus internalize everything when setting their off-net call price.

These results shed light on the debate of CPP versus RPP: the ability of charging customers for receiving calls does not imply that reception charges are actually used by operators. This question can be answered only if the level of the access charge is known, otherwise different regimes can occur as illustrated by Figure 1. Our findings can explain the empirical observation that consumers are often *not* charged for receiving calls in countries with high termination charges, even when there is nothing in principle to prevent operators from introducing such charges. This is in line, for instance, with the European experience of calls to mobile phones. On the contrary, if termination charges are set at or below cost then operators should start introducing reception charges if they have the ability to do so. This is in line with the North-American experience of mobile (cellular) telephony. Marcus (2004) discusses how the call termination system in the U.S. has a strong tendency toward symmetry in the rates charged for reciprocal compensation. In particular he shows that:

- ILEC-CLEC and ILEC-mobile reciprocal compensation rates are generally symmetric, and set at a rate that reflects the marginal cost of the ILEC.[19]
- ILEC-ILEC, CLEC-CLEC, CLEC-mobile, and mobile-mobile reciprocal compensation rates are determined through voluntary negotiations, and in many cases are set to zero ("bill-and-keep"), in particular for ILEC-ILEC and mobile-mobile interconnection.

The mobile sector is a particularly good candidate to test our positive findings since it is indeed a case of network-based competition where roughly symmetric mobile firms have to terminate calls on each other's network. The rate for reciprocal compensation is established through unregulated commercial negotiations. These agreements are generally on a "bill-and-keep" basis. Mobile operators also charge their customers for receiving calls (RPP): this is expected from our analysis in the presence of below-cost termination rates.[20]

5.1 A first-best result when the propagation factor is positive

The North-American evidence supports our findings in the last stage of the game, as it links reception charges to below-cost access (termination) charges. It also points to another empirical fact, namely that operators seem to agree on below-cost reciprocal termination charges in an environment where both outgoing and incoming charges are market-determined. This fact is related to our last step in the technical analysis of this paper. We now ask what level of reciprocal termination rates would be selected by unregulated negotiations in stage I.

Proposition 6. *When the propagation factor is positive, the following reciprocal access charge is a candidate to emerge from unregulated commercial negotiations:* $a^* = (t - \beta c)/(1+\beta)$. *Under this charge, if* $1 - x < \beta < 1/(1-x)$, *there is no connectivity*

[19] ILECs are the incumbent fixed-line operators, while CLECs are their fixed-line competitors. This is in contrast with European countries where asymmetries in termination rates exist between fixed-line operators (who are subject to regulation) and mobile operators (who historically have not been subject to regulation).
[20] FCC, *In the Matter of developing a Unified Intercarrier Compensation Regime*, CC Docket 01-92, §95.

breakdown and market-determined retail prices are efficient. If the propagation factor is zero or negative, there is always a connectivity breakdown.

This last result is quite striking. First of all, it points quite clearly to the importance of the propagation factor. In the absence of the propagation factor ($x = 0$) there would never be a possibility of reaching unregulated agreements over access charges for any value of β. Operators would be using extremely high charges either for sending or for receiving calls, according to the value the utility from receiving calls, to induce a break-down. Although they do not conduct such analysis, this would be the outcome in JLT of endogenizing reciprocal access charges. This would also occur if calls are substitutes in the information exchange ($x < 0$). On the contrary, if calls are complements in the information exchange – which we recall is the empirically-supported case – this breakdown does not arise in a suitable range of values of β. In the limit, if x tends to one, the range extends to all possible values of β.

The second aspect emerging from Proposition 6 is that, if negotiations do not induce a breakdown, they are efficient. From a normative point of view, this remarkable result implies that the industry can indeed "self-regulate" under a mild form of intervention such as reciprocity. Intuitively, operators recognize that, by choosing suitable access charges they can affect the intensity of retail competition. In our earlier analysis in Section 4, with regulated or zero reception charges, the access charge could just affect off-net call prices. As we showed in Proposition 3, operators would then try to be cheap off-net, in order not to compete too vigorously for customers. In the presence of market-determined reception charges a very low access charge is more problematic: if it is set too low ($a < a^*$), it would increase the reception charge and induce a break-down. In this case competition for customers would be very intense since all off-net exchanges would be cut down and customers would prefer to belong to the bigger network. Operators then face two options when selecting the reciprocal value for a. Either they ensure that off-net communications do take place, or they induce a break-down.

Proposition 6 finds that it is in the joint interest of firms to avoid a breakdown, not because allowing off-net traffic generates profits per se, but because competition for customers is less intense. In fact, both with and without breakdown off-net profits are zero (in the former case because there is no traffic, in the latter case because prices are efficient, so the sum of off-net prices is just equal to marginal costs). As also on-net profits are zero (on-net prices are efficient with and without a breakdown), profits are only raised via the fixed fee, which is lower in the presence of a breakdown due to more intense competition for customers.

This intuition, while being helpful to understand why the trade-off between allowing and not allowing off-net exchanges is resolved in favor of permitting them, does not precisely tell why efficiency is also reached. This last aspect can be understood starting from an equilibrium with sender sovereignty most of the time, i.e., reception charges do not bind most of the time. Then we have an intuition similar to Proposition 3 in this restricted range: operators are better off by reducing a. In fact, they reduce a precisely to the point $a = a^*$ where it does not violate sender sovereignty. They do not go below it otherwise they would trigger a breakdown. As found by Proposition 5, at this point prices are given by eq. (25), where off-net charges achieve efficiency.[21]

Notice that, in the range described by Proposition 6 when the first-best is attained, the equilibrium multi-part tariffs takes a simple expression: $\hat{p}^* = p^* = (t+c)/(1+\beta)$, $\hat{r}^* = \beta p^*$, $F^* = f + 1/(2\sigma)$ and firms do not engage in network-based price discrimination.[22] However this result has been reached by granting firms the possibility to engage in on-net/off-net price discrimination. Given that the first-best is attained, it does not seem relevant in this context to propose a policy that forbids operators to practice price discrimination.

6. Conclusions

This paper has developed a model of information exchange between calling parties and used it to analyze competition between network operators. We have found that a relevant role is played by the "propagation factor", which measures how return traffic is induced by sending calls. When inter-network (access) charges are taken as given, we have shown that the "connectivity breakdown" problem highlighted by JLT and HK is enhanced when calls made and received are substitutes in the information production function, and reduced when they are complements. When access charges are endogenized and chosen by negotiating operators, we have also shown that the breakdown problem can be totally eliminated.

The empirical literature has found that "calls generate calls", i.e., the propagation factor is positive. As mentioned by Taylor (2004), this earlier literature has used aggregate data

[21] Notice the value that the optimal $a^* = (t - \beta c)/(1+\beta)$ might take in practice. If the cost of termination and origination are similar, and if the value from receiving calls is sufficiently high, then a^* can be approximated by a "bill-and-keep", even before taking into account transaction and metering costs.

[22] HK find a related result that, when firms charge the same prices on-net and off-net, they also set the fixed subscription fee at zero. Recall that, in their model, $f = 0$ and networks are identical (the degree substitutability is infinity).

and might have captured both the propagation factor and network externalities. It is an important question for further empirical research to study the information exchange process and correctly identify the propagation factor using micro-data. At a theoretical level, it would be very interesting to consider a more strategic dynamic game between callers, to see under what conditions this could generate the reduced-form representation summarized by the information exchange framework.

Until better analysis is available, however, the evidence available to us so far points to a positive propagation factor, thus our results suggest that connectivity breakdowns should not be seen as a major source of concern under a mild form of regulation (reciprocity). Indeed, we have shown how, under some circumstances, the industry can self-regulate and achieve first-best allocations via negotiated access charges that internalize externalities.

Our results also imply that the debate over the merits of CPP versus RPP must be assessed against the level set for the access charge. We have shown that, in the presence of below-cost access charges, operators may introduce positive reception charges. When the benefit from receiving calls is sufficiently high, below-cost access charges (in the limit a "bill-and-keep" system) have good properties and may be selected by operators themselves. Given that optimal regulation is difficult and costly to achieve, our simplest policy message is to allow operators to negotiate over reciprocal access charges, while granting them full pricing flexibility.

Our findings can be applied when operators are symmetric and calling patterns balanced. If traffic were significantly imbalanced, voluntary negotiations of symmetric rates would be more difficult to achieve. Off-net price discrimination could also be used for anti-competitive purposes. We believe that further work is needed in this area to see if "light" regulation (e.g., reciprocity deals or non-discrimination requirements) could work also in asymmetric settings.

References

Armstrong, M. (1998), "Network Interconnection in Telecommunications," *Economic Journal*, 108: 545-564.

Armstrong, M. (2002), "The Theory of Access Pricing and Interconnection," in M. Cave, S. Majumdar and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland, Amsterdam.

Berger, U. (2004), "Access Charges in the Presence of Call Externalities", *Contributions to Economic Analysis & Policy*, 3(1): Article 21.

Carter, M. and J. Wright (2003), "Asymmetric Network Interconnection," *Review of Industrial Organization*, 22(1): 27-46.

Crandall, R. and J.G. Sidak (2004), "Should Regulators Set Rates to Terminate Calls on Mobile Networks?", *Yale Journal on Regulation*, 21: 1-46.

DeGraba, P. (2002), "Central Office Bill and Keep as a Unified Inter-carrier Compensation Regime," *Yale Journal on Regulation*, 19(1): 37-84.

DeGraba, P. (2003), "Efficient Interconnection Rates for Interconnected Competing Networks," *Journal of Economics and Management Strategy*, 12: 207-230.

Feller, W. (1971), *Probability Theory and its Applications*, vol. II, John Wiley, New York.

Gans, J. and S. King (2001), "Using 'Bill-and-Keep' Interconnection Arrangements to Soften Network Competition," *Economics Letters*, 71(3): 413-420.

Hermalin, B. and M. Katz (2004), "Sender or Receiver: Who Should Pay to Exchange an Electronic Message?" *RAND Journal of Economics,* 35(3): 423-448.

Hermalin, B. and M. Katz (2005), "Customer or Complementor: Intercarrier Compemsation with Two-sided Benefits," *mimeo*, UC Berkeley.

Kim, J.Y. and Y. Lim (2001), "An Economic Analysis of the Receiver Pays Principle," *Information Economic and Policy*, 13: 231-260.

Jeon, D.-S., J-J. Laffont, and J. Tirole (2004), "On the Receiver Pays Principle," *RAND Journal of Economics,* 35(1): 85-110.

Laffont, J.-J., P. Rey and J. Tirole (1998), "Network Competition: I. Overview and Nondiscriminatory Pricing; II. Discriminatory Pricing," *RAND Journal of Economics*, 29: 1-56.

Larson, A., Lehman D. and D. Weisman (1990), "A General Theory of Point-to-Point Long Distance Demand," in A. de Fontenay, M.H. Shugard and D.S. Sibley (eds.), *Telecommunications Demand Modelling*, North-Holland, Amsterdam.

Littlechild S.C. (2005), "Mobile Termination Charges: Calling Party Pays vs. Receiving Party Pays," CWPE n. 426, Cambridge University.

Marcus, S.J. (2004), "Call Termination Fees: the U.S. in Global Perspective," *mimeo*.

Rochet, J.-C. and J. Tirole (2003), "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1(4): 990-1029.

Rochet, J.-C. and J. Tirole (2004), "Two-Sided Markets: an Overview", *mimeo*, University of Toulouse.

Taylor, L.D. (2004), "Customer Demand Analysis," in M. Cave, S. Majumdar and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North-Holland, Amsterdam.

Valletti, T.M. and C. Cambini (2005), "Investments and Network Competition," *RAND Journal of Economics*, 36(2), 446-67.

Vogelsang, I. (2003), "Price Regulation of Access to Telecommunications Networks," *Journal of Economic Literature* 41(3): 830-62.

Appendix

Derivation of eq. (7).

Following JLT we study the optimization program of network i assuming that α_i is given. Let $\widetilde{F}_i = F_i + r_i\left[\alpha_i Q^{ii}(p_i,p_i) + \alpha_j Q^{ji}(p_j,p_i)\right]$ be a generalized fixed fee incorporating also the cost for receiving calls. Given eq. (2) we can write $w_i - w_j = (\alpha_i - 1/2)/\sigma$ and, using eq. (3), we obtain:

$$\widetilde{F}_i = \widetilde{F}_j + (1/2-\alpha_i)/\sigma + \alpha_i v(p_i,p_i) + \alpha_j v(p_i,p_j) - \alpha_j v(p_j,p_j) - \alpha_i v(p_j,p_i).$$

Network i's profit in eq. (6) can be rearranged in the following way:

$$\begin{aligned}\pi_i = \alpha_i &\left\{(p_i-c-t)\left[\alpha_i Q^{ii}(p_i,p_i) + \alpha_j Q^{ij}(p_i,p_j)\right] + \widetilde{F}_i - f\right\} + \\ &+ \alpha_i\alpha_j(a-t)\left[Q^{ji}(p_j,p_i) - Q^{ij}(p_i,p_j)\right]\end{aligned}$$

The FOC with respect to p_i is given by:

$$\text{(A1)}\quad \begin{aligned}\frac{\partial \pi_i}{\partial p_i} = \alpha_i &\left[\alpha_i Q^{ii}(p_i,p_i) + \alpha_j Q^{ij}(p_i,p_j) + (p_i-c-t)\left(\alpha_i \frac{\partial Q^{ii}}{\partial p_i} + \alpha_j \frac{\partial Q^{ij}}{\partial p_i}\right)\right] + \\ &+ \alpha_i \frac{\partial \widetilde{F}_i}{\partial p_i} + \alpha_i\alpha_j(a-t)\left(\frac{\partial Q^{ji}}{\partial p_i} - \frac{\partial Q^{ij}}{\partial p_i}\right)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \widetilde{F}_i}{\partial p_i} &= \frac{\partial \widetilde{F}_j}{\partial p_i} + \alpha_i \frac{\partial v(p_i,p_i)}{\partial p_i} + \alpha_j \frac{\partial v(p_i,p_j)}{\partial p_i} - \alpha_i \frac{\partial v(p_j,p_i)}{\partial p_i} \\ \frac{\partial \widetilde{F}_j}{\partial p_i} &= r_j \alpha_i \frac{\partial Q^{ij}}{\partial p_i}.\end{aligned}$$

For calls originated by network i and terminated on network j, the indirect utility is given by eq. (5). Thus it results:

$$\frac{\partial v(p_i,p_j)}{\partial p_i} = -Q^{ij}(p_i,p_i) + U_i' f_2^i \frac{\partial Q^{ji}}{\partial p_i}.$$

The effect of a price change on the indirect utility of customers connected to network i is given by the sum of a "standard" loss in utility on each originated call ($-Q^{ij}$) and a term that reflects the change in utility due to way the recipient j reacts to the information exchange, thus leading j to change the number of calls sent to customer i ($U_i' f_2 \partial Q^{ji}/\partial p_i$).

For calls originated by network j and terminated on network i, where $v(p_j,p_i) = U_j(f^j(Q^{ji}(p_j,p_i), Q^{ij}(p_i,p_j)) - p_j Q^{ji}(p_j,p_i)$, it results:

$$\frac{\partial v(p_j, p_i)}{\partial p_i} = U'_j f_2^j \frac{\partial Q^{ij}}{\partial p_i}.$$

In this case, the marginal change in indirect utility depends only on the impact that network i's price has on the calls received by users j.

For calls originated and terminated on the same network, where the indirect utility is given by eq. (4), it results:

$$\frac{\partial v(p_i, p_i)}{\partial p_i} = -Q^{ii}(p_i, p_i) + U'_i f_2^i \frac{dQ^{ii}}{dp_i}.$$

Note that in the last term the effect of p_i influences both call origination and call reception, $dQ^{ii}/dp_i = Q_1^{ii} + Q_2^{ii}$.

After substituting the previous expressions in eq. (A1), eq. (7) results. QED

Proof of Proposition 1.

In a symmetric equilibrium $p_i^* = p_j^* = p^*$, $r_i = r_j = r$, $U_i = U_j = U$, α = ½. We also have $Q_1^{ij} = Q_1^{ii}$ and $Q_2^{ij} = Q_2^{ii}$. Thus it results $\Phi|_{sym} = \frac{2}{2+x}, \Gamma|_{sym} = \frac{2x}{2+x}, \Theta|_{sym} = \frac{2x+1}{2+x}$, where x is a *propagation* factor defined as $x = \Gamma/\Phi = Q_2^{ji}/Q_1^{ij}$. Substitution into eq. (7) gives eq. (8).

The level of the fixed fee is determined by studying the derivative w.r.t. α_i:

$$\pi_i = \alpha_i \{(p_i - c - t)[\alpha_i Q^{ii} + \alpha_j Q^{ij}] + \widetilde{F}_j - f + (1/2 - \alpha_i)/\sigma\} +$$
$$\alpha_i \{\alpha_i v(p_i, p_i) + \alpha_j v(p_i, p_j) - \alpha_j v(p_j, p_j) - \alpha_i v(p_j, p_i)\} + \alpha_i \alpha_j (a - t)[Q^{ji} - Q^{ij}]$$

Since in $p_i = p_i^*$ it results $\frac{d\pi_i}{d\alpha_i} = \frac{\partial \pi_i}{\partial \alpha_i} + \frac{\partial \pi_i}{\partial p_i^*}\frac{\partial p_i^*}{\partial \alpha_i} = \frac{\partial \pi_i}{\partial \alpha_i}$, we have:

$$\frac{\partial \pi_i}{\partial \alpha_i} = \{(p_i - c - t)[\alpha_i Q^{ii} + \alpha_j Q^{ij}] + \widetilde{F}_j - f + (1/2 - \alpha_i)/\sigma\} +$$
$$+ \{\alpha_i v(p_i, p_i) + \alpha_j v(p_i, p_j) - \alpha_j v(p_j, p_j) - \alpha_i v(p_j, p_i)\} + (1 - 2\alpha_i)(a - t)[Q^{ji} - Q^{ij}] +$$
$$+ \alpha_i \{(p_i - c - t)[Q^{ii} - Q^{ij}] - r_j (Q^{jj} - Q^{ij}) - 1/\sigma + v(p_i, p_i) - v(p_i, p_j) + v(p_j, p_j) - v(p_j, p_i)\} = 0.$$

In a symmetric equilibrium the above expression reduces to:

$$\left.\frac{\partial \pi_i}{\partial \alpha_i}\right|_{sym} = (p^* - c - t)Q(p^*)/2 + \widetilde{F}^* - f - 1/2\sigma = 0$$

where $\widetilde{F}^* = F^* + rQ(p^*)$ and $Q(p^*) = Q^{ii}(p^*, p^*) + Q^{ij}(p^*, p^*)$, giving part b). Finally, substituting F* and p* in eq. (6) gives part c): $\pi^* = 1/4\sigma$.[23] QED

Proof of Proposition 2.
We follow the same procedure as in Section 3 with α_i given. Result a) is standard, as for on-net prices a network maximizes joint surplus with all customers, hence reaching the socially optimal value. To get result b) consider only the profit for off-net calls (both outgoing and incoming). After manipulations of eq. (10) we get:

$$\text{(A2)} \quad \begin{aligned} \hat{\pi}_i(\hat{p}_i;\alpha_i,\hat{r}_j) = \alpha_i \{\alpha_j [v(\hat{p}_i,\hat{p}_j) + (\hat{p}_i - c - a)Q^{ij}(\hat{p}_i,\hat{p}_j)] + \alpha_i \hat{r}_j Q^{ij}(\hat{p}_i,\hat{p}_j)\} + \\ + \alpha_i \{\alpha_j (a-t)Q^{ji}(\hat{p}_j,\hat{p}_i) - \alpha_i v(\hat{p}_j,\hat{p}_i)\} \end{aligned}$$

The FOC is:

$$\begin{aligned} \frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} &= \alpha_i \{\alpha_j [-Q^{ij}(\hat{p}_i,\hat{p}_j) + U'_i f_2^i Q_2^{ji} + Q^{ij}(\hat{p}_i,\hat{p}_j) + (\hat{p}_i - c - a)Q_1^{ij}] + \alpha_i \hat{r}_j Q_1^{ij}\} + \\ &+ \alpha_i [\alpha_j (a-t)Q_2^{ji} - \alpha_i U'_j f_2^j Q_1^{ij}] \end{aligned}$$

that can be rewritten as:

$$\frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} = \alpha_i Q_1^{ij} \{\alpha_j [\hat{p}_i - (c + a - xU'_i f_2^i - x(a-t))] + \alpha_i (\hat{r}_j - U'_j f_2^j)\}$$

where $x = Q_2^{ji} / Q_1^{ij}$ is the propagation effect. Solving $\frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} = 0$ we get:

$$\hat{p}_i^* = c + a - (a-t)x - U'_i f_2^i x - \frac{\alpha_i}{\alpha_j}(\hat{r}_j - U'_j f_2^j)$$

that, if a symmetric equilibrium exists, reduces to eq. (11). In order to show under what conditions this price is indeed a maximum, we have to determine the SOC. Denote with $\beta^i = f_2^i / f_1^i$ such that $U'_i f_2^i = \beta^i \hat{p}_i$. In the same way, define $\beta^j = f_2^j / f_1^j$. Since $\partial(U'_i f_2^i)/\partial \hat{p}_i = \beta^i + \hat{p}_i(\partial \beta^i / \partial \hat{p}_i)$ and $\partial(U'_j f_2^j)/\partial \hat{p}_i = \hat{p}_j(\partial \beta^j / \partial \hat{p}_i)$, it results:

$$\begin{aligned} \frac{\partial^2 \hat{\pi}_i}{\partial \hat{p}_i^2} &= \alpha_i Q_{11}^{ij} \{\alpha_j [\hat{p}_i - (c + a - x\beta^i \hat{p}_i - x(a-t))] + \alpha_i (\hat{r}_j - \beta^j \hat{p}_j)\} + \\ &+ \alpha_i Q_1^{ij} \left\{ \alpha_j \left[1 + x\left(\beta^i + \hat{p}_i \frac{\partial \beta^i}{\partial \hat{p}_i} + x'(\beta^i \hat{p}_i + a - t) \right) \right] - \alpha_i \hat{p}_j \frac{\partial \beta^j}{\partial \hat{p}_i} \right\} \end{aligned}$$

that reduces in a symmetric equilibrium to:

---

[23] As in JLT, one can show that a symmetric equilibrium exists if σ is small enough or $|a - t + r|$ is small enough.

$$\left.\frac{\partial^2 \hat{\pi}_i}{\partial \hat{p}_i^2}\right|_{sym} = \frac{1}{4} Q_1^{ij} \left\{ 1 + x\left(\beta + \hat{p}\frac{\partial \beta^i}{\partial \hat{p}_i} + x'(\beta\hat{p} + a - t) \right) - \hat{p}\frac{\partial \beta^j}{\partial \hat{p}_i} \right\}.$$

Consider first the case of JLT with independent outgoing and incoming call, i.e., $x = 0$. The SOC is $\left.\frac{\partial^2 \hat{\pi}_i}{\partial \hat{p}_i^2}\right|_{\substack{sym \\ x=0}} = \frac{1}{4} Q_1^{ij} < 0$. It is easy to find sufficient conditions that ensure a negative SOC more in general. Recalling the definition of the propagation factor, we can write:

$$\frac{\partial \beta^i}{\partial \hat{p}_i} = \frac{\left(\partial f_2^i / \partial \hat{p}_i\right) f_1^i - \left(\partial f_1^i / \partial \hat{p}_i\right) f_2^i}{\left(f_1^i\right)^2} = Q_1^{ij} \frac{f_{12}^i - \beta^i f_{11}^i + x(f_{22}^i - \beta^i f_{12}^i)}{f_1^i}$$

$$\frac{\partial \beta^j}{\partial \hat{p}_i} = Q_1^{ij} \frac{f_{22}^j - \beta^j f_{12}^j + x(f_{12}^j - \beta^j f_{11}^j)}{f_1^j}.$$

In a symmetric equilibrium the SOC becomes:

$$\left.\frac{\partial^2 \hat{\pi}_i}{\partial \hat{p}_i^2}\right|_{sym} = Q_1^{ij} \left\{ 1 + x\left[\beta + x'(\beta\hat{p} + a - t)\right] - \hat{p} Q_1^{ij} (f_{22} - \beta f_{12})(1 - x^2) / f_1 \right\} / 4.$$

Suppose that calls are perfect complements, i.e., $x = 1$. A sufficient condition to have $\left.\frac{\partial^2 \hat{\pi}_i}{\partial \hat{p}_i^2}\right|_{\substack{sym \\ x=1}} < 0$ is that $x' \approx 0$, i.e., the propagation factor is not affected by a change in prices.

Since we have assumed sender sovereignty, it must be that the reception charge is low enough, i.e., $U'f_2 \geq \hat{r}$ in a symmetric equilibrium. Finally, to determine the fixed fee of the multi-part tariff, we maximize network i's profit from eq. (10) w.r.t. to α_i, after substituting:

$$F_i = F_j + (1/2 - \alpha_i)/\sigma + \alpha_i v(p_i, p_i) + \alpha_j v(\hat{p}_i, \hat{p}_j) - r_i \alpha_i Q^{ii}(p_i, p_i) - \hat{r}_i \alpha_j Q^{ji}(\hat{p}_j, \hat{p}_i)$$
$$- \alpha_j v(p_j, p_j) - \alpha_i v(\hat{p}_j, \hat{p}_i) + r_j \alpha_j Q^{jj}(p_j, p_j) + \hat{r}_j \alpha_i Q^{ij}(\hat{p}_i, \hat{p}_j),$$

$$\begin{aligned} \frac{\partial \pi_i}{\partial \alpha_i} &= \{\alpha_i (p_i - c - t) Q^{ii} + \alpha_j (\hat{p}_i - c - t) Q^{ij} + F_j + (1/2 - \alpha_i)/\sigma + \alpha_i v(p_i, p_i) \\ &+ \alpha_j v(\hat{p}_i, \hat{p}_j) - \alpha_j v(p_j, p_j) - \alpha_i v(\hat{p}_j, \hat{p}_i) + r_j \alpha_j Q^{jj} + \hat{r}_j \alpha_i Q^{ij} - f\} \\ &+ \alpha_i \{(p_i - c - t) Q^{ii} - (\hat{p}_i - c - t) Q^{ij} - 1/\sigma + v(p_i, p_i) - v(\hat{p}_i, \hat{p}_j) + \\ &+ v(p_j, p_j) - v(\hat{p}_j, \hat{p}_i) - r_j Q^{jj} + \hat{r}_j Q^{ij}\} + (1 - 2\alpha_i)(a - t)\left(Q^{ji} - Q^{ij}\right) = 0. \end{aligned} \tag{A3}$$

In a symmetric equilibrium (A3) is: $\left.\frac{\partial \pi_i}{\partial \alpha_i}\right|_{sym} = \left(p^* - c - t\right) Q(p^*, p^*) + F^* - f + \hat{r} Q(\hat{p}^*, \hat{p}^*) - 1/2\sigma + \left(v(p^*, p^*) - v(\hat{p}^*, \hat{p}^*)\right) = 0$, which gives eq. (12). QED

Proof of Proposition 3.
The off-net price is given by eq. (11bis), and under the assumption $x' \approx 0$, we have:

(A6) $$\frac{\partial \hat{p}^*}{\partial a} = \frac{(1-x)[1-\beta(1-x)] + (1-x)\frac{\partial \beta}{\partial a}[c+a-(a-t)x-\hat{r}]}{(1-\beta(1-x))^2} = \frac{1-x}{1-\beta(1-x)}\left(1+\hat{p}^*\frac{\partial \beta}{\partial a}\right),$$

where $\frac{\partial \beta}{\partial a} = \frac{\partial \beta}{\partial \hat{p}}\frac{\partial \hat{p}^*}{\partial a}$. Notice how $\partial p^* / \partial a$ is positive under rather general conditions (it suffices that $\partial \beta / \partial a$ is positive or not "too" negative). From eq. (12) the effect on F^* is $\frac{\partial F^*}{\partial a} = \frac{\partial v(\hat{p}^*, \hat{p}^*)}{\partial a} - \hat{r}\frac{\partial Q(\hat{p}^*, \hat{p}^*)}{\partial a}$, where $\frac{\partial v(\hat{p}^*, \hat{p}^*)}{\partial a} = U'f_1\frac{\partial Q}{\partial a} + U'f_2\frac{\partial Q}{\partial a} - Q\frac{\partial \hat{p}^*}{\partial a} - \hat{p}^*\frac{\partial Q}{\partial a} = \left[-Q + \beta\hat{p}^* Q_1(1+x)\right]\frac{\partial \hat{p}^*}{\partial a}$. Therefore, it results:

(A7) $$\frac{\partial F^*}{\partial a} = \left[-Q(\hat{p}^*, \hat{p}^*) + (\beta\hat{p}^* - \hat{r})Q_1(\hat{p}^*, \hat{p}^*)(1+x)\right]\frac{\partial \hat{p}^*}{\partial a}.$$

Substituting eq. (14), (A6) and (A7) in eq. (13) we obtain:

(A8) $$\frac{\partial \pi^*}{\partial a} = \frac{1}{4}\frac{\partial \hat{p}^*}{\partial a}\left\{-Q(\hat{p}^*, \hat{p}^*) + Q_1(\hat{p}^*, \hat{p}^*)(1+x)\left[\hat{p}^*(1+2\beta) - c - t - \hat{r}\right]\right\} \leq 0$$

which gives eq. (15) at an interior solution when $x \neq 1$. In order to determine if the optimal access charge is a local maximum, we consider the sign of the SOC. Denoting by $\Psi = -Q + Q_1(1+x)\left[\hat{p}^*(1+2\beta) - c - t - \hat{r}\right]$, the SOC evaluated in a = a* reduces to $\left.\frac{\partial^2 \pi^*}{\partial a^2}\right|_{a=a*} = \frac{1}{4}\frac{\partial \hat{p}^*}{\partial a}\frac{\partial \Psi}{\partial a}$. Since $\partial p^* / \partial a$ is positive, we need to study the sign of $\partial \Psi / \partial a$:

$$\frac{\partial \Psi}{\partial a} = (1+x)\left\{\frac{\partial \hat{p}^*}{\partial a}(Q_{11} + Q_{12})\left[\hat{p}^*(1+2\beta) - c - t - \hat{r}\right] + Q_1\left[2\beta\frac{\partial \hat{p}^*}{\partial a} + \hat{p}^*(1+2\frac{\partial \beta}{\partial a})\right]\right\},$$

$$\left.\frac{\partial \Psi}{\partial a}\right|_{a=a^*} = (1+x)\left\{\frac{\partial \hat{p}^*}{\partial a}(Q_{11} + Q_{12})\frac{Q}{Q_1} + Q_1\left[2\beta\frac{\partial \hat{p}^*}{\partial a} + \hat{p}^*(1+2\frac{\partial \beta}{\partial a})\right]\right\}.$$

The sign of the last expression is not obvious a priori. If demand functions are linear, and $\partial \beta / \partial a$ is positive or not "too" negative, then it results $\partial \Psi / \partial a|_{a=a^*} < 0$ and the access charge in (15) is a local maximum. However, this is not a general property. QED

Proof of Proposition 4.
Parts a) and b) are already discussed in the main text. To complete part b) imagine a "bill-and-keep" system is in place, a = 0. From (17) the regulator can induce efficiency by choosing a reception charge equal to:

$$\hat{r}^W = \frac{\beta(c+t)(2-x)}{1+\beta} - t(1-x),$$

which satisfies the constraint $\hat{r} \le \frac{\beta}{1+\beta}(c+t)$ if $(\beta c - t)(1-x)/(1+\beta) \le 0$. In other words the first-best can be achieved if either the propagation factor is 1, or if $x < 1$ and $\beta c < t$. In particular, if β is sufficiently low, reception charges may not be needed at all ($\beta \le \frac{t(1-x)}{t+c(2-x)} \Rightarrow \hat{r} = 0$). Finally, part c) points to a potential problem that the regulator faces when trying to introduce a regulated termination charge in order to drive a^* towards efficiency. Such charge, if it exists, is given by eq. (19) when both a^* and a^W are positive. This charge has also to satisfy "sender sovereignty", which is written as $\hat{r} \le \frac{\beta}{1+\beta}(c+t)$ when the efficient retail price is induced. It is straightforward to see that this inequality is always violated by (19), unless the propagation factor is infinitely high. Thus the regulator can at most set the highest possible reception charge that does not violate the constraint. QED

Proof of Proposition 5.
We follow again the same procedure used earlier keeping market shares constant. From (A2), we denote with

$$\text{(A9)} \quad \psi(.) = \alpha_j \left(v^{ij}(.) + (\hat{p}_i - c - a)Q^{ij}(.)\right) + \alpha_i \hat{r}_j Q^{ij}(.) + \alpha_j (a-t)Q^{ji}(.) - \alpha_i v^{ji}(.)$$

the off-net per customer profit, where (.) depends on the different pricing regimes (sender, receiver, or network sovereignty). We consider the maximization of the profit given by eq. (20) as the noise vanishes while keeping a large support to avoid price indeterminacy.[24]

Network i's off net profit can be written as follows:

$$\text{(A10)} \quad \begin{aligned} \hat{\pi}_i = \alpha_i \Big\{ & (1-G(\hat{\varepsilon}_i))(1-G(\hat{\varepsilon}_j))\psi(\hat{p}_i,\hat{p}_j) + (1-G(\hat{\varepsilon}_j))\int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i Q^{ji}(\hat{p}_i,\hat{p}_j) g(\varepsilon_i) d\varepsilon_i - \\ & (1-G(\hat{\varepsilon}_i))\int_{\hat{\varepsilon}_j}^{\bar{\varepsilon}} \varepsilon_j Q^{ij}(\hat{p}_i,\hat{p}_j) g(\varepsilon_j) d\varepsilon_j + (1-G(\hat{\varepsilon}_j))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \psi(\hat{p}_i,\hat{r}_i-\varepsilon_i) g(\varepsilon_i) d\varepsilon_i + \\ & (1-G(\hat{\varepsilon}_j))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \varepsilon_i Q^{ji}(\hat{r}_i-\varepsilon_i,\hat{p}_i) g(\varepsilon_i) d\varepsilon_i - G(\hat{\varepsilon}_i)\int_{\hat{\varepsilon}_j}^{\bar{\varepsilon}} \varepsilon_j Q^{ij}(\hat{p}_i,\hat{r}_i-\varepsilon_i) g(\varepsilon_j) d\varepsilon_j + \\ & (1-G(\hat{\varepsilon}_i))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \psi(\hat{r}_j-\varepsilon_j,\hat{p}_j) g(\varepsilon_j) d\varepsilon_j + G(\hat{\varepsilon}_j)\int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i Q^{ji}(\hat{p}_j,\hat{r}_j-\varepsilon_j) g(\varepsilon_i) d\varepsilon_i - \\ & (1-G(\hat{\varepsilon}_i))\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \varepsilon_j Q^{ij}(\hat{r}_j-\varepsilon_j,\hat{p}_j) g(\varepsilon_j) d\varepsilon_j + \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \psi(\hat{r}_j-\varepsilon_j,\hat{r}_i-\varepsilon_i) g(\varepsilon_i) g(\varepsilon_j) d\varepsilon_i d\varepsilon_j + \\ & \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \varepsilon_i Q^{ji}(\hat{r}_i-\varepsilon_i,\hat{r}_j-\varepsilon_j) g(\varepsilon_i) g(\varepsilon_j) d\varepsilon_i d\varepsilon_j - \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \varepsilon_j Q^{ij}(\hat{r}_j-\varepsilon_j,\hat{r}_i-\varepsilon_i) g(\varepsilon_i) g(\varepsilon_j) d\varepsilon_i d\varepsilon_j \Big\} \end{aligned}$$

Consider first the case of sender sovereignty most of the time. The threshold levels of the noise are given by $\hat{\varepsilon}_i = \hat{r}_i - U_i' f_2^i(\hat{p}_i,\hat{p}_j)$ and $\hat{\varepsilon}_j = \hat{r}_j - U_j' f_2^j(\hat{p}_j,\hat{p}_i)$. Since we deal with the case of sender sovereignty most of the time, we solve the maximization problem for $\hat{\varepsilon}_i, \hat{\varepsilon}_j < 0$. Maximizing (A10) w.r.t. $\hat{p}_i$ and $\hat{r}_i$ and evaluating the corresponding FOCs in a symmetric equilibrium, after tedious calculations and simplifications, we obtain:

---

[24] To get the results we impose a regularity condition. Given a continuous function $F(\cdot)$, a regular sequence of distributions $G_n(\varepsilon)$ of the random variable ε with zero mean satisfies: $\lim_{n\to\infty} G_n(F(\varepsilon) | \varepsilon \ge \varepsilon_0) = F(\varepsilon = \varepsilon_0)$ for $\varepsilon_0 > 0$, $\lim_{n\to\infty} G_n(F(\varepsilon) | \varepsilon \le \varepsilon_0) = F(\varepsilon = 0)$ for $\varepsilon_0 > 0$ (see Feller, 1971).

$$
\begin{aligned}
\frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} = \frac{1}{2}\Bigg\{ & \left(1-G(\hat{\varepsilon}_i)\right)^2 \frac{\partial \psi(\hat{p}_i,\hat{p}_j)}{\partial \hat{p}_i} + \left(1-G(\hat{\varepsilon}_i)\right)\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \frac{\partial \psi(\hat{p}_i,\hat{r}_i-\varepsilon_i)}{\partial \hat{p}_i} g(\varepsilon_i)d\varepsilon_i + \\
& \left(1-G(\hat{\varepsilon}_i)\right)\left[\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \varepsilon_i Q_2^{ji}(\hat{r}_i-\varepsilon_i,\hat{p}_i)g(\varepsilon_i)d\varepsilon_i + \int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i Q_2^{ji}(\hat{p}_i,\hat{p}_j)g(\varepsilon_i)d\varepsilon_i\right] - \\
& \int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i\left[\left(1-G(\hat{\varepsilon}_i)\right)Q_1^{ij}(\hat{p}_i,\hat{p}_j) + G(\hat{\varepsilon}_i)Q_1^{ij}(\hat{p}_i,\hat{r}_i-\varepsilon_i)\right]g(\varepsilon_i)d\varepsilon_i + \\
& \frac{\partial \hat{\varepsilon}_j}{\partial \hat{p}_i} g(\hat{\varepsilon}_i)\left[\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \hat{\varepsilon}_i\left(Q^{ij}(\hat{p}_i,\hat{p}_j) - Q^{ij}(\hat{p}_i,\hat{r}_i-\varepsilon_i)\right)g(\varepsilon_i)d\varepsilon_i + \int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i\left(Q^{ji}(\hat{p}_j,\hat{r}_j-\varepsilon_j) - Q^{ji}(\hat{p}_j,\hat{p}_i)\right)g(\varepsilon_i)d\varepsilon_i\right] + \\
& \frac{\partial \hat{\varepsilon}_i}{\partial \hat{p}_i} g(\hat{\varepsilon}_i)\left[\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \hat{\varepsilon}_i\left(Q^{ji}(\hat{p}_j,\hat{r}_j-\varepsilon_j) - Q^{ji}(\hat{p}_j,\hat{p}_i)\right)g(\varepsilon_i)d\varepsilon_i + \int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i\left(Q^{ij}(\hat{p}_i,\hat{p}_j) - Q^{ij}(\hat{p}_i,\hat{r}_i-\varepsilon_i)\right)g(\varepsilon_i)d\varepsilon_i\right]\Bigg\}.
\end{aligned}
\tag{A11}
$$

and

$$
\begin{aligned}
\frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} = \frac{1}{2}\Bigg\{ & \left(1-G(\hat{\varepsilon}_i)\right)\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \frac{\partial \psi(\hat{p}_i,\hat{r}_i-\varepsilon_i)}{\partial \hat{r}_i} g(\varepsilon_i)d\varepsilon_i + \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \frac{\partial \psi(\hat{r}_j-\varepsilon_j,\hat{r}_i-\varepsilon_i)}{\partial \hat{r}_i} g(\varepsilon_i)g(\varepsilon_j)d\varepsilon_i d\varepsilon_j + \\
& \left(1-G(\hat{\varepsilon}_i)\right)\left[\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \varepsilon_i Q_1^{ji}(\hat{r}_i-\varepsilon_i,\hat{p}_i)g(\varepsilon_i)d\varepsilon_i\right] - G(\hat{\varepsilon}_i)\left[\int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i Q_2^{ij}(\hat{p}_i,\hat{r}_i-\varepsilon_i)g(\varepsilon_i)d\varepsilon_i\right] - \\
& G(\hat{\varepsilon}_i)\frac{\partial \hat{\varepsilon}_i}{\partial \hat{r}_i}\hat{\varepsilon}_i g(\hat{\varepsilon}_i)Q^{ji}(\hat{p}_j,\hat{p}_i) + \frac{\partial \hat{\varepsilon}_i}{\partial \hat{r}_i} g(\hat{\varepsilon}_i)\int_{\hat{\varepsilon}_i}^{\bar{\varepsilon}} \varepsilon_i\left(Q^{ij}(\hat{p}_i,\hat{p}_i) - Q^{ij}(\hat{p}_i,\hat{r}_i-\varepsilon_i)\right)g(\varepsilon_i)d\varepsilon_i + \\
& \frac{\partial \hat{\varepsilon}_i}{\partial \hat{r}_i} g(\hat{\varepsilon}_i)\left[\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i} \hat{\varepsilon}_i Q^{ji}(\hat{p}_j,\hat{r}_j-\varepsilon_i)g(\varepsilon_i)d\varepsilon_i\right] + \\
& \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \varepsilon_i Q_1^{ji}(\hat{r}_i-\varepsilon_i,\hat{r}_j-\varepsilon_j)g(\varepsilon_i)g(\varepsilon_j)d\varepsilon_i d\varepsilon_j - \int_{\underline{\varepsilon}}^{\hat{\varepsilon}_i}\int_{\underline{\varepsilon}}^{\hat{\varepsilon}_j} \varepsilon_j Q_2^{ij}(\hat{r}_j-\varepsilon_j,\hat{r}_i-\varepsilon_i)g(\varepsilon_i)g(\varepsilon_j)d\varepsilon_i d\varepsilon_j \Bigg\}
\end{aligned}
\tag{A12}
$$

As the noise vanishes, from the regularity condition of the distribution function when the threshold value $\hat{\varepsilon}$ is negative, it follows for a continuous function $F(.)$ that:

$$\int_{\underline{\varepsilon}}^{\hat{\varepsilon}} F(\varepsilon)dg(\varepsilon) = F(\hat{\varepsilon})G(\hat{\varepsilon}), \quad \int_{\hat{\varepsilon}}^{\bar{\varepsilon}} F(\varepsilon)dg(\varepsilon) = F(0)(1-G(\hat{\varepsilon})).$$

Using this property, in a symmetric equilibrium, (A11) and (A12) simplify to:

$$
\begin{aligned}
\frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} &= \frac{1}{2}\left(1-G(\hat{\varepsilon})\right)\left\{\left(1-G(\hat{\varepsilon})\right)\frac{\partial \psi(\hat{p}_i,\hat{p}_j)}{\partial \hat{p}_i} + G(\hat{\varepsilon})\left.\frac{\partial \psi(\hat{p}_i,\hat{r}_i-\varepsilon_i)}{\partial \hat{p}_i}\right|_{\varepsilon_i=\hat{\varepsilon}} + G(\hat{\varepsilon})\hat{\varepsilon}Q_2^{ji}\right\} \\
\frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} &= \frac{1}{2}G(\hat{\varepsilon})\Bigg\{\left(1-G(\hat{\varepsilon})\right)\left.\frac{\partial \psi(\hat{p}_i,\hat{r}_i-\varepsilon_i)}{\partial \hat{r}_i}\right|_{\varepsilon_i=\hat{\varepsilon}} + G(\hat{\varepsilon})\left.\frac{\partial \psi(\hat{r}_j-\varepsilon_j,\hat{r}_i-\varepsilon_i)}{\partial \hat{r}_i}\right|_{\varepsilon_i=\varepsilon_j=\hat{\varepsilon}} + \\
&\quad \left(1-G(\hat{\varepsilon})\right)\hat{\varepsilon}Q_1^{ji} + G(\hat{\varepsilon})\hat{\varepsilon}\left[Q_1^{ji} - Q_2^{ij}\right]\Bigg\}.
\end{aligned}
$$

Since when $\hat{\varepsilon} < 0$ it results that $G(.) \to 0$, we can neglect the terms that go to zero faster and thus can write:

$$\frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} \cong \frac{1}{2}\frac{\partial \psi(\hat{p}_i,\hat{p}_j)}{\partial \hat{p}_i} \tag{A13}$$

$$\frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} \cong \frac{1}{2}\left\{\left.\frac{\partial \psi(\hat{p}_i,\hat{r}_i-\varepsilon_i)}{\partial \hat{r}_i}\right|_{\varepsilon_i=\hat{\varepsilon}} + \hat{\varepsilon}Q_1^{ji}\right\}. \tag{A14}$$

The analysis of the RHS of eq. (A13) is identical to the analysis conducted in Proposition 2, thus the expression of the symmetric FOC w.r.t. $\hat{p}$ at equilibrium is:

(A15) $$\frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} = \frac{1}{4} Q_1^{ij} \{\hat{p}[1-\beta(1-x)] - (c + a - x(a-t)) - \hat{r}\}.$$

The analysis of the term $\psi(.)$ in eq. (A14) when network i is sovereign needs some more intermediate steps. As the traffic is only determined by i, the first-order conditions are $U_i' f_1^i = \hat{p}_i$ and $U_i' f_2^i = \hat{r}_i - \varepsilon_i$, which determine the optimal level of calls $Q^{ij}(\hat{p}_i, \hat{r}_i - \varepsilon_i)$ and $Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{p}_i)$. The expressions for the change in off-net quantities caused by a change in price are:

$$\begin{bmatrix} A^{ij} & B^{ij} \\ B^{ij} & C^{ij} \end{bmatrix} \begin{bmatrix} dQ^{ij} \\ dQ^{ji} \end{bmatrix} = \begin{bmatrix} d\hat{p}_i \\ d\hat{r}_i \end{bmatrix}$$

where $A^{ij} = U_i''(f_1^i)^2 + U_i' f_{11}^i < 0$, $C^{ij} = U_i''(f_2^i)^2 + U_i' f_{22}^i < 0$ and $B^{ij} = U_i'' f_1^i f_2^i + U_i' f_{12}^i$. The indirect utilities derived from the information exchange between i and j are:

$$\begin{aligned}
v^{ij} &= v(\hat{p}_i, \hat{r}_i - \varepsilon_i) = U_i(f^i(Q^{ij}(\hat{p}_i, \hat{r}_i - \varepsilon_i), Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{p}_i)) - \hat{p}_i Q^{ij}(\hat{p}_i, \hat{r}_i - \varepsilon_i) \\
v^{ji} &= v(\hat{r}_i - \varepsilon_i, \hat{p}_i) = U_j(f^j(Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{p}_i), Q^{ij}(\hat{p}_i, \hat{r}_i - \varepsilon_i)) - \hat{p}_j Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{p}_i), \\
\frac{\partial v^{ij}}{\partial \hat{r}_i} &= U_i' f_2^i \frac{\partial Q^{ji}}{\partial \hat{r}_i} \\
\frac{\partial v^{ji}}{\partial \hat{r}_i} &= \frac{\partial Q^{ji}}{\partial \hat{r}_i}(U_j' f_1^j - \hat{p}_j) + U_j' f_2^j \frac{\partial Q^{ij}}{\partial \hat{r}_i}.
\end{aligned}$$

The FOC w.r.t. $\hat{r}_i$ of (A9) is the following at a symmetric equilibrium:

(A16) $$\begin{aligned}
&\frac{\partial \psi(\hat{p}_i, \hat{r}_i - \varepsilon_i)}{\partial \hat{r}_i} = \frac{1}{2}\left[U_i' f_2^i \frac{\partial Q^{ji}}{\partial \hat{r}_i} + (\hat{p}_i - c - a + \hat{r}_j)\frac{\partial Q^{ij}}{\partial \hat{r}_i} + (a-t)\frac{\partial Q^{ji}}{\partial \hat{r}_i}\right] - \\
&\frac{1}{2}\left[\frac{\partial Q^{ji}}{\partial \hat{r}_i}(U_j' f_1^j - \hat{p}_j) + U_j' f_2^j \frac{\partial Q^{ij}}{\partial \hat{r}_i}\right] = \\
&\frac{1}{2}\frac{\partial Q^{ji}}{\partial \hat{r}_i}\left\{\left[U_i' f_2^i + a - t + (\hat{p} - c - a)x\right] - \left[(U_i' f_1^i - \hat{p}) + x(U_i' f_2^i - \hat{r})\right]\right\}
\end{aligned}$$

where $x = \dfrac{\partial Q^{ij}/\partial \hat{r}_i}{\partial Q^{ji}/\partial \hat{r}_i} = \dfrac{-B^{ij}}{A^{ij}}$ denotes the incoming propagation effect related to the calls received by network i who sovereigns the information exchange. After replacing (A16) in (A14) as the noise vanishes, when $\varepsilon_i = \hat{\varepsilon} = \hat{r} - U' f_2$, it results:

$$\frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} \cong \frac{1}{4} Q_1^{ji} \left\{ U'f_2 + a - t + (\hat{p} - c - a)x + \hat{r} - U'f_2 - \left[(U'f_1 - \hat{p}) + x(U'f_2 - \hat{r}) \right] \right\}.$$

Since we started by considering the case where senders are sovereign most of the time, i.e., $\hat{\varepsilon} < 0 \Leftrightarrow U'f_2 = \beta\hat{p} > \hat{r}$ for the receiver and $U'f_1 = \hat{p}$ for the sender, we finally get:

$$\text{(A17)} \quad \frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} = \frac{1}{4} \left\{ Q_1^{ji} \left[\hat{r} + a - t + (\hat{p} - c - a)x - x(\beta\hat{p} - \hat{r}) \right] \right\}.$$

The analysis when there is receiver sovereignty most of the time follows the same procedure. The threshold levels of the noise are given by $\hat{\varepsilon}_i = \hat{r}_i - U'_i f_2^i(\hat{r}_j, \hat{r}_i) > 0$ and $\hat{\varepsilon}_j = \hat{r}_j - U'_j f_2^j(\hat{r}_i, \hat{r}_j) > 0$. Maximizing (A10) w.r.t. $\hat{p}_i$ and $\hat{r}_i$, and using the regularity condition we obtain after simplifications:

$$\frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} = \frac{1}{2}(1 - G(\hat{\varepsilon})) \left\{ (1 - G(\hat{\varepsilon})) \frac{\partial \psi(\hat{p}_i, \hat{p}_j)}{\partial \hat{p}_i} + G(\hat{\varepsilon}) \left. \frac{\partial \psi(\hat{p}_i, \hat{r}_i - \varepsilon_i)}{\partial \hat{p}_i} \right|_{\varepsilon_i = \hat{\varepsilon}} + (1 - G(\hat{\varepsilon}))\hat{\varepsilon} Q_2^{ji}(\hat{p}_i, \hat{p}_j) - G(\hat{\varepsilon})\hat{\varepsilon} Q_1^{ij} \right\}$$

$$\frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} = \frac{1}{2} G(\hat{\varepsilon}) \left\{ (1 - G(\hat{\varepsilon})) \left. \frac{\partial \psi(\hat{p}_i, \hat{r}_i - \varepsilon_i)}{\partial \hat{r}_i} \right|_{\varepsilon_i = \hat{\varepsilon}} + G(\hat{\varepsilon}) \left. \frac{\partial \psi(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i)}{\partial \hat{r}_i} \right|_{\varepsilon_i = \varepsilon_j = \hat{\varepsilon}} - (1 - G(\hat{\varepsilon}))\hat{\varepsilon} Q_2^{ij} \right\}.$$

Since when $\hat{\varepsilon} > 0$ it results that $G(.) \to 1$, we can further approximate:

$$\text{(A18)} \quad \frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} \cong \frac{1}{2} \left\{ \left. \frac{\partial \psi(\hat{p}_i, \hat{r}_i - \varepsilon_i)}{\partial \hat{p}_i} \right|_{\varepsilon_i = \hat{\varepsilon}} - \hat{\varepsilon} Q_1^{ij} \right\}$$

$$\text{(A19)} \quad \frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} \cong \frac{1}{2} \left. \frac{\partial \psi(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i)}{\partial \hat{r}_i} \right|_{\varepsilon_i = \varepsilon_j = \hat{\varepsilon}}$$

The analysis of the term $\psi(.)$ in (A18) is the mirror of the analysis of the term $\psi(.)$ in (A14) as now the regime is one of "network *j*" sovereignty. The FOC w.r.t. $\hat{p}_i$ of (A9) is the following at a symmetric equilibrium:

$$\frac{\partial \psi(\hat{p}_i, \hat{r}_i - \varepsilon_i)}{\partial \hat{p}_i} = \frac{1}{2} \left[-Q^{ij} + U'_i f_2^i \frac{\partial Q^{ji}}{\partial \hat{p}_i} + Q^{ji} + (\hat{p}_i - c - a) \frac{\partial Q^{ij}}{\partial \hat{p}_i} + \hat{r}_j \frac{\partial Q^{ij}}{\partial \hat{p}_i} \right] +$$

$$\text{(A20)} \quad \frac{1}{2} \left[(a - t) \frac{\partial Q^{ji}}{\partial \hat{p}_i} - \frac{\partial Q^{ji}}{\partial \hat{p}_i} (U'_j f_1^j - \hat{p}_j) - U'_j f_2^j \frac{\partial Q^{ij}}{\partial \hat{p}_i} \right] =$$

$$\frac{1}{2} \frac{\partial Q^{ij}}{\partial \hat{p}_i} \left\{ \left[\hat{p}_i - (c + a - x\hat{r}_i - x(a - t)) \right] - \left[x(U'_j f_1^j - \hat{p}_j) + (U'_j f_2^j - \hat{r}_j) \right] \right\}$$

where $x = \dfrac{\partial Q^{ji}/\partial \hat{p}_i}{\partial Q^{ij}/\partial \hat{p}_i} = \dfrac{-B^{ij}}{C^{ij}}$ is the outgoing propagation effect related to the calls originated by network *i*. After replacing (A20) in (A18) as the noise vanishes, when $\varepsilon_i = \hat{\varepsilon} = \hat{r} - U'f_2$, it results:

$$\frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} \cong \frac{1}{4} Q_i^{ij} \left\{ \left[\hat{p}_i - \left(c + a - x\hat{r} - x(a-t) \right) \right] - x(U'f_1 - \hat{p}_j) \right\}.$$

Since we started by considering the case where receivers are sovereign most of the time, i.e., $\hat{\varepsilon} > 0$ and $U'f_1 = \hat{r}/\beta$ for the receiver, we finally get:

$$\text{(A21)} \quad \frac{\partial \hat{\pi}_i}{\partial \hat{p}_i} = \frac{1}{4} Q_i^{ij} \left\{ \left[\hat{p} - \left(c + a - x\hat{r} - x(a-t) \right) \right] - x(\hat{r}/\beta - \hat{p}) \right\}.$$

The analysis of the term $\psi(.)$ in (A19) corresponds to the case of "receiver sovereignty". The indirect utilities from the information exchange between *i* and *j* are:

$$v^{ij} = v(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i) = U_i(f^i(Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i), Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{r}_j - \varepsilon_j)) - \hat{p}_i Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i)$$
$$v^{ji} = v(\hat{r}_i - \varepsilon_i, \hat{r}_j - \varepsilon_j) = U_j(f^j(Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{r}_j - \varepsilon_j), Q^{ij}(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i)) - \hat{p}_j Q^{ji}(\hat{r}_i - \varepsilon_i, \hat{r}_j - \varepsilon_j).$$

The FOC w.r.t. $\hat{r}_i$ of (A9) is:

$$\frac{\partial \psi(\hat{r}_j - \varepsilon_j, \hat{r}_i - \varepsilon_i)}{\partial \hat{r}_i} = \left[\frac{\partial Q^{ij}}{\partial \hat{r}_i}(U_i' f_1^i - \hat{p}_i) + \hat{r}_i \frac{\partial Q^{ji}}{\partial \hat{r}_i} + \left(\hat{p}_i - c - a \right) \frac{\partial Q^{ij}}{\partial \hat{r}_i} \right] + \alpha_i \hat{r}_j \frac{\partial Q^{ij}}{\partial \hat{r}_i} +$$
$$\text{(A22)} \quad \alpha_j (a-t) \frac{\partial Q^{ji}}{\partial \hat{r}_i} - \alpha_i \left(\frac{\partial Q^{ji}}{\partial \hat{r}_i}(U_j' f_1^j - \hat{p}_j) + \hat{r}_j \frac{\partial Q^{ij}}{\partial \hat{r}_i} \right) =$$
$$= Q_1^{ji} \left\{ \alpha_j x \left[U_i' f_1^i - (c+a) \right] + \alpha_j \hat{r}_i + \alpha_j (a-t) - \alpha_i \left(U_j' f_1^j - \hat{p}_j \right) \right\}$$

where $\partial Q^{ij}/\partial \hat{r}_i = Q_2^{ij}, \partial Q^{ji}/\partial \hat{r}_i = Q_1^{ji}$ and $x = Q_2^{ij}/Q_1^{ji}$ is the propagation effect when the receiver is sovereign.[25] Denoting $\beta = f_2/f_1$, and recalling that $\hat{r} = U'f_2$ when receivers are sovereign as the noise vanishes, after replacing (A22) in (A19), the FOC at the symmetric equilibrium is:

$$\text{(A23)} \quad \frac{\partial \hat{\pi}_i}{\partial \hat{r}_i} = \frac{1}{4} Q_1^{ji} \left\{ \hat{r}\left[1 - (1-x)/\beta \right] + (a-t) - x(c+a) + \hat{p} \right\}.$$

Putting all the results together (eqs. (A15), (A17), (A21), (A23)), these are the final expressions of the FOCs at a symmetric equilibrium as the noise vanishes:

---

25 The study of the sign of the SOC is the same as in the proof of Proposition 2. Sufficient conditions for having $\left.\dfrac{\partial^2 \hat{\pi}_i}{\partial \hat{r}_i^2}\right|_{\substack{sym \\ x=0}} < 0$ are that either x is small, or that it is close to 1 and $x' \approx 0$.

$$4\frac{\partial\hat{\pi}}{\partial\hat{p}} = \begin{cases} Q_1^{ij}\{\hat{p}[1-\beta(1-x)] - (c+a-x(a-t)) - \hat{r}\} & \text{if } \beta\hat{p} \geq \hat{r} \\ Q_1^{ij}\{\hat{p} - (c+a-x(\hat{r}+a-t)) - x(\hat{r}/\beta - \hat{p})\} & \text{if } \beta\hat{p} \leq \hat{r} \end{cases}$$

$$4\frac{\partial\hat{\pi}}{\partial\hat{r}} = \begin{cases} Q_1^{ij}[\hat{r}+a-t+(\hat{p}-c-a)x - x(\beta\hat{p}-\hat{r})] & \text{if } \beta\hat{p} \geq \hat{r} \\ Q_1^{ij}\{\hat{r}[1-(1-x)/\beta] + (a-t) - x(c+a) + \hat{p}\} & \text{if } \beta\hat{p} \leq \hat{r} \end{cases}$$

Notice that these FOCs boil down to those of JLT (p. 109-110) when $x = 0$. Consider first the case where $\beta\hat{p} > \hat{r}$ and senders determine most volume in equilibrium. The relevant FOCs give eq. (21) as the interior solution. This solution is valid if $\beta < 1/(1 - x - x^2)$ and compatible with the regime of sender sovereignty if $\beta\hat{p}^* > \hat{r}^* \Rightarrow a > (t-\beta c)/(1+\beta)$. If the reception charge $\hat{r}^*$ must take positive values: $\hat{r}^* \geq 0 \Rightarrow a \leq t + x^2\dfrac{\beta(c+t)}{(1-x)(1+x-\beta)}$. If a is above this threshold, then the reception charge $\hat{r}^*$ is zero and $\hat{p}^*$ is obtained directly from eq. (11bis), resulting in eq. (22). In this case the range of validity is $\beta < 1/(1 - x)$. The case where $\beta\hat{p} < \hat{r}$ and receivers determine most volume in equilibrium can be analyzed in a similar manner, giving eq. (23) and (24). Finally, consider the case where $\beta\hat{p} = \hat{r}$. After substituting this into the FOCs, we have the following candidate pair of prices when $a = (t-\beta c)/(1+\beta)$: $\hat{p}^* = p^* = (c+t)/(1+\beta)$, $\hat{r}^* = \beta p^* = \beta(c+t)/(1+\beta)$. These prices are efficient. However, when $a = (t-\beta c)/(1+\beta)$, imagine to set the reception charge at the efficient level and to increase the call price above it: $\hat{p}^* = p^* + \delta$, $\delta > 0$, and $\hat{r}^* = \beta p^*$. In this range $\beta\hat{p}^* > \hat{r}^*$ and the FOC w.r.t. $\hat{p}$ simplifies to:

$$4\frac{\partial\hat{\pi}}{\partial\hat{p}} = Q_1^{ij}\delta[1-\beta(1-x)].$$

Similarly, imagine $\hat{p}^* = p^*$ and $\hat{r}^* = \beta p^* + \delta$; then the FOC w.r.t. $\hat{r}$ is:

$$4\frac{\partial\hat{\pi}}{\partial\hat{r}} = Q_1^{ij}\delta(\beta + x - 1)/\beta.$$

It follows that efficient prices can *never* be reached when the utility function is separable since, with $x = 0$, there is always an incentive to deviate and induce a break-down by setting either arbitrarily high call prices when $\beta > 1$ or arbitrarily high reception charges when $\beta < 1$. However, in the presence of a positive propagation factor there is a range $1 - x < \beta < 1/(1-x)$ such that the deviation is not profitable, thus prices are given by eq. (25) and are efficient. QED

Proof of Proposition 6.
We only consider the case of senders sovereignty most of the time. We also suppose that there is no problem associated with implementing negative reception charges. This means that retail prices are given by eq. (21) over the entire range $a > (t-\beta c)/(1+\beta)$, by eq. (25)

when $a = (t - \beta c)/(1+\beta)$, and induce an off-net breakdown when $a < (t - \beta c)/(1+\beta)$. As in Section 4.2, in a symmetric equilibrium the profit is given by:

$$\pi^* = \tfrac{1}{4}\left(p^* + r^* - c - t\right)Q(p^*, p^*) + \tfrac{1}{4}\left(\hat{p}^* + \hat{r}^* - c - t\right)Q(\hat{p}^*, \hat{p}^*) + \tfrac{1}{2}\left(F^* - f\right).$$

Since the sender is sovereign most of the time, the effect of a is as in the proof of Proposition 3, $\dfrac{\partial Q(\hat{p}^*, \hat{p}^*)}{\partial a} = (1+x)Q_1(\hat{p}^*, \hat{p}^*)\dfrac{\partial \hat{p}^*}{\partial a}$, $\dfrac{\partial v(\hat{p}^*, \hat{p}^*)}{\partial a} = \left[-Q + \beta\hat{p}^* Q_1(1+x)\right]\dfrac{\partial \hat{p}^*}{\partial a}$ while $\dfrac{\partial F^*}{\partial a} = -Q(\hat{p}^*, \hat{p}^*)\left(\dfrac{\partial(\hat{p}^* + \hat{r}^*)}{\partial a}\right) + \left(\beta\hat{p}^* - \hat{r}^*\right)Q_1(1+x)\dfrac{\partial \hat{p}^*}{\partial a}$.

The FOC w.r.t. the reciprocal access charge is given by the following condition:

$$\frac{\partial \pi^*}{\partial a} = \frac{1}{4}\left[\frac{\partial Q(\hat{p}^*, \hat{p}^*)}{\partial a}\left(\hat{p}^* - c - t + \hat{r}^*\right) + Q(\hat{p}^*, \hat{p}^*)\frac{\partial(\hat{p}^* + \hat{r}^*)}{\partial a}\right] + \frac{1}{2}\frac{\partial F^*}{\partial a}.$$

Under the assumption that x does not change for a small variation in prices, we have:

(A23)
$$\begin{aligned}\frac{\partial \pi^*}{\partial a} &= \frac{1}{4}\left\{-Q(\hat{p}^*, \hat{p}^*)\frac{\partial(\hat{p}^* + \hat{r}^*)}{\partial a} + Q_1(\hat{p}^*, \hat{p}^*)(1+x)\frac{\partial \hat{p}^*}{\partial a}\left[\hat{p}^*(1+2\beta) - c - t - \hat{r}^*\right]\right\} \\ &= \frac{1}{4}\left\{-Q(\hat{p}^*, \hat{p}^*)\frac{\partial(\hat{p}^* + \hat{r}^*)}{\partial a} + Q_1(\hat{p}^*, \hat{p}^*)\frac{(1-x)(3+2x)}{1-\beta+\beta x(1+x)}\frac{\partial \hat{p}^*}{\partial a}\left[a(1+\beta) - t + \beta c\right]\right\}\end{aligned}$$

with:

- $\dfrac{\partial(\hat{p}^* + \hat{r}^*)}{\partial a} = \dfrac{1-x}{1-\beta+\beta x(1+x)}\left((1+\beta) + \dfrac{\partial \beta}{\partial a}\left[a + c + (\hat{p}^* + \hat{r}^* - c - t)\dfrac{1 - x(1+x)}{1-x}\right]\right)$,
- $\dfrac{\partial \hat{p}^*}{\partial a} = \dfrac{1}{1-\beta+\beta x(1+x)}\left(2 - x(1+x) + \dfrac{\partial \beta}{\partial a}\hat{p}^*\left(1 - x(1+x)\right)\right)$,

where $\frac{\partial \beta}{\partial a} = \frac{\partial \beta}{\partial \hat{p}}\frac{\partial \hat{p}^*}{\partial a}$. Note that $\frac{\partial \hat{p}^*}{\partial a}$ is always positive when $\frac{\partial \beta}{\partial a}$ is small enough.

If (A23) is negative everywhere in the whole domain $a > \dfrac{t - \beta c}{1+\beta}$, when an interior equilibrium in retail prices exists, then the lowest threshold of a is chosen. At $a^* = \dfrac{t - \beta c}{1+\beta}$, eq. (25) is valid and $\left.\dfrac{\partial \pi^*}{\partial a}\right|_{a=a^*} = 0$. A sufficient condition for having (A23) negative everywhere is that both $\frac{\partial \hat{p}^*}{\partial a} > 0$ and $\frac{\partial(\hat{p}^* + \hat{r}^*)}{\partial a} > 0$. This is ensured by the following condition: $\frac{\partial \beta}{\partial a} \cong 0$, i.e., the access charge does not influence much the marginal utility in receiving calls. This condition is by no means necessary.[26]

In order to analyse if a^* is a global maximum, we have to compare the profit level reached when $a = a^*$ with the profit level that can be reached by setting $a < a^*$ and inducing a connectivity breakdown. The comparison is simple since we can note that:

- on-net retail prices are always efficient in both cases, thus they do not generate differences;
- off-net calls also do not generate differences: in case of connectivity breakdown there are no off-net calls, thus profits are zero, while when $a = a^*$ off-net calls are efficient but also generate zero profits;
- the only differences arise from the fixed fee F^* given by eq. (12).

In both cases profits are raised only through the fixed fee. In case of connectivity breakdown the fixed fee is:

$$\begin{aligned} F^* &= f - (p^* - c - t)Q(p^*, p^*) + 1/2\sigma - \hat{r}Q(\hat{p}^*, \hat{p}^*) - (v(p^*, p^*) - v(\hat{p}^*, \hat{p}^*)) \\ &= f - (p^* - c - t)Q(p^*, p^*) + 1/2\sigma - v(p^*, p^*). \end{aligned}$$

In case $a = a^*$ the fixed fee is:

$$\begin{aligned} F^* &= f - (p^* - c - t)Q(p^*, p^*) + 1/2\sigma - \hat{r}Q(\hat{p}^*, \hat{p}^*) - (v(p^*, p^*) - v(\hat{p}^*, \hat{p}^*)) \\ &= f - (p^* - c - t)Q(p^*, p^*) + 1/2\sigma - r^*Q(p^*, p^*). \end{aligned}$$

Thus it is better to set $a = a^*$ than to induce a breakdown if $-r^*Q(p^*, p^*) > -v(p^*, p^*) \Rightarrow U(p^*, p^*) - (p^* + r^*)Q(p^*, p^*) > 0 \Rightarrow U(p^*, p^*) - (c + t)Q(p^*, p^*) > 0$ that is always verified. QED

---

[26] As an example, take the same specification as in footnote 17 with a constant and positive propagation factor $x < 1$, $U^i = Q^{ij}(1 + xQ^{ji}) - Q^{ij^2}/2$. This generates the following demand functions when senders are sovereign most of the time: $Q^{ij} = [1 - \hat{p}_i + x(1 - \hat{p}_j)]/(1 - x^2)$. At a symmetric equilibrium the impact of the price on the ratio of marginal utilities is different from zero: $\partial\beta/\partial\hat{p} = -x/[\hat{p}^{*2}(1 - x)]$. However, after substitutions one also gets: $\partial\hat{p}^*/\partial a = (1 - x)^2(2 + x)/(1 - x^2 - x^3)$ and $\partial(\hat{p}^* + \hat{r}^*)/\partial a = (1 - x)(1 - 2x)/(1 - x^2 - x^3)$. Both derivatives are strictly positive for $0 < x < ½$, thus (A23) is negative. Outside this range of x, it is possible to show that (A23) is still negative everywhere by direct substitution of equilibrium values and by taking into account restrictions on parameters that ensure that equilibrium quantities are positive.

Monopoly Practices and Competitive Behavior in the French Satellite Pay-TV Market*

Michele Cincera[†] and Abdul G. Noury[‡]
Université Libre de Bruxelles

November 10, 2003

Abstract

This paper uses monthly data from a differentiated market dominated by a duopoly to analyze the nature of interactions between competitor firms. The French satellite pay-TV market, characterized by a monopoly until the end of 1997, has been dominated by the incumbent, Canal Satellite, and the entrant, TPS. This paper investigates the effects of the entry and tests for collusive behavior by means of non-nested methods in this duopoly with differentiated products. The main findings reject collusive behaviors in favor of Stackelberg competion with the incumbent as the leader. A preliminary comparison with the UK Satellite pay-TV market characterized by a monopoly, indicates that prices are substantially lower in France. In view of our results, we argue that the difference can be explained by more competition between firms in the French case.

Keywords: Pay-TV, media, competition, structural models

JEL codes: L22, L82

*We are grateful to Simon Anderson, Tom Coupé, Olivier Gergaud, Estelle Malavolti, Yannick Scaramozzino as well as participants at the 2nd Media Economics Workshop, Bergen, Norway for comments and helpful discussions. The authors are solely responsible for any remaining errors.

[†]ULB-DULBEA and CEPR, mcincera@ulb.ac.be

[‡]ULB-DULBEA and ECARES, anoury@ulb.ac.be

1 Introduction

Since the beginning of the 1990s satellite pay-TV has played an increasingly important role in the broadcasting industry. In almost all countries consumers have now access to several hundred programs and various interactive services. This process is only accelerating with substantial progress in digital technology. While most satellite pay-TV markets in Europe are characterized by monopolies, in some cases rival firms compete with each other. French pay-TV market is an example of such competitive situation. However, even in markets with several firms, there is a risk of anti-competitive conduct. Whether the fruit of the technological progress is realized, or eliminated by anti-competitive conducts, is a question of central importance to economics. Not surprisingly, authorities such as the European Commission focus with great interest on competitive behavior in industries like pay-TV[1].

Much has been said and written on the nature and consequences of competition between firms in various industries. The merits as well as the limits of competition are relatively well documented in the economics literature. As far as pay-TV industry is concerned, however, the bulk of the literature focuses on the theoretical arguments on competition (see for example Armstrong, 1999; Harbord and Ottaviani, 2001). In addition, studies devoted to empirical verifications of day-to-day interactions between firms are quite rare. To the best of our knowledge, the existing studies focused on vertical integration in the US pay-TV market (Rubinovitz, 1993; Waterman and Weiss, 1996; Chipty, 2001).

The goal of this paper is to provide an empirical assessment of the nature of competition between two dominant (and a priori rival) firms in the French satellite pay-TV industry. The French satellite pay-TV market has been characterized by a monopoly until the end of 1997 when a second firm entered this market. Both the incumbent, CanalSatellite, and the entrant, Télévision par Satellite (TPS), organize their digital

[1] See for instance, the European directive "Television without Frontiers" (89/552/EEC and 97/36/EC), which establishes the legal frame of reference for the free movement of television broadcasting services in the European Union.

offer around three main bundles containing channels relating to movies, sport and interactive services. This paper investigates the effects of the entry and tests for different (implicit or explicit) collusive behaviors by means of non-nested methods in this duopoly with differentiated products. Structural econometric models are estimated using monthly data from 1997 to 2002 on the number of subscriptions, subscription fees, advertising effort, discounts, number of channels and content as well as other exogenous variables characterizing the demand.

We find that despite the entry of a competitor the subscription price of satellite pay-TV in France did not decrease. However, the 'quality' (i.e. the number as well as the diversity) of programs substantially increased over time. Our empirical analysis show that an increase in the CanalSatellite price leads to a decrease in 'quantity' (i.e. the number of subscriptions) but for TPS an increase in price does not affect its quantity. Moreover, advertising expenses of each firm have a positive and significant impact on its own demand, but a negative (predatory) effect on its rival. Finally, the results of nonnested tests indicate that CanalSatellite is the leader and TPS the follower.

The rest of the paper is organized as follows. Section 2 provides a picture of the structure and the main actors of the French digital satellite pay-TV market. Section 3 describes the construction of the data set. Section 4 discusses the modelling strategy and presents the econometric frameworks used to estimate the model. Results of our empirical analysis and comments are provided in Section 5. Section 6 compares the French competitive situation with the British monopolistic satellite pay-TV industry. Section 7 concludes.

2 The French digital satellite Pay-TV market

With some oversimplifications the pay-TV industry can be described as having three sectors or layers (See Figure 1 in the Appendix). The production of programming constitutes the first sector (the upstream market). Producers in this sector obtains sports rights, produce movies as well as other "premium" programming. Typically, premium programming corresponds to first run movies and major sport events. Access to these inputs are widely viewed as essential to attract new subscribers[2]. Then comes retailing of programming to consumers (the intermediate market). In this sector programs are bought from producers at the wholesale level. And finally the distribution of programming to customers constitutes the third sector (the downstream market). The central ingredient of the industry is the encryption system and a set-top box in order to decode the scrambled signal. There are several encryption systems in the industry, and usually the system used for one encryption is not compatible with another encryption. The encryption system together with the set-top box is called the "Conditional Access System". Naturally, there may be vertical integration, between the first and the second sectors, between the second and the third, or between the three sectors. In this paper we only focus on competitive interactions in the distribution sector in the French case.

France is the only European market where three digital satellite television platforms coexist. The two dominant firms, CanalSatellite and TPS, together with a third but minor company, ABsat, share the French digital satellite market[3]. In addition, the French market includes cable television or terrestrial pay-TV. Such a competitive situation can be expected to promote technological progress and benefit consumers given the increased pay-TV offer and more advantageous subscription conditions.

The pay-TV market has been identified as a separate one from the free-access television market, by the European Commission in its decisions[4]. Since pay-TV broadcasters are usually financed through subscription, the pay-TV market is distinct from free access television, which is usually financed by advertising or by State contributions[5]. Moreover, for cultural and linguistic reasons, the French geographic market can be viewed

[2] A second run movie will be regarded as a lower quality movie by the satellite subscribers. As a result, pay-TV broadcasters may be forced to reduce their subscription prices to differentiate accordingly.

[3] In December 2000, the market share of ABsat was less than 2% whereas the combined market share of the two dominant firms was over 98%.

[4] See for instance, Commission decisions 94/922/EC (IV/M 469-MSG Media Service), OJ L364, paragraph 32 and 33.

[5] The economics literature on free access television is large. Recent theoretical contributions include Anderson and Coate (2000) and Gabszewicz et al. (2002).

as a national market. In this paper a further distinction is made between digital satellite broadcasters and cable television or terrestrial pay-TV. These two types of pay-TV broadcasting channels are assumed to form separate markets. Admittedly, this assumption is a relatively strong one since cable pay-TV services are somehow substitutes to satellite pay-TV services. Nonetheless, for at least two reasons, one may consider the two markets as separate. The first reason is the geographic landscape of France, which except major urban areas is not well equipped with TV cable infrastructure[6]. The second reason rests in the differences regarding the content and the quality of the programs broadcast by these two sources. The satellite digital service is not only of a better quality but also is characterized by a larger diversity. As a result, in this analysis we confine our attention to the competitive behaviors of the two main actors, CanalSatellite and TPS[7].

CanalSatellite launched a bouquet of analogue pay-TV channels broadcast by satellite since 1992. This service which ceased in October 1998 was replaced by a digital service launched in 1996. In 2002, Canal+, the first pay-TV firm in France since 1984, owns 66% of CanalSatellite and the remaining 34% are owned by Lagardère. 49% of Canal+ are owned by Group Canal+, fully-owned by Vivendi-Universal. TPS launched its digital platform for the distribution of satellite pay-TV programs and services in January 1997. In 2002, the ownership structure of TPS is divided into three shareholders: TF1 (50%), M6 (25%) and Lyonnaise satellite (25%).

Besides the basic bundle, each package of CanalSatellite and TPS offers a full range of specialist channels such as movies, children programs, sport events, music programs, and so on[8]. For CanalSatellite, consumers must subscribe to the basic offer or to the maximal offer in order to access to the movie offer. On the contrary, consumers can subscribe directly to TPS' movie offer. CanalSatellite and TPS do not employ the same

[6] In the MediaCabSat (2002) report, the share of subcriptions to cable-TV was about 35% for the period of January 2002 to May 2002 against 65% for Satellite pay-TV.

[7] As already mentioned, the third player, ABSat, only plays a minor role in the market given its marginal number of subscriptions.

[8] See Tables A2 and A4 for more details about the content and the number of channels.

encryption technology and do not market compatible decoder equipment. However, since the subscription and the renting of the set-top box are not expensive, and decoders compatible with both encryption technologies are more and more available, it is reasonable to assume that switching costs do not constitute a barrier for a consumer to churn[9]. As already mentionned, the quality of the movie offer depends on the number of the premium movies, considered as driver products to capture new consumers. Consequently we assume that the consumers' choice to subscribe to the movie offer is influenced by both the quality and the price. That is, on the downstream market, CanalSatellite and TPS are assumed to compete in (quality-adjusted) price. Their strategies consist of proposing to consumers a high quality movie offer at a low price.

CanalSatellite has two main advantages over its rival. First, it was the first company to offer a satellite digital bouquet. The initial investment constitutes a first mover advantage for CanalSatellite. Second, CanalSatellite also benefits from a reputational advantage, due to a substantial "catalogue" of French and American movies acquired by Canal+, which built its reputation on first transmissions of quality feature films. In addition, CanalSatellite has access to outside premium inputs, through its parent's network[10]. While this advantage in terms of consumers' preference should have decreased with the entry of TPS, the investments of the latter on the outside upstream market are negligible as compared to the incumbent to be taken into account as a relevant parameter. Quantitative information as regards the aquisition of exclusive pay-TV rights is not easily available. However, CanalSatellite appears to invest six more times in its programs' content as compared to TPS and in terms of the movie offer, the former has exclusive rights to broadcast 30 movies out of 32 listed on the box office (TéléSatellite, 143, p. 44). As noted by Harbord and Ottaviani (2001), acquiring exclusive rights to

[9] The rental cost of a set-top box varies from about 30% to 25% of the annual total subscription's costs at the end of the period and is the same for both firms. The entry costs (activation costs) at the end of the period represent about 11% the total subscription's costs for one year, 5.5% for two years, and so on.

[10] Vivendi Universal has co-financed a high number of lucrative movies with others majors Hollywood studios (Columbia, Disney, Warner, Fox and MGM).

premium programming (or sport events) gives competitive advantages over rival, which suffer a loss (negative externality).

The entry of TPS in the market was not considered by the European Commission as restricting competition. The European Commission noting the primarily pro-competitive effect of having a new operator, considered the entry of TPS on the market as not falling under Article 85(1) of the EC Treaty. However, the Commission noticed that two clauses included in the contract did restrict competition, in particular by limiting competitors' access to a certain type of contents, but that these clauses can be exempted under Article 85(3) of the EC Treaty for three years, which is the crucial period of TPS' launch[11].

Table A1 in the Appendix gives the market shares in terms of the number of subscriptions of CanalSatellite and TPS. The market share represents the penetration rate, which is the fraction of homes which have subscribed with a pay-TV operators. In practice, the European Commission has relied strongly on market shares in order to establish dominance. Firms are frequently declared dominant if they supply at least 40-45 per cent of the market[12]. Given this definition and the market share of CanalSatellite, this firm can be considered as dominant on the French Satellite pay-TV market.

[11]These clauses, which had originally been envisaged to remain in place for 10 years, grant to TPS: priority and the right of last refusal to broadcast channels and television services edited and controled by its parent companies; and exclusive rights to distribute digitally the four general content channels TF1, France 2, France 3 and M6. The Commission considers in particular that the exclusive availability of these four general content channels on TPS is a differentiating feature and an appealing product which are essential for this new entrant to enter the pay-TV market.

[12]The European Court of Justice has defined the dominance in the United Brands case as : "a position of economic strength enjoyed by an undertaking which enables it to prevent effective competition being maintained on the relevant market by giving it the power to behave to an appreciable extent independently of its competitors, customers and ultimately consumers".

3 Data

The data set has been constructed for the period of January 1997 to December 2002 from raw data published by TeleSatellite, the French reference magazine for Satellite pay-TV programs and news. Each month, this magazine provides a detailed table including the subscription prices of the different options, i.e. bundles, for CanalSatellite and TPS as well as their content in terms of program channels. It is important to note that both firms frequently propose sales promotions that consist of price rebates and free subscription periods (in general one to three months for a one year subscription contract). All in all, these promotions represent about 20% of the subscription price. The variables constructed are the monthly number of subscriptions for both firms, the price of the subscription net of bonus or free periods and the number of channels for each bundle (basic option, movie option, thematic channels, pay per view movies and sport events)[13]. Furthermore, for the basic bundle, 6 groups of programs have been created on the basis of the description of each channel, and for each group the number of channels has been calculated for each month[14].

Figure A2 and Table A3 in the Appendix give the evolution of the number and the net price of subscriptions over the period. Descriptive statistics for the main variables are reported in Table A4 in the Appendix. In order to take into account the level of advertising effort, a qualitative variable has been constructed. To that end, the number of advertisements for CanalSatellite and TPS has been counted and weighted by its size (1 for a full page, .5 for an half page, .25 for a quarter page, and so on) for each issue of the TeleSatellite magazine. Finally, additional macroeconomic variables such as the number of unemployed, an index of purchase power, and temporal dummies assumed to pick up seasonal effects have also been considered in the empirical analysis as additional control variables.

[13]TéléSatellite is the primary source for this information. Note that for the number of subscriptions, other sources on the internet (CanalSatellite and TPS websites, specialized reports, magazines,...) were consulted to check for consistency and to complete missing observations in a few cases.

[14]See Table A2 in the Appendix.

4 The Model[15]

4.1 Demand Functions

As already mentioned, the French digital pay-TV industry is dominated by two firms, the incumbent CanalSatellite and the entrant TPS together with a third player (ABsat) with a negligible, and declining over time, market share. Given that ABsat is clearly a competitive fringe, its market share was just slightly over 1 % during the period covered by this analysis, we focused on the two main firms.

CanalSatellite (resp. TPS) in the rest of the analysis is also denoted firm 1 (resp. firm 2). For analytical and empirical tractability, in this analysis we postulate a simple demand function. As a result, the residual demand function of firm i (resp. firm j) is assumed to have the following form (Gasmi and Vuong, 1991; Gasmi et al., 1992):

$$N_i = \gamma_{i0} + \alpha_{ii}p_i + \alpha_{ij}p_j + \gamma_{ii}A_i^{\frac{1}{2}} + \gamma_{ij}A_j^{\frac{1}{2}}, \qquad i,j = 1,2 \ i \neq j \qquad (1)$$

where N_i, p_i and A_i are firm i's number of subscribers or "quantity", price, and advertising expenditures, respectively, and α's and γ's are the parameters of the model to be estimated. According to economic theory we expect $\alpha_{ii} \leq 0$, $\alpha_{ij} \geq 0$, and $\gamma_{ii} \geq 0$. Note that for simplicity advertising expenditures are assumed to be exogenous. One justification for this simplifying assumption is the limited number of observations at hand. By considering advertising as endogenous we would have two additional equations and at least six more parameters to be estimated. Since the total number of observations per firm is about 70, estimation of those additional parameters would considerably reduce the number of degrees of freedom. In addition, we do not have monthly data on the firms' precise expenditures, we only use a proxy variable. Clearly, with a larger data set and with more detailed observations it would be interesting to relax this assumption and estimate advertising as endogenous. The sign of γ_{ij}, $i \neq j$, is unclear because advertising has two opposing effects. The first is negative (predatory effect) while the

[15] The model is based on the one proposed by Gasmi et al. (1992).

second is positive (spillover effect). Depending on which effect is dominating the sign of this variable may be negative or positive.

The demand of the pay-TV industry, as it is often the case, is affected by a number of other exogenous variables. These exogenous variables are usually related to the economic environment such as variation in family budget and available time for leisure. As a result, the demand equation (4) contains a term, γ_{i0}, that captures the effect of outside environment on the firms' "quantity", i.e., number of subscribers,

$$\gamma_{i0} = \beta_{i0} + \beta_{i1}Y + \beta_{i2}D + \beta_{i3}U, \qquad i = 1,2 \tag{2}$$

where Y, D, and U denotes income, seasonal dummies, and unemployment, respectively. Obviously, an increase in family income will have, ceteris paribus, a positive effect on pay-TV demand. As a result, we expect a positive sign for β_{i1}. Also, since for a given level of income, unemployment should increase the demand for the type of leisure considered in this paper, we expect $\beta_{i3} \geq 0$. As far as the seasonal dummy is concerned, we expect lower than average demand in Summer and higher in Winter.

4.2 Cost Functions

As far as the cost functions are concerned, we assume that both firms use similar technologies. Again, following the literature (and for simplicity) we assume a constant marginal cost function $C_i(q_i) = c_i N_i,\ i = 1,2$ where c_i is a function of input variables that are exogenous in the model.

$$c_i = \eta_{i1}B_i + \eta_{i2}F_i + \eta_{i3}S_i + \eta_{i4}O_i \qquad i = 1,2 \tag{3}$$

with B_i, F_i, S_i, and O_i denoting the number of basic channels, the number of premium movie channels, the number of sport channels and other channels such as music or shopping, respectively. Economic theory suggests $\eta_{ij} \geq 0\ i = 1,2$ and $j = 1,2,3,4$.

Given the data set at hand, we use a simple functional form for costs. Note, however, that the cost function can, in principle, be a complicated and/or non-linear function, though it is not clear how much better results one can obtain by complicating the analysis. Moreover, the econometric analysis of non-linear functions become very often non-tractable.

Given the demand and cost functions, firm i's profit can be written as:

$$\pi_i(p_i, p_j) = (p_i - c_i)(\gamma_{i0} + \alpha_{ii}p_i + \alpha_{ij}p_j + \gamma_{ii}A_i^{\frac{1}{2}} + \gamma_{ij}A_j^{\frac{1}{2}}) - K_i - A_i \qquad i = 1, 2\; i \neq j \quad (4)$$

where K_i is the fixed cost of operation for firm i. The two demand equations and the two first order conditions lead to a system of four equations to be estimated.

$$N_1 - \gamma_{10} - \alpha_{11}p_1 - \alpha_{12}p_2 - \gamma_{11}A_1^{\frac{1}{2}} - \gamma_{12}A_2^{\frac{1}{2}} = \epsilon_1 \quad (5)$$

$$N_2 - \gamma_{20} - \alpha_{21}p_1 - \alpha_{22}p_2 - \gamma_{21}A_1^{\frac{1}{2}} - \gamma_{22}A_2^{\frac{1}{2}} = \epsilon_2 \quad (6)$$

$$N_1 + \upsilon_{11}p_1 + \upsilon_{12}p_2 - \delta_{11}B_1 - \delta_{12}F_1 - \delta_{13}S_1 - \delta_{14}O_1 = \epsilon_3 \quad (7)$$

$$N_2 + \upsilon_{21}p_1 + \upsilon_{22}p_2 - \delta_{21}B_2 - \delta_{22}F_2 - \delta_{23}S_2 - \delta_{24}O_2 = \epsilon_4 \quad (8)$$

It is obvious that υ's and δ's are functions of the structural parameters α's, η's.

4.3 Strategic Behaviors

To analyze the nature of competition on the French pay-TV market, we confine our attention to three types of strategic behavior: Competitive, Stackelberg and Collusive. We thus estimate three models of noncollusive behavior. The first model assumes that the two firms play Nash and compete *à la* Bertrand (on prices). The second model considers the Stackelberg model with CanalSatellite as leader and TPS as follower. The third model is similar to the second one but with TPS as leader and CanalSatellite as follower. In addition, we estimate a model of full (tacit) collusion between the two firms.

4.3.1 Competitive Model

Here, we assume that firms compete on prices. That is, each firm simultaneously chooses its price given the price of the rival. In other words firm i $\max_{p_i} \pi_i(p_i, p_j)$. The strategy thus played by the firms leads to the competitive (Nash) equilibrium. We therefore impose the following restrictions to the model: $\upsilon_{ii} = \alpha_{ii}, i = 1, 2$ and $\upsilon_{12} = \upsilon_{21} = 0$[16].

4.3.2 Stackelberg Model

In this sequential game, first CanalSatellite chooses its strategy and then the follower, after observing the price of CanalSatellite, reacts. As usual, in this kind of games, the model is solved by backward induction. The first order condition of the follower is given as before. Firm 2 $\max_{p_2} \pi_2(p_1, p_2)$ which gives its reaction function to the leader price: $R_2 \equiv p_2 = \frac{1}{\alpha_{22}} c_2 - \frac{1}{\alpha_{22}} N_2$. The first order condition for the firm 2 ($FOC2$) is as before. On the other hand, the leader maximizes its profit which takes into account the reaction function of the follower $\max_{p_1} \pi_1(p_1, R_2(p_1))$

The first order condition for the leader ($FOC1$) is now different, and is given by $(p_1 - c_1)h_0 + h_0 p_1 + h_1 \gamma_0 + h_2 A_1^{1/2} + h_3 A_2^{1/2} + h_4 c_2 = 0$, or after solving for p_1:

$$p_1 = \frac{1}{2} c_1 - \frac{h_1}{2h_0} \gamma_0 - \frac{h_2}{2h_0} A_1^{1/2} - \frac{h_3}{2h_0} A_2^{1/2} - \frac{h_4}{2h_0} c_2 \tag{9}$$

where

$h_0 = \alpha_{11} - \frac{\alpha_{12}\alpha_{21}}{2\alpha_{22}}$

$h_1 = 1 - \frac{\alpha_{12}}{2\alpha_{22}}$

$h_2 = \gamma_{12} - \frac{\alpha_{12}\gamma_{21}}{2\alpha_{22}}$

$h_3 = \gamma_{22} - \frac{\alpha_{12}\gamma_{22}}{2\alpha_{22}}$

$h_4 = \frac{\alpha_{12}}{2\alpha_{22}}$

[16] The second order condition implies that the matrix $\begin{bmatrix} 2\alpha_{11} & \alpha_{12}\alpha_{21} \\ \alpha_{12}\alpha_{21} & 2\alpha_{22} \end{bmatrix}$ should be semi definite negative.

We thus have a system of four equations containing the two demand equations together with the two FOC's. Next, we consider the Stackelberg model with firm 2 as the leader.

4.3.3 Collusive Model

In the collusive model we assume that the cartel maximizes the joint profits of the firm $\max_{p_1,p_2} \frac{1}{2}\pi_1(p_1,p_2) + \frac{1}{2}\pi_2(p_1,p_2)$. Note that for simplicity, this cartel is assumed to give equal weight to firms. The profit maximization by the cartel yield the two following FOC's:

$$p_1 = \frac{1}{\alpha_{11}}c_1 - \frac{1}{\alpha_{11}}N_1 - \frac{\alpha_{21}}{\alpha_{11}}p_2 + \frac{\alpha_{21}}{\alpha_{11}}c_2 \tag{10}$$

$$p_2 = \frac{1}{\alpha_{22}}c_2 - \frac{1}{\alpha_{22}}N_2 - \frac{\alpha_{12}}{\alpha_{22}}p_1 + \frac{\alpha_{12}}{\alpha_{22}}c_1 \tag{11}$$

As before the two demand equations with these FOC's gives us a system of simultaneous equation that can be estimated using a linear, e.g. two-stage least squares (2SLS) or non-linear method, e.g. Full information maximum likelihood (FIML) or three-stage least squares (3SLS). Note that given the number of equations and exogenous variables one can easily show that the system is identified.

5 Econometric results

The four systems of structural equations, where some equations contain endogenous variables among the explanatory variables have been estimated by 3SLS. The main results are reported in Table A5 in the Appendix.

The trend variable is always highly statistically significant with a positive sign. This result indicates that the prices of both rivals is increasing over time, which can be explained by the fact that firms are in an early stage of their industrial development. Indeed, CanalSatellite started its business in 1992 and TPS was launched in 1997. This results confirms the patterns of the number of subscriptions shown in Figure A1 in the Appendix. The industrial production index is introduced as a control variable to capture the effects of business cycles. The estimated coefficients associated with this variable are never significant. Regarding seasonal dummies, only the third quarter is significant with a negative sign. Typically, this period corresponds to a lower economic activity combined with holidays and as a result the number of people subscribing to digital TV services is lower.

Interestingly, advertising variables are highly significant and have the expected sign across models and for both firms. That is, a firm's advertising effort always increases its own demand but decreases its rival's demand. More importantly, our main variable of interest, the prices of the subscription to digital programs is significant for CanalSatellite with the expected (negative) sign, that is an increase of CanalSatellite subscription fee has a negative effect on its demand of subscriptions. The subscription's price of TPS is not significant though it has the expected negative sign. This finding is not very surprising since this firm only recently entered this market and its prices have always been lower than the ones observed for CanalSatellite[17].

On the whole, the results reported in Table A5 show that the estimates of the different models are highly similar. The different goodness of fit's statistics (adjusted R-squared and F-stat) for the system of demand equations and first order conditions do not allow one to reject a model against a different one. As an alternative, the four alternative competitive behaviors have been tested against each other by applying tests for nonnested hypotheses proposed in Vuong (1989). These tests depart from the log-likelihood ratio statistic LR adjusted by a correction factor ξ_n that penalizes each model for complexity in terms of the number of estimated parameters and normalized by $\widehat{w}_n$.

Formally, the adjusted LR statistic is given by:

$$n_n^{-\frac{1}{2}}\widetilde{LR}_n\left(\widehat{\beta}_n, \widehat{\gamma}_n\right)/\widehat{w}_n \tag{12}$$

where

[17] A second explanation is a statistical one, the number of observations is rather low despite the fact that montly data are available.

$$\widehat{w}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{f\left(y_i/x_i;\widehat{\beta}\right)}{g\left(y_i/x_i;\widehat{\gamma}\right)}\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\log\frac{f\left(y_i/x_i;\widehat{\beta}\right)}{g\left(y_i/x_i;\widehat{\gamma}\right)}\right]^2} \tag{13}$$

$$\widetilde{LR}_n = LR_n\left(\widehat{\beta}_n, \widehat{\gamma}_n\right) - \xi_n \tag{14}$$

$$\xi_n = -\left[\left(\frac{p}{2}\right)\log n - \left(\frac{q}{2}\right)\log n\right] \tag{15}$$

p and q are the number of parameters in models f and g and n is the number of observations[18]. As shown by Vuong (1989), the test statistic is asymptotically standard normal under the null hypothesis, i.e. both models are equivalent.

The results of the Vuong tests are reported in Table A6 in the Appendix. It follows from this table that the Competitive Nash behavior is rejected in favour of the two Stackelberg models. The nonnested test for these two models reject the second model, i.e. the model where TPS is the leader and CanalSatellite the follower. As regards the tacit collusion, the test statistics is inconclusive as regards models 2 and 4, but at the 5% level, the test reject the collusive behavior in favor of the Stackelberg one where CanalSatellite is the leader and TPS the follower. Hence, both firms appear to behave according to a Stackelberg leader-follower setting, the leader being CanalSatellite and the follower TPS. Here again, this result is not unexpected given that TPS is a new entrant which reached its break-even point in October 2002[19].

CanalSatellite, on the other hand benefits from its longer presence in this market as an incumbent and from its reputation as well as from its parents' relationships.

[18]This correction factor for the number of estimated parameters is the one proposed in Schwartz (1978). Other penalty functions can be considered (Akaike, 1973 or Hannan and Quinn, 1979). These alternative correction factors have been estimated but do not change the conclusions as regards the models' selection process.

[19]Télé Satellite, number 155, p. 32. For CanalSatellite the break-event was reached in September 1999; source: Télé Satellite, number 118, p. 14.

6 A comparison with the British case

To have an idea about the consumers welfare one can compare the French competitive situation with a country characterized by a monopoly. Does such country exist? And if so, does a cross country static comparison make any sense? The answer to the first question is a qualified yes. Several countries in Europe has only a single digital satellite pay-TV operator. However, the answer to the second question is not that clear. The problem with a cross country comparison is that there are several factors that have an impact on the firms strategic behavior. Size for example is an important factor. A country with a large population size may benefit from higher return to scales. The benefit may be such that the consumers welfare is even larger than a small country with a high level of competition. The regulatory framework, e.g. competition policy, is another factor. Obviously the list of factors is long.

Nonetheless, we cannot resist in comparing the situation in France to that of the Great Britain. On several dimensions, the Great Britain is similar to France. But unlike France, the Great Britain is characterized by a single operator in the digital satellite pay-TV market.

The pay-TV industry in Britain started in 1989 when the two networks, Sky and British Satellite Broadcasting (BSB) were licensed.[20] The two companies used different satellites, and as a result, separate satellite dishes were required to receive both signals. Both companies tried to gain subscription by obtaining exclusive movie rights from Hollywood studios. In 1990 Sky merged with British Satellite Broadcasting, and the combined firm was named BskyB. BskyB had about 2.1, 3.1, and 3.4 million subscribers in 1994, 1996 and 1998, respectively.

A first comparison in Table 1 shows that the prices of basic packages are almost identical not only within a country (France) but also across countries. However, the prices of the premium packages are very different across firms. As expected the UK has the most expensive premium package. While it is true that the Sky's premium

[20] See Armstrong (1999) for more details on satellite pay-TV in Britain.

package contains a larger number of movie channel, but that fact alone cannot justify the price difference between Sky and CanalSatellite. Probably the quality of the premium programs offered by the Sky might be so high that can justify the price differential. In view of our analysis, we argue that the competitive force that characterize the French market can explain this difference[21].

Table 1. Comparison between France and UK

	BskyB (UK)		CS^a (FR)		TPS (FR)	
	$\min^b$	Max^c	$\min^d$	Max^e	$\min^f$	Max^g
Premium Movies[h]	0	12	0	11	0	7
Premium Kids[h]	0	4	0	9	0	3
Premium Sports[h]	0	4	0	7	0	7
Price (EUR)	17	54	17	35	17	33

Note : [a]CS is CanalSatellite; [b]Sky Entertainment Package; [c]Sky World; [d]CanalSatellite Thematiques; [e] Grand Spectacle; [f]TPS Théma; [g]TPS Maximal; [h]entries are the minimum/maximum number of channels in the beginning of 2003.

7 Concluding Remarks

The main objective of this paper is to provide an empirical assessment of the nature of competition between two dominant rival firms in the French satellite pay-TV industry, CanalSatellite, the incumbent, and TPS, the entrant. Structural econometric models are estimated using monthly data from 1997 to 2002 gathering information on the number of subscriptions, subscription fees, advertising effort, discounts, number of channels and content as well as other exogenous variables characterizing the demand. These models reflect different competitive behavior, i.e. Nash-Bertrand, Stackelberg and tacit collusion, and can be tested by means of tests for nonnested hypotheses. Such an exercise ask for a careful and modest interpretation of the main results obtained given a certain

[21] The UK may not be the best country for such a comparative exercise, mainly because of the exchange rates changes and the higher cost of living in this country. A first look at the Italian market, which is also characterized by a monoploy, Sky Italia, confirms our conclusions. In future work, it will be interesting to systemically extend the comparison with more countries and time periods.

number of assumptions on which they hinge. First, the pay-TV market in France has been assumed to form a separate one from the cable-TV market. Though the share of the latter is small as compared to the former, and both services are rather different in terms of their contents, such a hypothesis may appear as restrictive if consumers can substitute one service to the other. Second, given data constraints, it has not been possible to endogenize advertisement efforts. Such activities, however, are a main component of non-price competition. Third, the functional forms of the demand and the cost functions are rather basic. Here also, data availability constraints, in particular the number of observations, prevent investigating more general functional forms.

Despite these limitations, the main empirical findings indicate that the subscription price of satellite pay-TV in France did not decrease after the entry of TPS in this market. The substantial increase of programs suggests that competition is mainly driven by reactions in quality-adjusted quantity between the leader (CanalSatellite) and the entrant (TPS). Furthermore, advertising expenses of each firm have a positive and significant impact on its own demand, but a negative (predatory) effect on its rival. Finally the results indicate that an increase in the CanalSatellite price leads to a decrease in 'quantity' (i.e. the number of subscriptions) while for TPS an increase in price does not affect its quantity. We explain these findings by the history of both competitors. CanalSatellite started its business in 1992 in this industry and benefits from a reputation that allow this firm to maintain higher prices. On the other hand, TPS, through lower price and a continuous improvement of the quality and diversity of its programs, has been able to attract new subscribers and as a result to increase his market share. However, given its recent entry in this market, this firm has not yet reach its break-even point which prevent her to behave more aggressively by proposing lower prices.

In terms of welfare, a preliminary investigation of the French and the UK pay-TV markets, which is characterized by the presence of a single monopoly, clearly shows a price differential in favour of France, even after controlling for quality differences. This observation supports the idea that competitive pressures are higher in France. In our future research agenda, we will explore these results more systematically by collecting time series data on prices and quality for the British pay-TV firm and by implementing hedonic regressions to compare price differences in both markets. Another extension of the analysis is to take into account the vertical structure of this industry. As discussed before, CanalSatellite, through its parent network, appears to benefit from lower input prices given its vertical agreements with upstream movies suppliers. Here, the welfare effects are ambiguous, lower inputs' costs could be translated into lower subscriptions' fees given the existing level of competition. On the other hand, vertical restrictions, e.g. foreclosure, could lead to exclusionary practices, i.e. full access to high premium movies or sport programs denied to competitors. It would also be interesting to examine the existing regulations, e.g. The EU Audiovisual policy, with that respect and test how they interact with these practices.

References

[1] Akaike, H. (1983) "Information Theory and an Extension of the Likelihood Ratio Principle", in B.N. Petrov and F. Coaki (eds.), Proceedings of the Second International Symposium of Information Theory, Budapest: Akademiai Kiado, 257-281.

[2] Anderson, S.P. and S. Coate (2000) "Market Provision of Public Goods: The Case of Broadcasting", NBER Working Paper # 7513.

[3] Armstrong, M. (1999) "Competition in the Pay-TV Market", *Journal of the Japanese and International Economies*, 13, 257-280.

[4] Chipty, T. (2001) "Vertical Integration, Market Foreclosure, and Consumer Welfare in the Cable Television Industry", *American Economic Review,* 91(3), 428-53.

[5] Gabszewicz J., D. Laussel and N. Sonnac (2002) "TV Broadcasting Competition and Advertising", Working Paper, CORE, Belgium.

[6] Gasmi, F. and Q. Vuong (1991) "An Econometric Analysis of Some Duopolistic Games in Prices and Advertising", in G. Rhodes (ed.), Econometric methods and models for industrial organizations. Advances in Econometrics, vol. 9, Greenwich, Conn. and London: JAI Press, 225-54.

[7] Gasmi, F., J.J. Laffont and Q. Vuong (1992) "An Econometric Analysis of Collusive Behavior in a Soft-Drink Market", *Journal of Economics & Management Strategy*, 1(2), 277-311.

[8] Harbord, D. and M. Ottaviani (2001) "Contracts and Competition in the Pay-TV Market", Mimeo.

[9] Hannan, E.J. and B.G. Quinn (1979) "The Determination of the Order of an Autoregression", *Journal of the Royal Statistical Society*, Series B, 41, 190-195.

[10] MediaCabSat (2002) http://www.mediametrie.fr/

[11] Rubinovitz, R.N. (1993) "Market Power and Price Increases for Basic Cable Service since Deregulation", *Rand Journal of Economics*,*24*(1), 1-18

[12] Schwartz, G. (1978) "Estimating the Dimension of a Model", *Annals of Statistics*, 6, 461-464.

[13] Vuong, Q. (1989) "Likelihood Ratio Tests for Model Selection and Nonnested Hypotheses", *Econometrica*, 57(2), 307-335.

[14] Waterman, D. and A. Weiss (1996) "The Effects of Vertical Integration between Cable Television Systems and Pay-Cable Networks", *Journal of Econometrics*, 72(1-2), 357-395.

Appendix

Table A1. Market shares of CanalSatellite and TPS

	CanalSatellite Number of subscriptions (mio)	Market share	TPS Number of subscriptions (mio)	Market share	TOTAL Number of subscriptions (mio)
1993	0.1	100%			0.10
1994	0.21	100%			0.21
1995	0.3	100%			0.30
1996	0.45	100%			0.45
1997	0.78	70.9%	0.32	29.09%	1.12
1998	1.1	64.3%	0.61	35.67%	1.71
1999	1.37	62.3%	0.83	37.73%	2.20
2000	1.62	61.8%	1.00	38.17%	2.62

Source: AVICAM/CSA – Direction des études et de la prospective (Lettre du CSA, n°146-1, Décembre 2001, p.2)

Table A2. Content by category of programs of the basic bundle, CanalSatellite and TPS as of end of 2002

CanalSatellite		TPS	
Entertainment			
13e Rue	Movies	AB Moteurs	Motors
AB 1	Movies	Bandiagara	Games
Canal Festival Dauville	US film festival	BBC Prime	Entertainment
Canal Jimmy	Children	Escales	Tourism
CanalJ	Children	Festival	Movies
Canalsat Jeux	Games	France Supervision	Sport
Cartoon Network	Children	Game one	Games
Comédie	Humor	Gourmet TV	Culinar
Fashion TV	Fashion	Histoire	History
Forum Planète	Debates	knowledge TV	Education
Fox Kids	Children	KTO	Religion
Game one	Games	La Chaîne Sud	Family
Gourmet TV	Culinar	Motors TV	Motors
Match TV	People	Odyssée	Documentary
Monte Carlo TMC	Family	Paris Première	Art
Paris Première	Art	Série Club	Series
Planète	Documentary	Télétoon	Children
Planète Future (2)	Documentary	Télétoon +1	Children
Planète Thalassa	See	Téva	Women
Santé	Health	TF6	Entertainment
Téva	Women	TPS Hors Série	Events
Tiji	Children		
TV Festival	Cannes Festival		
Voyage	Tourism		

Table A2. Content by category of programs of the basic bundle, CanalSatellite and TPS as of end of 2002 (con't)

News			
BBC World	British news	Arte	Cultural
Bloomberg TV France	Finanical news	BBC World	British news
CNBC-NBC	US news	CNN International	US news
Cuisine.TV	Culinar	Fox News Channel	US news
Euronews	European new	France 2	Generalist
France 2	Generalist	France 3	Generalist
France 3	Generalist	I Télévision	News
Free One	Generalist	La Chaîne Parlementaire	Parlamentiary
Itélévision	News	La Cinquième	Education
LCI	News	LCI	News
RFO Sat	Regional news	M6	Generalist
RTL 9	Generalist	Régions	Regional news
RTPI	Generalist	RFO Sat	Regional news
TV Breizh	Regional	RTL 9	Generalist
Zaléa TV	Citizens	RTM	Generalist
		TF1	Generalist
		TFJ	Jewish
		TV Breizh	Regional
		TV5	Generalist
		TV7	Generalist
Music			
MCM		Fun TV	
MCM 2		M6 Music	
MCM Africa		Mezzo	
MTV		MTV	
Nostalgie la Télé		VH-1	
RFM la Télé			
Zik			
Interactive services			
Allociné TV	News	Annonces et Shopping	Shoping
Auto Moto	Motors	Astrologie (Equilibre)	Astrology
Bonjour.fr	Ads	Canal Auto	Motors
Canal Club	Shoping	Chronobourse	Financial
Canalsat Boutique	Shoping	Ciné info	News
Canalsat Finances	Financial	Club Téléachat	Shoping
Club Téléachat	Shoping	Espace annonces	Ads
CTV	Ads	Fi (CATV/Vega/Echos)	Financial
Demain!	Job	France courses	Hippism
Fiches Elle	Culinar	Info Express	News
Forum Boutiques	Shoping	Infoscore	Sport news
France courses	Hippism	Le journal de chez vous	News
I.Sport	Sport news	L'oeil du hibou	Cultural
Journal de chez vous	Local news	Météo Express	Weather
La chaîne météo	Weather	Mon Shopping	Shoping
Météo interactive	Weather	Pari à domicile	Hippism
Parents	Young parents	Shopping Avenue	Shoping
Pari à domicile	Hippism	TPS Boutique	Shoping
Première	Cinema	TPS Mail	E-mail
Spectacle	Shoping		
Zoé	Job		
Sport			
AB Sports		Equidia	
Equidia		ESPN Classic Sport	
ESPN Classic Sport		Eurosport Breizh	
Eurosport France		Eurosport France	
L'équipe TV		Infosport	
Pathé Sport		Superfoot	
Sport+			

Table A3. Net subscription price (euros), CanalSatellite (CS) and TPS, 1996-2002

December of	TPS	CanalSatellite
1996	237.82	228.67
1997	256.11	246.36
1998	272.88	272.58
1999	283.56	301.85
2000	300.02	311.00
2001	300.02	311.00
2002	300.02	334.78

Source: TéléSatellite, own calculations.

Table A4. Descriptive statistics of main variables

Variable		Mean	Std. error	Min	Max
net annual subscription price (euros)	p1	288.13	29.12	228.67	334.78
	p2	269.38	37.05	158.55	300.02
# of subscribers	q1	1212789	508735	201200	1998400
	q2	726392	361634	0	1172000
advertising	a1	1069	268	447	1414
	a2	810	268	0	1083
total # of channels	nt1	73.68	19.80	29	108
	nt2	49.36	10.30	16	59
total # of channels in the basic bundle	nb1	37.83	14.00	15	56
	nb2	40.00	9.53	21	52
total # of channels in the movie bundle	nc1	6.53	1.51	5	11
	nc2	3.76	0.78	3	5
total # of bundles	o1	9.25	1.00	5	11
	o2	7.49	2.12	2	11
industrial production index (OECD)	y	112.59	4.72	101.30	118.30

Notes:

72 observations

1: CanalSatellite

2: TPS

Source: TéléSatellite, own calculations.

Table A5. ¨Parameter estimates (3SLS) of different strategic behaviours

	Model 1 Nash		Model 2 Leader TPS		Model 3 Leader CanalSat		Model 4 Tacit collusion	
	Q TPS	Q CS	Q TPS	Q CS	Q TPS	Q CS	Q TPS	Q CS
Constant	-205	-399	-208	-429	-197	-427	-190	-414
	(1.16)	(2.20)**	(1.18)	(2.35)**	(1.11)	(2.35)**	(1.07)	(2.27)**
Qt-1	0.555	0.182	0.550	0.186	0.549	0.171	0.556	0.182
	(7.80)***	(3.05)***	(7.72)***	(3.12)***	(7.69)***	(2.90)***	(7.79)***	(3.06)***
P TPS	-18.087	14.162	-16.309	9.192	-16.560	15.670	-18.644	13.273
	(1.15)	(0.81)	(1.03)	(0.52)	(1.05)	(0.90)	(1.18)	(0.76)
P CS	18.773	-63.245	18.532	-60.606	-6.159	-95.005	11.163	-69.272
	(0.67)	(2.01)**	(0.66)	(1.92)*	(0.22)	(3.04)***	(0.40)	(2.20)**
A TPS	515	-270	518	-281	515	-291	517	-267
	(5.65)***	(2.94)***	(5.68)***	(3.04)***	(5.62)***	(3.17)***	(5.65)***	(2.90)***
A CS	-182	1,214	-182	1,212	-186	1,228	-183	1,205
	(2.68)***	(12.25)***	(2.68)***	(12.14)***	(2.73)***	(12.43)***	(2.69)***	(12.13)***
S2	-9,339	-7,726	-9,411	-7,100	-10,198	-10,003	-9,082	-7,514
	(1.40)	(1.01)	(1.41)	(0.92)	(1.52)	(1.31)	(1.36)	(0.98)
S3	-15,645	-20,637	-15,568	-21,388	-16,424	-22,367	-15,657	-20,046
	(2.31)**	(2.66)***	(2.30)**	(2.74)***	(2.42)**	(2.89)***	(2.31)**	(2.58)**
S4	-288	1,785	-420	1,336	-298	1,319	-13.265	2,311
	(0.04)	(0.22)	(0.06)	(0.17)	(0.04)	(0.17)	(0.00)	(0.29)
Y	1,779	1,156	1,785	1,510	2,109	1,926	1,762	1,441
	(0.94)	(0.58)	(0.94)	(0.75)	(1.11)	(0.96)	(0.93)	(0.72)
T	3,045	7,580	3,082	7,580	3,297	7,953	3,087	7,647
	(4.29)***	(10.23)***	(4.34)***	(10.19)***	(4.63)***	(10.81)***	(4.34)***	(10.33)***
Observations	69	69	69	69	69	69	69	69
R^2a	0.9278	0.8948	0.9281	0.8947	0.9274	0.8964	0.9285	0.8972
F stat	96.53	66.22	96.51	66.38	96.23	66.12	96.21	65.87

Notes:

- ∞ Absolute value of t statistics in parentheses.
- ∞ * Significant at 10%; ** significant at 5%; *** significant at 1%.
- ∞ The estimates of each model's first order conditions are not reported.
- ∞ Model 1: Nash competition; model 2: Stackelberg model with CS as leader; Model 3: Stackelberg model with TPS as leader; Model 4: Tacit collusion on prices.
- ∞ Advertising expenditures are assumed to be exogenous.
- ∞ Q is number of subscriptions; P is subscription's fee (price), A is firm's advertising expenditures; S2-S4: Quarterly dummies; Y = industrial production index; T: trend.

Table A6. Vuong nonnested tests for alternative competition models

	2.	3.	4.
1.	3.98	5.12	2.57
2.		2.11	-.27
3.			-2.44

Notes:

1: Nash competition; 2: Leader TPS; 3: Leader CanaSatellite; 4: Tacit collusion.

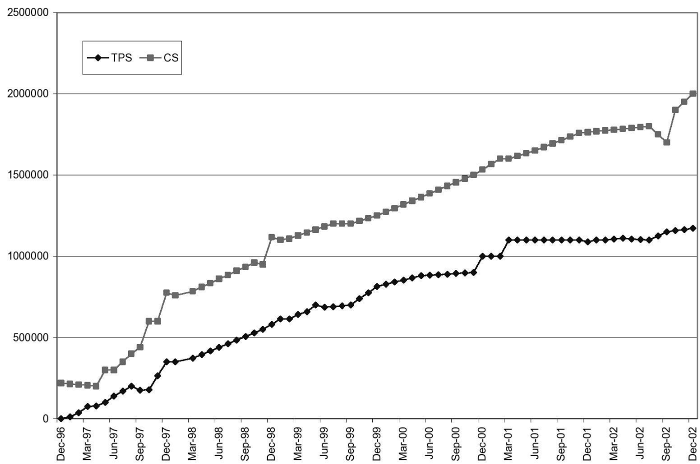
Values for the Vuong test (column j vs. line i) above 2 reject the model in line (i), and below –2 favor the model in column (j) at the 1% level.

Figure A1. The French satellite pay-TV market

Upstream market	**Movie and other Programming production** Outside production in-house / independent — Domestic production in-house / independent
	⇩
Intermediate market	**Movie channels = intermediate firms** The main players in this market are, on the one hand, firms fully owned by TPS such as Cinéstar 1, Cinéstar 2, Cinétoile, Cinefaz and, on the other hand, firms like Ciné Cinéma 1, Ciné Cinéma 2, Ciné Cinéma 3, and Ciné Classic owned by Multithématique, which is a property of Groupe Canal + (27%), Liberty media (27%), Lagardere (27%), Havas Image (9%) and Groupe Tectis (10%)[a].
	⇩
Downstream market	**Distributors** CanalSatellite — TPS — ABsat ⇩ ⇩ ⇩ Final subscribers

Note : a) Extract from CNC Mai 2001, Liens financiers.

Figure A2. Number of subscription CanalSatellite (CS) and TPS, 1997-2002



Source: TéléSatellite, own calculations.

Commercial or Open Source Software? Winner-Take-All Competition, Partial Adoption and Efficiency

Eric DARMON ♣ Thomas LE TEXIER ♣♦

Dominique TORRE ♣

♣ *GREDEG - CNRS - University of Nice Sophia-Antipolis*
250, bd. Albert Einstein, F-06560 Valbonne Cedex
E-mail: {darmon ; letexier ; torre}@idefi.cnrs.fr

♦ *CREA – Centre of Research of the French Airforce Academy*
CREA 10.401- BA 701, F-13661 Salon Air
E-mail: tletexier@cr-ea.net

Third Draft – Please do not quote without authors' permission
March 2006

Abstract

In this paper, we study the conditions ruling the diffusion of an Open Source software as opposed to a commercial one. The two software differ according to their usage costs (adoption costs, existence of customized functionalities). We distinguish two categories of users, namely "end-users" and "users-developers", according to their ability to contribute to the development of the OS software. We characterize the Nash equilibria in a sequential game where players are first the producer of the commercial software which chooses a price and a quality level, and second users who chose between adopting the OS, the commercial software or not adopting. Due to the externalities of use and development, the final level of adoption of the commercial software and of the OS solution depend on initial conditions relative to users expectations. Starting from identical initial conditions (adoption costs, magnitude of the externalities generated by the community of developers and users, etc.), a winner-takes-all competition may arise between the two software, which may lead to the crowding out of one of them. Other cases exhibit separating equilibria where users are distributed between the two software according to their aptitude to develop or to adopt. Such multiple equilibria being only imperfectly controlled by the commercial firm, we show that the strategy of the commercial firm can be understood as a balance between a low price – high quality strategy and a high price – low quality strategy. We analyse the qualitative properties of pooling and separating equilibria. The sole existence of a credible OS alternative improves the utility of end-users, even if only the commercial solution is finally adopted. The diffusion of the OS software may sometimes generate conflicts of interests. In some cases, the division is not only between the users and the firm but the interests of some users can also be aligned to that of the firm.

Key words: Open Source – Commercial Software – Increasing returns – Community of developers
***JEL* Codes :** D3-D4-L1

1 Introduction

Unlike its proprietary counterparts, open source (OS) software is based on particular license terms which allow any user to freely "*use, modify and redistribute*"[1] the original software[2]. From an economic point of view, one key issue is to find out whether this original production and distribution mode is viable in the long run and could establish itself as a credible alternative to proprietary software. Empirical studies reveal that the success of OS software is very diverse: in the last few years, the OS model has rapidly spread in technical areas (cf. Apache, Sendmail, PHP) where most users are expert (*e.g.* developers, system administrators, webmasters). In other end-users oriented areas, the OS model shows ambiguous performances. For example, Linux has started to establish itself as a competitive alternative to proprietary operating systems, while LATEX (an open source word processor) has not managed to expand to a large community of users although it was introduce before other proprietary solutions (cf. Gaudeul [2003a]). In all these examples, we can identify several disparate causal factors. Some of them are structural (type of users, existence and quality of a proprietary counterpart). Others are more historical (unsuited leadership of the open source software, lock-in effects). All these factors partially explain why open source software may succeed in some cases and fail in others. One could integrate these factors into two inter-related topics: *i*) *production* of the OS software *i.e.* developers' individual incentives to participate in OS communities and *ii*) *diffusion* of the OS software *i.e.* users' individual incentive to adopt open source software.

On the developers' side, contributors' motivational aspects have been widely analyzed by both economists and sociologists. Indeed, according to Lerner and Tirole [2001], the apparent "non-profit" and altruistic character of the contributions has to be downplayed and partially replaced by more selfish objectives (career concerns, prestige as evidenced by the hierarchical structure of communities). Following these authors, Mustonen and Leppämäki [2003] present a situation in which programmers' involvement in open source projects is considered as a signal for their potential employers and the development of the open source project can either harm the proprietary software (substitute products) or benefit from it (complementary products). This signal hypothesis has been empirically tested and corroborated in the case of Apache developers (cf. Hann *et al.* [2002]). Furthermore, Dalle and Jullien [2003] emphasize the crucial role of institutions (as e.g. the *copyleft* license) to provide a credible commitment able to preserve such incentives. Indeed, developers' incentives are highly dependent on the type of license under which the OS is released (cf. Gaudeul [2003b] for a theoretical analysis of developers' and project leaders' incentives under the GPL, BSD and proprietary licenses). Incentives are also sensitive to the way the OS project is managed. In particular, the type of leadership (cf. Gaudeul (2003a)) and the organization of the software (modular versus centralized, see e.g. Hertel et al. (2003)) should be stressed.

On the users' side, the adoption issue has also attracted attention. Dalle and Jullien [2003] develop a simulation model: assuming that all potential adopters initially use the proprietary software and that network externalities are both local and global, they show that the diffusion of the 'libre' software goes through a phase of transition associated with a lock-in effect. As a consequence, they point out the crucial role of early adopters on the success or the failure of the open source project. However, their analysis is based on the premise that, everything being equal (*i.e.* when the two software have the same share of the global and local markets), the quality of the OS software is better than that of the proprietary one. In a close framework, Bonaccorsi and Rossi [2003] add the possibility for users either to adopt the two kinds of software simultaneously or separately. While the network externality is strictly global, they differentiate users according to the benefit they derive from the adoption of the open source software. The intrinsic utility attached to the open source software is probabilistic and several distributions are tested (normal, uniform, exponential biased beliefs against the open source software, exponential biased beliefs against the proprietary software). With 'moderate' network effects, they conclude that both type of software coexist with all forms of distribution (except for the exponential distribution biased against the open-source software, which blocks the diffusion of open source). However, the coexistence of the two technologies may be partly caused by the fact that agents can maintain the two types of software simultaneously with no additional cost. This feature does not hold in the long run for software that are not

[1] See Cremer and Gaudeul [(2004)] and Von Krogh and Von Hippel [2003] for a general presentation of open source.

[2] GPL (*General Public License*) and BSD (*Boston Software Development*) are the most famous open-source licenses but it can be shown that there exists a continuum of license ranging from pure open source software to proprietary one (see West [2003], Muselli [2004] and Zimmermann and Jullien [2002] for an analysis of such mixed developement strategies and Lerner and Tirole [2002] for an empirical survey). GPL and BSD licenses differ on the following point: any software which draws on GPL software, has to preserve the original licence terms, while BSD license terms allow a developer to freely change the type of the license, e.g. transforming it into a proprietary or a GPL license.

perfectly compatible[3]. In such a dynamic context, Valimaki and Oksanen [2005] underline the role of switching costs, that may lead to the non-adoption of the OS software in spite of its possibly better quality compared to the proprietary one. In a static context, Mustonen [2003] studies the interaction between open source and proprietary software through the allocation of developers' time: developers can choose to allocate their time either by being hired by the proprietary software company (direct revenue) or by participating in an OS project and thus acquiring experience and signalling their quality to future employers (indirect revenue). Depending on the magnitude of the open source adoption cost incurred by users, the commercial firm can crowd out the diffusion of the open-source software by proposing higher wages (which discourages developers to participate in the open-source community) or by increasing the quality of its software (vertical differentiation). However, this model assumes that developers are not users and consequently that developers' participation is only motivated by signalling motives. As it should be stressed, another important motive to participate in OS community is to develop a software customized to the developers' own needs (innovation-led users, cf. e.g. Bitzer and Schröder [2005]). Bitzer [2004] thus underlines the role of the products' heterogeneity in the adoption of the OS software, as the more similar these products are, the more intense the price competition between the two software is.

Tackling the adoption issue, it seems important to stress the heterogeneity of users' needs. Users are first heterogeneous regarding their preferences *i.e.* two users may not use a software to perform the same task. To capture this element, we develop a model which is able to take into account two motives for OS adoption: i) difference in license cost when a user can substitute a proprietary software to an OS software and ii) adoption of OS software as the proprietary software does not provide specialized functionalities. Along with this heterogeneity in preferences, users are also different according to their contribution to the open source software. We should at least consider two distinct populations: *developers-users* (cf. Franke and Von Hippel [2003]) and *end-users*. End-users can orient the future development (*e.g.* by asking for new functionalities) but are not able to contribute directly to the development effort. Developers-users may use the software for their own needs (or similarly for the needs of their company) but are able to write new lines of code and customize the software as they need new functionalities. The relative size of these two populations obviously depends on the type of software: highly technical software may primarily concern users-developers while operating systems or office suites may concern a larger set of final-users. In turn, this distinction affects the economic incentives that drive each population to adopt the OS software or not. In other terms, the presence or the absence of one type of users alters the benefits (development of customized functionalities, compatibility externality) derived from the adoption of the software. In this paper, we consider the adoption issue by explicitly taking into account the coexistence of heterogeneous users (*i.e.* users with different programming abilities and differentiated preferences). As a twin issue, we propose to evaluate the welfare generated by the introduction of OS software. This last question seems particularly relevant considering the recent initiatives to promote open source adoption[4]. Indeed, individual and collective welfare may diverge in the presence of network externalities which are typical of software goods (cf. Shapiro and Varian [1999])[5]. To be fully appropriate, such policies then need a comprehensive framework able to account for such externalities at a more macroscopic level. Schmidt and Schnitzer [2003] survey some of these policy-related topics. With a simple model (horizontal product differentiation and three populations of users, a 'mobile' population and two captive ones), they show that increasing the share of OS captive users can harm the welfare of other populations. This model shares several features with our model, though we do not suppose here that some users are exogenously captive to one kind of software and we add the important distinction between end-users and developer-users.

This paper is organized as follows: Section *2* presents the model. Section *3* analyzes the Nash equilibria it generates and discusses its qualitative properties. Section 4 concludes.

[3] For example, while switching from the proprietary suite *Microsoft Office* to the GPL *Open Office*, a user may lose functionalities (e.g. macro-commands). It may then be more costly for him to maintain two versions of the same file with the same quality than to adopt only one of the two software (either open-source or proprietary).

[4] See, e.g. China's plan to equip its central administration or Taiwan's initiative to promote the development of Open Source programs.

[5] See Schmidt and Schnitzer [2003] for a survey of all policy-related topics. With a simple model (horizontal product differentiation and three populations of users, a 'mobile' population and two captive ones), they show that increasing the share of OS captive users can harm the welfare of other populations.

2 The Model

The model analyses the interactions between a firm producing the commercial software and two categories of users, namely "end-users" and "users-developers", according to their ability to contribute to the development of the OS software. The firm chooses and implements a strategy in prices and quality at a first step of the game, before the decisions of adoption and development of users are taken and implemented at a second step of the game.

2.1 End-Users and Developers

There are two categories of users: *developers* (indexed by *D*, in proportion μ of the total population) and *end-users* (indexed by *EU,* in proportion $1-\mu$ of the total population). The mass of the total population is normalized to 1. End-users and developers differ according to two elements: first, developers are able to develop new functionalities and customize an OS software by adding new lines of codes, while end-users are not. In other words, as the source code is open, a developer is able to develop new functionalities for his own benefit. Once these functionalities are developed, there is no cost in releasing them to the whole community[6]. Besides, an end-user may freely benefit from the experiences and releases of developers. Secondly, end-users and developers do not face the same adoption costs. Developers are generally expert users (engineers, computer scientists) and are used to adopting new software. On the contrary, end-users face high initial costs while adopting a new software. Besides, it seems reasonable to assume that these costs are generally higher when adopting an OS software than those incurred when adopting a commercial one. One reason for this is that OS software are primarily developed and OS are led by expert users who cannot devote much time to make extensive tutorials/FAQs, numerous user-friendly interfaces, etc[7].

For that reasons, let us suppose that developers do not face any adoption cost while end-users do. Even inside the population of end-users, some agents are more expert than others. That is why we also assume that the adoption cost is heterogeneous among this category of agents End-users are heterogeneous in this respect and face an adoption cost uniformly distributed between 0 and $\overline{c}^{EU}$.

2.2 Commercial and Open Source software

The two software need to fulfil two types of functionalities *i.e.* generic and specialized functionalities. To be eligible, a software needs to develop all generic functionalities. Users (both end-users and developers) are heterogeneous regarding the second component. We will thus suppose that they are uniformly distributed on a unit circle: each agent is located on a point and needs the functionality related to that point. Users may choose to adopt *i)* a proprietary (closed source) software; *ii)* a GPL-open source software; or *iii)* neither of the two software. In the last case, we can suppose that users can manage to do the same task either "manually" or with a combination of previously acquired software (*i.e.* without costs). By convention, this default option leads to a null payoff.

The strategy of the commercial firm

We suppose that a single commercial firm produces the proprietary software. Because the software shares many characteristics with digital goods, its reproduction costs are negligible and we will thus further assume them to be null (and independent from the number of adopters). That is why this development cost can be analyzed as a fixed cost. Conversely, the extent of the specific needs covered by the proprietary software is

[6] We restrict our analysis to GPL (General Public License) licenses. Other types of licenses (as e.g. BSD) allow any user or developer to commercialize its own software derived from the initial software. Consequently, this other type of license raises an additional incentive problem for developers which may be inclined to "close" the code of their customized software instead of making it freely available to other users. This specific case could be the subject of an extension of the present work.

[7] It could be argued that users may incur an adoption cost even when adopting a commercial software. However, because user-friendliness is an important factor for adoption and purchase of a software, commercial firms devote important marketing efforts to maximize the usability of their software. To clearly delineate OS from commercial software, we can thus overlook the adoption costs for the proprietary software.

controlled by the firm. By devoting more or less important research effort on R&D, it can choose to develop more or less specialized functionalities in order to cover a more important fraction of these specialized competencies on the unit circle (noted q and called "quality"). As a consequence, the production costs of the proprietary software are fixed according to the number of users (digital goods, no reproduction cost), but variable relatively to the extent of the provided functionalities q. This is one of the control variables for the firm. Because the firm needs to invest in R&D to develop these functionalities, we will thus assume that there is a cost $g(q)$ (with $g'(q)>0$ and $g(0)=0$) to provide them.

Let us hereby denote by q $(q \in [0,1])$, the extent of the functionalities offered by the firm, defined as a proportion of the unit circle covered by the selected software quality, and by c_L ($c_L > 0$), the license fee paid to the firm by each user. The commercial firm chooses a couple (q^*, c_L^*) maximizing its profit given by Equation (1) :

$$\pi = n^P c_L - g(q) \tag{1}$$

where n^P $(n^P \in [0,1])$ denotes the proportion of users of the proprietary software

The utility generated by the proprietary software

The ability of the commercial software (and of the OS software too) to fulfil the generic functionalities strongly depends on the total number of adopters (see e.g. Shapiro and Varian [1999]). This network externality is motivated by several factors, such as the compatibility between users, the availability of complementary services and software, *etc.* To take it into account, we will suppose that the utility associated to such functionalities is closely dependent on the number of its users. Consequently, we assume that the utility generated by the general needs $f_1(n^P)$ positively depends on the adoption level expressed by n^P (with $f_1'(.)>0$, $f_1(0)=0$ and $f_1(1)=k_1,(k_1>0)$.

As the source code of the proprietary software is closed, no user can develop his own specialized functionalities. The utility generated by the commercial software then depends on whether the commercial firm has developed the desired specific functionality. Users whose functionalities are covered by the commercial software get an additional utility equal to d (with $0 < d$), 0 otherwise. Because software is an experience good, a user cannot fully anticipate at adoption time, which specialized functionality she will need in the future, and assess its *ex post* utility. Consequently, she can *ex ante* only consider the extent of specificity covered by the commercial software and try to anticipate the future utility of the commercial software. By convenience, we will also suppose that users are risk-neutral. As a consequence, the expected utility of potential users relatively to the specific uses of proprietary software is the weighted average of their respective *ex post* utilities according to their location, inside or outside the set of the functionalities effectively covered by the commercial software. Let d corresponds to the utility generated by the specific uses of proprietary software. Once considered that the firm develops a proportion q of all the specialized functionalities and if agents are risk-neutral, the expected utility derived from the specialized functionalities is thus equal to dq.

Once deduced the license cost c_L charged by the firm, the net utility derived from the adoption of the proprietary software can be expressed as follows (equation 2):

$$u_i^P = f_1(n^P) + dq - c_L \tag{2}$$

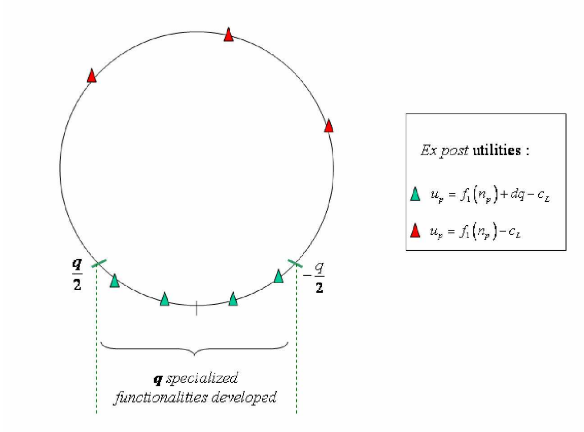


Fig. 1: The utility of the Commercial Software

The utility generated by the OS software

When adopting the OS software, there is no license fee. Similar to the commercial software, we will suppose that the OS software needs to fill all the generic functionalities to be eligible[8]. As previously, these generic functionalities are subject to a compatibility externality (again noted f_1). Besides, developers may, in the case of OS software, freely modify and customize the initial software. The modular organization of most recent OS software favours such customization so that open source developers may add functionalities that will then benefit all users. As the OS software is maintained and managed by developers, the utility derived from specialized functionalities is highly dependent on the number of contributors (noted n_D^{OS})[9]. In other terms, the larger the number of developers who contribute to the code is, the more functionalities are added, and the higher the utility of the OS software is for all users (*i.e.* for both developers and end users). This positive link will be described by a positive externality (noted f_2) that benefits the whole population (users and developers). We will capture this influence of the size of developers' population on the OS individual utility by a positive externality noted $f_2\left(n_D^{OS}\right)$ with n_D^{OS} denoting the size of the community of developers ($f_2'(.) > 0$, $f_1(0) = 0$ and $f_2(0) = 0, f_2(1) = k_2, (k_2 > 0)$.

The adoption cost c_i^{EU} $\left(0 \le c_i^{EU} \le \overline{c}^{EU}\right)$ of any user $i\left(i \in [0,1]\right)$ is depicted by Equation (3). Developers and end-users are ordered according to the size of their adoption costs. Consequently, as $i \in [0,\mu]$, we will conventionally set that user i is a developer and her adoption cost is null. As $i \in [\mu,1]$, user i is an end-user and is defined in such a way that $c_i^{EU} = \left(\frac{i-\mu}{1-\mu}\right)\overline{c}^{EU}$ (cf. figure 2).

[8] We may suppose that these functionalities are *ex ante* developed by "kernel" developers or by the project manager, so that the OS exists, once all these functionalities are fully available (release of the "1.0 version").
[9] We will assume that once a developer adopts the open source software, she actively contributes to the code and does not act as an end-user. This simplification implies that n_D^{OS} is equal to the size of the population of developers who adopt the OS.

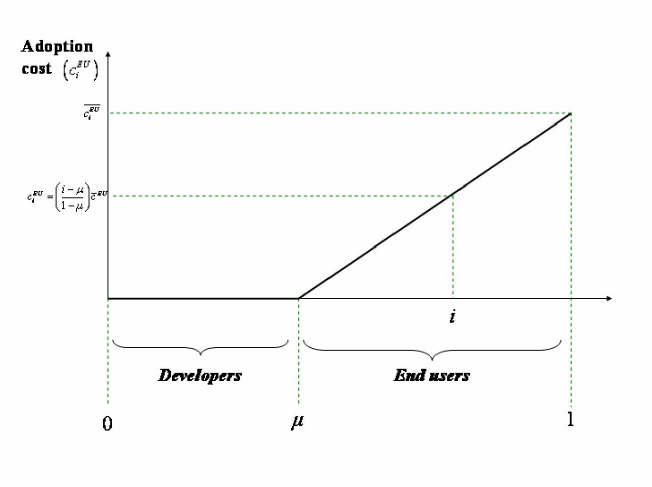


Fig. 2: Adoption costs of the OS Software

Thus, we express the utility generated by the OS software for agent i ($i \in [0,1]$) as follows:

$$u_i^{OS} = f_1\left(n^{OS}\right) + df_2\left(n_D^{OS}\right) - c_i^A \text{ with } c_i^A = \begin{cases} 0 \text{ when } i \in [0,\mu] \\ c_i^{EU} \text{ when } i \in [\mu,1] \end{cases} \qquad (3)$$

2.3 The Sequential Game

The firm has complete but imperfect information about the composition and the possible strategies of the potential adopters of its software: it knows the different characteristics of its potential users $(\mu, \alpha_1, \alpha_3, \ldots)$ but can imperfectly anticipate their behaviour. Since we can suppose that the price and the quality of a software cannot be frequently altered, we assume that firms' and users' choices are sequential: the firm makes its choices initially and users reply subsequently to firms' strategy by formulating their own adoption choice. The structure of the sequential game is depicted by Figure 2:

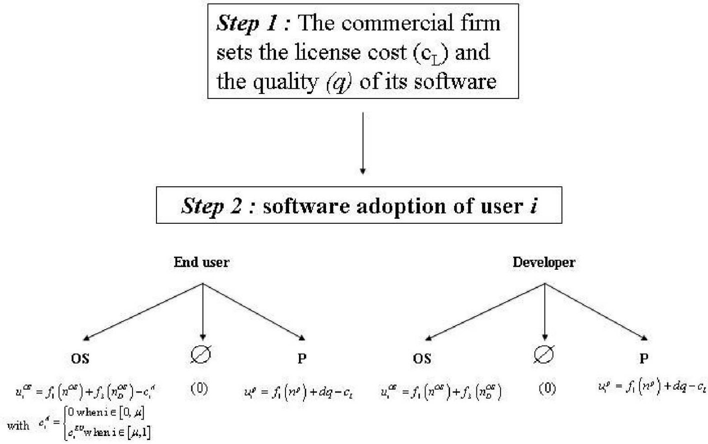


Fig. 3: Structure of the sequential game (OS = Open Source Software adoption ; P = Proprietary Software adoption ; ∅ = No adoption)

At *Step 1*, the firm chooses the extent of functionalities q^* and the level of its license fees c_L^* that maximize its expected profit. At *Step 2*, each user observes these two variables, compares its respective utilities and chooses between the proprietary software (Strategy "*P*"), the OS software (Strategy "*OS*") or no adoption (Strategy "∅"). Individual utilities of users depend at this stage on each user's expected level of adoption of each software. As usual in such cases and in order to avoid complexities in the description of the expectation dynamics, we suppose that these expectations (be it wrong or right) are the same for all users. An eductive process generates a convergence of wrong expectations towards the locally stable equilibrium of the area of stability to which they belong.

3 Nash Equilibria of the Sequential Game

We solve the game by backward induction: We first consider the second sub-game (adoption decisions of developers and end-users for a given strategy of the firm). We then determine the profit-maximizing strategy of the firm

3.1 The Second Step

3.1.1 General case

Let us thus consider a couple (q, c_L). Once we note the expectations n^P, n^{OS}, n_D^{OS} relative to the respective sizes of the commercial software customers population, the total population of OS end-users and the population of OS developers, the choice of any user $(i,\ i \in [0,1])$ is as follows:

Choosing P if $f_1(n^P)+dq-c_L \geq \sup\{0, f_1(n^{OS})+df_2(n_D^{OS})-c_i^A\}$

Choosing OS if $f_1(n^{OS})+df_2(n_D^{OS})-c_i^A \geq \sup\{0, f_1(n^P)+dq-c_L\}$

Choosing $\varnothing$ if $0 > \sup\{f_1(n^P)+dq-c_L, f_1(n^{OS})+df_2(n_D^{OS})-c_i^A\}$

Each agent associates her own optimal choices to each expected level of n^P, n^{OS} and n_D^{OS}. Let us conventionally assume that, if they expect the level n^{OS} for the OS users, they will consistently suppose that this population is only formed by developers if $n^{OS} \leq \mu$. If $n^{OS} > \mu$, this population also integrates in supplement the end-users whose adoption costs are the lowest[10]. From the aggregation of these optimal decisions, the effective proportions of proprietary software users, of developers and end-users of the OS software (n^P, n^{OS} and n_D^{OS} respectively) are then deduced as function of the expected levels of each populations (see appendix 1). Hence, for each couple (q, c_L), the condition $(n^P, n^{OS}, n_D^{OS}) = (n^P, n^{OS}, n_D^{OS})$ defines the Nash equilibria of the sub-game corresponding to *Step 2* of the game.

Without any specification of the functions f_1, f_2 and g, the following propositions can be established in the general case:

Proposition 1: Whatever the pair (q, c_L) selected by the firm at the Step 1 of the game, there exists at least one Nash equilibrium associated with Step 2 sub-game. *(Proof, see Annex 1)*

Without any specifications on the functional forms of $f_1(.), f_2(.)$ and of the development costs $g(.)$, we have no precision about the nature of the equilibrium position(s) of the second step sub-game. However, such precisions can be obtained with a slightly weak assumption of the economic "viability" of the firm.

A0. There exists an economically viable pair (q^+, c_L^+), *i.e.* a quality / price solution for the commercial software such that i) for this pair (q^+, c_L^+), its exists an expected size level n^{P+}, $(0 \leq n^{P+} \leq 1)$ of the commercial software users (P-users) sufficient induce users to adopt the commercial solution when there is no alternative to this technology and ii) $\pi = c_L^+ - \beta(q^+)^2 \geq 0$.

Proposition 2: Under A0, the distribution of users $\{n^P{*} = 1, n_D^{OS}{*} = 0, n^{OS}{*} = 0\}$ corresponding to the full adoption of the commercial software is always a Nash equilibrium of Step 2 sub-game. *(Proof, see Appendix 1)*

Proposition 2 establishes that under a very weak assumption, there *always* exists at least one equilibrium of the second step sub-game such that the open source software is crowded out. At this equilibrium, all users adopt the commercial software. The limitation introduced by the assumption is that there exists at least one pair "quality / price" of the commercial software which insures that its implementation is i) profitable for the firm and ii) utility-improving (against inaction) for users, at least in the most favourable circumstance, that is when all agents expect a general adoption of the commercial software.

Other equilibria may exist. Consider for instance the (limit) case where $c_i^A = 0, \forall i$. In this case there exists a range of variation of q^+ and c_L^+ (q^+ sufficiently low and c_L^+ sufficiently high) such that $f_1(1)+dq^+-c_L^+ \geq 0 \geq f_1(0)+dq^+-c_L^+$. We then verify that $f_1(1)+df_2(\mu)-c_i^A \geq \sup\{0, f_1(0)+dq-c_L\}$ whatever i. The outcome $\{n^P{**} = 0, n_D^{OS}{**} = \mu, n^{OS}{**} = 1\}$ is then a second equilibrium of the second sub-game. By continuity and by the fact that OS is for all users strictly preferred to P when $\{n^P = 0, n_D^{OS} = \mu, n^{OS} = 1\}$, the same result is maintained when $c_i^A = \varepsilon_i\,(\forall i)$ with ε_i as close as possible to 0 (yet $\neq 0$). Then, the equilibrium

[10] This kind of assumption is frequently used in adoption models (see for instance Cremer [2000]).

$\{n^P{**}=0, n_D^{OS}{**}=\mu, n^{OS}{**}=1\}$ exists for a not-empty range of variation of $c_i^A\,(\forall i)$ .However, as $c_i^A\,(\forall i)$ is far from 0, the result is not maintained while other equilibria may as well emerge. The nature of the equilibria (when they exist) other than $\{n^P{*}=1, n_D^{OS}{*}=0, n^{OS}{*}=0\}$ then depends on the forms of $f_1(.), f_2(.)$ and $g(.)$.

Hence, the multiplicity of the second step sub-game Nash equilibria seems to be the norm in the general case. An exhaustive topology of these equilibria is however only possible after the forms of $f_1(.), f_2(.)$ and $g(.)$ are specified. We will select the linear-quadratic case as an illustration.

3.1.2 The linear-quadratic specification

Suppose the functions $f_1(.)$ and $f_2(.)$ respectively capturing the compatibility and development externalities to be linear. This linearity can be considered as a choice to favour the median occurrence between the contrasting cases of exponential network effects on the one hand and of congestion effects on the other hand. Also assume also that $g(.)$ is quadratic. This last assumption implies that it is more and more costly for the firm to develop the specific functionalities of the software. Given these specifications, profit and utilities can be re-expressed as follows:

$$\pi = n^P c_L - \beta q^2 \tag{1'}$$

$$u_i^P = \alpha_1 n^P + dq - c_L \tag{2'}$$

$$u_i^{OS} = \alpha_1 n^{OS} + \alpha_2 d\,\frac{n_D^{OS}}{\mu} - c_i^A \text{ with } c_i^A = \begin{cases} 0 \text{ when } i \in [0,\mu] \text{ (developers)} \\ c_i^{EU} = \frac{(i-\mu)}{(1-\mu)}\overline{c}^{EU} \text{ when } i \in [\mu,1] \text{ (end users)} \end{cases} \tag{3'}$$

To improve the tractability of the linear-quadratic case, moreover introduce two additional (and mild) assumptions:

A1. Suppose that $\overline{c}^{EU} \le (\alpha_1 + d\alpha_2)$. This condition provides an upper bound to the adoption costs. Put differently, we suppose that the end-user exhibiting the largest adoption cost $\left(\overline{c}^{EU}\right)$ always obtains a non-negative utility when she adopts the OS software, given that all other users are expected to adopt the OS software. This assumption is a sort of counterpart of the assumption that the commercial software is economically viable. It rules out possible inaction and equilibria such that $\{n^P, n_D^{OS}, n^{OS}\} = \{0,0,0\}$

A2. We need an assumption relative to the initial expectations of agents. It could be considered as an extension of the assumption of symmetry to the case where all agents are heterogeneous. As usually in this kind of model, we first suppose that - right or wrong - expectations are the same for all agents but we also exclude total inconsistency between expectations and choices. More precisely, as agents are ordered according to their cost of adoption of the OS software, we suppose *that agent i cannot simultaneously adopt OS and expect that a proportion of agents less than i will adopt it*: agents are aware of their position in the "implicit hierarchy" of users and integrate this information in their plans by excluding some expectations inconsistent with this hierarchy.

Given this last assumption, we can exhaustively analyze the nature and stability of equilibria in the linear-quadratic specification of the model.

Then consider the second step sub-game under such specifications and assumptions. This step begins while the firm has already fixed c_L and q at the first stage of the game. The remaining variables to be determined by users are $n^P{*}$ (the Nash equilibrium proportion of agents adopting the proprietary software), $n^{OS}{*}$ (the Nash equilibrium proportion of agents adopting the OS software) and $n_D^{OS}{*}$ (the Nash equilibrium proportion of developers among agents adopting the OS software). According to our specifications of $f_1(.), f_2(.)$ and $g(.)$, given assumptions A0,A1 and A2, the typology of Nash equilibria of the second step sub-game can be exhaustively analyzed considering the possible values of parameters α_1, α_2, $\overline{c}^{EU}$, μ and d ,and the first step determined variables q and c_L. These cases are presented graphically in *appendix 1* and summarized in the following table:

	Number of equilibria	Number of stable equilibria	Type of stable equilibria	Conditions of prevalence
Case 1	3	2	commercial software only [P] OS software only [OS]	$\alpha_1(1-\mu) > \overline{c}^{EU}$, $dq - c_L < \alpha_1(2\mu-1) + d\alpha_2$
Case 4	3	2	commercial software only [P] OS software only [OS]	$\alpha_1(1-\mu) < \overline{c}^{EU}$, $dq - c_L < \alpha_1(2\mu-1) + d\alpha_2$ $dq - c_L < \alpha_1 + d\alpha_2 - \overline{c}^{EU}$
Case 2	3	2	commercial software only [P] OS software only [OS]	$\alpha_1(1-\mu) > \overline{c}^{EU}$ $dq - c_L > \alpha_1(2\mu-1) + d\alpha_2$ $dq - c_L < \alpha_1 + d\alpha_2 - \overline{c}^{EU}$
Case 5	3	2	commercial software only [P] OS software only [OS]	$\alpha_1(1-\mu) < \overline{c}^{EU}$ $dq - c_L > \alpha_1(2\mu-1) + d\alpha_2$ $dq - c_L < \alpha_1 + d\alpha_2 - \overline{c}^{EU}$
Case 3	1	1	commercial software only [P]	$\alpha_1(1-\mu) > \overline{c}^{EU}$ $dq - c_L > \alpha_1 + d\alpha_2 - \overline{c}^{EU}$
Case 6	1	1	commercial software only [P]	$\alpha_1(1-\mu) < \overline{c}^{EU}$ $dq - c_L > \alpha_1(2\mu-1) + d\alpha_2$ $dq - c_L > \alpha_1 + d\alpha_2 - \overline{c}^{EU}$
Case 7	3	2	commercial software only [P] coexistence of the two software [OS-P]	$\alpha_1(1-\mu) < \overline{c}^{EU}$, $dq - c_L < \alpha_1(2\mu-1) + d\alpha_2$ $dq - c_L > \alpha_1 + d\alpha_2 - \overline{c}^{EU}$

In some cases (cases 1, 4, 2 and 5), there are three Nash Equilibria and only two are stable. In these cases, either the commercial or the OS software is adopted at equilibrium but the two software never simultaneously coexist at the end of eductive process. This case exhibits a "winner-takes-all" situation. The area of attraction of each "pooling" equilibria (see appendix 1) is defined by the place of the unstable "separating" equilibrium (E3): when initial agents' expectations about the adoption of the OS software are below a critical level, the proprietary software will finally be adopted by all at the end of the eductive process. On the contrary, when expectations are beyond this critical level, the OS software will finally prevail. Other cases correspond to one single stable "pooling" equilibrium (cases 3 and 6) where either the commercial software or the OS software is adopted. The last case (case 7) corresponds to two stable equilibria, one pooling and corresponding to a full adoption of the commercial software and the other separating where one part of the population stabily adopts the commercial software and the other the OS software.

These equilibria stand for one given (q, c_L) strategy of the firm. Solving backward, we then need to establish the equilibria in the first sub-game.

3.2 The first step

The first step of the game corresponds to the choice of the pair (q, c_L) maximising the expected profit of the firm, given the possible choices of users when they consider the selected levels of these variables. We consider this step in the general case, then with the linear-quadratic specification.

3.2.1 The general case

At this stage, the firm has to choose the couple (q, c_L) maximizing its objective function. As the second sub-game equilibrium is generally not unique, the firm has to integrate in this objective function each possible stable outcome associated with the second stage of the game ensuing from the pair (q, c_L). The occurrence of these stable equilibria and their corresponding payoffs are also to be considered. The behaviour of the firm towards risk can at last influence the form of its objective function. By simplification, we suppose that the firm is risk-neutral and only considers the expected profit generated by the different outcomes of the second step sub-game as answers to its choices of step 1[11]. Suppose that there exists a maximum of n equilibria corresponding to the profits $\pi_1(q, c_L), \pi_2(q, c_L), \ldots \pi_n(q, c_L)$ and that the occurrence of each of them is given by the probability distribution $p_1(q, c_L), p_2(q, c_L), \ldots p_n(q, c_L)$ with $p_1(.,.) + p_2(.,.) + \ldots p_n(.,.) = 1$. In this case, the expected profit to be maximized can be expressed as:

$$\pi^e = p_1(q, c_L)\pi_1(q, c_L) + p_2(q, c_L)\pi_2(q, c_L) + \ldots + p_n(q, c_L)\pi_n(q, c_L)$$

The probability distribution $p_1, p_2, \ldots p_n$ may have diverse origins (subjective or objective probabilities, associated or not with previous observations of the distribution of users' expectations and beliefs…). We consider at this stage the case where the firm has a complete knowledge of the possible outcomes and the stability areas of each of them, but no information about the possible consistent expectations that users may have at the beginning of step 2 (according to A2). We then suppose that the expected level of OS-adoption n^{OS} is uniformly distributed on $[0,1]$. Given A2, the size of the respective stability areas of each equilibria then provides the required distribution of probability generating the expected profit of the firm. Given the above assumptions, we are able to prove the following general result:

Proposition 3: An optimal outcome always exists for the firm at the first step of the game *(proof: see Appendix 2)*

Corollary: An equilibrium of the two-step game always exists. When this equilibrium is not unique, equilibria correspond (except for cases of vanishing measure), to a single level of functionalities and price (or license cost) of the commercial software but to different levels of adoption of the commercial software and the OS software

3.2.2 A numerical analysis of the linear-quadratic specification

Proposition 3 holds as well in the linear-quadratic case. However, as seen in section 3.1.2., things are simpler in this case much simple, since there exists only one or two stable equilibria. The number of equilibria of the second step sub-game however depends on the pair (q, c_L) chosen by the firm. For instance, cases 1, 2 and 3 can be generated by the same parameters α_1, α_2, $\overline{c}^{EU}$, μ and d ,and by different values of the variables q and c_L (see *Appendix 1*). However, cases 1 and 2 correspond to 2 stable pooling equilibria while only equilibrium of full adoption of the commercial software corresponds to case 3, as depicted in Figure 4.

[11] We can make this assumption since the probability for an unstable equilibrium to occur is of null mass.

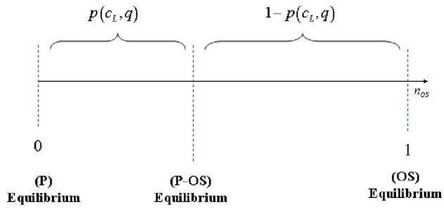


Fig. 4: Case 1: zone of attraction of the two stable equilibria

In this context, given the parameters generating cases 1, 2 and 3, the choice of the firm will be between *i)* a (relatively) low level of quality and a (relatively) high level of prices, and *ii)* a (relatively) high level of quality and a (relatively) low level of prices. The first case will provide high profits at the "good" equilibrium from the firm's point of view but the occurrence of this equilibrium is not guaranteed. The second case will provide relatively low profits at the "good" equilibrium but this equilibrium is guaranteed (the OS software is crowded out). Define as π_1 the profit corresponding to the "good" equilibrium for the firm (*i.e.* the equilibrium with a full adoption of the commercial software) and π_2 the profit corresponding to the "bad" equilibrium for the firm (*i.e.* the equilibrium with a full adoption of the commercial software). The expected profit of the firm is then given by the following expression:

$$\pi^e = p(q,c_L)(n^P c_L - \beta q^2) + [1 - p(q,c_L)](-\beta q^2) = p(q,c_L)(n^P c_L) - \beta q^2$$

where $n^P c_L - \beta q^2$ figures the profit associated with the "good" equilibrium and $-\beta q^2$ the loss incurred with the bad equilibrium (costs must be supported even if the commercial software is not adopted). $p(q,c_L)$ is deduced from the size of the stability area of the "good" equilibrium, with $p(q,c_L) = 1$ when there exists only 1 (stable) equilibrium.

In a first analysis, we can illustrate the previous arguments by some numerical examples. Let us consider for instance a typical case of "Case 1-2-3" $\left(\alpha_1 = 1, d = 1, \alpha_2 = 1, \beta = 0.2, \mu = 0.2, \overline{c}^{EU} = 0.1\right)$, when the compatibility externality (depicted in the linear case by Coefficient α_1) outweighs the adoption cost incurred by the marginal user. In this case, let us bear in mind that the final outcome is always a winner-takes-all one, ending in the crowding out of one of the two software. We can then plot the expected profit of the firm for different (q, c_L) strategies[12].

[12] For each strategy, we consider the relative positions of the U^P and U^{OS} curves. From this comparison, we deduce the case (1-2-3) and compute the expected profit and the expected diffusion of the OS project accordingly. If $U^P < 0\left(\forall i \in [0,1]\right)$, no user is willing to adopt the commercial software. Knowing that, the firm does not make any investment on quality. Hence the firm's profit is null in this case.

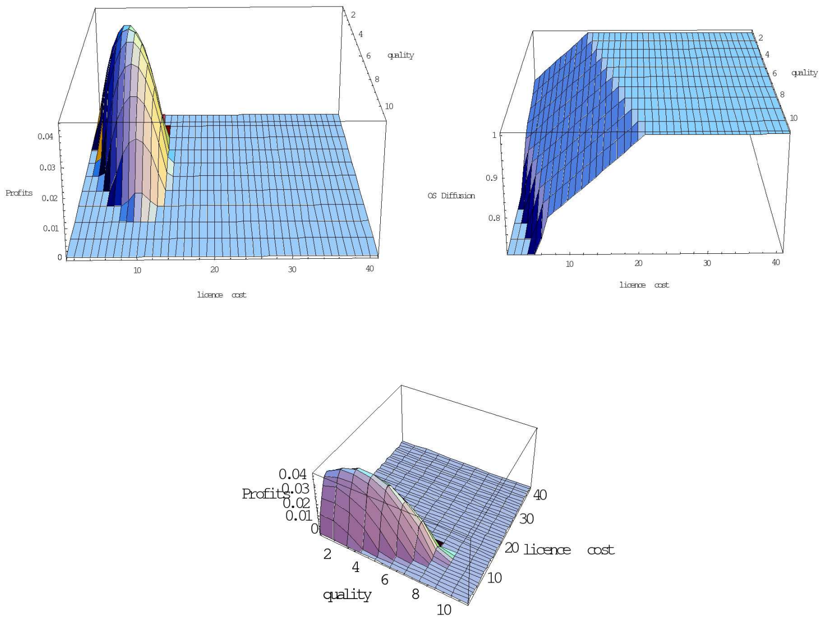


Fig. 5: Firm's expected profit (left) and OS expected diffusion (right) as a function of the quality and license cost charged by the firm $\left(\alpha_1 = 1, d = 1, \alpha_2 = 1, \beta = 0.2, \mu = 0.2, \overline{c}^{EU} = 0.1\right)$; $\pi^{\max} = 0.043 > 0$ reached for $\left(c_L^* = 0.6 \text{ and } q^* = 0.2\right)$.

As one can see in the previous figure, the maximum expected profit $\left(\pi^{\max}\right)$ is strictly positive and reached for an interior value of the strategy set $\left(c_L^* = 0.6 \text{ and } q^* = 0.2\right)$. The right figure depicts the expected diffusion of the OS software *i.e.* here the probability of the OS software to diffuse to the whole population. It shows how the commercial firm can indirectly (and partially) control the diffusion of the OS software by setting different strategies (price and quality). When it chooses a "low price – high quality" strategy, it can crowd out the diffusion of the OS software (expected OS diffusion close to 0). However, we see in this example that the firm has no incentive to do so due to the development cost. Instead, the *expected* diffusion of the OS software corresponding to firm's optimal choice is equal to 0.91. The firm faces a trade-off: either it selects a strategy that crowds out the OS project by setting a high quality and/or low license cost, or it has the "opposite" strategy (low quality and/or high price). In the last case, it obtains a maximal profit when the "P" equilibrium (no OS adoption) is selected but faces a risk of massive adoption of the OS project, being then itself crowded out. In the first case, this risk vanishes but the profit is realized when the "P" equilibrium decreases. In the case presented, the optimal behaviour of the firm has to be comprehended as a balance between these two strategies. We could also mitigate this result by taking into account the firm's risk aversion. However, one could bet that such a change would not change our results qualitatively. Intuitively, a more risk-adverse firm would likely minimize the risk of being crowded out so that the optimal strategy would be more conservative (quality increase and/or price decrease).

Let us now consider the case of a more "specific" software *i.e.* a software where specialized functionalities are more prevalent in users' utilities ($d = 5$ instead of 1 in the previous case):

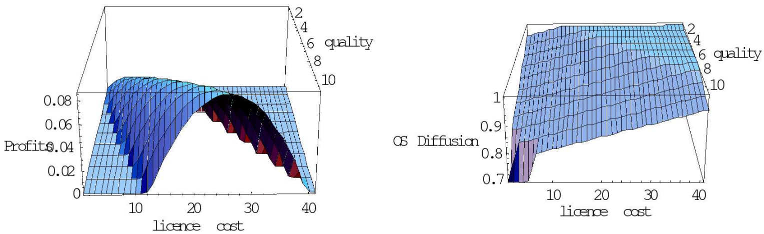


Fig. 6: Firm's expected profit (left) and OS expected diffusion (right) as a function of the quality and license cost charged by the firm $\left(\alpha_1 = 1, d = 5, \alpha_2 = 1, \beta = 0.2, \mu = 0.2, \overline{c}^{FU} = 0.1\right)$; $\pi^{max} = 0.08 > 0$ reached for $\left(c_L^* = 2.5 \text{ and } q^* = 1\right)$.

We can see that the profit-maximizing strategy is here to develop all specialized functionalities $\left(q^* = 1\right)$ all other parameters being constant. Hence, the firm has an incentive to invest in quality in order to differentiate its software from the OS software. In this case, we can notice that the expected diffusion of the OS software is equal to 0.88. Again the firm's optimal strategy is not to crowd out the OS software.

Let us consider a third case characterized by opposite software features: generic functionalities are now prevalent in agents' decisions ($d = 0.1$).

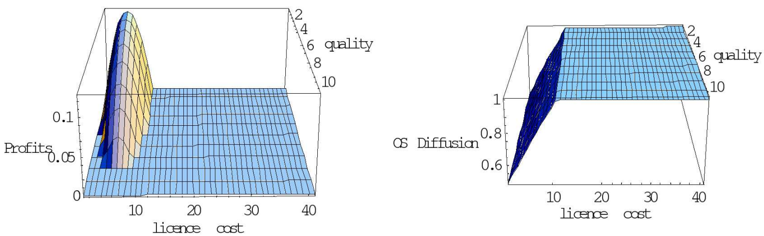


Fig. 7: Firm's expected profit (left) and OS expected diffusion (right) as a function of the quality and license cost charged by the firm $\left(\alpha_1 = 1, d = 0.1, \alpha_2 = 1, \beta = 0.2, \mu = 0.2, \overline{c}^{FU} = 0.1\right)$; $\pi^{\max} = 0.124 > 0$ reached for $\left(c_L^* = 0.4 \text{ and } q^* = 1\right)$.

The firm does not to develop any specific functionality $\left(q = 0\right)$. Since users give more importance to specific functionalities, the optimal strategy of the firm is to focus on generic functionalities and not to invest in specialized ones. Again, one interesting feature is that the firm has again no interest in crowding out the OS software.

3.3 Market failures and the welfare properties of the introduction of the OS software

The multiple equilibria of the second step sub-game may generate market failures that we first analyze. We then provide some elements to evaluate the welfare gains or prejudice resulting from the existence and "competitive pressure" generated by the OS software.

When the optimal quality / price pair (q^*, c_L^*) involves the occurrence of multiple equilibria, market failure emerges. Consider for instance in the linear-quadratic specification, the case where (q^*, c_L^*) generates two stable pooling equilibria, $\{n^P* = 1, n_D^{OS}* = 0, n^{OS}* = 0\}$ and $\{n^P** = 0, n_D^{OS}** = \mu, n^{OS}** = 1\}$. The comparison of these pooling equilibria provides the following result in the linear-quadratic case:

Proposition 4: Under the linear-quadratic specification and when the optimal price / quality pair (q^*, c_L^*) is associated with two pooling stable equilibria at the second step sub-game,
i) if there is a coordination failure, the high level equilibrium is always the 'all-commercial' one. All things equal, this case corresponds to low development externalities.
ii) if there is conflict of interest, the conflict is between developers and part of or all end-users on the one hand and the firm, with - or without - some other end-users on the other hand. All things equal, this case corresponds to high development externalities. *(proof: see Appendix 3)*

The first part of the proposition points out cases where the low level of expectations relative to the adoption of the commercial software may finally involve an "all OS" solution, which is not desirable even for developers. The second part of the proposition establishes that when externalities of development are rather high, the "all OS" solution is the more desirable for developers and part of end-users, while the other part of end-users and the firm would have preferred, if possible, the "all commercial" equilibrium.

These results may be complemented by an equivalent proposition in the case where one of the stable solutions is a separating equilibrium:

Proposition 5: Under the linear-quadratic specification and when the optimal price / quality pair (q^*, c_L^*) is associated with one 'all-commercial' and one separating equilibrium at the second step sub-game,
i) if there is coordination failure, the high level equilibrium is always the 'all-commercial' one. All things equal, this case corresponds to low development externalities.
ii) if there is conflict of interest, the conflict is between developers and part of all end-users on the one hand, and the firm and the other end-users on the other hand. All things equal, this case corresponds to high development externalities. *(proof: see Appendix 3)*

Proposition 5 confirms proposition 4. Due to the existence of externalities in each technology and in opposition to the 'general result', the separating equilibrium is not efficient, even if the firm is not considered in the analysis of welfare. Propositions 4 and 5 tend to disqualify the OS technology since it is possible to derive some result asserting the equilibrium with the more OS than possible is always the best for the economy. This observation however contrasts with the following result:

Proposition 6: Suppose that a (unique or not) pooling 'all commercial' equilibrium is associated with the optimal price / quality pair (q^*, c_L^*) in a setting in which agents can choose between the commercial software and the OS solution. Then this equilibrium is preferred by all users to the 'all commercial' equilibrium existing the same setting without any possibility to choose the OS software. *(proof: see Appendix 3)*

Proposition 6 counterbalances propositions 4 and 5. Even if the OS solution is not actually selected and adopted, the existence of a 'potential competitor' for the commercial solution provides incentives for the producer to lowering prices or improving quality of the commercial technology. The outcome is always positive for all users. Propositions 4 and 5 compare the properties of the two possible equilibria when the OS is already present. Put together, they show that the magnitude of the development externality is crucial to determine the agents that benefit or are harmed when a particular equilibrium is selected. Interestingly, they show that, when a conflict of interest arises, in some cases, this conflict is not only between the firm and the whole population of users. Instead, it establishes that the interests of some users (those with high adoption cost) are associated to that of the firm.

4 Discussion and further research

This paper depicts the competition and coordination issues arising from the existence of OS projects: in a framework where externalities favour lock-in and winner-takes-all situations, the competition between OS and commercial software have been analyzed using a two-step game: in a first step, the firm sets the quality and the price of its software. In a second step, end-users and developers react by adopting the commercial software or engaging into the OS project. We have shown that such a situation often leads to multiple equilibria. Because the commercial firm can influence but cannot entirely select the most favourable equilibrium, it faces a trade-off: either it adopts a "low price – high quality" strategy and deters the OS project; or it adopts a "high price – low quality" strategy and bears the risk of being crowded out by the OS software. We show that the strategy of the commercial firm can be understood as a balance between these two basic strategies. Depending on the preferences of the users, we have pointed out some cases where *i)* the firm has no interest to invest in quality or *ii)* on the contrary, it has an incentive to develop all functionalities to maximize its profit.

These first findings can be discussed in several directions. First, we need to go further into the numerical exploration of different scenarii. One issue is how the efficiency of the organization of the OS project (captured by α_2 in our model) impacts on its potential diffusion. This impact can be analyzed within our framework, by studying the reaction of the firm to alternative degrees of efficiency. Secondly, we could extend the current analysis by considering an additional step during which the firm could revise its price-quality strategy. This could be done using a repeated game. However, such extension would raise many additional problems since many choices are not easily reversible (adoption of users, quality of the firm).

Further, one could consider alternative strategies concerning the quality policy of the commercial firm. For example, Lerner and Tirole [2001] argue that, in letting their own developers implicated in OS projects, firms can understand competition better, by innovating as well as by using OS ideology to improve their production as detecting skillful developers. Dahlander and Magnusson [2005] stress that the firms may directly benefit from the OS communities by managing their activities, as the knowledge required to develop a software is not only controlled by the firm but also by the surrounding communities. As the addition of a specific functionality may sometimes be too costly, Krishnamurthy [2003] points out that firms may be encouraged to support OS projects to lower their production costs. Such arguments leave room for a further analysis of new forms of cooperation between OS and commercial software.

References

BITZER J. [2004], Commercial versus open source software: the role of product heterogeneity in competition, *Economic Systems*, 28:4, 369-381.

BITZER J. and SCHRÖDER P. J. H. [2005], Bug-fixing and code-writing: The private provision of open source software, *Information Economics and Policy*, 17:3, 389-406.

BONACCORSI A. and ROSSI C. [2003], Why Open Source Software Can Succeed, *Research Policy*, 32:7, 1243-1258.

CREMER J. [2000], Network externalities and universal service obligation in the internet, *European Economic Review*, 44:4-6, 1021-1031.

CREMER J. and GAUDEUL A. [(2004)], Some economics of the open-source software, *Réseaux*, (forthcoming),

DAHLANDER L. and MAGNUSSON M. G. [2005], Relationships between open source software companies and communities: Observations from Nordic firms, *Research Policy*, 34:4, 481-493.

DALLE J.-M. and JULLIEN N. [2003], 'Libre' Software: Turning Fads into Institutions?, *Research Policy*, 32:1, 1-11.

FRANKE N. and VON HIPPEL E. [2003], Satisfying Heterogeneous User Needs via Innovation Toolkits: The Case of Apache Security Software, *Research Policy*, 32:7, 1199-1215.

GAUDEUL A. [2003a], The (LA)TEX project: a case study of open source software, *GREMAQ Working Paper*.

GAUDEUL A. [2003b], Open Source Software Development Patterns and License Terms, *GREMAQ WP*.

HANN I.-H., ROPERS J., SLAUGHTER S. and FIEDING R. [2002], Delayed Returns to Open Source Participation: An Empirical Analysis Of the Apache HTTP Server Project, *GREMAQ-IDEI 2002 Conference on Software and Media Industry*,

KRISHNAMURTHY S. [2003], A managerial overview of open source software, *Business Horizons*:sept.-oct. 2003, 47-56.
LERNER J. and TIROLE J. [2001], The Open Source Movement: Key Research Questions, *European Economic Review*, 45:4-6, 819-826.
LERNER J. and TIROLE J. [2002], The Scope of Open Source Licensing, *NBER Working Paper*:w9363.
MUSELLI L. [2004], La licence informatique: un outil stratégique puissant pour les éditeurs de logiciels ?, *Second Workshop of the Research Consortium "ICT & Society"*, january 2004 28-29th.
MUSTONEN M. [2003], Copyleft--The Economics of Linux and Other Open Source Software, *Information Economics and Policy*, 15:1, 99-121.
MUSTONEN M. and LEPPÄMÄKI M. [2003], Spence Revisited - Signalling with Externality: The Case of Open Source Programming, *5th Berlecon Conference on the Economics of I.T.*, June, 13-14th.
SCHMIDT K. and SCHNITZER M. [2003], Public Subsidies for Open Source ? Some Economic Policy Issues of the Software Market, *Harvard Journal of Law and Technology*, 16:2,
SHAPIRO C. and VARIAN H. R. [1999], *Information rules: A strategic guide to the network economy*, x, 352.
VALIMAKI M. and OKSANEN V. [2005], The impact of free and open source licensing on operating system software markets, *Telematics and Informatics*, 22:1-2, 97-110.
VON KROGH G. and VON HIPPEL E. [2003], Special Issue on Open Source Software Development: Editorial, *Research Policy*, 32:7, 1149-1157.
WEST J. [2003], How Open Is Open Enough? Melding Proprietary and Open Source Platform Strategies, *Research Policy*, 32:7, 1259-1285.
ZIMMERMANN J.-B. and JULLIEN N. [2002], Le logiciel libre : une nouvelle approche de la propriété intellectuelle, *Revue d'Economie Industrielle*, 99:Special Issue on "Intellectual Property Rights",

Appendix 1: Nash equilibria of the Step 2 sub-game

Proof of Proposition 1: Given that $n_D^{OS} = \inf \; n^{OS}, \mu$, consider the choice made by agent i when she observes the pair (q, c_L) at the second step of the game:

$$\text{Choosing } P \text{ if } f_1\left(n^P\right) + dq - c_L \geq \sup\left\{0, f_1\left(n^{OS}\right) + df_2 \;\; \inf\left(n^{OS}, \mu\right) \;\; - c_i^A \;\right\}$$

$$\text{Choosing } OS \text{ if } f_1\left(n^{OS}\right) + df_2 \;\; \inf\left(n^{OS}, \mu\right) \;\; - c_i^A \;\; \geq \sup\left\{0, f_1\left(n^P\right) + dq - c_L \;\right\}$$

$$\text{Choosing } \varnothing \text{ if } 0 \geq \sup\left\{f_1\left(n^P\right) + dq - c_L, f_1\left(n^{OS}\right) + df_2 \;\; \inf\left(n^{OS}, \mu\right) \;\; - c_i^A \;\right\}$$

Let $n^P = t^P\left(n^P, n^{OS}\right)$, $n^{OS} = t^{OS}\left(n^P, n^{OS}\right)$ and $n^{\varnothing} = t^{\varnothing}\left(n^P, n^{OS}\right)$ defined as

$$n^P = \{i\} / f_1\left(n^P\right) + dq - c_L \geq \sup\left\{0, f_1\left(n^{OS}\right) + df_2 \;\; \inf\left(n^{OS}, \mu\right) \;\; - c_i^A \;\right\}$$

$$n^{OS} = \{i\} / f_1\left(n^{OS}\right) + df_2 \;\; \inf\left(n^{OS}, \mu\right) \;\; - c_i^A \;\; \geq \sup\left\{0, f_1\left(n^P\right) + dq - c_L \;\right\} \text{ and}$$

$$n^{\varnothing} = \{i\} / 0 \geq \sup\left\{f_1\left(n^P\right) + dq - c_L, f_1\left(n^{OS}\right) + df_2 \;\; \inf\left(n^{OS}, \mu\right) \;\; - c_i^A \;\right\}$$

By construction, $n^P + n^{OS} + n^{\varnothing} = 1$. It follows that $t^{\varnothing}(.)$ is not independent but derived from $t^P(.)$ and $t^{OS}(.)$. Let $t = \left\{t^P, t^{OS}\right\}$ the transformation defined from $([0,1],[0,1])$ to $([0,1],[0,1])$ with t^P and t^{OS} such that $n^P = t^P\left(n^P, n^{OS}\right)$ and $n^{OS} = t^{OS}\left(n^P, n^{OS}\right)$. The transformation $t = \left\{t^P, t^{OS}\right\}$ then exhaustively describes the choices of users when they expectations are given by n^P and n^{OS}. The utility functions $u_i^{OS}\left(n^P, n^{OS}\right)$, $u_i^P\left(n^P, n^{OS}\right)$ and $u_i^{\varnothing}\left(n^P, n^{OS}\right)$ are continuous from their arguments and then $t = \left\{t^P, t^{OS}\right\}$ is also continuous. The transformation being defined from a compact set on itself, it admits at least a fixed point on this compact set $([0,1],[0,1])$. This (these) fixed point(s) are Nash equilibrium (equilibria) of the second step sub-game. Suppose that $\left\{n^P{*}, n^{OS}{*}\right\}$ is a (one of the) fixed point(s) of this transformation. The size of the population of developers adopting the open source software $n_D^{OS}*$ can then be deduced elementarily from the relation of definition $n_D^{OS}* = \inf\ n^{OS}*, \mu$

Proof of Proposition 2: Let (q, c_L) an economically viable pair and and $n^{P+}, (0 \le n^{P+} \le 1)$ and definition, $f_1(n^{P+}) + dq^+ - c_L^+ \ge 0$. Now consider the expected distribution of agents $\{n^P, n_D^{OS}, n^{OS}\} = \{1,0,0\}$. Individual utilities generated by this distribution are $u^P = f_1(1) + dq^+ - c_L^+ \ge f_1(n^{P+}) + dq^+ - c_L^+ \ge 0$ for the commercial software and $u_i^{OS} = f_1(0) + df_2(0) - c_A^i \le 0, \forall i$ for the *OS* software. Whatever the agent *i* considered, the inequality $f_1(n^{P+}) + dq^+ - c_L^+ \ge \sup\{0, f_1(0) + df_2(0) - c_A^i\}$ holds and $\{n^P, n_D^{OS}, n^{OS}\} = \{1,0,0\}$, proving that $\{n^P*, n_D^{OS}*, n^{OS}*\} = \{1,0,0\}$ is a Nash equilibrium and demonstrating *Proposition 2.*

Derivation of the Second Step Sub-Game Nash Equilibria: Given the linearization of functions $f_1(.), f_2(.)$ and the specification of the cost of use c_i^A, the possible outcomes of the second step sub-game can be described exhaustively, according to the values of parameters α_1, α_2, $\overline{c}^{EU}$, μ and d, and the first step determined variables q and c_L. For each set of values of these parameters and variables, we simultaneously use two related figures which capture the role of expectations on the individual utilities associated with the three terms of the choice (adopting the commercial software, the OS software or choosing reservation).

0n the upper chart, we display the utility of the i^{th} agent respectively as P-user and OS-user while agents $j \in [0, i]$ adopt the OS software and agents $k \in [i, 1]$ after her adopt the commercial one (we designate agent *i* as the "marginal agent"). Intersections determine interior solutions, *i.e.*, equilibria such that one part of agents adopts OS and the other one the commercial software. The choice of the commercial software by all agents is always a solution when u_i^P, *i.e.* the utility of the marginal agent as P-user is non-negative when $i = 0$. The choice of the OS by all agents is also a solution when the utility u_i^{os} of the marginal agent when, as $i = 1$, is such that $u_i^{os} \ge u_i^P, 0$.

First consider case 1 defined by the conditions $\alpha_1(1-\mu) > \overline{c}^{EU}$, $dq - c_L > -\alpha_1$ and $dq - c_L < \alpha_1(2\mu - 1) + d\alpha_2$. The equation of the curve providing the utility of the marginal agent if she chooses the commercial software is $(\alpha_1(1-i) + dq - c_L)$. The slope of the u_i^P-curve of this marginal agent represented on *figure 1a* does not depend on the type of agent (developer or end-user) examined but only on the level of externalities generated by the commercial software, *i.e.*, given A2, on the rank of the marginal agent. This explains the linearity of u_i^P. The equation corresponding to the utility generated by the OS software for the marginal agent is $\alpha_1 i + \dfrac{\alpha_2 di}{\mu}$ if $i \in [0, \mu]$ and $\left(\alpha_1 i + \alpha_2 d - \frac{i-\mu}{1-\mu}\overline{c}^{EU}\right)$ if $i \in [\mu, 1]$. As a consequence, the slope of u_i^{OS} depends on the rank of this agent: this slope is $\alpha_1 + \dfrac{\alpha_2 d}{\mu}$ as $i \in [0, \mu]$ and $\left(\alpha_1 - \frac{\overline{c}^{EU}}{1-\mu}\right)$ as $i \in [\mu, 1]$. According to the sign of this last term, we are able to distinguish two series of cases: *i)* cases where the adoption costs only dampen the effect of adoption externalities and *ii)* cases where the opposite relation is verified (adoption costs are higher than adoption externalities for the end-user). The curves u_i^P and u_i^{OS} may exhibit no, one or two intersections within the interval $[0,1]$. Let us first consider the case of a single intersection. In this case, the intersection of the curves u_i^P and u_i^{OS} corresponds to the critical level of adoption of the OS software sufficient to induce the adoption of the OS software. Supposing that all agents $j \in [0, i]$ adopt the OS software, this intersection indicates that the i^{th} indifferently adopts the OS software or the commercial one and the subsequent agents will prefer the commercial software.

Figure 1b provides the expected size of the OS software users' population as a function of its expected level. Given A1, $\forall i, u_i^{OS} \geq u_i^{\varnothing} = 0$. As a consequence, A1 excludes that in any case, $n^{\varnothing} \neq 0$. Hereby, $n^P = 1 - n^{OS}$ and the transformation $n^{OS} = t^{OS}\left(n^P, n^{OS}\right)$ can be expressed as single variable function of n^{OS}. It is then the representation of the transformation $n^{OS} = t^{OS}\left(n^P, n^{OS}\right)$ that *Figure A-1b* depicts. According to A2, agents are supposed to adopt the OS software in the reverse order of their adoption costs (end users endowed with the lowest adoption costs, adopt first). Thus, the intersections of this function with the 45° curve can be regarded as levels the OS population such that if they are expected by the population, they are also confirmed by their rational choices; these positions are Nash equilibria of the second step sub-game.

Figures 1a and *1b* can be related in a simple way: as long as the i^{th} agent does not wish to adopt the OS software when agents $j \in [0, i]$ are expected to have adopted it, the function associating the desired level of adoption with the expected level one remains below the 45° curve. The 45° curve is cut up on *figure 1b* with a vertical intersection, since the (first) intersection of u_i^P and u_i^{OS} is attained on *figure 1a* in an increasing part of function u_i^{OS} before the point $i = \mu$. This intersection defines a separating Nash equilibrium E3, in the sense that the agents with low adoption costs adopt OS in E3 while the others use the commercial software. After this intersection, the slope of the function u_i^{OS} is still positive, even after $i = \mu$. There then exists for a certain value of i, a second level of OS users such that if the size of OS users is expected by agents greater than or equal to this level, all agents choose to adopt OS. In this case (where the second equilibrium of section 3.2. rules as well), the curve connecting the OS adoption level to the expected OS adoption level has the form depicted in *figure 1b.* There is a Nash equilibrium E2 at the NE corner of the box. At last, when no agent is supposed to adopt OS, the equilibrium E1 of Proposition 2 is at the SW corner of the box.

The local stability of the Nash equilibria can also be analyzed in the case of static expectations with the help of *figure 1b*. Suppose thus that agents expect the size of the OS software users to be somewhere between equilibria E1 and E3. Then, their answer will be for all to adopt the commercial software. We conclude that all initial expectations of a relatively low level of OS software users rapidly drive to a Nash equilibrium where the OS software is fully crowded out (E1). Now suppose now that the initial expected size of the OS software users is somewhere between E3 and E2. Then, agents declaring choosing the OS software are more numerous than expected; some of the agents having initially chosen to be proprietary software users change their strategy and join agents declaring choosing the OS software. By successive iterations, the number of the OS software users increases until reaching equilibrium E2, where only the OS software remains.
Other cases are possible, according to the values of parameters and first step variables. They are listed on figures 2 to 7. Note that the assumption that the commercial software is economically viable excludes all case where the curve u_i^P of he marginal agent has no positive intersection with the ordinate axis, excluding, jointly with A1, the case where there is one single equilibrium with no adoption.

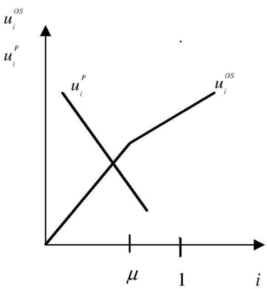


figure 1a

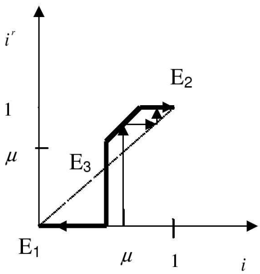


figure 1b

Case 1

Condition of appearance :

$\alpha_1(1-\mu) > \overline{c}^{EU}$

$dq - c_L > -\alpha_1$

$dq - c_L < \alpha_1(2\mu - 1) + d\alpha_2$

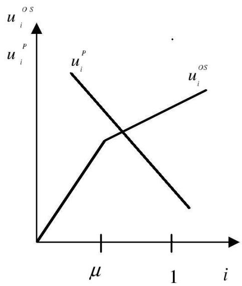


figure 2a

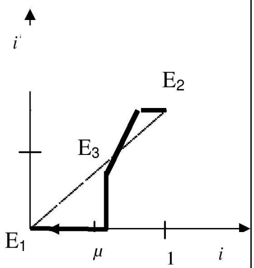


figure 2b

Case 2

Condition of appearance :

$\alpha_1(1-\mu) > \overline{c}^{EU}$

$dq - c_L > \alpha_1(2\mu - 1) + d\alpha_2$

$dq - c_L < \alpha_1 + d\alpha_2 - \overline{c}^{EU}$

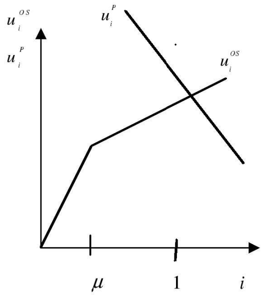


figure 3a

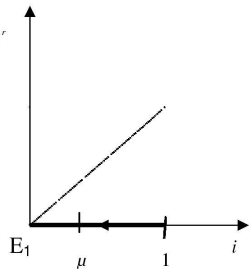


figure 3b

Case 3

Condition of appearance :

$\alpha_1(1-\mu) > \overline{c}^{EU}$

$dq - c_L > \alpha_1 + d\alpha_2 - \overline{c}^{EU}$

<u>Equilibria of the second step sub-game: cases 1 to 3.</u>

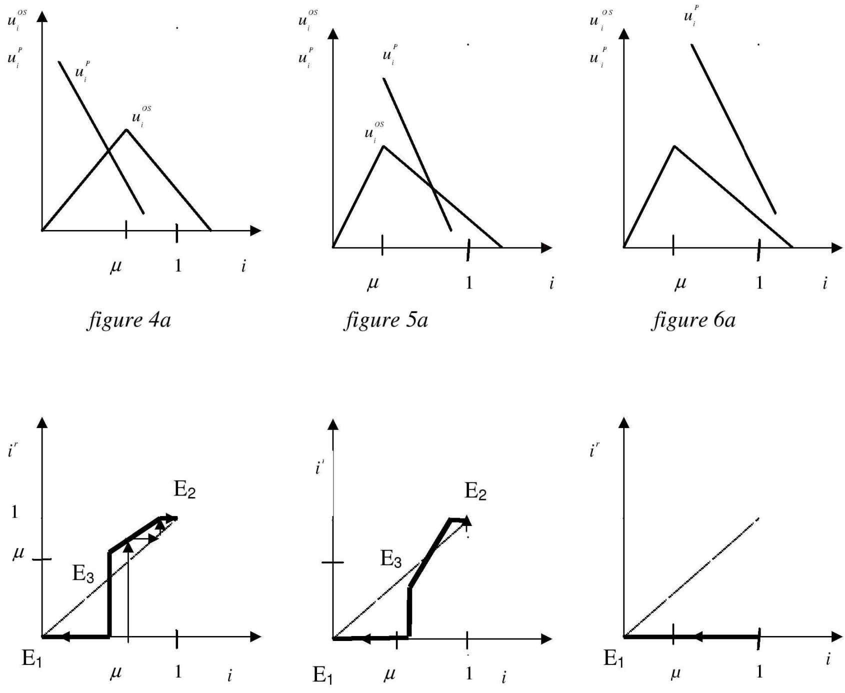


figure 4a *figure 5a* *figure 6a*

figure 4b *figure 5b* *figure 6b*

Case 4

Condition of appearance :

$\alpha_1(1-\mu) < \overline{c}^{EU}$

$dq - c_L > -\alpha_1$

$dq - c_L < \alpha_1(2\mu - 1) + d\alpha_2$

$dq - c_L < \alpha_1 + d\alpha_2 - \overline{c}^{EU}$

Case 5

Condition of appearance :

$\alpha_1(1-\mu) < \overline{c}^{EU}$

$dq - c_L > \alpha_1(2\mu - 1) + d\alpha_2$

$dq - c_L < \alpha_1 + d\alpha_2 - \overline{c}^{EU}$

Case 6

Condition of appearance :

$\alpha_1(1-\mu) < \overline{c}^{EU}$

$dq - c_L > \alpha_1(2\mu - 1) + d\alpha_2$

$dq - c_L > \alpha_1 + d\alpha_2 - \overline{c}^{EU}$

Equilibria of the second step sub-game: cases 4 to 6.

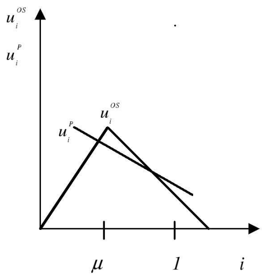


figure 7a

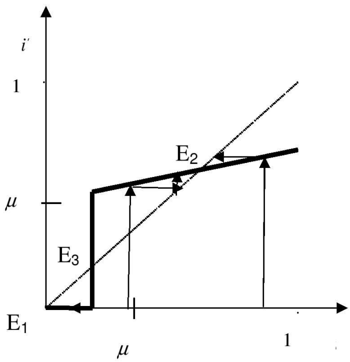


figure 7b

Case 3

Condition of appearance :

$$\alpha_1(1-\mu) < \overline{c}^{EU}$$
$$dq - c_L > -\alpha_1$$
$$dq - c_L < \alpha_1(2\mu - 1) + d\alpha_2$$
$$dq - c_L > \alpha_1 + d\alpha_2 - \overline{c}^{EU}$$

Equilibria of the second step sub-game: case 7.

In some cases (cases 1, 4, 2 and 5), there are three Nash Equilibria and only two of them are stable. In these cases, either the commercial or the OS software is adopted at equilibrium but the two software never finally coexist at the end of the eductive process. This case exhibits a “winner takes all” situation. The area of attraction of E1 and E2 is then defined by the place of t E3. Other cases correspond to one single stable "pooling" equilibrium (cases 3 and 6) where either E1 or E3 are the stable equilibria. The last case (case 7) corresponds to two stable equilibria, E1and E2.

Appendix 2: Nash equilibria of the first step sub-game

Proof of Proposition 3: Consider the expected profit to be maximized when agents have no alternative other to the adoption or the non adoption of the commercial software. This profit is then given by $\pi = \overline{n}^P c_L - g(q)$ where $\overline{n}^P$ figures the equilibrium level of adoption of commercial software by users when there is only one possible solution for adoption. Given the identity of agents when their choices are limited to adopting the commercial software or not adopting, the possible values for $\overline{n}^P$ when we exclude non-symmetric equilibria are $\overline{n}^P = 0$ or $\overline{n}^P = 1$. Constraints for the firm are $0 \le q \le 1$, $c_L \ge 0$ and $k_1 + dq \ge c_L$ (the reservation utility is associated for users to the possibility of not adopting). Given that for $\overline{n}^P = 0$, the reservation constraint of users is no longer satisfied and the profit of the firm vanishes, the optimal pair $(\overline{q}, \overline{c}_L)$ corresponds to $\overline{n}^P = 1$ and to a level of profit such that the whole 'social surplus' is taken by the firm, *i.e.*, $k_1 + d\overline{q} = \overline{c}_L$ (which is the case where all the "social surplus" is redistributed to the firm and corresponds to the monopoly equilibrium in the absence of the OS-competitor), the profit of the firm is $\pi = \overline{c}_L - g(\overline{q}) = k_1 + d\overline{q} - g(\overline{q})$. The optimal quality is given by $\overline{q} = \inf[1, \hat{q}]$ with $\hat{q}$ such that $g'(\hat{q}) = d$. The optimal price is then $\overline{c}_L = \inf[k_1 + d, k_1 + d\hat{q}]$. The profit associated at the pair $(\overline{q}, \overline{c}_L)$ works as an upper limit to the profit of the firm when there exists a competitor, since there is, in this last case, no shared externality between the commercial software and the OS solution, *i.e.*, $\pi^e = \overset{n}{\underset{i=1}{}} p_i(q, c_L)\pi_i(q, c_L) \le \inf[k_1 + d, k_1 + d\hat{q}] - g[\inf(1, \hat{q})]$ where $\pi_i(q, c_L)$ denotes the profit associated to the i^{th} equilibrium and $p_i(q, c_L)$ the probability associated with this equilibrium when the OS-solution 'competes' with the commercial one. As $\pi^e(\cdot)$ is continuous regarding to all arguments, there exists also a pair (q^*, c_L^*) which components are finite, associated with the profit $\pi^e(q, c_L)$, that demonstrates Proposition 3.

Appendix 3: Market Failures and Welfare Properties

Proof of Proposition 4: Suppose that the optimal quality / price pair (q^*, c_L^*) involves case, the occurrence of two stable equilibria, E$_1$ = $\{n^P* = 1, n_D^{OS}* = 0, n^{OS}* = 0\}$ and E$_2$ = $\{n^P** = 0, n_D^{OS}** = \mu, n^{OS}** = 1\}$ in the linear-quadratic. At E$_1$, individual utility of users is uniform and given by $u_i^P = \alpha_1 + dq^* - c_L^*$. At E$_2$, utility is $u_D^{OS} = \alpha_1 + \alpha_2 d$ for the developers and $u_i^{OS} = \alpha_1 + \alpha_2 d - \overline{c}_i^{EU}$ for the "last" end-users. For developers, the result of the comparison between the two equilibria is such that E$_1$ ^{D} E$_2$ when $\alpha_2 \le q^* - (c_L^*/d)$ and E$_2$ ^{D} E$_1$ when $\alpha_2 \ge q^* - (c_L^*/d)$. If (q^*, c_L^*) is an optimal solution for the firm, the expected profit is non-negative and $u_i^P = \alpha_1 + dq^* - c_L^* \ge 0$. For the "last" end-users, the result of the comparison is E$_1$ $^{EU_{i=1}}$ E$_2$ when $\alpha_2 \le q^* - \frac{(c_L^* - \overline{c}_i^{EU})}{d}$ and E$_2$ $^{EU_{i=1}}$ E$_1$ when $\alpha_2 \ge q^* - \frac{(c_L^* - \overline{c}_i^{EU})}{d}$. One concludes that the "all commercial" equilibrium E_1 is preferred by all users (and the firm) when $\alpha_2 \leq q^* - (c_L^*/d)$ and that there exists conflict of interest in the other cases.

Proof of Proposition 5: Suppose that the optimal quality / price pair (q^*, c_L^*) involves case, the occurrence of two stable equilibria, E_1 = $\{n^P{}^* = 1, n_D^{OS}{}^* = 0, n^{OS}{}^* = 0\}$ and E_3 = $\{n^P{}^{**} = 0, n_D^{OS}{}^{**} = \mu, n^{OS}{}^{**} = \eta\}$ with $\mu < \eta < 1$ in the linear-quadratic. At E_1, individual utility of users is uniform and given by $u_i^P = \alpha_1 + dq^* - c_L^*$. At E_3, utility is $u_D^{OS} = \alpha_1\eta + \alpha_2 d$ for the developers, $u_i^{OS} = \alpha_1\eta + \alpha_2 d - \frac{(i-\mu)}{(1-\mu)}\overline{c}_i^{EU}$, $\mu < i \leq \eta$ for the "first" end-users, and $u_i^P = \alpha_1(1-\eta) + dq^* - c_L^*, \eta \leq i \leq 1$, for the "last" end-users. For developers, the result of the comparison between the two equilibria is such that $E_1 \;^{D}\; E_3$ when $\alpha_2 \leq q^* + \frac{\alpha_1(1-\eta)}{d} - \frac{c_L^*}{d}$ and $E_3 \;^{D}\; E_1$ when $\alpha_2 \geq q^* + \frac{\alpha_1(1-\eta)}{d} - \frac{c_L^*}{d}$. For the i-end-users such that $\mu < i \leq \eta$, the result of the comparison between the two equilibria is such that $E_1 \;^{EU_{\mu<i\leq\eta}}\; E_3$ when $\alpha_2 \leq q^* + \frac{\alpha_1(1-\eta)}{d} - \frac{c_L^* - \frac{(i-\mu)}{(1-\mu)}\overline{c}_i^{EU}}{d}$ and $E_3 \;^{EU_{\mu<i\leq\eta}}\; E_1$ when $\alpha_2 \geq q^* + \frac{\alpha_1(1-\eta)}{d} - \frac{c_L^* - \frac{(i-\mu)}{(1-\mu)}\overline{c}_i^{EU}}{d}$. For the i-end-users such that $\eta \leq i \leq 1$, the result of the comparison between the two equilibria is always $E_1 \;^{EU_{\mu<i\leq\eta}}\; E_3$. As a conclusion, if $\alpha_2 \leq q^* + \frac{\alpha_1(1-\eta)}{d} - \frac{c_L^*}{d}$, there is coordination failure and E_1 is the high level equilibrium. If $\alpha_2 \geq q^* + \frac{\alpha_1(1-\eta)}{d} - \frac{c_L^*}{d}$, there is conflict of interest: developers and a part of end-users prefer E_1 while the firm and the other end-users prefer E_3

Proof of Proposition 6: Consider the pair $(\overline{q}, \overline{c}_L)$ corresponding to the optimal price / quality solution when users have no other alternative than adopting or not the commercial software. This solution is such (see *proof of proposition 3*) that the firm takes the whole surplus and that users' utility vanishes, *i.e.* $\forall i,\ u_i^P = \alpha_1 + d\overline{q} - \overline{c}_L = 0$. The profit is then $\pi = \overline{c}_L - \beta\overline{q}^2$. If (q^*, c_L^*) represents the optimal price / quality solution when users have the choice between the OS and commercial software and $\pi(E_1) = c_L^* - \beta q^{*2}$ the profit of the firm in the case where the second step equilibrium is E_1. Suppose that $\pi(E_1) = c_L^* - \beta q^{*2} = \pi = \overline{c}_L - \beta\overline{q}^2$. Then (see *proof of proposition 3*), $c_L^* = \overline{c}_L$, $q^* = \overline{q}$ and $\forall i,\ u_i^P(E_1) = \alpha_1 + dq^* - c_L^* = \alpha_1 + d\overline{q} - \overline{c}_L = 0$. Consider the agent $i = 0$, *i.e.* the first developer. At E_1, the utility of this agent as an OS-user is $u_i^{OS}(E_1) = 0$ in such a way that this agent is indifferent between the two technologies. For expected amount of n^{OS} larger that 0, this utility is $u_0^{OS} = \alpha_1 n^{OS} + \alpha_2 d \ \frac{n_D^{OS}}{\mu} > 0$, $n_D^{OS} = \inf\ n^{OS}, \mu$. One concludes that agent $i = 0$ will choose OS against the commercial software, except when $n^{OS} = 0$. It is then excluded that E_1 will be a locally stable equilibrium if $(q^*, c_L^*) = (\overline{q}, \overline{c}_L)$ and $\pi^e(\overline{q}, \overline{c}_L) = 0$ which disqualifies $(\overline{q}, \overline{c}_L)$ as an equilibrium solution. One conclude that E_1 is always such that $\forall i,\ u_i^P(E_1) = \alpha_1 + dq^* - c_L^* > \alpha_1 + d\overline{q} - \overline{c}_L = 0$, that demonstrates proposition 6

Bundling in Vertically Differentiated Communication Markets

Thierno Diallo*

May 2006

Abstract

We look at the competition and the welfare effects of bundling in the context of vertically differentiated communication services (e.g. Television, Telephone and Internet). We consider a two-stage game with two asymmetric firms (e.g. Telecom and Cable Operator). In the first stage firms simultaneously commit to adopt bundling or component pricing. These decisions give four possible configurations: (i) a configuration where both firms use component pricing; (ii) a configuration where both firms use bundling; and finally (iii) the two configurations where one firm use bundling and the other firm does not. In the second stage firms set simultaneously prices. We show that bundling is a dominant strategy equilibrium for both firms. The reason is that bundling increases the differentiation of services and reduces the intensity of price competition. We also find that although the bundling-bundling equilibrium reduces consumers' surplus, total economic welfare is higher than when both firms use component pricing.

* For comments and discussions I thank Abraham Hollander, Lars Ehlers and seminar participants at the Competition Bureau, Industry Canada. Correspondance: Thierno Diallo, Assistant Professor, Département des sciences économiques et administratives, Université du Québec à Chicoutimi, 555, boulevard de l'Université, Chicoutimi (Québec) G7H 2B1, email : thierno_diallo@uqac.ca

1. Introduction

Today consumers are offered telephone, high speed Internet and television services by cable operators and telecom companies. Cable operators supply broadband Internet access and voice telephony in addition to their "traditional" video services. Similarly, telecom companies supply telephone, video images and high speed Internet[1]. Typically cable operators and telecoms require subscribers to take their traditional service and they offer add-on services for an extra payment that is lower than the stand alone price of these services.

The purchase of all services from a single supplier is said to be convenient for buyers. It is also said to be a deterrent to churn [2] because disappointment with one service can be compensated by satisfaction with another service[3].

It is believed that the traditional telephone operator provides better telephone service than the cable operator, whereas the latter provides better television. Both offer a similar quality of high speed Internet[4]. Consumers therefore have to choose, between lower quality telephone combined

[1] In Quebec the dominant cable operator, Videotron provides digital television, telephone and high-speed Internet services with the coaxial cable technology while the dominant telecommunication company, Bell Canada provides the same services with satellite transmission and twisted pair. Coaxial cable is the kind of cable used by cable TV companies between the community antenna and the user homes and businesses. It carries broadband services for a great distance. To offer high speed Internet services, a cable operator creates a data network that operates over its hybrid fiber/ coax (HFC) plant. A twisted pair is an ordinary copper wire that connects home and business computers to the telephone company. DSL (Digital Subscriber line) Internet access provided by the local telephone company convert existing twisted-pair telephone lines into access paths for multimedia and high-speed data communications. So with satellite and twisted pair technologies, a local telecommunication company can also supply same kind of services as a local cable operator.

[2] The rate at which customer discontinues service (in order to shift to competitor) - among high usage customers, at the expense of profit margins: Keith Damsell "Telecom bundling seem luring customers. Grouping services together for lower prices builds loyalty, turn "churn" low. Study says "The Globe and Mail, 29 September 2003, at p.138, citing Convergence Consulting Group ltd study: The Battle for the North American Couch Potatoes, and referring to Cox Communications, extremely low churn rate with the triple play services of digital television, high speed internet access, and local telephone services.

[3] On the other hand consumers can drop all services when they are disappointed with one of them.

[4] Indeed cable telephony has some limitations: e.g. it doesn't work when there is power failure and drop out when broadband demand (the ability of the user to view content across the internet that includes large files, such as video, audio and 3D) is high. Also not all areas are served by the POC since hybrid fiber/ coax (HFC) plants are expensive to install. Consequently additional costs of providing services to additional customers are higher for co-ax (HFC) technology than twisted pair technology. On the other hand the television service of telecoms has also severe

with high quality television offered by the cable operator, and higher quality telephone with lower quality television offered by the telephone company.

This paper addresses the following questions: (i) under what conditions do suppliers bundle? that is under what conditions does it sell two or more services as a package only? ; (ii) how does bundling compare to component selling in terms of welfare? ; and (iii) what attitude should competition authorities adopt toward such bundling?

There are no clear-cut results pertaining to the profitability and welfare effect of bundling as opposed to separate selling of components. Adams and Yellen (1976), Schmalensee (1982, 1984), Mc Afee et al (1989), and Whinston (1990) show that under monopoly bundling raises profits when variable costs are zero. However, the vast majority of consumer services are supplied in non-monopolistic environments. Only few papers [(Matutes and Regibeau, 1989; Economides, 1993; Anderson and Leruth, 1993; Kopalle and al, 1999)] examine the non-monopolistic case where firms have the option of bundling. Theses papers assume horizontal differentiation of services and their conclusions are numerical.

Economides (1993) considers a two-stage game and shows that the Nash equilibrium is mixed bundling[5] rather than component selling. Because competition is more intense under mixed bundling, a prisoner's dilemma arises, that is firms would be better off if they could commit not to bundle. Anderson and Leruth (1993) show in a two-stage model that the Nash equilibrium is both firms offer components selling. The reason is that firms fear the extra degree of competition intrinsic to mixed bundling. Kopalle and al (1999) reconcile the result of Economides (1993) and Anderson and Leruth (1993) by incorporating the role of market expansion on equilibrium bundling strategies. They show that for complementary components mixed bundling dominates component selling only when it creates a new market for the bundle.

---

limitations in competing with cable TV. It encounters some constraints due to broadband transmission, to cities' architectures, to weather conditions and it needs an installation of a non esthetical device, the dish.

[5] Mixed bundling means that the packages as well as the individual components of the package are available.

Matutes and Regibeau (1992) consider a game where in the first stage there is a choice between compatibility versus incompatibility[6]. In the second and the third stage of the game firms choose the selling strategy and prices respectively. Matutes and Regibeau ask whether firms would choose to make their products compatible and whether they would sell their products as a bundle. For compatible components, they find that, depending on consumer's reservation price there can be two kinds of equilibria. In the first, one firm bundles and one firm does not. In the second, both firms bundle[7].

None of the aforementioned papers (i) is concerned with vertically differentiated services; (ii) They do not give clear results about the welfare effect of bundling. The underlying motivation of this paper is to analyze the competition and the welfare effect of bundling in the communication market within the context of vertically differentiated services. We consider a two-stage game with two asymmetric firms. In the first stage firms simultaneously commit to use bundling or component pricing. These decisions give four possible configurations: (i) a configuration where both firms use component pricing; (ii) a configuration where both firms use bundling; and finally (iii) the two configurations where one firm use bundling and the other firm does not. In the second stage firms set prices simultaneously.

We show that bundling is a dominant strategy equilibrium for both firms. The reason is that bundling increases the differentiation of services and reduces the intensity of price competition[8]. We also find that although the bundling-bundling equilibrium reduces consumers' surplus, the total economic welfare is higher than when both firms use component pricing.

---

[6] A component is incompatible with components sold by other firms', if it cannot be assembled with them to form a usable system. The economic consequence of compatibility versus incompatibility have been examined by Matutes and Regibeau (1988), Economides (1989, 1991), and Einhorn (1992). They have looked at the case where each firm supplied all the necessary goods.

[7] The first equilibrium occurs when consumer's reservation price is low, while the second one occurs when it is high.

[8] Chen (1997) also analyzes bundling as differentiation tool. He studies the case where two sellers compete in a first market, and both also sell another product in a second competitive market. Absent bundling, Bertrand competition drives both sellers' profits in the first market to zero. If one seller uses bundling and the other does not, however, both can earn positive profits since the bundle and the individual first-market product are effectively differentiated products.

The paper is structured as follows. In section 2, we present a duopoly model, where each firm has the choice to sell its services either separately or as a bundle. In section 3, we analyze the game. In section 4, we derive the equilibrium selling strategy of each firm. In section 5, we analyze the welfare consequences of bundling. In section 6, we provide an application of the model to the communication market and we conclude.

2. The Model

There are two firms, denoted *h* and *l*. E.g. firm *h* is a telecom company and firm *l* is a cable operator. Each firm sells two services, denoted *A* and *B*. E.g. service *A* is a telephone service and service *B* is an Internet service. The service *A* comes in two qualities, a_h and a_l supplied respectively by firm *h* and firm *l*, $a_h > a_l > 0$. The quality b of service *B* is the same for both firms. Both the variable cost and the fixed cost are zero for each service. Every consumer demands one or zero unit of service of *A* and/or *B*.

A consumer with a parameter θ derives a utility θa_i from quality a_i of service *A*, $i = h, l$. Similarly, a consumer with a parameter γ derives a utility γb from quality *b* of service *B*. If the consumer chooses not to buy a service, she receives her reference utility which is normalized to zero. A consumer with preference indices (θ, γ) who buys one unit of *A* of quality a_i at price p_i and one unit of B at price p_B receives a net surplus:

$$U = (\theta a_i - p_i) + (\gamma b - p_B), i = h, l$$

Each consumer makes her purchase decision to maximize her consumer surplus. Consumer preference indices θ and γ are independently and uniformly distributed on $[0,1] \times [0,1]$.

We model the competition as a two-stage game. In the first stage, firms decide whether to bundle, or not to bundle; in the second stage[9] they set prices. There are four possible subgames at stage 2: (i) (C_h, C_l) denotes the game where both firms sell components separately; (ii) (B_h, B_l) denotes

[9] Firms observe the choices made in the first stage.

the game where both firms bundle; (iii) (B_h, C_l) denotes the game where firm *h* bundles, and firm *l* sells its components separately; and (iv) (C_h, B_l) denotes the game where firm *h* sells its components separately, and firm *l* bundles. We will examine under what conditions each of the subgames is an equilibrium.

3. Price determination

3.1 Case (i): (C_h, C_l), Pure Components by both Firms

Since both firms produce the same quality of service *B*, Bertrand competition insures that its price is driven down to marginal cost, which is zero. With regard to service *A,* we know[10] that an equilibrium with two active firms requires[11]: $\frac{p_h}{a_h} \Diamond \frac{p_l}{a_l}$. We designate by $\bar{\theta}$ the consumer indifferent between not purchasing and purchasing one unit of A_l, and by $\tilde{\theta}$ the consumer who is indifferent between purchasing A_l and A_h. Figure 1 displays market shares as $(1-\tilde{\theta})$ for firm *h* and $(\tilde{\theta}-\bar{\theta})$ for firm *l* when firms compete in prices.

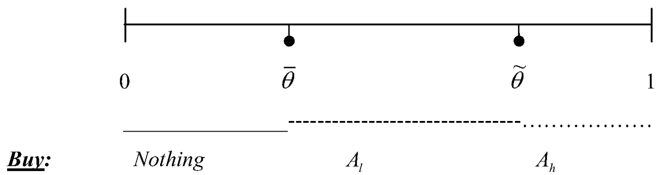


Figure 1. Market areas under the regime (C_h, C_l)

---

[10] See Tirole (1988)

[11] The condition states that the *price per unit of quality* is higher for a_h than for a_l. This means that low quality is not dominated by high quality. If low quality is dominated by high quality then the firm with the low quality exits the market.

The firms' profits are:

$$\pi_h^{C_h,C_l} = p_h(1-\tilde{\theta}) \text{ for firm } h,$$

$$\pi_l^{C_h,C_l} = p_l(\tilde{\theta}-\bar{\theta}) \text{ for firm } l,$$

where $\bar{\theta} = \frac{p_l}{a_l}$ and $\tilde{\theta} = \frac{p_h - p_l}{a_h - a_l}$. Prices are chosen optimally when they satisfy the conditions below

$$p_h = \frac{2(a_h^2 - a_h a_l)}{4a_h - a_l}, \text{ and } p_l = \frac{(a_h - a_l)a_l}{4a_h - a_l}.$$

Then:

$$\pi_h^C = \frac{4a_h^2(a_h - a_l)}{(4a_h - a_l)^2}, \text{ and } \pi_l^C = \frac{a_h a_l (a_h - a_l)}{(4a_h - a_l)^2}.$$

3.2 Case (ii): (B_h, B_l), Bundling by both Firms:

Denote by p_{Gh} and p_{Gl} the prices of bundles $A_h B$ and $A_l B$ respectively. The individual-rationality constraints are

for consumers of $A_h B$: $\theta a_h + \gamma b - p_{Gh} \lozenge 0$, (R_h)

for consumers of $A_l B$: $\theta a_l + \gamma b - p_{Gl} \lozenge 0$. (R_l)

Self-selection constraints are

for consumers of $A_h B$: $\theta a_h + \gamma b - p_{Gh} \lozenge \theta a_l + \gamma b - p_{Gl}$, (S_h)

for consumers of $A_l B$: $\theta a_l + \gamma b - p_{Gl} \lozenge \theta a_h + \gamma b - p_{Gh}$. (S_l)

The condition $\theta a_h + \gamma b - p_{Gh} \lozenge Max(0, \theta a_l + \gamma b - p_{Gl})$ must be satisfied by buyers of $A_h B$. The condition $\theta a_l + \gamma b - p_{Gl} \lozenge Max(0, \theta a_h + \gamma b - p_{Gh})$ must be satisfied by buyers of $A_l B$.

We find again that market areas depend on the ranking of *price per unit of quality* of service *A*. To see how, we define the preference parameter of the consumer indifferent between the bundles $A_h B$ and $A_l B$ by $\theta^* \copyright \frac{p_{Gh} - p_{Gl}}{a_h - a_l}$. We distinguish three cases.

Case 1: $\frac{p_{Gh}}{a_h} \leq \frac{p_{Gl}}{a_l}$.

Case 2: $\frac{p_{Gh}}{a_h} \lozenge \frac{p_{Gl}}{a_l}$ and $\theta^* < 1$.

Case 3: $\frac{p_{Gh}}{a_h} \lozenge \frac{p_{Gl}}{a_l}$ and $\theta^* > 1$

In case 1 the *price per unit of quality* of $A_h B$ is lower than *price per unit of quality* of $A_l B$. In case 2 and case 3 *the price per unit of quality* of $A_h B$ is higher than the *price per unit of quality* of $A_l B$. The difference between case 2 and case 3 is that in the latter there is no consumer indifferent between $A_h B$ and $A_l B$.

Case 1: $\frac{p_{Gh}}{a_h} \leq \frac{p_{Gl}}{a_l}$.

The price of bundle A_hB per unit of quality of service A is lower than the price of bundle A_lB per unit of quality of service A. The lines labelled R_h and R_l in Figure 2 are the individual-rationality constraints of high and low quality buyers.

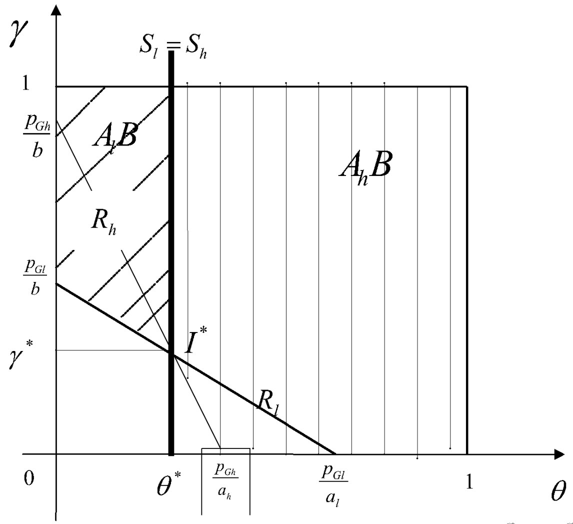


Figure 2: Market areas under the regime (B_h, B_l) when $\frac{p_{Gh}}{a_h} \leq \frac{p_{Gl}}{a_l}$

The lines S_h and S_l represent[12] the self-selection constraint faced by consumers. Consumers with preference parameters above R_h derive positive utility from A_hB. Consumers with preference parameters above R_l derive positive utility from A_lB . We note that S_h, S_l, R_h and R_l intersect at $I^* © \theta^* © \frac{p_{Gh} - p_{Gl}}{a_h - a_l}; \gamma^* © \frac{a_h p_{Gl} - a_l p_{Gh}}{a_h - a_l}$.

[12] The constraints S_h and S_l yield the same line.

The high quality firm serves consumers with preference parameters $\theta \in [\theta^*,1]$ and above R_h ; the low quality firm serves consumers with preference parameters $\theta \in [0,\theta^*]$ and above R_l. The size of market served by the firm *h* is $D_{Gh} = 1-\theta^* - \frac{(\gamma^*)^2}{2a_h}$ and the size of market served by the firm *l* *is* $D_{Gl} = \theta^*(1-\frac{p_{Gl}}{2b}-\frac{\gamma^*}{2})$. Firm *h* and firm *l* profit functions' are respectively:

$$\pi_h^{PB} = p_{Gh}D_{Gh} \text{ and } \pi_l^{PB} = p_{Gl}D_{Gl}.$$

In contrast to the standard model of single differentiated good, we find that there can be two active firms even when $\frac{p_{Gh}}{a_h} \leq \frac{p_{Gl}}{a_l}$[13]. The Difference is as follows: in standard models of a single differentiated good, consumers make the comparison on a service by service basis. If A_l is dominated by A_h, all consumers obtain more surplus from A_h than from A_l. Nobody purchases A_l. We also know that for service *B*, Bertrand competition and zero marginal cost imply that consumers obtain *B* for free from both firms. Thus firm *l* is excluded from market *A*, but remains in market *B*. In the regime of (B_h, B_l) there is a competition for vertically differentiated system goods. Therefore, the best available alternative for consumers who wish to purchase only service *B* is to purchase the low quality bundle A_lB. In that case, if A_l is dominated by A_h the low quality firm can survive in both markets because it serves the bundle A_lB to consumers who care very little about service *A,* while the high quality firm serves the bundle A_hB to consumers who care for service *A* and for service *B*.

[13] That is the low quality system is dominated by the high quality system.

Case 2: $\frac{p_{Gh}}{a_h} \Diamond \frac{p_{Gl}}{a_l}$ and $\theta^* < 1$.

The price of bundle $A_h B$ per unit of quality of service A is higher than the price of bundle $A_l B$ per unit of quality of service A and there exist a consumer who is indifferent between the bundles. Market areas are shown in Figure 3.

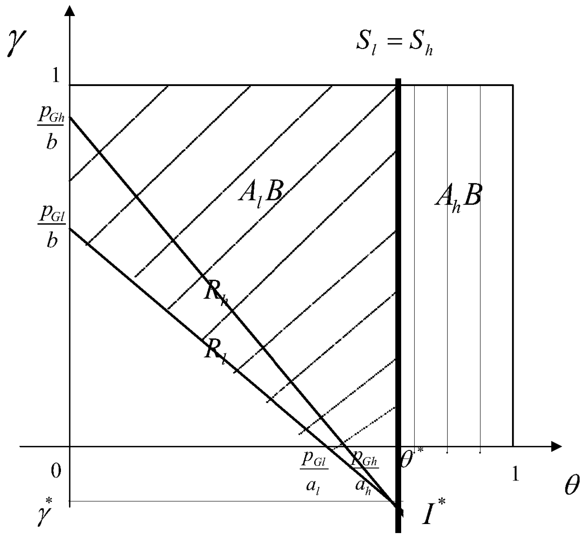


Figure 3: Market areas under the regime (B_h, B_l) when $\frac{p_{Gh}}{a_h} \Diamond \frac{p_{Gl}}{a_l}$ and $\theta^* < 1$.

The high quality firm serves consumers with preference parameters $\theta \rfloor [\theta^*, 1]$ and above R_h. The low quality firm serves consumers with preference parameters $\theta \rfloor [0, \theta^*]$ and above R_l. Consumers with preference parameters below R_l do not purchase at all. The market areas for $A_h B$ and $A_l B$ are respectively: $D_{Gh} = 1 - \theta^*$ and $D_{Gl} = \theta^* \; 1 - \frac{(p_{Gl})^2}{2ba_l}$.

Case 3: $\frac{p_{Gh}}{a_h} \lozenge \frac{p_{Gl}}{a_l}$ and $\theta^* \lozenge 1$.

The price of bundle A_hB per unit of quality of service A is higher than the price of bundle A_lB per unit of quality of service A and nobody is indifferent between the bundles. Market areas are displayed in Figure 4.

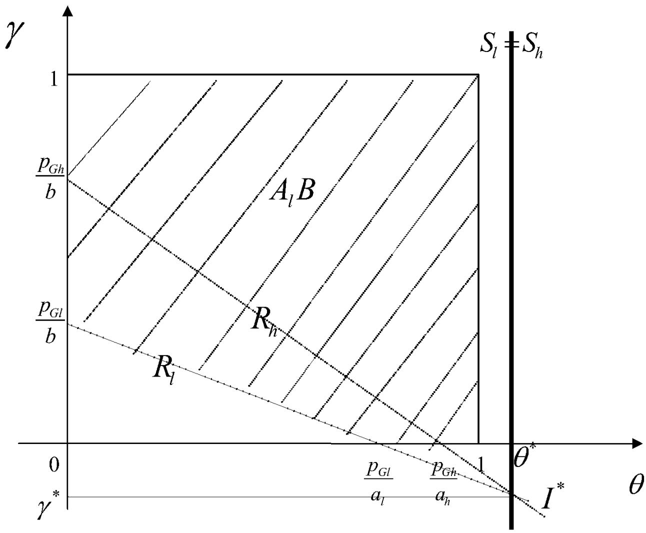


Figure 4: Market areas under the regime (B_h, B_l) when $\frac{p_{Gh}}{a_h} \lozenge \frac{p_{Gl}}{a_l}$ and $\theta^* \lozenge 1$.

Because $\theta^* \lozenge 1$, the demand for the bundle A_hB is zero. The low quality firm serves consumers with preference parameters θ ⌡ [0,1] and above R_l [14]. Consumers with preference parameters below R_l do not purchase at all. The market areas for A_hB and A_lB are respectively: $D_{Gh} = 0$ and $D_{Gl} = 1 - \frac{(p_{Gl})^2}{2ba_l}$

[14] In that case , firm h makes zero profit and it easy to see that this is not an equilibrium because firm h is always better off (makes positive profits) by choosing its price such that $\theta^* \leq 1$.

3.2.1 Determination of the equilibrium prices in (B_h, B_l)

In case 1 the first order conditions are:

$$\frac{f\pi_h^B}{fp_{Gh}} = 1 - \frac{2p_{Gh}}{a_h - a_l} - \frac{3}{2a_h}\left(\frac{a_l p_{Gh}}{a_h - a_l}\right)^2 + \left(\frac{a_l}{a_h - a_l} + \frac{2a_l p_{Gh}}{(a_h - a_l)^2}\right) p_{Gl} - \frac{a_h}{2}\left(\frac{p_{Gl}}{a_h - a_l}\right)^2 = 0$$

$$\frac{f\pi_l^B}{fp_{Gl}} = p_{Gh} + \frac{a_l (p_{Gh})^2}{2(a_h - a_l)} - 2p_{Gl} - \left(\frac{1}{b} + \frac{a_h + a_l}{a_h - a_l}\right)(p_{Gh} p_{Gl}) + \left(\frac{3}{2b} + \frac{3a_h}{2(a_h - a_l)}\right)(p_{Gl})^2 = 0$$

We can see that the first order conditions are quite complex. We obtain similar complicated first order conditions for all others cases. For this reason, we search the price equilibria numerically for a range of values of a_h and a_l.

In case 1, the equilibrium prices that we obtain always satisfy $\frac{p_{Gh}}{a_h} \leq \frac{p_{Gl}}{a_l}$. Figure 5 shows the equilibrium values of $\frac{p_{Gh}}{a_h}$ and $\frac{p_{Gl}}{a_l}$ for different values of a_h.

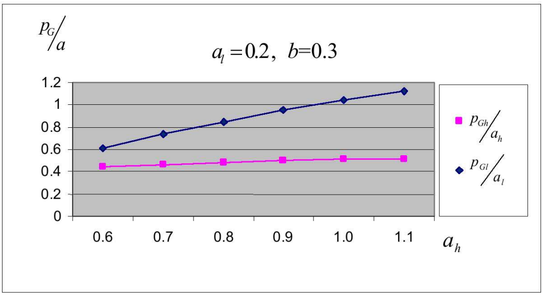


Figure 5: Comparison of prices per unit of quality under regime (B_h, B_l)

In case 2 and 3, the equilibrium prices that we obtain do not satisfy $\frac{p_{Gh}}{a_h} > \frac{p_{Gl}}{a_l}$. Therefore, we will only look at case 1.

It is interesting to compare profits in (B_h, B_l) to profits in (C_h, C_l). The profits under the regime (C_h, C_l) and the regime (B_h, B_l) for different values of a_h are shown in Figure 6. $\pi_h^{B_h,B_l}$ is always higher than $\pi_h^{C_h,C_l}$. Similarly, $\pi_l^{B_h,B_l}$ is always higher than $\pi_l^{C_h,C_l}$. We obtain similar results for various values of a_l.

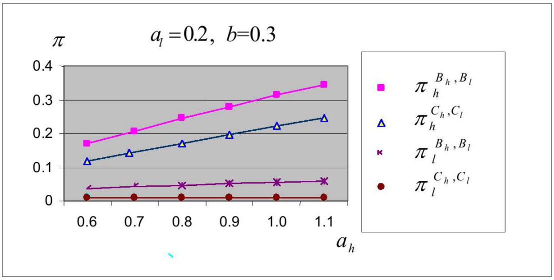


Figure 6: Comparison of firms' profits under the regimes (C_h, C_l) and (B_h, B_l).

We see that both firms are better when they bundle. The reason is that bundling affects the intensity of competition via two channels: (i) it reduces the intensity of the competition for service B by increasing differentiation; and (ii) it increases the intensity of the competition for service A by reducing differentiation. The net effect of bundling is a decrease of competition between the two firms because the competition for B under component pricing is extreme (Bertrand competition). Therefore each makes more profit in the subgame (B_h, B_l) than in the subgame (C_h, C_l).

3.3 Case (iii): (B_h, C_l), Bundling by firm *h*, Pure Component Selling by firm *l*

The individual-rationality constraints are now

for consumers of $A_h B$: $\theta a_h + \gamma b - p_{Gh} \lozenge 0$, (R_h)

for consumers of A_l : $\theta a_l - p_l \lozenge 0$, (R_l)

for consumers of B : $\gamma b - p_B \lozenge 0$, (R_b)

for consumers of $A_l + B$ [15] : $\theta a_l + \gamma b - p_l - p_B \lozenge 0 . (R_m)$

The self-selection constraints are

for consumers of $A_h B$: $\theta a_h + \gamma b - p_{Gh} \lozenge Max(\theta a_l + \gamma b - p_l - p_B , \theta a_l - p_l , \gamma b - p_B) , (S_h)$

for consumers of A_l : $\theta a_l - p_l \lozenge Max(\theta a_h + \gamma b - p_{Gh} ; \theta a_l + \gamma b - p_l - p_B ; \gamma b - p_B)$, (S_l)

for consumers of B : $\gamma b - p_B \lozenge Max(\theta a_h + \gamma b - p_{Gh} ; \theta a_l + \gamma b - p_l - p_B ; \theta a_l - p_l)$, (S_b)

for consumers of $A_l + B$: $\theta a_l + \gamma b - p_l - p_B \lozenge Max(\theta a_h + \gamma b - p_{Gh} ; \gamma b - p_B ; \theta a_l - p_l)$. (S_m)

The condition $\theta a_h + \gamma b - p_{Gh} \lozenge Max(0, \theta a_l + \gamma b - p_l - p_B , \theta a_l - p_l , \gamma b - p_B)$ must be satisfied by buyers of $A_h B$.

The condition $\theta a_l - p_l \lozenge Max(0, \theta a_h + \gamma b - p_{Gh} ; \theta a_l + \gamma b - p_l - p_B ; \gamma b - p_B)$ must be satisfied by buyers of A_l alone.

The condition $\gamma b - p_B \lozenge Max(0, \theta a_h + \gamma b - p_{Gh} ; \theta a_l + \gamma b - p_l - p_B ; \theta a_l - p_l)$ must be satisfied by buyers of B alone.

And finally the condition $\theta a_l + \gamma b - p_l - p_B \lozenge Max(0, \theta a_h + \gamma b - p_{Gh})$ must be satisfied by buyers of both A_l and B .

From S_h , we derive that the preference index of the consumer indifferent between purchasing $A_h B$ and purchasing A_l and B separately is $\hat{\theta} \copyright \dfrac{p_{Gh} - p_l - p_B}{a_h - a_l}$. Note that S_l , R_h , and R_l intersect at $\hat{I} \copyright \dfrac{p_l}{a_l} ; \hat{\gamma} \copyright \dfrac{1}{b}(p_{Gh} - \dfrac{a_h}{a_l} p_l)$. $\hat{\gamma}$ can be understood as the implicit *price per unit of quality of B*, when the quality of *A* is valued at the price set by firm *l*. Note also that R_h , R_b , and

[15] Consumers of $A_l + B$ means consumers of both A_l and B but each component is purchased separately.

S_b intersect at $\frac{p_{Gh}-p_B}{a_h};\frac{p_B}{b}$ [16]. The market areas depend on whether $\hat{\gamma}$, the *implicit price per unit of quality* of B is greater, lower or equal to $\frac{p_B}{b}$, the *explicit price per unit of quality* of B set by firm l. We now distinguish three cases.

Case 1: $p_{Gh} > p_l + p_B$ and $\hat{\gamma} > \frac{p_B}{b}$.

Case 2: $p_{Gh} > p_l + p_B$ and $\hat{\gamma} \leq \frac{p_B}{b}$.

Case 3: $p_{Gh} \leq p_l + p_B$.

In case 1 and case 2 the price of the bundle A_hB is higher than the price of both A_l and B. The difference between the two cases is that in case 1 the *implicit price per unit of quality* of B *is* greater than the *explicit price per unit of quality* of B set by firm l. While in case 2 the *implicit price per unit of quality* of B *is* lower than the *explicit price per unit of quality* of B set by firm l. In case 3 the price of the bundle A_hB is lower than the price of both A_l and B.

[16] Service B is sold as part of the bundle A_hB. Remark that $\frac{p_B}{b} < \frac{1}{b}(p_{Gh} - \frac{a_h}{a_l} p_l)$ ©$\hat{\gamma}$ can be written as: $\frac{p_{Gh}-p_B}{a_h} > \frac{p_l}{a_l}$. The ratio $\frac{p_{Gh}-p_B}{a_h}$ represents the *implicit price per unit of quality* of A_h when it is sold as part of the bundle A_hB . Thus market areas depend on whether this *implicit price per unit of quality* is greater, lower or equal to the explicit *price per unit of quality* of A_l.

Case 1: $p_{Gh} > p_l + p_B$ and $\hat{\gamma} > \dfrac{p_B}{b}$

It is the case where the implicit *price per unit of quality of B* is larger than the *explicit price per unit of quality* of *B*. Figure 7 shows that parameter space divides into five segments[17]

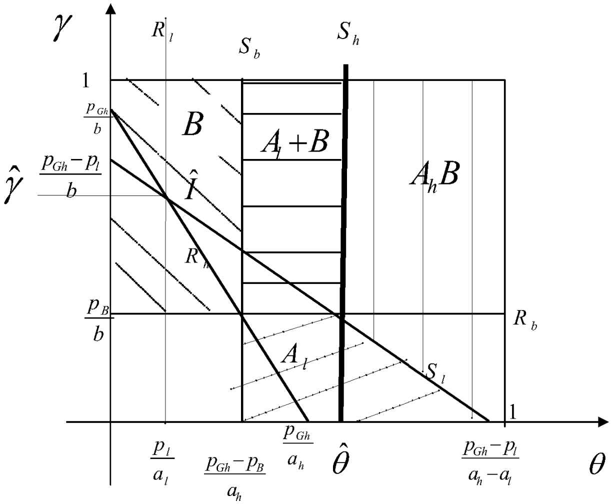


Figure 7: Market areas under the regime (B_h, C_l) when $p_{Gh} > p_l + p_B$ and $\hat{\gamma} > \dfrac{p_B}{b}$

Consumers with preference parameters to the right of S_h and above S_l purchase the bundle A_hB. The market area of those consumers is $D_{Gh} = 1 - \hat{\theta} - \dfrac{(p_B)^2}{2b(a_h - a_l)}$. Consumers with preference parameters between S_b and S_l and below R_b purchase A_l alone. The market area of those consumers is $D_{A_l} = \dfrac{p_B}{b}\left(\hat{\theta} - \dfrac{p_{Gh} - p_B}{a_h} + \dfrac{p_B}{2(a_h - a_l)}\right)$. Consumers with preference parameters to the left of S_h, to the right of S_b, and above R_b purchase A_l and *B*. The market area of those

[17] We recall that $\hat{\theta}$ © $\dfrac{p_{Gh} - p_l - p_B}{a_h - a_l}$ and $\hat{\gamma}$ © $\dfrac{1}{b}(p_{Gh} - \dfrac{a_h}{a_l} p_l)$.

consumers is $D_{A_l+B} = \hat{\theta}-\frac{p_{Gh}-p_B}{a_h} \quad 1-\frac{p_B}{b}$. Finally consumers with preference parameter to the left of S_b and above R_b purchase B alone. The market area of those consumers is $D_B = \frac{p_{Gh}-p_B}{a_h} \quad 1-\frac{p_B}{b}$. Others don't purchase.

Case 2: $p_{Gh} > p_l + p_B$ and $\hat{\gamma} \le \frac{p_B}{b}$

It is the case where the implicit *price per unit of quality of B* is lower than the explicit *price per unit of quality* of *B*. Figure 8 shows that parameter space divides into four segments.

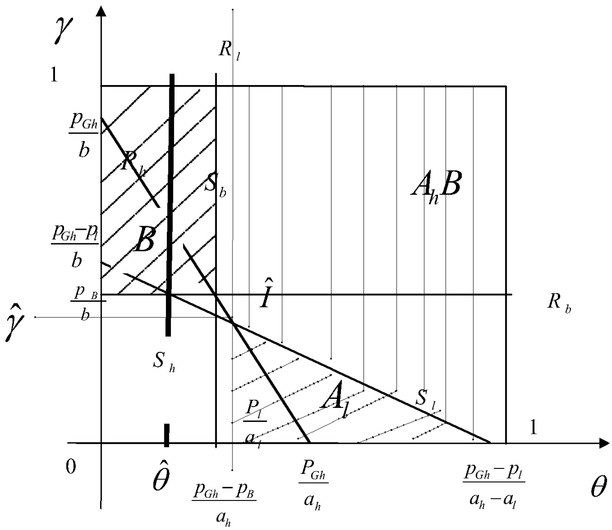


Figure 8: Market areas under the regime (B_h, C_l) when $p_{Gh} > p_l + p_B$ and $\hat{\gamma} \le \frac{p_B}{b}$

The allocation of consumers is as follows: consumers with preference parameters to the right of S_b, above R_h and above S_l purchase A_hB. Those with preference parameters to the left of S_b

and above R_b purchase B alone. Finally consumers with preference parameter to the right of R_l and below S_l purchase A_l alone. Others don't purchase.

We see in Figure 8 that $\hat{\theta} < \frac{p_{Gh} - p_B}{a_h}$ and S_b is to the right of S_h. Therefore nobody buys $A_l + B$, that is no consumer purchases B separately when she also purchases A_l [18]. The reason is that the explicit price of B is higher than the implicit price of B when purchased as part in the bundle $A_h B$.

For the particular case where $\hat{\gamma}$ © $\frac{p_B}{b}$, we have S_h © S_l © R_l and the allocation of consumers in the parameter space is the same as above.

[18] Recall that buy $A_l + B$ means buy both A_l and B.

Case 3: $p_{Gh} \leq p_l + p_B$

It is the case where consumers who purchase B separately, also purchase A_l and pay for both services a price higher than the price of the bundle $A_h B$. In this case the demand for both A_l and B is zero since consumers can purchase $A_h B$, that is they can get a better bundle at a lower price. But consumers purchase A_l alone and B alone. Market areas are shown in Figure 9.

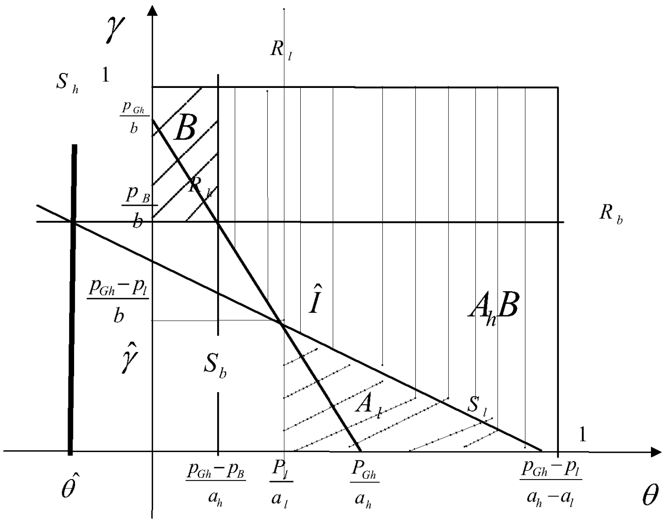


Figure 9: Market areas under the regime (B_h, C_l) when $p_{Gh} \leq p_l + p_B$

Figure 9 depicts a similar pattern as for Figure 3.8. Therefore case 2 and case 3 give the same allocation of consumers in the parameter space.

3.3.1 Determination of the equilibrium prices in (B_h, C_l)

Here also the first order conditions are quite complex. We search the price equilibria numerically for a range of values of a_h and a_l. Figure 10 shows the equilibrium values of $\hat{\gamma}$ and $\frac{p_B}{b}$ for a range of values of a_h [19] in case 1. The equilibrium prices that we obtain always satisfy $\hat{\gamma} > \frac{p_B}{b}$. In case 2 and 3, the equilibrium prices that we obtain do not satisfy $\hat{\gamma} \leq \frac{p_B}{b}$. Therefore, we will only look at case 1.

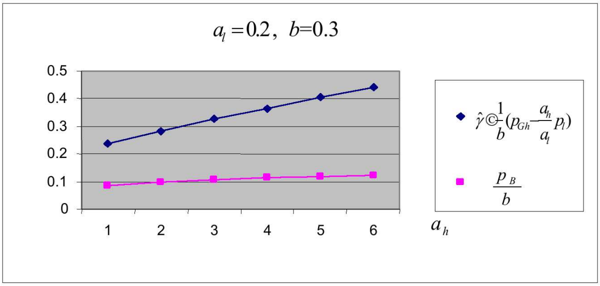


Figure 10: Comparison of prices per unit of quality under the regime of (B_h, C_l)

Note that gap between the explicit price of service *B* and the implicit price of service *B* becomes larger when a_h becomes larger.

[19] The equilibrium condition $\frac{p_B}{b} \Diamond \hat{\gamma}$ is not satisfied for the parameter values that we have chooses for simulation.

3.4. Case (iv): (C_h, B_l), Pure Component Selling by firm *h*, Bundling by firm *l*

The participation constraints are

for consumers of $A_l B$: $\theta a_l + \gamma b - p_{Gl} \lozenge 0$, (R_l)

for consumers of A_h : $\theta a_h - p_h \lozenge 0$, (R_h)

for consumers of B : $\gamma b - p_B \lozenge 0$, (R_b)

for consumers of $A_h + B$[20] : $\theta a_h + \gamma b - p_h - p_B \lozenge 0$. (R_m)

The self-selection constraints are

for consumers of $A_l B$: $\theta a_l + \gamma b - p_{Gl} \lozenge Max(0, \theta a_h + \gamma b - p_h - p_B, \theta a_h - p_h, \gamma b - p_B)$, (S_h)

for consumers of A_h : $\theta a_h - p_h \lozenge Max(\theta a_l + \gamma b - p_{Gl}; \theta a_h + \gamma b - p_h - p_B; \gamma b - p_B)$, (S_l)

for consumers of B : $\gamma b - p_B \lozenge Max(\theta a_l + \gamma b - p_{Gl}; \theta a_h + \gamma b - p_h - p_B; \theta a_h - p_h)$, (S_b)

for consumers of $A_h + B$: $\theta a_h + \gamma b - p_h - p_B \lozenge Max(\theta a_l + \gamma b - p_{Gl}; \gamma b - p_B; \theta a_h - p_h)$. (S_m)

The condition $\theta a_l + \gamma b - p_{Gl} \lozenge Max(0; \theta a_h + \gamma b - p_h - p_B; \theta a_h - p_h; \gamma b - p_B)$ must be satisfied by buyers of $A_l B$.

The condition $\theta a_h - p_h \lozenge Max(0; \theta a_l + \gamma b - p_{Gl}; \theta a_h + \gamma b - p_h - p_B; \gamma b - p_B)$ must be satisfied by buyers of A_h alone.

The condition $\gamma b - p_B \lozenge Max(0; \theta a_l + \gamma b - p_{Gl}; \theta a_h + \gamma b - p_h - p_B; \theta a_h - p_h)$ must be satisfied by buyers of B alone.

And finally $\theta a_h + \gamma b - p_h - p_B \lozenge Max(0; \theta a_l + \gamma b - p_{Gl}; \theta a_h - p_h; \gamma b - p_B)$ must be satisfied by buyers of both A_h and B.

[20] Consumers of $A_h + B$ means consumers of both A_h and B but each component is purchased separately.

From S_h and R_b, we derive that the preference index of the consumer indifferent between purchasing $A_l B$ and purchasing A_h and B separately is $\hat{\theta} \copyright \frac{p_{Gl} - p_h - p_B}{a_h - a_l}$. Now S_l, R_h, and R_l intersect at $I = \frac{p_h}{a_h}; \hat{\gamma} \copyright \frac{1}{b}(p_{Gl} - \frac{a_l}{a_h} p_h)$. $\hat{\gamma}$ can be understood as the implicit *price per unit of quality of B*, when the quality of *A* is valued at the price set by firm *h*. Note also that R_l, R_b, and S_b intersect at $\frac{p_{Gl} - p_B}{a_l}; \frac{p_B}{b}$ [21]. The market areas depend on whether this *implicit price* is greater, lower or equal to the explicit *price per unit of quality* of *B* set by firm *h*. There are now four cases to consider that correspond to number of cells in Table 1.[22]

Table 1: Different cases under the regime of (C_h, B_l)

	$\hat{\gamma} < \frac{p_B}{b}$	$\hat{\gamma} \lozenge \frac{p_B}{b}$
$p_h \leq p_{Gl} < p_h + p_B$	Case 1	Case 2
$p_{Gl} < p_h$	Case 3	Case 4

In all cells except cell 3 we have $\hat{\theta} < 1$, i.e. the preference parameter of the consumer indifferent between $A_l B$ and both A_h and *B* is lower than one. For cell 3, we consider separately the case where $\hat{\theta} < 1$ and the case where $\hat{\theta} \lozenge 1$.

---

[21] Also remark that $\frac{p_B}{b} < \frac{1}{b}(p_{Gl} - \frac{a_l}{a_h} p_h) \copyright \hat{\gamma}$ can be written as: $\frac{p_{Gl} - p_B}{a_l} > \frac{p_h}{a_h}$. The ratio $\frac{p_{Gl} - p_B}{a_l}$ represents the *implicit price per unit of quality* of A_l when it is sold as part of the bundle $A_l B$. Thus market areas depends on whether this *implicit price per unit of quality* is greater, lower or equal to the explicit *price per unit of quality* of A_h.

[22] There is no equilibrium with $p_{Gl} > p_h + p_B$ because nobody buys $A_l B$.

Case 1: $p_h \le p_{Gl} < p_h + p_B$ and $\hat{\gamma} < \frac{p_B}{b}$

Figure 11 shows a segmentation of the space of preference parameters into five segments. $\hat{\theta} = \frac{p_h + p_B - p_{Gl}}{a_h - a_l}$ is derived from S_h.

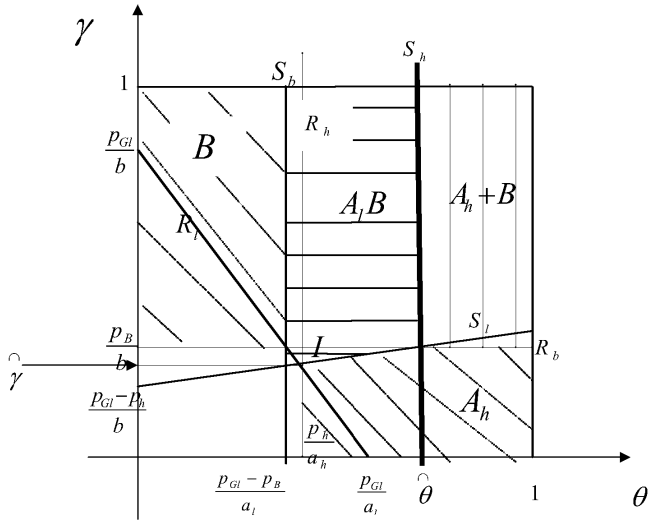


Figure 11: Market areas under the regime (C_h, B_l) when $p_h \le p_{Gl} < p_h + p_B$ and

$\hat{\gamma} < \frac{p_B}{b}$

Consumers with preference parameters to the left of S_h, to the right S_b and above S_l and R_l purchase A_lB. The market area of those consumers is:

$$D_{Gl} = (1 - \frac{p_B}{b})(\hat{\theta} - \frac{p_{Gl} - p_B}{a_l}) + \frac{1}{2}(\frac{p_B}{b} - \hat{\gamma})(\hat{\theta} - \frac{p_{Gl} - p_B}{a_l}),\ D_{Gl} = (\hat{\theta} - \frac{p_{Gl} - p_B}{a_l})(1 - \frac{p_B}{2b} - \frac{\hat{\gamma}}{2}).$$

Consumers with preference parameters to the right of S_h and above S_b purchase both A_h and B. Consumers with preference parameters to the left of S_b and above R_b, purchase B alone.

Finally consumers with preference parameters to the right of R_h and below S_l purchase A_h only. Demands D_{A_h} for A_h and D_B for B are respectively:

$$D_{A_h} = 1 - \overset{\frown}{\theta} + \frac{1}{2}(\overset{\frown}{\theta} - \frac{p_h}{a_h})(\frac{p_B}{b} + \overset{\frown}{\gamma}), \text{ and } D_B = (1 - \frac{p_B}{b})(1 - \overset{\frown}{\theta} + \frac{p_{Gh} - p_B}{a_l}).$$

The others consumers do not purchase.

Case 2: $p_{Gl} < p_h$ and $\overset{\frown}{\gamma} < \frac{p_B}{b}$

When $\overset{\frown}{\theta} < 1$

Figure 12 displays also a segmentation of the space of preference parameters into five segments. We obtain the same segmentation of preference parameters space than in case 1. Therefore demands in case 2 when $\overset{\frown}{\theta} < 1$ are similar to demands in case 1.

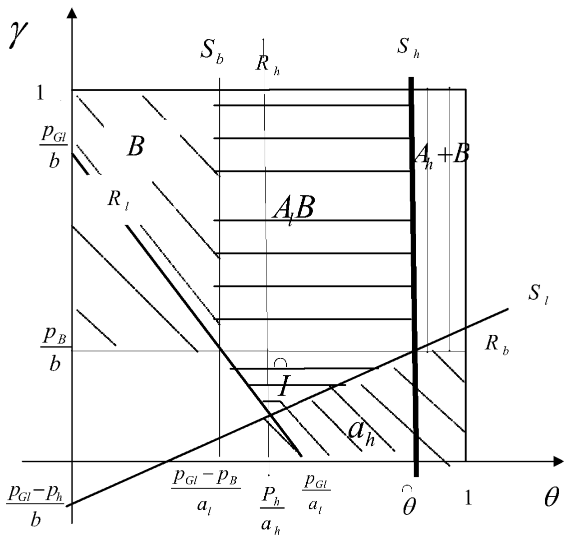


Figure 12: Market areas under the regime (C_h, B_l) when $p_{Gl} < p_h$, $\overset{\frown}{\gamma} < \frac{p_B}{b}$ and $\overset{\frown}{\theta} < 1$.

When $\hat{\theta} \Diamond 1$

Figure 13 shows a segmentation of the space of preference parameters into three segments.

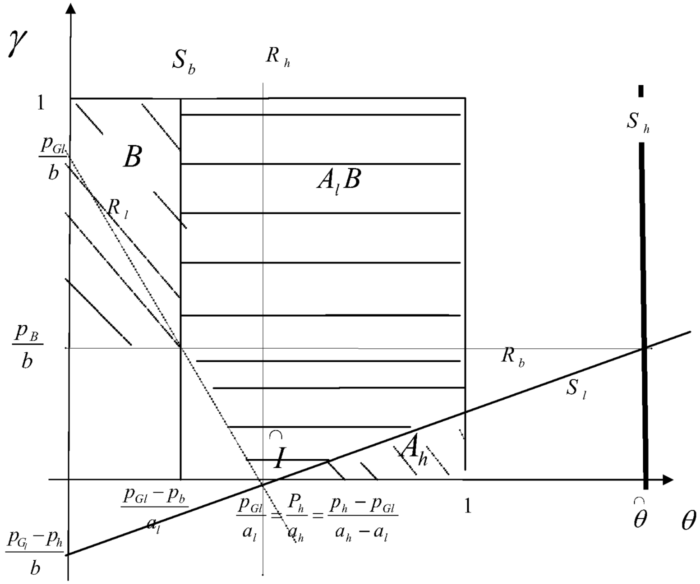


Figure 13: Market areas under the regime (C_h, B_l) when $p_{Gl} < p_h$, $\hat{\gamma} < \frac{p_B}{b}$ and $\hat{\theta} \Diamond 1$

Consumers with preference parameters to the right of S_b, above R_l and above S_l purchase $A_l B$. Those with preference parameters to the left of S_b and above R_b , purchase B alone. Consumers with preference parameters below S_l purchase A_h only, to the right S_b and above S_l and R_l purchase $A_l B$. Others don't purchase. No one purchases both A_h and B.

Case 3: $p_h \leq p_{Gl} < p_h + p_B$ and $\hat{\gamma} \Diamond \frac{p_B}{b}$

Figure 14 shows segmentation of the space of preference parameters into four segments.

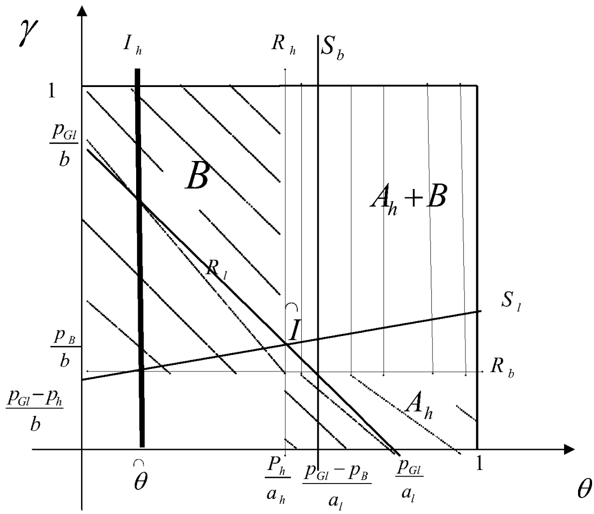


Figure 14: Market areas under the regime (C_h, B_l) when $p_{Gl} < p_h + p_B$, $\hat{\gamma} \Diamond \frac{p_B}{b}$ and $p_{Gl} > p_h$

Consumers with preference parameters to the right R_h and above R_b purchase both A_h and B. Those with preference parameters to the left of R_h and above R_b , purchase B alone. Consumers with preference parameters to the right R_h and below R_b purchase A_h only. Others don't purchase. Demand for $A_l B$ is zero.

Case 4: $p_{Gl} < p_h$ and $\hat{\gamma} \Diamond \frac{p_B}{b}$

Figure 15 depicts the same segmentation of preference parameters space than in case 3. Therefore demands in cases 4 are similar to demands in case 3.

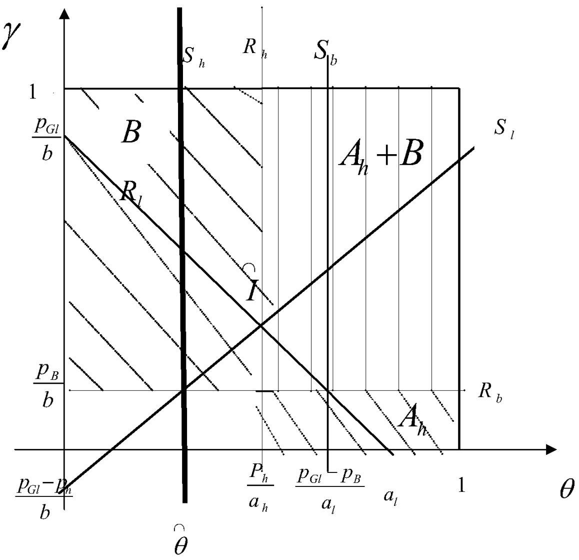


Figure 15: Market areas under the regime (C_h, B_l) when $p_{Gl} < p_h$ and $\hat{\gamma} \Diamond \frac{p_B}{b}$

3.3.2 Determination of the equilibrium prices in (C_h, B_l)

We note that there is no duopoly equilibrium with $\overset{\cap}{\gamma} \Diamond \frac{p_B}{b}$ because this condition entails zero sales by the low quality firm. Therefore, the remaining case to analyze is the case where $\overset{\cap}{\gamma} < \frac{p_B}{b}$ and $\overset{\cap}{\theta} < 1$ or $\overset{\cap}{\theta} \Diamond 1$. So the question is what subcase constitutes a Nash equilibrium in prices?

As in the previous sections, analytical difficulties lead us to search the price equilibria numerically for a range of values of a_h and a_l. Figure 16 displays the value of $\overset{\cap}{\gamma}$ and $\frac{p_B}{b}$ for a range of values of a_h[23]. We find that the only equilibrium prices that we ever get always are consistent with $\overset{\cap}{\gamma} < \frac{p_B}{b}$. Also at equilibrium $\overset{\cap}{\theta} < 1$. Therefore we look only at case 1 when $\overset{\cap}{\theta} < 1$.

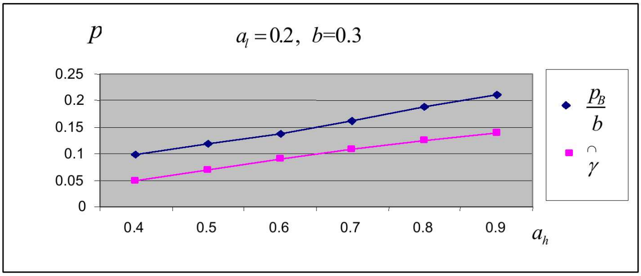


Figure 16: Equilibrium characteristic of the subgame (C_h, B_l)

Here also the gap between the explicit price of service *B* and the implicit price of service *B* becomes larger when a_h becomes larger.

[23] The equilibrium condition $\overset{\cap}{\gamma} \Diamond \frac{p_B}{b}$ is not satisfied for the parameter values that we have chooses for simulation.

4. Equilibrium Strategy of the Game

We now compare profit under each of the four possibilities. We let a_h and a_l vary in the interval $[0.2,+\times[$. Figure 17 shows firms' profit in each possibility. We always find that:

$$\pi_h^{B_h,B_l} \lozenge \pi_h^{C_h,B_l} \lozenge \pi_h^{B_h,C_l} \lozenge \pi_h^{C_h,C_l}$$

$$\pi_l^{B_h,B_l} \lozenge \pi_l^{B_h,C_l} \lozenge \pi_l^{C_h,B_l} \lozenge \pi_l^{C_h,C_l}$$

We conclude that (B_h, B_l) is an equilibrium in dominant strategy.

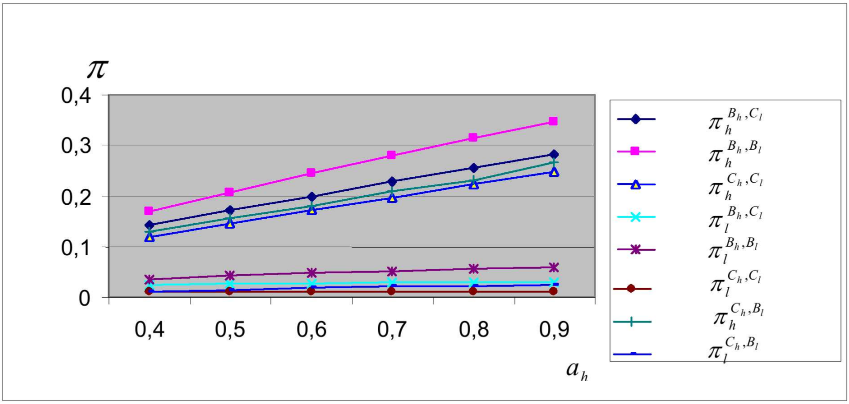


Figure 17: Equilibrium Strategy of the Game

Thus we state the following result:

(i) *Pure bundling is a dominant strategy equilibrium for both firms.*

Bundling is a dominant strategy for both firms because it reduces the intensity of the competition between the two firms by increasing the differentiation of services. Therefore firms' profits are higher under (B_h, B_l) than under the other subgames where one of two firms at least sells its services separately.

5. Welfare Implications

Now let us see how consumers' surplus and social welfare are affected when the regime shifts from (C_h, C_l) to (B_h, B_l).

For (C_h, C_l), consumers' surplus denoted by CS^S is:

$$CS^C = \int_{\tilde{\theta}}^{1} (\theta a_h - P_h) d\theta + \int_{\bar{\theta}}^{\tilde{\theta}} (\theta a_l - P_l) d\theta + \int_{0}^{1} (\gamma - 0) d\gamma .$$

The first, second and third expression of CS^S are respectively the surplus of purchasers of A_h, A_l and B. We obtain:

$$CS^C = a_h \left(\frac{1-\tilde{\theta}^2}{2}\right) - p_h (1-\tilde{\theta}) + a_l \left(\frac{\tilde{\theta}^2 - \bar{\theta}^2}{2}\right) - p_l (\tilde{\theta} - \bar{\theta}) + \frac{1}{2}$$

The social welfare denoted by SW^C gives:

$$SW^C = CS^C + \pi_h^{C_h, C_l} + \pi_l^{C_h, C} .$$

For (B_h, B_l), the surplus of consumers of $A_h B$ denoted by CS_h^B is:

$$CS_h^B = \int_{\theta^*}^{\frac{p_{Gh}}{a_h}} \int_{p_{Gh} - \theta a_h}^{1} (\theta a_h + \gamma - p_{Gh}) d\gamma d\theta + \int_{\frac{p_{Gh}}{a_h}}^{1} \int_{0}^{1} (\theta a_h + \gamma - p_{Gh}) d\gamma d\theta$$

$$CS_h^B = \frac{a_h}{2} + (1 + \theta^* - 2\frac{p_{Gh}}{a_h})(\frac{1}{2} - p_{Gh}) - a_h \frac{(\theta^*)^2}{2} + \frac{(a_h \theta^* - p_{Gh})^3}{6a_h}$$

While the surplus of consumers of $A_l B$ is:

$$CS_l^B = \int_{\gamma^*}^{p_{Gl}} \int_{\frac{p_{Gl} - \gamma}{a_l}}^{\theta^*} (\theta a_l + \gamma - p_{Gl}) d\theta d\gamma + \int_{p_{Gl}}^{1} \int_{0}^{\theta^*} (\theta a_l + \gamma - p_{Gl}) d\theta d\gamma$$

$$CS_l^B = \frac{\theta^*}{2} \left[1 + (p_{Gl})^2 - 2\gamma^* + a_l (1 - \gamma^*) \right] - \frac{(\gamma^* - p_{Gl})^3}{6a_l}$$

Thus, consumers' surplus under (B_h, B_l) is:

$$CS^B = CS_h^B + C_l^B$$

The social welfare under (B_h, B_l) denoted by SW^B is:

$$SW^B = CS^B + \pi_h^{B_h,B_l} + \pi_l^{B_h,B_l}$$

We now compare social welfare under (C_h, C_l) and (B_h, B_l) for a_h and a_l varying in the interval $[0.2,+\times[$. Figure 18 displays the total economic welfare under (C_h, C_l) and (B_h, B_l). It shows that:

$$SW^B > SW^C.$$

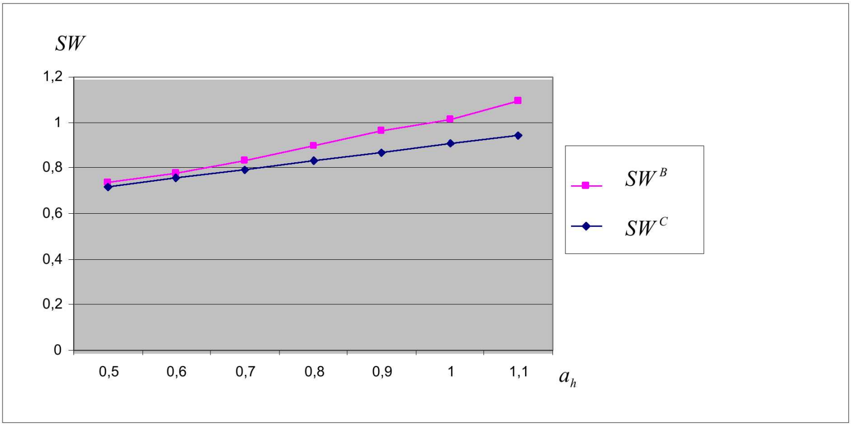


Figure 18: Total welfare under the regimes of (C_h, C_l) and (B_h, B_l)

Thus we state the following result:

(ii) *A shift from (C_h, C_l) to (B_h, B_l) results in a decrease in social welfare when there is a small differentiation between services.*

It is obvious that bundling in this context reduces consumers'surplus. Indeed bundling increases prices and there are more constraints under the regime of (B_h, B_l) because to obtain B consumers must purchase $A_l B$ even though they don't want A. So for welfare to increase, aggregate profits must raise enough to offset the reduction of aggregate consumer surplus and then result in a potential efficiency gain. For all the parameters we have chooseen we find that total welfare increase.

6. Concluding Remarks

In the traditional literature on bundling by duopolists, the conclusion is bundling is a dominant strategy equilibrium for both firms but it is not a profitable strategy for both firms. We find also that bundling is a dominant strategy equilibrium. But contrary to other studies, we find that bundling is a profitable strategy for both firms. The reason is that in the context of vertically differentiated services, bundling can be used as differentiation tool. It then reduces the intensity of competition between the firms and then they make more profits under bundling than they would under component selling. We also find that bundling increases total welfare.

Cable operators and telecommunication companies offer the combination of telephone, television, and Internet as a bundled service. Our results suggest that they would compete more vigorously and would realize less profit if there were restriction on bundling. If buyers care for the quality of services their surplus is reduced under bundling. However, the fact that a single supplier offers all services makes bundling convenient for buyers. So the benefit of the convenience must be balanced with the reduction of consumers' surplus to obtain the effects of bundling on buyers' welfare. Also authorities must decide what weight they give to the buyers' surplus and to the suppliers' profits to obtain the net welfare effects of bundling.

Our result should be interpreted under the assumptions that both the variable cost and the fixed cost are zero for each service and services are vertically differentiated. It would be interesting to analyze the case where the fixed costs are positive. Also since we know that firms in communication markets are both horizontally and vertically differentiated, we can study both differentiations to see how the results of this paper are robust to these assumptions.

References

Adams, W., J., and Yellen, J., L., 1976, "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, 90, pages 475 –98.

Anderson, S and Leruth, L., 1993, "Why firms may prefer not to price discriminate viamixed bundling," *International Journal of Industrial Organization*, Elsevier, vol. 11(1), pages 49-61.

Economides, N., 1993, "Mixed Bundling in Duopoly. Stern School of Business," *working paper*, *Stern School of Business, N.Y.U.*

Kopalle, P., Krishna, A., and Assuncao, **J.,** 1999, "The Role of Market Expansion on Equilibrium Bundling Strategies", *Managerial and Decision Economics*, 20, 365-377.

Matutes, Carmen and Regibeau, **Pierre**, 1992, "Compatibility and Bundling of Complementary Goods in a Duopoly," *Journal of Industrial Economics*, 40,37-54.

McAfee .R. Preston, McMillan, John, and Whinston, Michael D, 1989, "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *Quarterly Journal of Economics,* 84, 271 –284.

Mussa, Michael and Sherwin Rosen, 1978, "Monopoly and Product Quality," *Journal of Economic Theory,* 18, 301-317.

Schmalensee, Richard, 1982, "Commodity Bundling by Single-Product Monopolies,"*Journal of Law and Economics*, 25, 67 –71.

Schmalensee, Richard, 1984, "Gaussian Demand and Commodity Bundling," *Journal of Business*, 57, 58 –73.

Tirole, Jean, 1988, "*The Theory of Industrial Organization*," The MIT Press.

Whinston, Michael .D., 1990, "Tying Foreclosure, and Exclusion," *American Economic Review* 80, 837 –59.

Conference on the economics of Information and Communication Technologies
Paris, June 23-24, 2006

Plenary Session 5:
"Antitrust Policy, the Theory of Market Leaders and the New Economy" by Federico Etro

Comments by David Encaoua, Centre Economie Sorbonne, Université de Paris 1

There are three possible lectures of this interesting paper by Federico Etro.

1. It can be seen as an economic defense of Microsoft in its case against the European Commission.
2. It has also a more general objective as it presents a *new* economic argument in favour of business practices by dominant firms in the New Economy that were considered abusive from an *old* antitrust perspective.
3. Finally it presents a contribution in theoretical IO, which overcomes the difficulties raised by the usual interpretation of the commitment strategies made by Stackelberg leaders, according to whether they are aggressive or accommodating.

I will concentrate my comments on discussing the theoretical contribution and presenting some remarks on the links between competition and innovation in the New Economy.

I/ The theoretical contribution.

The central contribution of Etro (Aggressive Leaders, *RJE*, 2006) states that a Stackelberg leader, who commits to an investment that affects the outcome of a free entry competition, chooses an investment level that makes him always aggressive rather than accommodating, **independently of the nature of market competition**.
In other words, in a free-entry equilibrium, the Stackelberg leader is either a Top Dog (i.e. he over invests) when the investment increases his marginal profit or a Lean and Hungry Look (i.e. he under invests) when the investment decreases his marginal profit.
This important result must be contrasted to the usual outcome predicted by Fudenberg and Tirole (The Fat Cat effect, the Puppy Dog Ploy and the Lean and Hungry Look, *AER*, *PP*, 1984), according to which an aggressive behaviour holds when competition in the market occurs in strategic substitutes (quantities), whereas, with strategic complements (prices), the strategic leader is accommodating (he is a Puppy Dog when the investment increases the marginal profit and a Fat Cat when the investment decreases the marginal profit).
The main difference between Fudenberg and Tirole and Etro results lie in the assumption of **free-entry** that Etro introduces.

Introducing **Stackelberg with free-entry** appears therefore as an interesting and fruitful idea and one must recognize that this framework gives insightful results to analyze some business practices by market leaders.

Let me make three comments on the theoretical contribution.

First comment: The framework assumes that the external effect in the profit function is **additively separable and perfectly symmetric**.

a/ While this assumption covers a wide array of situations, it leaves away some others. For instance, linear demands derived from a quadratic utility or from a vertical differentiation framework do not fit this assumption whenever price competition prevails. It would be interesting to examine how the main results are affected by relaxing this assumption. Particularly, if the external effect depends on the identity of each competitor, it seems unlikely that the leader's investment affects only the number of competitors (in the free-entry equilibrium) and not their own strategies, as it happens under the present assumption.

b/ Moreover, the presence of heterogeneous competitors may be an important determinant of the intensity of competition. In a recent paper, Jan Boone (Balance of power, mimeo, Tilburg U. 2004) argues that the efficiency distribution of players in a game determine how aggressively these players interact. Players fight *aggressively inefficient players* but play *softly against equally or more efficient players*. This idea leads to predictions concerning the role of entry that are different from Etro. While Etro finds that free entry of equally efficient competitors leads to an aggressive behaviour of the market leader, Boone's finds that entry by new firms leads to a less aggressive outcome if it creates a *balance of power* with the incumbent.
How can one reconcile these two antagonistic results: in one hand entry by equally efficient firms pushes incumbents toward an accommodating behaviour (Boone's perspective) and in the other hand entry by equally efficient firms pushes market leaders to an aggressive behaviour that results in less entry (Etro's perspective)?

c/ Third, it is quite possible that a symmetric model leads to an asymmetric equilibrium outcome. For instance, this would arise if profit functions were convex rather than concave. In this case, corner solutions arise. An asymmetric outcome would also prevail in presence of network externalities that lead to tipping to a peculiar technology, which is the only stable equilibrium.

d/ Finally, many other questions arise that are not addressed by Etro. For instance, what happens whenever entry is not simultaneous as in Nash equilibrium, but rather sequential? Who supports the cost of the strategic investment that reduces the number of entrants when coexist in the market more than one leader?

It would be interesting to address these questions or at least some of them to test the robustness of the outcome of the Stackelberg equilibrium with free-entry. Once more, the idea of introducing this equilibrium is very interesting but the results depend so far on the specific assumption on the external effect.

Second comment: The more or less explicit objective of the aggressive leader's theory is to give some new economic foundations to antitrust policy, based on the following principle: while business practices by market leaders are aggressive against competitors, they do not harm consumers and therefore they should not be considered as being abusive.

I think that this corresponds presumably to an excessive interpretation of the aggressive leader's theory. Let me clarify this point.

The distinction between *normal* competition and *abusive* competition is not evident for an economist. I agree completely with Etro's claim that preserving the competitive process must not be confused with protecting competitors and I also agree that the application by the Commission of Article 82 related to abuse of dominance has been misleading. This is likely a consequence of the German *ordo-liberal* doctrine according to which the favoured competition law standard should be to constrain dominant firms to act as if they were constrained by competition (or in another way as if they were not dominant!) The practical implication of this doctrine has been to disallow *impediment competition,* hindering competitor's ability to offer better deals to consumers, while allowing *performance competition* based on competition on the merits. Impediment competition is *abusive* while performance competition is *normal.*

However, the distinction between *normal competition* and *abusive competition* is very difficult to reach and may lead to mistakes betrayed by the Coase famous criticism: "*If an economist finds something – a business practice of one sort or another – than he does not understand, he looks for a monopoly explanation. And as in this field we are very ignorant, the number of un-understandable practices tends to be very large, and the reliance on monopoly explanations, frequent*". (Industrial Organization: a proposal for research, in Fuchs ed. NBER, 1972)

In some sense, one objective of Etro's analysis is to try to give a more precise meaning on what is competition on the merits and what could be achieved under it. According to Etro, *normal* competition between market incumbents and potential competitors should recognize that market leaders have a first mover advantage and that they know that entrants compete until achieving any profit opportunity, all of this being common knowledge. The consequence is that some practices that are considered on a priori grounds as being abusive appear to be rather aggressive and in favour of consumers when they are examined with this precise perspective in mind. I agree.

However, it does not mean that **all** business practices emanating from market leaders are necessarily free of charge. In my view, the only practical policy recommendation that can be derived is that the market leaders practices must be evaluated according to their consequences (*economic effects-based* approach) rather than assessed according to a priori categorizations (*form-based* approach), or in other words that application of a *rule of reason* must be definitely preferred to a *per se* prohibition.

The underlying principles by reference to which *abusive conduct* can be distinguished from *normal competition* on the merits have to be clarified.

Three approaches (John Vickers, Abuse of market power, *EJ*, 2005) have been proposed to solve this question: the *sacrifice test*, the *as-efficient competitor test* and the *consumer harm test.*

The *sacrifice test* asks whether the dominant firm business practice in question would be profitable independently of its intent to eliminate or lessen the number of competitors. Clearly, this test does not satisfy the Stackelberg equilibrium with free entry equilibrium. This test can also be rejected on the logical grounds that eliminating or lessening competitors must not be discarded since it is a form of normal competition.

The *as-efficient competitor test* asks whether the dominant firm business practice in question eliminates competitors that are as efficient as the dominant firm. While this test appears to better accord with protecting competition rather than competitors, it does not satisfy any more the Stackelberg equilibrium with free entry framework. Moreover, the as-efficient test is difficult to enforce, particularly when products are differentiated.

Finally, the *consumer harm test* asks whether the dominant firm business practice in question excludes competitors whose presence would enhance consumer surplus. It is probably the most coherent test. It is important to remark that it may lead to an outcome that is quite different from the one reached in a Stackelberg equilibrium with free-entry. The main reason is that free entry may involve either excessive or insufficient entry relatively to the social optimum.

This is why, in my view, it is very difficult to transform the aggressive leader's theory into an economic principle on which antitrust policy could rely.

Third comment: A deeper discussion of **barriers to entry**, putting more emphasis on the distinction between economic and antitrust barriers to entry, would be useful. The recent contribution by McAfee, Mialon and Williams (What is a barrier to entry? *AER*, *PP*, 2004) could be used as a starter. An economist's view of barriers to entry seems more in line with Stigler's perspective according to which barriers to entry exist when there is some cost that is borne by the entrants but not by the incumbents. The antitrust perspective seems to be slightly different. It identifies a barrier to entry to any cost that simply *delays* entry. Therefore, while Stackelberg equilibrium with free entry is in line with Marshall Analysis and corresponds to long-term equilibrium, it is not clear what time horizon should be taken in an antitrust perspective. If free entry involves a very long delay, the antitrust authority could be interested by the short or medium term consequences of a business strategy. This is another reason why Stackelberg with free entry may be an inappropriate framework for antitrust purpose.

II Antitrust and innovation in the New Economy.

Does the New Economy, with its focus on innovative markets, require a new antitrust policy approach? This is a largely debated question to which Etro seems to argue for a definitely Yes answer. Let me develop here some arguments (Encaoua and Hollander, *OXREP*, 2002).

Competition policy in innovative industries is liable to affect welfare through three distinct channels. First is the intensity of the race to innovate. In innovative industries, competition *for* the market is crucial. Second is the competition for knowledge diffusion or in other words *competition in the licensing and technology markets*. Third is the *competition in the new product markets* that affect variables such as prices and qualities. These three channels are closely intertwined and raise interesting questions, both at a conceptual level and at a policy perspective level.

Conceptual level: Clearly, competition *for* the market and not only competition *in* the market matters in a dynamic framework. In the Schumpeterian endogenous growth literature, it has been generally assumed that the market life of a technological leader is equal to the expected life of its innovation. As soon as a new innovation appears, even if non-drastic, the previous leader is replaced by the new innovator and this process of *permanent leapfrogging* or creative destruction is supposed to be at the

heart of economic growth. Clearly, this *no leadership persistency* assumption contradicts or at least does not fit the real word observation.
In one of his papers, Etro (Innovation by Leaders, *EJ*, 2004) shows that in a patent race involving a technological incumbent and an endogenous number of competitors determined by the free-entry condition, the incumbent has a higher incentive to innovate when endowed with a first mover advantage in choosing his R&D effort. This result, which goes against the classical *replacement effect*, shows once more the strength of the notion of Stackelberg equilibrium with free entry[1]. There is now a growing part of the endogenous growth literature that tries to depart from the leapfrogging assumption.
Can we draw some robust conclusions from these analyses concerning the respective role of incumbents and entrants in the process of innovation? I think it would be premature to do it. The welfare comparisons of the leapfrogging and persistent-leadership equilibria are still incomplete. Even if it is justified to abandon the idea that leadership persistency plays no role in the innovation and growth process, it would be un-wise to abandon the idea that the path-breaking avenues opened by new entrants and the Darwinian selection mechanisms remain prominent in the process of competition *for* the market. It seems difficult to derive from these models firm conclusions prescribing neither to exempt technological leaders from antitrust scrutiny nor to exempt newcomers from antitrust law.
Therefore, it seems to me that antitrust policy does not go in a bad or wrong direction when it enlarges its scope by trying to maintain rivalry not only in the product market but also in the *technology market* and the *innovation market*.

Policy level: Two important questions arise.
The first question concerns the links, more precisely the strong complementarities that exist between technology policy and competition policy in the New Economy.
The second question concerns the relationship between competition and intellectual property. In my view, these questions are among the most challenging policy questions of the New Economy. I have no time to develop them in detail here, but I would just make two remarks[2].

The problem of the right balance between integrated technology policy and competition policy.

In the New Economy, we need a right balance between competition policy and technology policy, because, contrary to what is sometimes believed, especially among antitrust enforcers, competition policy alone, with its emphasis on free market principles, does not provide sufficient incentives to innovate, even in the presence of intellectual property rights. Many reasons can be invocated.
- One of them is that the incentive to innovate is not perfectly correlated with the intensity of market competition. Theoretical and empirical analyses (Aghion and Griffith, Competition and Growth, 2005, MIT Press) show that there exists an inverted U relationship between competition and innovation. After some threshold, the

[1] Denicolo (Growth with non-drastic innovations and the persistence of the leadership, *EER*, 2001) reaches also a similar conclusion. When the current leader has a move advantage in the next patent race, the value of being a leader exceeds temporary monopoly profit, as the leader reaps an extra-profit in the next patent race with free entry by outsiders.
[2] More details can be found in the report by Encaoua – Guesnerie (2006) "Les Politiques de la Concurrence" that can be downloaded: http://www.cae.gouv.fr/

competition pressure is no more associated to a higher incentive to innovate[3]. Moreover, competition intensity does not appear as the most important determinant of the incentive to innovate.

- Another reason is that in global markets, intellectual property rights are less protective as they would be if leaders of industrialized countries were absent from countries where IP is less stringent. Even in industrialized countries, patents are not ironclad property rights[4].

- A third reason is that the complementarities between different dimensions of structural economic policies, such as competition, regulation and innovation, are not always exploited. For instance, it is not sufficient to rely on environmental objectives alone to promote the introduction of cleaner technologies. The achievement of environmental as well as innovation goals requires a carefully crafted combination of environmental and technology policies.

For all these reasons, a right balance between different public policy instruments to stimulate innovation and complement individual incentives is needed. In a global world, such coordination is unavoidable, even if it is very complicated to enforce. For instance, it is now largely admitted that private partnerships to form joint research ventures exempted from antitrust enforcement may be beneficial. Private-public partnerships involving science-industry cooperation are also one form of complementarities that has to be encouraged.

It has been argued that one of the reasons behind the current technological gap between Europe and US is the unbalanced development of the structural policies devoted respectively to technology and competition in Europe. While technological policy remains largely un-integrated at the European level, an integrated competition policy has been used as the main instrument of the European market. By contrast, the complementarities between technology policy and competition policy have been largely perceived and exploited in the US[5].

Intellectual property rights and competition rules.

It is now largely admitted that the reinforcement of patent rights and the extension of the patentability subject matters are not necessarily coincident with higher R&D effort. One of the reasons is that patents are not ironclad property rights but only probabilistic rights. Other forms of protection, such as trade secrecy, first mover advantages and exploitation of lead-time are very often preferred forms of protection, at least in industries where knowledge is not a blue print easily imitable. However, even if patents do not always appear as the best form of protection, innovators may nevertheless prefer to patent their innovations for many reasons, one of them being that holding a patent offers the possibility to avoid a trial and settle a dispute against an alleged infringer through a licensing agreement. Alleged infringers may also prefer to avoid a litigation process not only because litigation is costly and uncertain but also because winning the lawsuit against a patent holder involves a free-riding aspect, insofar as other competitors benefit from the asserted patent's invalidity. Therefore,

[3] For a challenging view, see Encaoua – Ulph (2004), Catching-up or Leapfrogging ? The effects of competition on innovation and growth, mimeo, EUREqua, Paris I.

[4] See Encaoua – Lefouili (2006), Choosing Intellectual Protection : Imitation, Patent Strength and Licensing, CESifo WP 1715

[5] The chapter 8 (Innovation Today: A Private-Public Partnership) of the recent Scotchmer's book (Innovation and Incentives, MIT 2004) is entirely devoted to illustrate the role of such complementary instruments as public procurement, R&D incentives given to small enterprises, licensing and antitrust exemptions, spin offs between military and civil projects, etc.

private settlements taking the form of a licensing agreement offer an alternative to litigation and are very frequent. However private settlements raise serious concerns to antitrust enforcers[6].

Once again, IPR, knowledge diffusion and competition issues appear at the forefront of the New Economy and it would be too simplistic to reduce these questions to the aggressive leader's theory, despite the big interest of this theory.

[6] Farrell and Shapiro (How strong are weak patents? Mimeo Berkeley, 2005) explain why the royalty rate paid by the licensee may not depend on the patent strength, as a natural benchmark would command. Licensing very bad quality patents may involve as-high royalty rate as if the patent was undoubtful. Moreover the patentee pays a fixed fee to the licensee to compensate its loss in the market. While the two parties maximise their joint profits, such settlements harm consumers.

Antitrust Policy, the Theory of Market Leaders and the New Economy

by Federico Etro[1]
June 2006

The New Economy, characterized by dynamic, global and innovative markets, requires a new way to approach many economic issues and also a new way to approach policymaking. This work will analyse a new approach toward competition policy based on recent progress in the theory of market leaders and discuss its implications with special reference to markets in the New Economy, whose distinctive features, namely high fixed costs of R&D, less relevand marginal costs of production and network effects, require a different approach from traditional markets. Close attention will be goven to the software market, whose market leader has been (and still is) the subject of the attention of antitrust authorities around the world. For a more comprehensive discussion on these theoretical and applied issues see Etro (2007).

The work is organized as follows. In Section1 I will present a brief overview of antitrust policy in US and EU and I will try to motivate the need for a new approach to competition policy, especially for the markets in the New Economy. Section 2 will survey the traditional approaches to competition policy, while Section 3 will present the innovations associated with the theory of market leaders. Section 4 will apply the new approach to general issues of abuse of dominance with particular reference to the software market and to the Microsoft case. Section 5 will deal with bundling issues again with reference to the software market. Sections 6 will move to competition for the markets and to interoperability issues which are crucial for the dynamic markets of the New Economy. Section 7 concludes, while the Appendix contains some more technical results on the behaviour of market leaders.

1 Competition Policy in US and EU

In the United States the main federal antitrust statute is the Sherman Act of 1890, which was developed in reaction to the widespread growth of large scale business trusts. Section 1 prohibits restraints of trade in general, while Section 2 deals with monopolization stating that:

[1]The author is Associate Professor of Economics at the University of Milan (Bicocca). Email: fetro@intertic.org.

> "Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony"

Enforcement is shared by the Antitrust Division of the Department of Justice and by the Federal Trade Commission. The current interpretation of US antitrust law associates abusive conduct with predatory or anticompetitive actions having the specific intent to acquire, preserve or enhance monopoly power distinguished from acquisition through a superior product, business acumen or historical accident. It is generally accepted that an action is anticompetitive when it harms consumers.

In Europe competition policy has a more recent history which is mostly associated with the creation of the European Union and its coordination of policies for the promotion of free competition in the internal market. The main provisions of European Competition Law concerning abuse of dominance are contained in the Article 82 of the Treaty of the European Communities which states that:

> "Any abuse by one or more undertakings of a dominant position within the common market or in a substantial part of it shall be prohibited as incompatible with the common market in so far as it may affect trade between Member States. Such abuse may, in particular, consist in: (a) directly or indirectly imposing unfair purchase or selling prices or other unfair trading conditions; (b) limiting production, markets or technical development to the prejudice of consumers; (c) applying dissimilar conditions to equivalent transactions with other trading parties, thereby placing them at competitive disadvantage; (d) making the conclusion of contracts subject to acceptance by other parties of supplementary obligations which, by their nature or according to commercial usage, have no connection with the subject of such contracts."

This article (as Article 81 on horizontal and vertical agreements and the Merger Regulation) is part of the law of each member state and is enforced by the European Commission (in particular the Directorate General for Competition) and by all the National Competition Authorities. The application of EU competition law on abuse of dominance involves the finding of a dominance position and of an abusive behaviour of the dominant firm, usually associated with excessive pricing or with exclusionary practices as predatory pricing, rebates, tying or bundling, exclusive dealing or refusal to supply. However, the analysis of both dominance and abusive behaviour entails complex economic considerations and is the subject of an on going process of revision. A recent document, European Commission (2005), has proposed a new approach to exclusionary abuses under Article 82 which gives an important indication as to how the Commission may approach antitrust cases in the future. The purpose of Article 82 is defined as "the protection of competition on the market as a means of enhancing consumer welfare and of ensuring an efficient allocation of resources". This implies that antitrust should protect competition and not competitors and be based on an economic approach aiming at the maximization of consumer welfare and allocative efficiency rather than based on a legalistic approach, something which appears much more in line with the US approach.

Many economists have pointed out the necessity of a closer focus on consumer welfare in the implementation of competition policy with specific reference to abuses of dominance. While antitrust legislation was written with this objective in mind, its concrete application, especially within the post-Chicago approach, has often been biased against market leaders and in defence of their competitors rather than toward the defence of competition and of the interests of consumers. The two objectives do not necessarily overlap. The development of the New Economy, characterized by very dynamic and innovative markets, has increased the pressure for a new approach, already somewhat developed in the United States, but just in progress in the European Union.

A new approach to competition policy should be based on rigorous economic analysis, from both a theoretical and an empirical point of view. In an important EU Report, Rey *et al.* (2005) emphasize this element in the antitrust procedure: "a natural process would consist of asking the competition authority to first identify a consistent story of competitive harm, identifying the economic theory or theories on which the story is based, as well as the facts which support the theory as opposed to competing theories. Next, the firm should have the opportunity to present its defence, presumably to provide a counter-story indicating that the practice in question is not anticompetitive, but is in fact a legitimate, perhaps even pro-competitive business practice." Moreover, any theory of the market structure able to provide guidance in detecting abuses of dominant positions should: 1) take into account the role and the strategies of dominant firms; 2) describe the equilibrium outcomes taking into account the role of barriers to entry and of fixed costs of entry (which can endogenously determine entry of competitors) and in function of the demand and supply conditions; and 3) provide welfare comparisons under alternative set-ups.

In this paper I will try to argue that, while the Chicago school and the post-Chicago approach failed to provide a unified framework which matches these requirements, the theory of market leaders formalized for instance in Etro (2004, 2006a,b, 2007) has provided a possible alternative. The general principle proved in this new research is that dominant firms may behave in an anti-competitive way, accommodating or predatory, in markets where the number of firms is exogenous, while they always behave in an aggressive way when entry into the market is endogenous, which should be the relevant case in many situations: in these, a large market share of the market leader is a consequence of its aggressive strategies and of the endogenous entry, and not the consequence of its market power. Hence, markets with high concentration due to the presence of a market leadership are perfectly consistent with efficiency. This has major implications for competition policy: while the old approach to abuses of dominant positions needs to verify dominance through structural indicators and the existence of a certain abusive behaviour, we suggest that there is not a well founded reason to associate high levels of concentration with market power and a consistent approach to abuse of dominance would just need to verify the existence of harm to consumers. As Rey *et al.* (2005) correctly point out, "the case law tradition of having separate assessments of dominance and of abusiveness of behaviour simplifies procedures, but this simplification involves a loss of precision in the implementation of the legal norm. The structural indicators which traditionally serve as proxies for 'dominance' provide an appropriate measure of power in some markets, but not in others", notably in markets where entry is an important factor (a concentration index is uniquely concerned with actual competition and ignore potential competition) and when innovation is important (a concentration index can deal with competition *in* the market, not *for* the market).

The main policy implication of the theory of market leaders emerges under (imperfect) competition in prices. In this typical situation, while the traditional approach (the so-called post-Chicago approach) tends to associate aggressive pricing strategies with predatory (and hence anti-competitive and welfare-decreasing) purposes, in Etro (2006a) we proved that, whenever entry of firms in the market is endogenous (as often it is), an aggressive pricing strategy is a pro-competitive strategy which generally does not have an exclusionary purpose, but rather can enhance allocative efficiency and consumer welfare. The same holds for other strategies which typically have exclusionary motivations in the traditional view, like bundling strategies. Clearly, when entry is not endogenous and the leader and its followers cannot be threatened by further entry, the behaviour of the leader could be anti-competitive in line with the post-Chicago approach, but antitrust should intervene only in these cases (beyond its usual role against collusive behaviour).

Finally, notice that what matters is not only welfare of current consumers but also that of future ones. Rey *et al.* (2005) provide a simple example on the problem of monopoly pricing: "One response to the problem might be for the competition authority to intervene, citing excessive pricing by a monopolist as an infraction of the abuse-of-dominance prohibition in Article 82 of the Treaty. Another response might be to leave the matter alone, hoping that the profits that the monopolist earns will spur innovation or imitation and entry into the market, so that, eventually, the problem will be solved by competition." What the theory of market leaders suggests on this matter, as shown in Etro (2004), is that the dynamic gains in efficiency due to a leadership position in innovative markets can be quite high as long as entry in the market for innovation is endogenous: the leadership of a firm may persist because of its high incentives to invest in R&D under the threat of entry; nevertheless, this should not be seen as a signal of abusive conduct, but, oddly enough, as the result of competitive pressure.

The recent document of the European Commission (2005) has inspired a wide debate on the proper aims and methods of antitrust policy in Europe. While the aim of this proposal is to enhance consumer welfare and to protect competition and not competitors, we have some concern that these principles are not fully carried through into certain aspects of the analytic framework. As of now, the approach of the European Commission appears partly in line with outdated views, for instance when it stresses an excessive reliance on market shares in determining dominance. The novel part on the efficiency defences for dominant firms appears to be going in the right direction since it allows otherwise abusive strategies if they create a net efficiency gain (which benefits consumers).[2] Nevertheless, the effectiveness of these rules in safeguarding consumer welfare is weakened when it is stated that some firms are virtually excluded from the possibility of an efficiency defence. In particular, a strange concept of market position "approaching that of a monopoly" is introduced and associated with market shares above 75%, something without any justification in economic theory: a firm is a monopoly or is not (in which case, its behaviour is constrained by competitors), but it cannot be an "almost monopoly" or a "near monopoly". From an economic point of view, the real missing concept, which defines firms with high market shares but not monopolizing the all market, is that of a Stackelberg leader with endogenous entry, which is the subject of the analysis of the theory of market leaders (see ICC, 2006, for a more extensive discussion of such an approach).

2 Chicago and post-Chicago Approaches

In this section I am going to review the traditional approaches to antitrust policy on abuse of dominance and start comparing them with the insights of the recent theoretical attempts to build a comprehensive theory of market leadership and competition policy. In our view, a fully fledged model of the behaviour of market leaders is a necessary toolkit for deriving implications for antitrust policy, but it is not necessarily part of the endowment of the traditional theories.

The traditional "pre-Chicago" approach was mostly based on basic models of imperfect competition associating market power, high market shares and abusive conduct with the typical behaviour of monopolists. Such a *naïve* view has been challenged in the 60s and 70s by the "Chicago approach" whose main merit has been to show that, when there are potential entrants in a given sector,

[2]This can happen in two ways: through an *objective necessity defence* "where the dominant company is able to show that the otherwise abusive conduct is actually necessary conduct on the basis of objective factors external to the parties involved and in particular external to the dominant company" or a *meeting competition defence* "where the dominant company is able to show that the otherwise abusive conduct is actually a loss-minimising reaction to competition from others".

aggressive strategies that would be suspect, such as bundling, price discrimination and exclusive dealing, are not necessarily anti-competitive but may instead have a strong efficiency rationale. More recent theories, often associated with the so-called "post-Chicago" approach, have however shown that in the presence of pervasive market imperfections, the above strategies can be anti-competitive because they are aimed at deterring entry in the short run and protect monopolistic rents in the long run. Broadly speaking, US antitrust authorities have been highly influenced by all these approaches over time, while it is hard to claim that the same is true of the EU antitrust authorities. As has recently been pointed out by Ahlborn, Evans and Padilla (2004), "in Europe it has taken longer for new developments in economic theory to affect competition policy. While U.S. antitrust has been influenced by Chicago school and post-Chicago school theories, pre-Chicago school considerations still play a role in Europe, albeit at times dressed up in post-Chicago clothing".

I believe that these traditional approaches gave important insights into many antitrust issues, but they failed to provide a complete understanding of the behaviour of market leaders. The Chicago approach limited most of its analysis to either monopolistic or perfectly competitive markets, and in a few cases, to markets characterized by a monopolist and a competitive fringe of potential entrants. For instance, according to the Chicago school there is not such a thing as predatory pricing, that is reducing prices below costs to induce exit by the competitors so as to compensate the initial losses with future profits: if the incumbent can sustain such initial losses, also any other competitor can do it as long as credit markets are properly working, hence predatory pricing would not be effective to start with. This approach failed to provide results that were robust enough to withstand fully-fledged game-theoretical analysis of dynamic competition between incumbents and entrants. Somewhat related with it are the theory of contestable markets of Baumol, Panzar and Willig (1982) and the initial literature on entry deterrence associated with the socalled Bain-Modigliani-Sylos Labini framework. However, even if the initial theoretical contributions took into consideration the effects of entry on the behaviour of market leaders, these were not developed in a coherent game theoretic framework and were substantially limited to the case of competition with perfectly substitutable goods and constant or decreasing marginal costs.

In the 80s and 90s, post-Chicago research studied more complex market structures within a solid game-theoretic framework and introduced welfare considerations so as to derive sound normative implications, which represents one of the main contributions of this approach. However, in most cases, this literature studied the behaviour of incumbent monopolists facing a single potential entrant. To cite the most famous works with strong relevance for antitrust issues, this was the case of the Dixit model of entry deterrence, of the models by Kreps and Wilson (1982) and Milgrom and Roberts (1982) of predatory pricing,[3] of

[3]The post-Chicago school as shown that in presence of asymmetric information between

those by Fudenberg and Tirole (1984) and by Bulow *et al.* (1985) of strategic investment, of the Bonanno and Vickers (1988) model of vertical restraints, of the Whinston (1990) model of bundling for entry deterrence purposes, and of many other works, often based on analysis of Stackelberg duopolies (that is, markets with one leader and one follower).[4] Also most of the standard results on the behaviour of incumbents in terms of pricing, R&D investments, quality choices, vertical and horizontal differentiation are derived in models of Stackelberg duopoly, where the incumbent chooses its own strategies in competition with a single entrant. While this analysis simplifies the interaction between incumbents and competitors, it can be highly misleading, since it assumes away the possibility of endogenous entry, and hence limits its relevance to situations where the incumbent has already an exogenous amount of market power.

It is not surprising that the results of the post-Chicago approach are systematically biased toward an anti-competitive role by incumbents: these engage in aggressive pricing, threaten or undertake overinvestments in complementary markets, impose exclusive dealing contracts, or bundle their goods with the sole purpose of deterring competitor entry. Otherwise they engage in accommodating pricing, underinvest in product improvements and differentiation, and stifle innovation. In such a simple world, what antitrust authorities should do is unambiguously to fight against incumbents: punish their aggressive pricing strategies as predatory, and their accommodating pricing strategies as well (but in this case as monopolistic strategies), punish investments in complementary markets as attempts to monopolize them, forbid bundling strategies, and so on. The bottom line is that antitrust authorities should sanction virtually all behaviours of the incumbents which do not conform to those of their competitors.

The fallacy of this line of thought, in my view, derives from a simple fact: it is based on a partial theory which does not take into account that, at least in most cases, entry by competitors is not an exogenous fact, but an endogenous decision. Whether entry is more or less costly, entry is an endogenous decision by the potential competitors, especially in global markets as most markets in the New Economy. There are two different kinds of constraints on entry. The definition of *barriers to entry* has been quite debated in the literature. Bain associated them with the situation in which established firms can elevate their selling prices above minimal average costs of production without inducing entry in the long run, Stigler with costs of production which must be borne by firms which seek to enter an industry but not borne by incumbents. A similar approach has been prevailing more recently (Baumol, Panzar and Willig, 1982), so that we can talk of barriers to entry as sunk costs of entry for the competitors which are above the corresponding costs of the incumbent (or have been already paid by the incumbent). On the contrary, simple *fixed costs of entry* are equally

firms, of credit market imperfections and of strategic commitments to undertake preliminary investments the above argument breaks down and predatory pricing can be an equilibrium strategy for the incumbent and deter entry.

[4] See Motta (2004) for a survey.

faced by the incumbent and the followers to produce in the market, but they also constrain entry. Actually, while there is a fundamental difference between the two concepts, their role in constraining entry, and hence in endogenizing it, is basically the same. In our view, only a comprehensive understanding of the behaviour of incumbents when entry is endogenous and when it is not can provide the required tools to judge real world markets.

3 The Theory of Market Leaders

The theory of market leaders clarifies the role of market leaders under more general conditions than the post-Chicago approach. In this section we will discuss its results and compare its implications for antitrust with those of the traditional approach.

Let us start from a general model of Stackelberg competition in which a main difference between the new and the old approach emerges. Consider a market with price competition between firms offering imperfectly substitutable goods, zero marginal costs of production, for simplicity, and a sunk cost of entry S which the market leader does not bear. When a firm sets the price p_i and the other firms set their prices p_j, demand is $D\left[p_i, \sum_{j\neq i} g(p_j)\right]$. We assume that demand is decreasing in the first argument: a higher price by i reduces i's demand. Moreover demand is also decreasing in the second argument while the function $g(p)$ is decreasing in the price, so that a higher price by any firm j increases demand for firm i, as reasonable. This demand function generalizes most common demand functions used by economists, as those derived by isoelastic utilities, Logit demands, constant expenditure demands and the Dixit-Stiglitz demand (see the Appendix). Finally, profits for the leader are:

$$\pi_L = p_L D\left[p_L, \sum_{j\neq L} g(p_j)\right]$$

while profits for a follower j are:

$$\pi_i = p_i D\left[p_i, \sum_{j\neq i} g(p_j)\right] - S$$

Consider first the simple case with just a single follower, which was the subject of analysis of most literature in the post-Chicago approach. When the sunk cost S is small enough, the follower is active in equilibrium and chooses its own price, say p_F, as a function of the price of the leader p_L, according to the optimality condition:

$$D_1[p_F, g(p_L)]p_F + D[p_F, g(p_L)] = 0$$

Under standard conditions (strategic complementarity), this implies a price increasing in the price of the leader. Then the leader chooses its own price taking this into account, according to the optimality condition:

$$D_1[p_L, g(p_F)]p_L + D[p_L, g(p_F)] + D_2[p_L, g(p_F)]g'(p_F)\frac{\partial p_F}{\partial p_L} = 0$$

Since the last term is positive, we can conclude that $p_L > p_F$. The intuition of this outcome is quite simple. The leader is aware that the higher will be its own price, the higher will be the price chosen by the follower, and both firms will have large profits. Hence the leader exploits its first mover advantage to set a high price.

A fundamental contribution by Dixit (1980) and Fudenberg and Tirole (1984), at the origins of the post-Chicago approach, was to show that, when the sunk cost S is high enough, the optimal strategy for the leaders may be an entry deterring strategy which requires a low enough price of the leader. The policy implication was immediate: any low pricing strategy by a leader must be associated with predatory pricing, otherwise a market leader would prefer to be accomodating setting high prices.

Unfortunately, this story has a simple but pervasive problem. Let us go back to the hypothetical situation in which the sunk cost S is small enough, so that the follower is active in the duopoly equilibrium. In such a situation it may well be the case that one or more other firms could also find profitable to enter in this market after bearing the small sunk cost. If this is the case, and this must be the case when S is small enough, the right equilibrium concept for this market has to take endogenous entry in consideration. In Etro (2002) I have solved for this equilibrium, showing that it is characterized by the same optimality condition for each follower as above:

$$D_1(p_F, \cdot)p_F + D(p_F, \cdot) = 0$$

an endogenous entry condition equating gross profits of each follower to the sunk cost of entry:

$$D(p_F, \cdot)p_F = S$$

and the following optimality condition for the leader:[5]

$$D_1(p_L, \cdot)p_L + D(p_L, \cdot) - D_2(p_L, \cdot)g'(p_F)p_L = 0$$

[5]Proof: Equilibrium demand for the followers is $D[p_F, g(p_L) + (n-2)g(p_F)]$ where n is the number of firms active in the market. The equilibrium first order condition and the endogenous entry condition for the followers pin down both arguments p_F and $\beta = g(p_L) + (n-2)g(p_F)$ independently from the leader'strategy p_L: only the number of firms changes with the price of the leader. Hence, the profits of the leader can be written as:

$$\pi_L = p_L D[p_L, (n-1)g(p_F)] = p_L D[p_L, \beta + g(p_F) - g(p_L)]$$

whose maximization provides the equilibrium first order condition in the text.

Now the last term is negative, suggesting that $p_L < p_F$. When entry is endogenous, whether just one or many followers end up entering in the market, the leader is always aggressive setting a lower price than each one of the entrant. Hence, a low pricing strategy is associated with a normal strategy of the market leader when entry is endogenous, and not with entry deterrence purposes.

This is the general principle emerging from the theory of market leaders, at the basis of our critique to the post-Chicago approach to antitrust. It actually emerges in more general situations, as shown by Etro (2006a): whenever the leader can engage in preliminary investments, while in duopoly it will bias them strategically to increase its price in the market, under endogenous entry it will bias them in the opposite way, to decrease its price in the market. For instance, if a cost reducing technology exists, Fudenberg and Tirole (1984) have shown that a duopoly leader will underinvest in it, while Etro (2006a) has shown that the same leader will overinvest in it a long as entry in the market is endogenous. This outcome emerges in many other contexts with surprising results with respect to investments in R&D and exploitation of network effects, important factors in markets of the New Economy, in case of bundling of goods and many other situations: some of these are examined in detail in the Appendix. The general point is that in any market where entry is endogenous, the leader always overinvests to gain a strategic advantage and conquer a larger market share; however, this results in a reduction in prices with a net gain for consumers.

Until now we considered a market structure which is quite standard because goods are imperfect substitute. However, our results become even more drammatic under certain conditions, in particular with quantity competition and homogenous goods. To see why this happens, imagine a market of homogeneous products where production requires again a fixed sunk cost S and a zero marginal cost of production. Moreover, imagine that firms choose their production level and the market price just equates demand and supply. Such a simple structure approximates the situation in many sectors where product differentiation is not very important but there are high costs to starting production: this is typical of energy and telecommunication industries and many other high-tech sectors of the New Economy. The assumption of constant (and indeed zero) marginal cost matches the conditions of some markets, as the software one, where variable costs are negligible compared to R&D expenditure.

Imagine that inverse demand is a decreasing function of total production, $p(\sum x_j)$. Profits of the leader are now:

$$\pi_L = x_L p\left(\sum x_j\right)$$

while the profits of each follower are:

$$\pi_i = x_i p\left(\sum x_j\right) - S$$

As well known, a Stackelberg duopoly would imply a division of the market between the leader and the follower. Stackelberg equilibrium with endogenous entry, however, generates a different result. If the supply of the leader x_L is small enough followers will be active and their supply x will satisfy the first order condition:

$$p(X) + xp'(X) = 0$$

where $X = \sum x_j$ is total production. The endogenous entry condition will be:

$$xp(X) = S$$

It is clear that these two conditions will pin down both the strategy of the followers x, and total production X, and hence the price $p(X)$. Hence the perceived profit of the leader becomes:

$$\pi_L = x_L p(X)$$

which is always trivially increasing in its supply. Hence the leader will always produce as much as possible. No other firm will find convenient to enter in the market, but nevertheless the price will be determined by the free entry condition and hence it will be the same as in absence of the leader. Notice that the same result would emerge with any kind of constant or decreasing marginal cost function.

Etro (2002) also proves that this outcome is better for consumers than the free entry equilibrium without a leadership, but the point here is simpler: under certain technological conditions, it is natural for the leader to conquer a large market share while supplying it at a competitive price determined by a zero profit condition. These technological conditions amount to high sunk costs or fixed costs of production (which may be R&D costs) and constant marginal costs of production. Of course network effect would even strenghten the result, as shown in the Appendix. On the other side, introducing imperfect substitutability or increasing marginal costs would allow the followers to conquer market shares. Nevertheless, the general principle would always hold: the leader would be aggressive and price below its followers retaining a larger market share because of the competitive pressure.

This discussion implies two main conclusions. First, a leading market position associated with aggressive strategic investments can be the consequence of a competitive market environment and not the result of market power. In other words, the theory of market leaders suggests that it would be better to differentiate market leaders from dominant firms: market leaders have some strategic competitive advantage over their competitors, but only when they can use it to prevent effective competition and harm consumers should they be considered to be dominant and their behavoiur potentially abusive. The point is to understand when market leaders can prevent effective competition and when they cannot. Second, whenever firms engage in price competition, the post-Chicago approach associates aggressive pricing or other aggressive strategies with a predatory purpose, while the theory of market leaders provides arguments that an aggressive strategy is generally pro-competitive and without exclusionary purposes.

4 The Software Market and the Microsoft Case

As a case study, consider the software market, one of the markets of the so-called New Economy, developed in the very last decades through progress in the Information & Communication Technology. In the 1960s, the computer industry was dominated by IBM, which manufactured expensive mainframe computers that were used by large enterprise customers; at the time, very few consumers had access to computers. Apart from IBM, mainframes were offered by firms such as Bull, Burroughs, Data General, Fujitsu, ICL, Nixdorf and Sperry-Rand. There was little or no interoperability among mainframes from different vendors. For the most part, an enterprise customer was required to choose an all IBM solution or an all Nixdorf solution. In the 1970s, Digital Equipment achieved considerable success with a line of less expensive minicomputers that were well-suited to engineering and scientific tasks. Again, however, there was little or no interoperability between these minicomputers and mainframes offered by IBM and others. The structure of the industry at that time was still largely vertical. By 1980, a number of companies had started offering less expensive microcomputers which were not interoperable with one another: early PCs by Tandy, Apple, Commodore and Atari ran their own operating systems, meaning that applications written for one brand of PC would not run on any other brand: the industry was fragmented. In mid-1980, IBM announced plans to introduce an IBM personal computer. The first one was offered with a choice of three operating systems: CP/M-86 from Digital Research, UCSD-P System and MS-DOS from Microsoft, a company founded by Bill Gates, a young software architect who dropped Harvard to create what was going to become a symbol of market leadership.

To understand the peculiarities of the software market in general it is convenient to focus briefly on the main functions of PC operating systems (OSs). The main one is to serve as a platform on which applications (such as spreadsheets or word processors) can be created by software developers. OSs supply different types of functionality, referred to as system services, that software developers can call upon in creating their applications. These systems services are made available through Application Programming Interfaces (APIs). When an application calls a particular API, the operating system supplies the system service associated with that API by causing the microprocessor to execute a specified set of instructions. Software developers need well-defined platforms that remain stable over time. They need to know whether the system services on which their applications rely will be present on any given PC. If they did not, then software developers would have to write the software code to provide equivalent functionality in their own applications, generating redundancy, inefficiency and a lack of interoperability.[6] Moreover, modern OSs provide a user interface, the means by which a user interacts with his computer. User interfaces for computers have evolved dramatically over the last decades, from punch card readers, to teletype terminals, to character-based user interfaces, to Graphical User Interfaces, first introduced by Apple with Macintosh. Finally, operating systems enable users to find and use information contained in various storage devices: local ones, such as a floppy diskette, a CD-ROM drive or the hard drive built into a PC, or remote, such as local area networks that connect computers in a particular office, wide area networks that connect computers in geographically separated offices, and the Internet.

Over time, the OSs of Microsoft became the most popular because Microsoft continually added new functionality to the operating system and licensed it to a wide range of computer manufacturers with extremely aggressive pricing strategies. Microsoft recognised early on that an OS that served as a common platform for developing applications and could run on a wide range of PCs would provide substantial benefits to consumers. Among other advantages, development costs would fall and a broader array of products would become available because products could be developed for the common platform rather than for a large number of different platforms. By providing a single operating system that ran on multiple brands of PCs, Microsoft enabled software developers to create applications, confident that users could run those applications on PCs from many different computer manufacturers. In addition, applications developed for a single platform were more easily interoperable because they were relying on the same functionality supplied by the underlying OS. In other words, network effects were created.[7]

In 1981, Microsoft released its first operating system, MS-DOS, which had a character-based user interface that required users to type specific instructions to perform tasks. In 1985, Microsoft introduced a new product called *Windows* that included a GUI, enabling users to perform tasks by clicking on icons on the screen using a pointing device called a mouse. Windows 3.0, shipped in 1990, was the first commercially successful version of Windows. In 1995, Microsoft released Windows 95, which integrated the functionality of Windows 3.1 and

[6]For instance, the UNIX operating system, developed by Bell Laboratories in the 1970s was not preserved as a common platform but was instead allowed to fragment. IBM, Siemens, Silicon Graphics and many other firms created their own versions of UNIX, which were all different from one another. As a result, applications written for one version of UNIX frequently will not run on other versions.

[7]Nowadays, computer manufacturers benefit because their PCs can run the many applications written for Windows and because users are familiar with the Windows user interface. Software developers benefit because their applications can rely on system services exposed by Windows via published APIs and because they can write applications with assurance that they will run on a broad range of PCs. Consumers benefit because they can choose from among thousands of PC models and applications that will all work well with one another and because such broad compatibility fosters intense competition among computer manufacturers and software developers to deliver improved products at attractive prices.

MS-DOS in a single operating system. In 2000, Microsoft shipped Windows 2000 Professional, a new generation of PC operating system built on a more stable and reliable software code base than earlier versions of Windows. Windows XP and the forthcoming Vista represent furthers evolution of the operating system, with a range of added functionality for both business and home users. Even if official and unanimous data are unavailable, consistent evidence suggests that the market share of Windows on sales of OSs for PCs rapidly increased toward 80% in the first half of the 90s to gradually arrive at 92% in 1996, 94% in 1997, 95% in 1998 and remained basically at this level since then:[8] meanwhile the average consumer price of Windows (calculated as average revenue per licence in OEM channel based on Microsoft sales) was constant around 44-45$.

Beyond OSs, Microsoft produces very successful applications. Some essential applications have been freely bundled with the operating system: for instance a basic word processing software, *WordPad*, a browser to access Internet and media player functionalities have been gradually added for free to subsequent versions of Windows when they became standard components of a modern OS. Other more sophisticated applications are supplied separately. Most notably this is the case of the Office Suite consisting of the advanced word processor *Word*, the spreasheet *Excel*, the software for presentations *PowerPoint* and more. The main two applications, Word and Excel, have been successfully competing against alternative products like *WordPerfect*, *WordStar*, *AmiPro* and others on one side and *Lotus*, *Quattro* and others on the other side. Liebowitz and Margolis (1999) have shown convincing evidence for which a better quality/price ratio together with network effects were at the basis of this success (it is important to notice that Microsoft achieved leadership in the Macintosh market, hence without exploiting the presence of its own OS, considerably earlier than in the PC market).[9] In the market for word processing applications, Microsoft's market share was hardly above 10% at the end of the 80s, to gradually increase at 28% in 1990, 40% in 1991, 45% in 1992, 50% in 1993, 65% in 1994, 79% in 1995, 89 in 1996, 94% in 1997 and to arrive at 95% in 1998, meanwhile the average consumer price of Word (calculated as average revenue per license) decreased from 235$ in 1988 to 39$ in 2001.[10] In the market for spreadsheet applications, Microsoft followed a similar progress, with a market share of 18% in 1990, 34% in 1991, 43% in 1992, 46% in 1993, 68% in 1994, 76% in 1995, 84% in 1996, 91% in 1997 and 94% in 1998, with minor progress in the following years, while the average consumer price of Excel was decreasing from 249$ in 1988 to 42$ in 2001.

[8] Of course the market share of Windows is lower if we take in consideration Macintosh computers as well.

[9] Microsoft did not achieve large market shares for other important applications, for instance for personal finance software.

[10] The market for word processing software includes many other nihe products. For instance, this paper is written with *Scientific Word*, a software that is particularly useful for scientific writing and academic purposes in general, but which is also largely more expensive than Word.

The leading position of Microsoft induced large opposition in the industry and the emergence of multiple antitrust cases with importance at a global level. In the main Microsoft vs US case, the software company was accused of monopolizing the PC operating systems market for Intel-compatible computers, tying its Windows operating system with the *Internet Explorer* browser with predatory purposes and to engage in anti-competitive contractual agreeements with computer manufacturers and Internet service providers. After an initial decision which imposed heavy behavioural and structural remedies on Microsoft, including the break up in a operating system and an application company (the socalled "Baby Bills"), the November 2002 ruling of the District of Court decided only to impose behavioural remedies aimed at preventing Microsoft from adopting exclusionary strategies against firms challenging its market power in the market for operating systems.

The Microsoft vs EU case was developed on very similar issues and, at the time of writing, it is still unresolved. In the March 2004 decision, the European Commission imposed the largest fine in the history of antitrust, required Microsoft to issue a version of its Windows operating system without *Media Player*, and mandated the licensing of intellectual property to enable interoperability between Windows PCs and work group servers and competitor products. Microsoft's Appeal of the decision was heart by the European Court of First Instance in April 2006 and a decision is expected by the end of the year.

A common element in both cases has been the substantial involvement of competitors of Microsoft on the side of the antitrust authorities, something that usually can create suspicion on the fact that a firm is really behaving as a monopolist rather than as an aggressive competitor. In a neat article on *Business Week*, Robert Barro (1998) noticed that "[a] sad sidelight in the Microsoft case is the cooperation of its competitors, Netscape, Sun and Oracle Corp., with the government. One might have expected these robust innovators to rise above the category of whiner corporations... The real problem is that whining can sometimes be profitable, because the political process makes it so. The remedy requires a shift in public policies to provide less reward for whining. The bottom line is that the best policy for the government in the computer industry is to stay out of it." Nevertheless in the European case Sun, Oracle, Novell, IBM and the Free Software Movement are active sides against Microsoft.

The technological conditions in the software market are well known. Producing software (whether it is an operating system or a particular application) takes a very high up-front investment and a constant marginal cost which, as is well known, is close to zero. The entry conditions in this market are more debated, but there are good reasons to believe that even though entry into the software market may entail large costs, it is substantially open, i.e. endogenous. First of all, there are already many firms producing OSs (as IBM, Red Hat, Oracle, Sun, Apple, HP, Compaq, Data General,...), and even more potential entrants – think of the giants in adjacent sectors of the New Economy (hardware and telecommunications in particular). Second, it is hard to think of a market which is more "global" than the software market: demand comes from all over the world, transport costs are virtually zero, the knowledge required to build software is easily accessible worldwide and competition is global. Nevertheless, it has been claimed that in the market for PC (or client) operating systems, the high number of applications developed by many different firms for Windows represents a substantial barrier to entry. Unfortunately, such a claim usually leads to misleading conclusions. It is true that competitors need to offer (and some do offer already) a number of standard and technologically mature applications upon entry to match the high quality of the Windows package, but the cost of offering these applications is unlikely to be prohibitive compared to the global size of this market. There are at least two reasons for this. First, notice that the alleged "applications barrier to entry" is often erroneously associated with thousands of applications written for Windows, while it is actually limited to a handful of applications such as word processing, spreadsheet, graphics and communications software, which really satisfy the needs of most active computer users (McKenzie, 2001). Second, the competitors of Microsoft should not (and the existing ones do not) even finance the development of all the needed applications: as Microsoft did in most cases, they should just fund and encourage other firms to write applications for their operating system (or have old applications originally written for other operating systems "ported to" theirs). Finally, it is important to emphasize that if we look at competition in the software market in a dynamic sense, that is competition for the market (as opposed to competition in the market), there is no doubt that the opportunity to invest in innovations for future, better software is widely open not only to large companies in the New Economy, but even to smaller ones.

Summarizing, the software market is characterized by high entry costs, constant marginal costs close to zero and substantially open access by competitors able to create new software. According to the new theory of market leaders these are the ideal conditions under which we should expect a leader to produce for the whole market with very aggressive (low) prices. Hence, it should not be surprising that, at least in the market for operating systems, a single firm, Microsoft, has such a large market share. We can see the same fact from a different perspective: since entry into the software market is endogenous, the leader has to keep prices low enough to expand its market share to almost the whole market. Notice that network externalities require these prices to be even lower because competitors could (and indeed try to) offer their alternative software at even lower prices to build their own network effects. Notice that low prices in presence of network effects are very common and often extreme: most email services as *Yahoo*, search engines as *Google* and social networks as *aSmallWorld* are free because this is the best strategy available for their leading suppliers under the constraint of effective competition. All these market leaders gain from collateral services, and, for sure, their leaderships have nothing to do with dominance.

The extremely low price of Windows represents a double proof of our arguments above. Assume for simplicity that the marginal cost of producing Windows is zero, and that the price of hardware is constant and independent from the price of Windows. Standard economic theory implies that the monopolistic price for an operating system should be the price of the hardware divided by $\epsilon - 1$, where ϵ is the elasticity of demand for PCs (including both hardware and software):[11] it means that a 1% increase in the price of PCs reduces demand by ϵ%. Now, the above relationship tells us that, if the basic price of the hardware is 1000 Euros, which is about the current average price for PCs, the monopolistic price for Windows would be 1000 Euros if $\epsilon = 2$, 500 Euros if $\epsilon = 3$, 333 Euros if $\epsilon = 4$ and so on. It would take really unreasonable values of demand elasticity to even get close to the real price of Windows, which is around 50 Euros. Moreover, this is a very conservative estimate of the monopolistic price. In the real world, we can imagine that the price of hardware is not independent from the price of Windows: if the latter would double tomorrow, hardware producers would be forced to reduce somewhat their prices (eventually switching to lower cost techniques and/or lower quality products). Even if this effect may be limited by the high level of competition in the hardware sector, it goes in the direction of increasing further the monopolistic price of Windows, that is, even beyond the real price of Windows.[12]

What does all this tell us? Simply that Microsoft is not an unconstrained price-setter, while its prices are limited well below the monopolistic price to compete aggressively with the other firms active in the operating system market and with the potential entrants in it. Economides (2001) concludes in a similar fashion: "Microsoft priced low because of the threat of competition. This means that Microsoft believed that it could not price higher it if were to maintain its market position." McKenzie (2001) supports this view: "some firms with high market shares might act more like competitors than other firms in markets where they have much smaller market shares. The reason is that the threat posed by potential competitors in a highly concentrated market can be more constraining than the competitive threat of actual competitors in less-concentrated markets".

What the post-Chicago approach suggested about leaders in markets with price competition was that they should be accommodating and exploit their market power, setting higher prices than competitors, or otherwise engage in

[11]See Etro (2006b).

[12]It has been claimed that low Windows pricing may be explained with the higher pricing of the complementary applications, as the Microsoft Office suite. However, the combined price of Windows and the average application package sold with it is still below the monopolistic price. Moreover, these applications are not sold at lower prices for other operating systems. Finally, as Economides (2001) pointed out, "Windows has the ability to collect surplus from the whole assortment of applications that run on top of it. Keeping Windows' price artificially low would subsidize not only MS-Office, but also the whole array of tens of thousands of Windows applications that are *not* produced by Microsoft. Therefore, even if Microsoft had a monopoly power in the Office market, keeping the price of Windows low is definitely not the optimal way to collect surplus."

predatory pricing and, after having conquered the whole market, increase prices. But in the last 10-15 years of global leadership, Microsoft has done neither of these things. It has been constantly aggressive, as any firm under the threat of competitive pressure would be. The theory of market leaders has shown that a market leader in these conditions would price above marginal cost in such a way to compensate for the fixed costs of investment and obtain a profit margin (over the average costs of production) thanks to the economies of scale derived from the large (worldwide in the case of Microsoft) scale of production. Its (quality adjusted) price should be slightly below that of its immediate competitors or just low enough to avoid that they can exploit profitable opportunities increasing their prices. Where other theories cannot, the theory of market leaders can make perfect sense of Microsoft's large market share, large profits and relatively low prices in a global and open market.

5 Bundling

One of the issues where the new theory of market leaders applies and provides new insights for antitrust policy is bundling, that is, the combination of two separate products in a single one sold alone. Notice that tying refers to selling one product (the tying product) conditional on the purchase of another one (the tied product), but there will not be any substantial difference between the two for our purposes. Virtually any product is a bundle since it combines multiple basic products which could be or are sold separately: a car bundles many separate components, shoes bundle shoes without laces and shoelaces, a computer bundles hardware, a operating system and basic software applications of general interest. In some cases bundling is just a contractual restriction used to force customers to purchase an ancillary product in an aftermarket for goods or services, while in other cases bundling improves a finished product by integrating new components or features into it: of course, only the first situation should be subject to antitrust investigation.

The Chicago school has advanced efficiency rationales in favour of bundling with positive, or at worst ambiguous, consequences on welfare, including production or distribution cost savings, reduction in transaction costs for customers, protection of intellectual property, product improvements, quality assurance and legitimate price responses. Moreover, according to the so-called "single monopoly profit theorem", as long as the secondary market is competitive, a monopolist in a separate market cannot increase its profits in the former by tying the two products. Actually, in the presence of complementarities, it can only gain from having competition and high sales in the secondary market to enhance demand in its monopolistic market. A similar theory has been advanced by Economides (2001) to explain the tying strategies of Microsoft. With particular reference to the US case, Economides (2001) notes that Microsoft could not have been interested in the browser market when this was perfectly competitive, but only when this market became dominated by Netscape for two main reasons. "First, Netscape had a dominant position in the browser market, thereby taking away from Microsoft's operating system profits to the extent that Windows was used together with the Navigator. Second, as the markets for Internet applications and electronic commerce exploded, the potential loss to Microsoft from not having a top browser increased significantly... Clearly, Microsoft had a pro-competitive incentive to freely distribute IE since that would stimulate demand for the Windows platform." The very same point could be made for the free distribution of Media Player with Windows, the subject of the tying part of the EU case.

However, the post-Chicago approach, starting with Whinston (1990), has shown that, when the bundling firm has some market power, bundles can have a predatory purpose, that is, they can deter entry in the tied product market to expand monopolistic power and reduce consumer welfare, at least in the long run. Summarizing the past economic research in the field, Tirole (2005) has concluded that tying should be submitted to a rule-of-reason standard, since it can have both efficiency and anti-competitive purposes. As formally shown in the Appendix, the theory of market leaders emphasizes that bundling by the incumbent is just an aggressive (pro-competitive) strategy of the incumbent for a competitive tied product market, may not have a specific entry deterrence purpose, and may increase welfare even without taking efficiency reasons into account.

Here, to derive the intuitions of these results, let us adopt the strongest bias against the bundling firm, imagining that this is a monopolist in a primary market which is also active in a secondary market, and that there are no technological efficiencies emerging from bundling goods in the two markets. The Chicago school has studied such a situation when the secondary market is perfectly competitive, that is, firms price at marginal cost and earn no extra profits: in such a case, the monopolist has no incentives to bundle because this could only reduce demand in the primary market. The post-Chicago approach has studied the same situation when the secondary market is not perfectly competitive and there is actually one single firm active strategically and no possibility for other firms to enter: then, the only reason why the monopolist would adopt a bundling strategy is to induce exit of the rival in the secondary market. The new theory of market leaders has studied again the same situation but with an imperfectly competitive secondary market, where firms decide "endogenously" whether to enter or not. In this case, the purpose of bundling has nothing to do with entry deterrence, it is just an aggressive strategy (but not a predatory one) which has pro-competitive effects: it reduces the combined price level and increases welfare. Technically, the market leader can exploit a larger scale of production for the bundle to offer it at a competitive price: bundling the two products works as a commitment device to be aggressive, that is to produce more for the secondary market and hence to be able to adopt a lower price. As a consequence, the leader can exploit larger scale economies, reduce the average price level for consumers and hence increase welfare.

Summarizing, when approaching a bundling case we need to verify the entry conditions of the secondary market. If there is a dominant firm in this market as well, the main problem is not the bundling strategy, but the lack of competition in the secondary market, and it should be addressed within that market: punishing the bundling strategy would just guarantee the monopolistic (or duopolistic) rents of the dominant firm in the secondary market. However, things are different when the secondary market is not monopolized but open to endogenous entry (even if it is not perfectly competitive, in the sense that firms do not price at marginal cost). In such a case bundling is a pro-competitive strategy and punishing it would hurt consumers.[13]

In the case of Microsoft, we have the impression that in both bundling cases, that of Windows with Internet Explorer and that of Windows with Media Player, the tied market was and still is characterized by endogenous entry: just think of new successfull browsers as Mozilla or Firefox and media player software as RealPlayer, Quick or, more recently, Macromedia Flash. Consequently the bundling strategy of Microsoft could be simply seen as an aggressive and competitive strategy of a market leader active in a secondary market with endogenous entry.[14]

6 Innovation and Interoperability

Competition in high-tech markets is dynamic in the Schumpeterian sense that it takes place as competition for the market in a so-called winner-takes-all-race, and such an element requires an even deeper rethinking of antitrust policy than suggested in the analysis of the previous sections, which were mostly focused on a static concept of competition in the market.

Economic research has emphasized the positive relationship linking patents to investments in innovation and these investments to technological progress and growth. In the New Economy and high-tech sectors in general firms compete mainly by innovating. This is possible as long as there are well defined IPRs, and especially patents, protecting their innovations and investments, which is

[13]Looking at the approach of the European Commission (2005), it appears that its positive principles are not fully carried through into the discussion on bundling. For instance, the standard of proof the Commission is required to meet to establish harmful foreclosure effects is too low, particularly in light of the fact that the analysis of foreclosure effects can be speculative in nature. In the case of bundling, actual market foreclosure effects are not required by the Discussion Paper: it is enough that such effects are "likely" to occur. In other words, the mere risk of foreclosure can result in a finding against a dominant company. A standard of proof that requires convincing evidence will help ensure that companies will not be deterred from bringing new products to market as a result of concerns about remote, potential foreclosure effects.

[14]Notice that the defensive strategy of Microsoft in the EU case appears to be unrelated with this point, but based on the fact that a modern OS must include mediaplayer functionalities and hence there should not even be an issue about bundling.

ultimately what leads to technological progress in our economies. Moreover, even if most economists are used to thinking about market leaders as firms with weaker incentives to invest in R&D, recent theoretical and empirical research has also found that market leaders play a crucial role in the innovative activity. The recent theories of market leadership have clarified the mechanics of these results. In a sense, patents drive competition through innovation in these markets and induce technological progress led by market leaders. For instance, in Etro (2004) and in a simpler model in the Appendix, I have shown that dominant firms have more incentives to invest in innovation than the outsiders when the patent race is characterized by endogenous entry, as long as the dominant firms have a leadership. The crucial thing here is that dominant firms often remain dominant thanks to their investments, but this should not be seen as evidence of inefficiency or of monopolistic power, but rather as a proof of the opposite: the competitive environment spurs investment by leaders and consequently induces a chance that their dominance persists. According to a recent Economic Focus of the Economist on this research (May 20th 2004, "Slackers or Pace-setters? Monopolies may have more incentive to innovate than economists have thought"), competitive pressure is the key to why:

> "a market leader has a greater incentive than any other firm to keep innovating and thus stay on top. Blessed with scale and market knowledge, it is better placed than potential rivals to commit itself to financing innovations. Oddly—paradoxically, if you like—in fighting to maintain its monopoly it acts more competitively than firms in markets in which there is no obviously dominant player...
>
> The most important requirement for this result is a lack of barriers to entry: these might include, for example, big capital outlays to fund the building of new laboratories, or regulatory or licensing restrictions that make it hard for new firms to threaten an incumbent. If there are no such barriers, a monopolist will have an excellent reason to innovate before any potential competitor comes up with the next new thing. It stands to lose its current, bloated profits if it does not; it stands to gain plenty from continued market dominance if it does... the fact that a dominant firm remains on top might actually be strong evidence of vigorous competition.
>
> But what if there are barriers to entry? These tend to make the dominant firm less aggressive in investing in new technologies—in essence, because its monopoly with the existing technology is less likely to be challenged. Over time, however, other companies can innovate and gradually overcome the barriers - "leapfrogging", as Mr Etro calls it. Meanwhile, the monopolist lives on marked time, burning off the fat of its past innovations."

Clearly, this has strong implications for industrial policy. What the above theory suggests is that dominant firms in high-tech sectors investing a lot in innovation may create an efficient situation.[15]

New ideas are often protected with patents, but these are not the only form of protection for innovations. Not all inventive and innovative activities fall under the scope of patentability and it is not always in the interest of a firm to patent every single innovation. In most high-tech sectors, firms adopt a combination of patents and trade secrets to protect products which are the result of multiple innovations. Defending (intellectual or material) property rights is one of the fundamental conditions for a proper functioning of the market economy: defending trade secrets has not a minor role in this context. Things get even more complicated in high-tech sectors of the New Economy. In these sectors trade secrets often cover fundamental innovations and protecting them amounts to promote new fundamental innovations, which are the main engine of growth. In some fields, however, there maybe, at least apparently, a trade-off

[15]The EU approach to antitrust deals with issues concerning innovation and IPRs in the discipline on refusals to supply, that is, situations where a dominant company denies a buyer access to an input in order to exclude that buyer from participating in an economic activity. Four conditions have to be fulfilled in order to find a termination of such a supply relationship to be abusive: i) the behaviour must be properly characterised as a termination of the supply arrangement; ii) the refusing undertaking must be dominant; iii) the refusal must be likely to have a negative effect on competition; and iv) the refusal must not be justified objectively or by efficiencies. Only when the dominant supplier has not previously supplied the input to a potential buyer, as for IPRs, an additional criterion is added: the input must be "indispensable" to carry on normal economic activity in the downstream market (a so-called "essential facility"). Nevertheless, the European Commission (2005) correctly pointed out that "to maintain incentives to invest and innovate, the dominant firm must not be unduly restricted in the exploitation of valuable results of the investment. For these reasons the dominant firm should normally be free to seek compensation for successful projects that is sufficient to maintain investment incentives, taking the risk of failed projects into account. To achieve such compensation, it may be necessary for the dominant firm to exclude others from access to the input for a certain period of time. The risks facing the parties and the sunk investment that must be committed may thus mean that a dominant firm should be allowed to exclude others for a certain period of time in order to ensure an adequate return on such investment, even when this entails eliminating effective competition during this period".

The proposal clearly states the priority of IPR protection, saying that "[i]mposing on the holder of the rights the obligation to grant to third parties a licence for the supply of products incorporating the IPR, even in return for a reasonable royalty, would lead to the holder being deprived of the substance of the exclusive right". Hence, another more restrictive criterion is added in the case of a refusal to license IPRs: the undertaking which requests the licence should intend to produce new goods or services not offered by the owner of the IPRs and for which there is a potential consumer demand. This additional criterion is in line with established case-law, but the Commission introduces an exception to this criterion. It states that a refusal to license IPR-protected technology which is indispensable for follow-on innovation may be abusive even if the license is not sought to directly incorporate the technology in clearly identifiable new goods and services, since the refusal to license an IPR-protected technology "should not impair consumers' ability to benefit from innovation brought about by the dominant undertaking's competitors". However, this exception is not motivated by economic analysis and is inconsistent with mainstream theories. There are no serious economic arguments supporting the view that weakening IPRs would strengthen innovation in the long run: such an approach on this matter may have negative consequences for EU innovation in the long run.

between trade secret protection and "interoperability" between products, which is, broadly speaking, the ability to exchange and use information and data, especially in networks. For instance, take in consideration the leading on line search engine in the world, Google. We may look at its patented innovations, but after that, we would need to know its trade secrets to fully discover the mechanism of its precious algorithms. This would help many software companies and websites to interoperate with Google even better than they already do, as it would allow other search engines to improve their performances compared to that of the leading search engine. But after that, we can bet, few companies would invest huge resources and take substantial risks to create a leading search engine or other brilliant ideas like Google when they can just free ride on others' ideas. The same argument would apply for the trade secrets of Microsoft on the source codes of its successful operating system Windows and to many other trade secrets of innovative leading companies. Any forced disclosure of similar trade secrets represents an expropriation of legitimate investments and establishes inappropriate legal standards with perverse effects on the incentives to innovate.

Fortunately, giving up to the precious role of trade secrets, or other IPRs, in promoting innovations is not the only way to solve interoperability challenges. The market can do it much better: valuable ideas can be selectively commercialized on a voluntary basis through licenses. Coase (1960) has clarified that whenever there is social value to generate, the market will properly allocate all property rights, including intellectual ones, insuring the accessibility of the information that fuels interoperability and acknowledging legitimate ownership rights of the innovators, and hence enhancing R&D investments. Finally, in presence of network effects, dynamic market forces can do even more: as long as IPRs are well protected and firms can invest with the safe confidence that successful innovations will be rewarded, market forces can select the best standard when multiple standards are available and interoperability is only partial. In a famous book, Liebowitz and Margolis (1999) have shown that this was the case in many episodes. For instance, in the adoption of the QWERTY keybord (so-called from the first five letters on the top left): for years it has been claimed that the allocation of letters of this keyboard was an inefficient standard, while these researchers found out that all the evidence suggests that the Qwerty keyboard, somehow selected by the market, is not worse than any other alternative. In conclusion, also in this field, markets can properly balance the short run and long run interests of the consumers better than policymakers: promote innovation, enable an efficient degree of interoperability and select the best standards.[16]

[16]The approach of the European Commission (2005) on this subject in the proposed guidelines for the application of Art. 82 is quite ambiguous. In the section on refusals to supply, they state that although "there is no general obligation even for dominant companies to ensure interoperability, leveraging market power from one market to another by refusing interoperability information may be an abuse of a dominant position". It is added that even if such information may be considered a trade secret "it may not be appropriate to apply to such

To go back once again to our case study of the software market and the Microsoft case, a lot of the residual contrast between Microsoft and the European Commission depends on the approach to interoperability. The Commission's March 2004 antitrust decision mandated the licensing of intellectual property to enable interoperability between Windows PCs and work group servers and competitor products. This point has turned out to be the most problematic in the case. In reality, as we have suggested earlier, the degree of interoperability in the computer industry is uncomparably high compared to just a few years ago. At the time of writing, Microsoft's offer of access to Windows source code, including for technologies that are covered by patents and trade secrets, seems to have not convinced the European Commission. Nevertheless, its case appears weaker than in 2004: Microsoft was also forced to licence more than a hundred technologies and in Europe not one of its competitors has taken out a license, a sign that the existing level of interoperability was not as low as it was depicted.

7 Conclusion

In conclusion I would like to briefly point out the main message of this paper. Recent progress in the theory of market leaders suggests that the post-Chicago approach to abuse of dominance can be problematic for markets characterized by endogenous entry conditions. In particular when investments in R&D represent a large portion of the costs of production and constrain entry, marginal costs are approximately constant and small, and network effects are present, equilibrium market structures are naturally characterized by large marker shares for the leaders. These are the results of their aggressive pricing and investments strategies which are forced by competitive pressure in the market and for the market. Hence, antitrust authorities should be more careful in associating large market shares and aggressive strategies with abuse of dominance in the dynamic markets of the New Economy. Hopefully, these results may contribute to the current debate on the reform of the EU approach to competition policy.

References

Ahlborn, Christian., David Evans and Jorge Padilla, 2004, The Antitrust Economics of Tying: a Farewell to per se Illegality, *The Antitrust Bulletin*, Spring-Summer, 287-341

refusals to supply information the same high standards for intervention" as those regarding IPRs. However, there is no guidance on the lower standards that the Commission should apply and on the definition of "information needed for interoperability" and this statement appears to open doors to a systematic possibility that innovative firms are forced to reveal trade secret. In my own view, even the uncertainty induced by this ambiguous wording is likely to jeopardize the incentives to invest in R&D with dangerous consequences for (future) consumer welfare.

Barro, Robert, 1998, Why the Antitrust Cops should Lay off Hight Tech, *Business Week*, August 17, 20

Baumol, William, John Panzar and Robert Willig, 1982, *Contestable Markets and the Theory of Industry Structure*, San Diego, Harcourt Brace Jovanovich

Bonanno, Giacomo and John Vickers, 1988, Vertical Separation, *Journal of Industrial Economics*, 36, 257-265

Bulow, Jeremy, John Geanakoplos and Paul Klemperer, 1985, Multimarket Oligopoly: Strategic Substitutes and Complements, *Journal of Political Economy*, 93 (3) 488-511

Coase, Ronald, 1960, The Problem of Social Cost, *Journal of Law and Economics*, 3, 1-44

Dixit, Avinash, 1980, The Role of Investment in Entry-Deterrence, *The Economic Journal*, 90, 95-106

Economides, Nicholas, 2001, The Microsoft Anti-trust Case, *Journal of Industry, Competition and Trade*, 1, 7-39

Economist, The, 2004, *Slackers or Pace-setters? Monopolies may have more incentives to innovate than economists have thought*, Economic Focus, 22nd-28th May, 84

Etro, Federico, 2002, Stackelberg Competition with Endogenous Entry, mimeo, Harvard University

Etro, Federico, 2004, Innovation by Leaders, *The Economic Journal*, 114 (April), 495, 281-303

Etro, Federico, 2006a, Aggressive Leaders, *The Rand Journal of Economics*, Vol. 37 (Spring), 146-54

Etro, Federico, 2006b, Competition Policy: Toward a New Approach, *European Competition Journal*, Vol. 2 (March), 29-55

Etro, Federico, 2007, *Market Leaders and Industrial Policy*, Springer-Verlag New York Inc., New York, forthcoming

European Commission, 2005, *DG Competition Discussion Paper on the Application of Article 82 of the Treaty to Exclusionary Abuses*, Bruxelles

Fudenberg, Drew and Jean Tirole, 1984, The Fat Cat Effect, the Puppy Dog Ploy and the Lean and Hungry Look, *The American Economic Review, Papers and Procedeengs*, 74 (2), 361-68

ICC, 2006, *ICC Comments on the European Commission Discussion Paper on the Application of Article 82 of the Treaty to Exclusionary Abuses*, International Chamber of Commerce, Paris

Kreps, David and Wilson, Robert, 1982, Reputation and Imperfect Information, *Journal of Economic Theory*, 7, pp. 253-79

Liebowitz, Stan and Stephen Margolis, 1999, *Winners, Losers & Microsoft: Competition and Antitrust in High Technology*, The Independent Institute

McKenzie, Richard, 2001, *Trust on Trial: How the Microsoft Case is Reframing the Rules of Competition*, Cambridge, Perseus Publishing

Milgrom, Paul and Roberts, John, 1982, Predation, Reputation and Entry Deterrence, *Journal of Economic Theory*, 27, 280-312

Motta, Massimo, 2004, *Competition Policy. Theory and Practice*, Cambridge, Cambridge University Press

Rey, Patrick, Jordi Gual, Martin Hellwig, Anne Perrot, Michele Polo, Klaus Schmidt and Rune Stenbacka, 2005, *An Economic Analysis to Article 82, Report of the Economic Advisory Group for Competition Policy*, European Union

Tirole, Jean, 2005, The Analysis of Tying Cases: A Primer, mimeo, University of Toulouse

Whinston, Michael, 1990, Tying, Foreclosure and Exclusion, *The American Economic Review*, 80, 837-59

Appendix

In this appendix I will present a general model of strategic investment and Nash competition based on Etro (2006a). Consider n firms choosing a strategic variable $x_i > 0$ with $i = 1, 2, ..., n$. They all compete in Nash strategies, that is taking as given the strategies of each other. These strategies deliver for each firm i the net profit function:

$$\pi_i = \Pi\left(x_i, \beta_i, k\right) - F \tag{1}$$

where $F > 0$ is a fixed cost of production. The first argument is the strategy of firm i and I assume that gross profits are quasiconcave in x_i.

The second argument represents the effects (or spillovers) induced by the strategies of the other firms on firm i's profits, summarized by $\beta_i = \sum_{k=1,k\neq i}^{n} h(x_k)$ for some function $h(x)$ which is assumed positive, differentiable and increasing. These spillovers exert a negative effect on profits, $\Pi_2 < 0$. In general, the cross effect Π_{12} could be positive, so that we have strategic complementarity (SC), or negative so that we have strategic substitutability (SS). I will define strategy x_i as aggressive compared to strategy x_j when $x_i > x_j$ and accommodating when the opposite holds. Notice that a more aggressive strategy by one firm reduces the profits of the other firms.

The last argument of the profit function is a profit enhancing factor ($\Pi_3 > 0$) which for all firms except the leader is constant at a level $\bar{k}$. Only the leader is able to make a strategic precommitment on k in a preliminary stage. The cost of its strategic investment is given by the function $f(k)$ with $f'(k) > 0$ and $f''(k) > 0$. Our focus will be exactly on the incentives for this firm to undertake such an investment so as to maximize its total profits:[17]

$$\pi_L(k) = \Pi^L\left(x_L, \beta_L, k\right) - f(k) - F \tag{2}$$

where x_L is the strategy of the leader and $\beta_L = \sum_{j\neq L} h(x_j)$. We will say that the investment makes the leader tough when $\Pi_{13}^L > 0$, that is an increase in k increases

[17] To avoid confusion, I will add the label L to denote the profit function, the strategy and the spillovers of the leader.

the marginal profitability of its strategy, while the investment makes the leader soft in the opposite case ($\Pi_{13}^L < 0$).

Most of the commonly used models of oligopolistic competition in quantities and in prices are nested in our general specification.[18] For instance, consider a market with quantity competition so that the strategy x_i represents the quantity produced by firm i. The corresponding inverse demand for firm i is $p_i = p\left[x_i, \sum_{j\neq i} h(x_j)\right]$ which is decreasing in both arguments (goods are substitutes). The cost function is $c(x_i)$ with $c'(\cdot) > 0$. It follows that gross profits for firm i are:

$$\Pi(x_i, \beta_i) = x_i p(x_i, \beta_i) - c(x_i) \tag{3}$$

Examples include linear and isoelastic demands and other common cases. This set up satisfies our general assumptions under weak conditions and can locally imply SS (as in most cases) or SC.

Consider now models of price competition where p_i is the price of firm i. Any model with direct demand:

$$D_i = D\left[p_i, \sum_{j=1, j\neq i}^{n} g(p_j)\right] \qquad \text{where } D_1 < 0,\ D_2 < 0,\ g'(p) < 0$$

is nested in our general framework after setting $x_i \equiv 1/p_i$ and $h(x_i) = g(1/x_i)$. This specification guarantees that goods are substitutes in a standard way since $\partial D_i/\partial p_j = D_2 g'(p_j) > 0$. Examples include models of price competition with isoelastic demand, Logit demand, constant expenditure demand[19] and other demand functions as in the general class due too Dixit and Stiglitz. Adopting, just for simplicity, a constant marginal cost c, we obtain the gross profits for firm i:

$$\Pi(x_i, \beta_i) = \left(\frac{1}{x_i} - c\right) D\left(\frac{1}{x_i}, \beta_i\right) = (p_i - c) D(p_i, \beta_i) \tag{4}$$

[18] Other models of oligopolistic interaction such as patent races and contests are also nested in my general framework, but I have discussed them elsewhere (Etro, 2002; 2004). In the following examples I omit the variable k for simplicity.

[19] For instance, consider a isoelastic utility like $u = \left[\sum_{j=1}^n C_j^\theta\right]^\gamma$, where $\theta \in (0,1]$ and $\gamma \in (0, 1/\theta)$. Demand for good i can be derived as:

$$D_i \propto \frac{p_i^{-\frac{1}{1-\theta}}}{\left[\sum_{j=1}^n p_j^{-\frac{\theta}{1-\theta}}\right]^{\frac{1-\gamma}{1-\gamma\theta}}}$$

which is nested in our framework after setting $g(p) = p^{-\theta/(1-\theta)}$. The Logit demand is

$$D_i = \frac{e^{-\lambda p_i}}{\sum_{j=1}^n e^{-\lambda p_j}}$$

which requires $g(p) = e^{-\lambda p}$. Notice that linear demands are not nested in our model.

which is nested in our general model and, under weak conditions assumed through the paper, implies SC.

We can now note that a more aggressive strategy corresponds to a larger production level in models of quantity competition and a lower price under price competition. In these models, we can introduce many kinds of preliminary investments, as we will see later on.

Strategic investment by the leader

We will now solve for the equilibrium in the two-stage model where the leader chooses its preliminary investment in the first stage and all firms compete in Nash strategies in the second stage.

For a given preliminary investment k by the leader, the second stage where firms compete in Nash strategies is characterized by a system of n optimality conditions. For the sake of simplicity, I follow Fudenberg and Tirole (1984) by assuming that a unique symmetric equilibrium exists and that there is entry of some followers for any possible preliminary investment. Given the symmetry of the model, in equilibrium each follower chooses a common strategy x and the leader chooses a strategy x_L satisfying the optimality conditions:

$$\Pi_1\left[x, (n-2)h(x) + h(x_L), \bar{k}\right] = 0 \tag{5}$$

$$\Pi_1^L\left[x_L, (n-1)h(x), k\right] = 0 \tag{6}$$

where I use the fact that in equilibrium the spillovers of each follower is $\beta = (n-2)h(x) + h(x_L)$ and of the leader is $\beta_L = (n-1)h(x)$.

Before analyzing the model with free entry, it is convenient to briefly summarize the results in the presence of barriers to entry. The system above provides the equilibrium values of the strategies as functions of the preliminary investment, $x(k)$ and $x_L(k)$, whose comparative statics can be easily derived. In the first stage the leader chooses its investment k to maximize:

$$\pi_L(k) = \Pi^L\left\{x_L(k), (n-1)h\left[x(k)\right], k\right\} - f(k) - F$$

and it is immediate to obtain the optimality condition:

$$\Pi_3^L + \frac{h'(x_L)\Pi_{13}^L\Pi_2^L\Pi_{12}}{\Omega} = f'(k) \tag{7}$$

where the second term on the left hand side represents the strategic incentive to commit to k.[20] The sign of this incentive is the opposite of the sign of $\Pi_{12}\Pi_{13}^L$. Hence, we have the following traditional result: under barriers to entry: 1) when the leader is tough ($\Pi_{13}^L > 0$), strategic over (under)-investment occurs under SS (SC), inducing

[20]Here

$$\Omega = \frac{\Pi_{11}^L}{(n-1)h'(x)}\left[\Pi_{11} + (n-2)\,h'(x)\Pi_{12}\right] + \Pi_{12}^L\Pi_{12}$$

is positive by the assumption of the stability of the system.

a "top dog" ("puppy dog") strategy; 2) when the leader is soft ($\Pi_{13}^L < 0$), strategic under (over)-investment occurs under SS (SC), inducing a "lean and hungry" ("fat cat") strategy.

The intuition behind this result is important for what follows. Basically, under SS the leader gains from committing to an aggressive behaviour in the market and can accomplish such a task by overinvesting or underinvesting strategically when the investment promotes aggressive or accommodating behaviour. Otherwise, under SC the leader tries to commit to accommodating behaviour in the market and can achieve this by adopting the opposite kind of strategy. The ultimate behaviour of the leader in the market depends on whether strategies are substitutes or complements.

I will now consider the case of endogenous entry assuming that the number of potential entrants is great enough that a zero profit condition pins down the number of active firms, n.[21] The equilibrium conditions in the second stage for a given preliminary investment k are the optimality conditions derived before and the zero profit condition for the followers:

$$\Pi\left[x, (n-2)h(x) + h(x_L), \bar{k}\right] = F \tag{8}$$

We can now prove that a change in the strategic commitment by the leader does not affect the equilibrium strategies of the other firms, but it reduces their equilibrium number. Let us use the fact that $\beta_L = \beta + h(x) - h(x_L)$ to rewrite the three equilibrium equations in terms of x, β and x_L:

$$\Pi\left(x, \beta, \bar{k}\right) = F, \quad \Pi_1\left(x, \beta, \bar{k}\right) = 0, \quad \Pi_1^L\left[x_L, \beta + h(x) - h(x_L), k\right] = 0$$

This system is block recursive and stable under the condition $\Pi_{11}^L - h'(x_L)\Pi_{12}^L < 0$. The first two equations provide the equilibrium values for the strategy of the followers and their spillovers, x and β, which are independent of k, while the last equation provides the equilibrium strategy of the leader $x_L(k)$ as a function of k with $x_L(\bar{k}) = x$ and:

$$x_L'(k) = -\frac{\Pi_{13}^L}{\Pi_{11}^L - h'(x_L)\Pi_{12}^L} \gtreqless 0 \quad \text{for } \Pi_{13}^L \gtreqless 0 \tag{9}$$

In the first stage the optimal choice of investment k for the leader maximizes:

$$\pi_L(k) = \Pi^L\left\{x_L(k), \beta + h(x) - h\left[x_L(k)\right], k\right\} - f(k) - F$$

and hence it satisfies the optimality condition:

$$\Pi_3^L + \frac{h'(x_L)\Pi_2^L\Pi_{13}^L}{\Pi_{11}^L - h'(x_L)\Pi_{12}^L} = f'(k) \tag{10}$$

where the sign of the second term is just the sign of Π_{13}^L. This implies that the leader has a positive strategic incentive to invest when it is tough ($\Pi_{13}^L > 0$) and a negative one when it is soft.

[21] As customary in the literature, I will assume n is a real number.

Since our focus is on the strategic incentive to invest, I will normalize the profit functions in such a way that, in absence of strategic motivations, the leader would choose $k = \bar{k}$ resulting in a symmetric situation with the other firms.[22] Consequently we can conclude that a tough leader overinvests compared to the other firms, in the sense that $k > \bar{k}$, while a soft leader underinvests. We also noticed that a tough leader is made more aggressive by overinvesting and a soft leader is made more aggressive by underinvesting. Finally, the strategy of the other firms is independent of the investment of the leader. Hence, we can conclude that the leader will be always more aggressive in the market than any other firm. Summarizing, we have:

Under Nash competition with endogenous entry, when the strategic investment makes the leader tough (soft), over (under)-investment occurs, but the leader is always more aggressive than the other firms.

Basically, under free entry, the taxonomy of Fudenberg and Tirole (1984) boils down to two simple kinds of investment and an unambiguous aggressive behaviour in the market: whenever $\Pi_{13}^L > 0$, it is always optimal to adopt a "top dog" strategy with overinvestment in the first stage so as to be aggressive in the second stage; while when $\Pi_{13}^L < 0$ we always have a "lean and hungry" look with underinvestment, but the behaviour in the second stage is still aggressive. Strategic investment is always used as a commitment to be more aggressive in a market with free entry, and this does not depend on the kind of competition or strategic interaction between the firms. As we will see in the applications of the next section, the result is particularly drastic for markets with price competition. In these markets, leaders are accommodating in the presence of entry barriers (choosing higher prices than their competitors), but they are aggressive under free entry (choosing lower prices). This difference may be useful for empirical research on barriers to entry and may have crucial implications for anti-trust policy.

Applications

We will now apply the above results to a number of basic industrial organization situations, with particular reference to specific features of markets of the New Economy: R&D investments, network externalities and bundling issues.

R&D Investments

Our first application is to a standard situation where a firm can adopt preliminary R&D investments to improve its production technology and hence reduce its cost function. Traditional results on the opportunity of these investments for market leaders are ambiguous when the number of firms is exogenous, but, as I will show, they are not when entry is endogenous. From now on, I will assume for simplicity that marginal

[22]This requires $\Pi_3^L\left(x, \beta, \bar{k}\right) = f'(\bar{k})$. Such a normalization does not affect qualitatively the incentives to adopt strategic investments and has a realistic motivation. We can imagine that all firms choose k but only the leader can do it before the others and commit to it, hence only a strategic motivation can induce the leader to choose a different investment.

costs are constant. Here, the leader can invest k and reduce its marginal cost to $c(k) > 0$ with $c'(k) < 0$, while the marginal cost is constant for all the other firms.

Consider first a model of quantity competition. The gross profit of the leader becomes:

$$\Pi^L(x_L, \beta_L, k) = x_L p(x_L, \beta_L) - c(k)x_L \tag{11}$$

Notice that in such a model, Π^L_{12} has an ambiguous sign, but $\Pi^L_{13} = -c'(k) > 0$, hence the leader may overinvest or underinvest in R&D under barriers to entry, but, will always overinvest in R&D and produce more than the other firms when entry is free. For instance assuming inverse demand $p = a - \sum x_i$ with $c(k) = c - dk$ and $f(k) = k^2/2$, for d small enough the leader invests:

$$k = \frac{2d\sqrt{F}}{1 - 2d^2}$$

in cost reductions and produces:

$$x_L = \frac{\sqrt{F}}{1 - 2d^2}$$

while all entrants produce $x = \sqrt{F}$ (while for a large enough d the leader would invest more to deter entry and to remain alone in the market).

Consider now the model of price competition where the leader can invest to reduce its marginal costs in the same way and its profit function becomes:

$$\Pi^L(x_L, \beta_L, k) = \left[\frac{1}{x_L} - c(k)\right] D\left(\frac{1}{x_L}, \beta_L\right) \tag{12}$$

where $\Pi^L_{13} = c'(k)D_1/x_L^2 > 0$. Hence, underinvestment in R&D emerges when there are barriers to entry, but overinvestment is optimal when there is free entry. Whenever entry is endogenous, the leader wants to improve its cost function to be more aggressive in the market by selling its good at a lower price. Summarizing, we have:[23]

Under both quantity and price competition with endogenous entry, a firm has always an incentive to overinvest in R&D to reduce costs and to be more aggressive than the others in the market.

Network externalities and learning by doing

Consider now dynamic models where profitability depends on past strategies. For instance, learning by doing implies that the cost function is decreasing in past production. This is the typical case of the aircraft industry (Boeing, Airbus), the production of chips (Intel) and many other sectors with a fast technological progress. Network

[23] Welfare analysis is beyond the scope of this paper, but in this case one can show that a leadership improves the allocation of resources. This is not due only to the cost reduction but also to the reduction in the number of firms since, as well known, Cournot and Bertrand equilibria with free entry are characterized by excessive entry.

externalities imply that demand is enhanced by past production and the consequent diffusion of the product across customers. This may be the case of the markets for operatying systems and general softwares (Microsoft), computers (IBM, Hewlett Packard) or wireless and broadband communications (Nokia, Motorola).

In these contexts it is natural to think in terms of quantity competition and, for simplicity, following Bulow et al. (1985), I will focus on two period models with the leader alone in the market in the first period and facing free entry in the second period. In case of learning by doing the leader will always overproduce initially to exploit the learning curve. In case of demand externalities the leader will overproduce initially to create network effects, which broadly matches pricing strategies by leaders in high tech sectors characterized by network externalities.

To formalize these results in the simplest setting, assume perfectly substitute goods. Imagine that in the first period the leader produces k facing the inverse demand $p(k)$ and a marginal cost c. In the second period other firms compete in quantities and the leader faces the inverse demand $p(x_L + \beta_L)\phi(k)$ where $\phi(k)$ is some increasing function of past production, which is a measure of the diffusion of the product across consumers (and induces the network externality), while the marginal cost $c(k)$ is decreasing in past production (because of learning by doing). The profit function for the leader becomes:

$$\Pi^L(x_L, \beta_L, k) = kp(k) - ck + \delta\left[p(x_L + \beta_L)\phi(k) - c(k)\right]x_L \qquad (13)$$

where $\delta < 1$ is the discount factor. In this case in equilibrium we have $\Pi^L_{13} = \delta\left[\phi'(k)/c(k)\phi(k) - c'(k)\right] > 0$ which already suggests that the initial monopolist will overproduce to be more aggressive when the market opens up. Moreover, the choice of initial production will satisfy:

$$p(k) + kp'(k) = c - \delta x_L\left[p\phi'(k) - c'(k)\right] - \delta x_L c(k)\left[\phi'(k)/\phi(k) - c'(k)/c(k)\right]$$

which equates marginal revenue to effective marginal cost. The latter includes the myopic marginal cost c, a second term which represents the direct benefit due to the network effects on future demand and costs and a last term representing the indirect (strategic) benefits due to the commitment to adopt a more aggressive strategy in the future. Summarizing:

Under learning by doing and network externalities a firm has always an incentive to overproduce initially so as to be more aggressive when endogenous entry takes place in the future.

Notice that the leader may engage in dumping (pricing below marginal cost) in the first period (if the discount factor is large enough), but this may well be beneficial to consumers in both periods (see Economides, 2001, for a related analysis of the software market).

Bundling

There has been a lot of attention in the economic literature on the rationale for bundling products rather than selling them separately. A fundamental reason for this is that many antitrust cases have focused on such a practice as an anti-competitive device. This paper tries to derive some general results on why firms bundle their products and some welfare implications.

According to the traditional leverage theory of tied good sales, monopolists would bundle their products with others for competitive or partially competitive markets to extend their monopolistic power. Such a view as been criticized by the Chicago school because it would erroneously claim that a firm can artificially increase monopolistic profits from a competitive market. Bundling should have different motivations, as price discrimination or creation of joint economies, whose welfare consequences are ambiguous and sometimes even positive.

Whinston (1990) has changed the terms of the discussion trying to verify how a monopolist can affect strategic interaction with competitors in another market by bundling. His main finding is that the only reason why a monopolist could bundle is to deter entry (as in Dixit, 1980), which has typically negative effects on welfare. His analysis is based on price competition between two firms, hence strategic complementarity holds, and it can be extended in many directions, especially including complementarities between products.

We depart from this analysis and consider a more general model where there may be more firms and alternative market structures. In particular, under free entry, bundling may become the optimal aggressive strategy. In this case, bundling does not need to have an exclusionary purpose as assumed by the leverage theory, and the reduction in the price of the two bundled goods together can also benefit consumers. Such an analysis may apply to the bundling of Windows with Internet Explorer and Media Player at no extra price (see Economides, 2001), which, nevertheless, has been harshly treated by the US and EU antitrust authorities.

To make our point in a neat way, let us follow the example by Whinston (1990), who has shown that a monopolist in one market does not have incentives to bundle its product with another one sold in a duopolistic market (unless this deters entry in the latter), and that this corresponds to a "puppy dog" (accomodating) strategy (see Fudenberg and Tirole, 1984). However, under free entry, bundling may become the optimal "top dog" (aggressive) strategy.

Imagine that a monopolistic market is characterized by zero costs of production and unitary demand at price v, which corresponds to the valuation of the good. For simplicity, there are no complementarities with a good produced in another market which is characterized by standard price competition, a fixed cost F and a constant marginal cost c.

Gross profits for the monopolist without bundling are:

$$\Pi^M(p_M, \beta_M) = v + (p_M - c) D(p_M, \beta_M) - F \tag{14}$$

while profits for the other firms are $\Pi^i(p_i, \beta_i) = (p_i - c) D(p_i, \beta_i) - F$. In Bertrand equilibrium with free entry the monopolist enjoys just the profits $\Pi^M = v$.

Under bundling, demand for the monopolist is constrained by demand for the other good, which is assumed less than unitary. Given a bundle price corresponding to $P_M = v + p_M$, profits for the monopolist become $\Pi^{MB} = (P_M - c)\, D(P_M - v, \beta_M) = (p_M + v - c)\, D\,(p_M, \beta_M)$, while the other firms have the same objective function as before. In Bertrand equilibrium the monopolist chooses the price P_M satisfying:

$$(p_M + v - c) D_1 \left[p_M, (n-1)g(p)\right] + D\left[p_M, (n-1)g(p)\right] = 0 \tag{15}$$

while each one of the other firms chooses p satisfying:

$$(p - c) D_1 \left[p, g(p_M) + (n-2)g(p)\right] + D\left[p, g(p_M) + (n-2)g(p)\right] = 0 \tag{16}$$

If endogenous entry holds, the number of firms satisfies also:

$$(p - c) D\left[p, g(p_M) + (n-2)g(p)\right] = F \tag{17}$$

so that the profit of the monopolist becomes:

$$\Pi^{MB} = (p_M + v - c)\, D\left[p_M, (n-1)g(p)\right]$$

Notice that if we define $\beta = g(p_M) + (n-2)g(p)$ the equilibrium spillovers received by the entrants as a consequence of the price chosen by their competitors, the equilibrium conditions jointly determine the price of the entrants p and β independently from the price of the monopolist. Hence, using $\beta_M = \beta + g(p) - g(p_M)$ we can rewrite the equilibrium first order condition of the monopolist as an implicit expression for $p_M = p_M(v)$, and it is immediate to derive that the equilibrium price of the secondary good decided by the monopolist has to be decreasing in v.[24] As well known from the theory of market leaders (Etro, 2006a), even under price competition, any strategic commitment is undertaken with the aim of being aggressive on the market. Nevertheless, when v is small enough, the equilibrium does not imply exclusion of other firms. Clearly, if the profit in the primary market is large enough, the monopolist may find convenient to offer such a large discount on the bundle that all its competitors will have to exit the market, but this only happens under retrictive conditions.

Clearly bundling is optimal if $\Pi^{MB} > \Pi^M$, and we need to verify under which conditions this happens. The first element to take in consideration is the way in which bundling changes the strategy of the monopolist. Since $\Pi_1^{MB} - \Pi_1^M = vD_1 < 0$, bundling makes the monopolist tough. This implies that the monopolist is led to reduce the effective price in the other market by choosing a low price of the bundle.

[24]In particular we have:

$$\frac{dp_M}{dv} = \frac{-D_1\left[p_M, \beta + g(p) - g(p_M)\right]}{\Delta} < 0$$

where $\Delta \equiv 2D_1 + (p_M + v - c)[D_{11} - g'(p_M)D_{12}] - g'(p_M)D_2 < 0$ by the stability of the equilibrium system. In other words, the price of the bundle increases less than proportionally with v or the monopolist offers the bundle with a discount on the secondary good compared to its competitors.

Since strategic complementarity holds, a price decrease by the monopolist induces the other firms to decrease their prices. Under barriers to entry, as in the Whinston (1990) model with two firms, this reduces profits of all firms in the other market, hence bundling is never optimal unless it manages to deter entry.

Under free entry, however, result can change: bundling can now be an effective device to outplace some of the other firms without deterring entry but creating some profits for the monopolist in the other market through an aggressive strategy. In particular, bundling is optimal if the low price of the bundle increases profits in the competitive market more than it reduces them in the monopolistic one. It is easy to verify that bundling is optimal if:

$$(p_M - c)D\left[p_M, (n-1)g(p)\right] - F > v\left\{1 - D\left[p_M, (n-1)g(p)\right]\right\}$$

whose left hand side is the gain in profits in the competitive market and whose right hand side is the loss in profits in the monopolistic market.

Moreover, in this case, bundling does not need to have an exclusionary purpose as assumed by the leverage theory of tied good sales. The reduction in the price of the two bundled goods together can also benefit consumers. This is even more likely when they are complements.

Usage and Diffusion of Cellular Telephony, 1998-2004†

Michał Grajek*

Wissenschaftszentrum Berlin (WZB) and *Humboldt University*

Tobias Kretschmer**

London School of Economics and *Centre for Economic Performance*

March 2006

PRELIMINARY, PLEASE DO NOT QUOTE. COMMENTS WELCOME.

Abstract: Like many other successful technologies, cellular telephony has diffused in an s-shape in most countries. Theory suggests a number of reasons for this diffusion pattern, including word-of-mouth, consumer heterogeneity and network effects. Using quarterly data from 41 countries over 6 years, we study both diffusion speed and usage intensity to uncover the impact of consumer heterogeneity, network effects, and fixed-mobile substitutability. We find that the consumer heterogeneity effect dominates the network effect, that fixed-line and cellular telephony share elements of both complementarity and substitutability, and that the complementarity particularly occurs in the early stages of cellular diffusion.

Keywords: Cellular telephony, diffusion, network effects, consumer heterogeneity, fixed-mobile substitutability.

JEL Codes: L1, L52, O38.

---

† We are grateful to Dr Jan Krancke from T-Mobile International and Mark Burk from Informa Telecoms & Media for making the data available to us and to seminar participants at the University of Munich, Wissenschaftszentrum Berlin, London School of Economics, and HEC School of Management, Paris, for helpful comments. Michał Grajek gratefully acknowledges financial support from the German Federal Ministry of Education and Research (project 01AK702A).

* Wissenschaftszentrum Berlin, Competitiveness and Industrial Change Research Unit. Reichpietschufer 50, 10785 Berlin, Germany. Tel: +49 30 25 49 14 19. Email: grajek@wz-berlin.de.

** London School of Economics and Centre for Economic Performance. Houghton Street, London WC2A 2AE, United Kingdom. Tel.: +44 20 79 55 60 41. Email: t.kretschmer@lse.ac.uk.

I Introduction

Nearly every new technology diffuses in an s-shape. Not surprisingly then, every theory of technology diffusion will be able to replicate an s-shaped pattern. The underlying mechanisms of different theories however are quite varied (see Geroski, 2000 for an extensive discussion of models of technology diffusion). In empirical work therefore it has been difficult to discriminate between different theories of diffusion, precisely because they all generate the same s-shape.

The literature proposes three main mechanisms by which a consumer technology can diffuse in an s-shape. First, consumers may learn about the product at different times. If information spreads by word-of-mouth, we obtain a slow start (because there are few people with experience of the new technology), a period of rapid diffusion (after critical mass is reached), and a tapering-off (because the few non-adopters are hard to reach). A second plausible mechanism is that adopters may have different preferences for the new technology an adopt in their order of preference, and assuming that the distribution of preferences is unimodal (e.g. normal), the highest density of adopter preferences corresponds to phase of most rapid diffusion. Finally, network effects can also lead to an s-shaped diffusion path – early adopters have little reason to adopt due to a lack of network size, late adopters are likely to have a low preference for the new technology and are unlikely to be swayed by others using them, while the mass market will adopt as soon as a sufficient network size is reached, which creates a snowball effect as more consumers adopt because more consumers have adopted.

One of the main problems in empirical economics is the difficulty of disentangling these effects. We propose a simple method that allows us to assess the relative strength of network effects and heterogeneity effects and find that consumer heterogeneity appears much more important in determining usage patterns than network effects. Further, the substitutability between old and new technological generations is an important issue in the study of new technologies and in cellular telephony in particular. We shed some light on this question by considering relationship between the fixed-line and cellular telephony in different stages of the diffusion curve and find elements of both complementarity and substitutability between these two services.

The paper proceeds as follows: In Section II, we briefly describe the global cellular telecommunications industry and then briefly review the existing literature on cellular telecommunications in Section III. We then describe our data and give some descriptive statistics in Section IV. Our empirical results are presented in Section V and a discussion follows in Section VI. Section VII concludes.

II The global cellular telecommunications industry

The recent history and general features of the cellular telecommunications industry are discussed in detail in Grajek (2004), Koski and Kretschmer (2005a) and Gruber and Verboven (2001). Here, we only give a brief history of the technological improvements and the corresponding generation changes over time.

In most countries, cellular phones were first available to end consumers in the eighties. The technology initially used was based on analogue signal transmission, which was relatively inefficient and unreliable. In some countries, first-generation (1G) cellular networks reached their capacity relatively quickly, leading to lower service quality in general and congestion for initiating calls in particular. As soon as digital technology (second generation, 2G) had matured enough to present a credible alternative to analogue cellular, it was introduced gradually across the world (Dekimpe et al, 2000). Several different technological standards were in existence in different countries – for example, GSM in Europe, PDCS in Japan – and some countries – most notably the US – even introduced several standards in one country.

Technological competition between standards within countries has been suggested to have slowed down overall diffusion (Koski and Kretschmer, 2005a), but may have had the long-term effect of fostering technological progress for future generations (Cabral and Kretschmer, forthcoming). In addition to 2G's improved reliability and network capacity, 2G phones also had SMS functionality, which enabled users to send short text messages to each other and was a huge success among younger users, especially in Asia and Europe.[1] Following the success of 2G, a third generation with more advanced data transmission facilities was developed and is currently being rolled out.

For our sample period 1998- 2004 however, 2G cellular was dominant. Second generation telephony itself displayed significant technological progress, with handsets becoming smaller and containing an increasing number of additional functions (Koski and Kretschmer, 2005b). Apart from the ongoing technological innovations on the product side, there were also significant innovations on the pricing and services side.

First-generation cellular phones were mainly targeted at business customers for several reasons: First, most handsets were rather heavy and thus used predominantly in cars,[2] which, in combination with very high tariffs, was attractive only to business customers. With the introduction of digital cellular telephony, however, the focus was on capturing the mass market in order to make the technology succeed commercially.

Penetration pricing: Early attempts by second-generation cellphone operators were targeted at gaining a critical mass of consumers. Since later adopters would be basing their adoption decisions on those of early adopters, operators were willing to take a loss, or at least price aggressively to grow their installed base. With lock-in contracts over one, sometimes two years, this penetration pricing was profitable (Farrell and Klemperer, 2005).

[1] In the US, text messaging had not caught on and the average user was sending 203 text messages a year compared to 651 in China (http://www.newsmax.com/archives/articles/2005/8/11/135257.shtml).

[2] One of the largest cellphone stores in the UK, founded in 1989, has only recently been renamed "The Phone House" from the original "Carphone Warehouse".

Handset subsidies: Most cellular handsets were, and still are, heavily subsidized. This was a strategy to get consumers to adopt in the first place, as handsets were typically the most expensive part of getting a cellphone connected.[3] Quite frequently then, basic handsets are given away "for free"[4] if the consumer signed up for a long-term contract. This is a particular form of product cross-subsidization to overcome the installed-base problem.

Prepaid contracts (Pay-as-you-go): Possibly the most successful strategy of moving cellular telephony into the mass market was the introduction of pay-as-you-go contracts. These contracts involve no monthly fee, but a higher cost per call. Such contracts are especially attractive for low-frequency users for whom this fee would be too high to warrant the few calls they make or who do not have access to a bank account to set up a monthly debit. The introduction of pay-as-you-go tariffs coincided with a rapid increase in diffusion speed, and most of the growth in later stages of diffusion came from prepaid users.

Tariff proliferation: Finally, with an increase in competition and increasingly fine market segmentation, the number of tariffs has proliferated enormously. This has two effects: First, it could serve as a collusive device by confusing consumers (Hörnig, 2005), and second, it could enable consumers to make more fine-grained decisions based on their expected calling patterns (Miravete and Röller, 2004, Narayanan, Chintagunta and Miravete, 2005). The fact that consumers seem to switch quite readily between contracts to optimize their behavior (Miravete and Röller, 2004) suggests that consumers will have some degree of uncertainty about their future calling patterns, but eventually settle on the contract that suits their consumption behavior best.

[3] Initially, operators charged a connection fee similar to the fixed-line market, but competition among operators forces and an unwillingness of consumers to pay for the privilege of going to a shop and having a shop assistant "activate" the connection by clicking a button put paid to this practice.
[4] Most commonly, operators would charge £/$/€ 1 for a handset that typically cost about $100 to produce. There are instances however of "paying" consumers to buy a handset: In France, a Siemens S35 was sold in connection with a contract for FFR190 and contained a voucher for a FFR200 reimbursement if sent to the mobile operator.

III Studies on cellular diffusion

An increasing number of studies have looked at the diffusion of cellular telephony in several countries. Single-country studies (e.g. Grajek, 2004, Doganoglu and Grzybowski, 2004) typically impose more structure on the diffusion process and try to discriminate between the price effect on diffusion and the network effect. On the other hand, multi-country studies often focused on identifying regulatory, economic, or competitive covariates that affect diffusion (e.g. Gruber and Verboven, 2001, Koski and Kretschmer, 2005a). What most studies have in common is that s-shaped diffusion (or segments of the s-curve) is either assumed from the outset or generated through lagged dependent variables. In most existing studies however, the a-priori assumption of an s-shaped curve, while realistic, leaves out the question of what mechanisms generate it. In particular, by imposing an s-shape ex-ante it is difficult to make any inferences about the underlying distribution of preference parameters in the population.

This is not necessarily a problem if the researcher is predominantly concerned with the identification and forecasting of macro-trends of the existing or similar future technologies. For example, Portugal entered the 2G cellular market in 1992, and early results on diffusion speed may have been helpful to predict the take-up in geographically (e.g. Spain – 2G service started in 1994) or culturally close (e.g. Brazil – 1996 start date) countries. Controlling for other determinants of technology diffusion, researchers can also attempt to identify the strength of network effects by comparing the slope of an actual diffusion curve with one that would have been generated if there were no network effects.

The mere shape of a diffusion curve is less helpful however for actors that actively want to influence the diffusion curve. For example, regulators or governments interested in accelerating the speed of technology diffusion will want to know what the effect of a subsidy for early adopters versus general information provision will be (Stoneman and David, 1986). If diffusion is driven by an epidemic process, information provision is likely to be more efficient, while a network technology will benefit from building up a critical mass of adopters. Similarly, firms planning to introduce a new technology will assess the benefits of a penetration pricing strategy in terms of their short-term losses versus the long-term expected benefits, both of which are affected by the distribution of preferences and the extent of network effects. An empirical procedure that permits discrimination between the different causes for technology diffusion is therefore likely to be beneficial to both policymakers and firms.

Another factor influencing the cellular telephony diffusion that has drawn researcher' attention is substitution between different generations of telecommunications technologies. Intergenerational effects have been studied both between the cellular and the fixed-line technologies (Sung et al., 2000) and between different generations of the cellular technology (Liikanen et al., 2004). Our contribution to that literature is twofold: first we report some results on fixed-mobile substitution usage in terms of usage, second we consider possible change in the substitution pattern depending on the stage of cellular diffusion.

IV Data

We draw our data predominantly from two sources: The Informa Telecoms & Media's World Cellular GSM Datapack and the Merrill Lynch's Global Wireless Matrix. The Informa T&M data has been used in previous studies (e.g. Koski and Kretschmer, 2005a) and covers the number of subscribers for individual mobile operators, average prices and technological standards in considerable detail. Informa Telecoms and & Media is a provider of market and business intelligence to commercial entities in the mobile and media industries. Buyers of this data base commercial and marketing decisions on the data, thus ensuring a high level of accuracy. Merrill Lynch, a US-based investment bank, publishes a quarterly report on the development of the global cellular telephony market as a service to their clients and industry observers. ML reports, among other data, the total number of called minutes per operator, which can be used to construct an average usage per consumer.

Obtaining data often involves a tradeoff between the level of detail (which is often higher in commercial datasets) and the reliability of the data (which is generally regarded higher for data collected by non-commercial organizations). In order to minimize these problems, we triangulated the data with available public data sources (OECD's Communications Outlook, ITU's Telecommunications Indicators) and found that the variables common to both private and public data were comparable, so that we are quite confident that our data is accurate.

To complement our main data sources, we additionally use IMF's International Financial Statistics (for GDP) and World Bank's World Development Indicators (for population, telephone mainlines, and average cost of a local call). The disadvantage of the WDI database is that it only provides yearly time series. To arrive at the quarterly series we therefore linearly interpolated the variables.

Descriptive statistics

Table 1 gives descriptive statistics of our variables. Looking at time trends of our variables of interest in table 2, we can see that the increased penetration of cellphones in our sample coincides with a significant increase in the share of prepaid consumers. We also find a clear downward trend in cellular service prices in our sample and an upward trend in average usage. Contrary to the cellular telephony, fixed-line telephony experienced a slight decrease in subscription over the period studied. Fixed-line prices exhibited no significant trend.

Our data reflects some of the interesting dynamics in the cellular phone industry in the late 1990 and early 2000s. Diffusion was rapid – penetration rates increased almost four-fold over six years – and prepaid usage went from being an option chosen by one cellphone subscriber in four to the option preferred by 50% of users. Of course, a look at our sample average is likely to hide a number of idiosyncrasies, in particular some of the effects we are interested in. We therefore consider the diffusion and usage patterns of two individual countries.

Two examples – Chile and Malaysia

Our goal is to link usage patterns to different stages of the diffusion curve. The following figures plot diffusion and average usage in Chile and Malaysia, respectively.

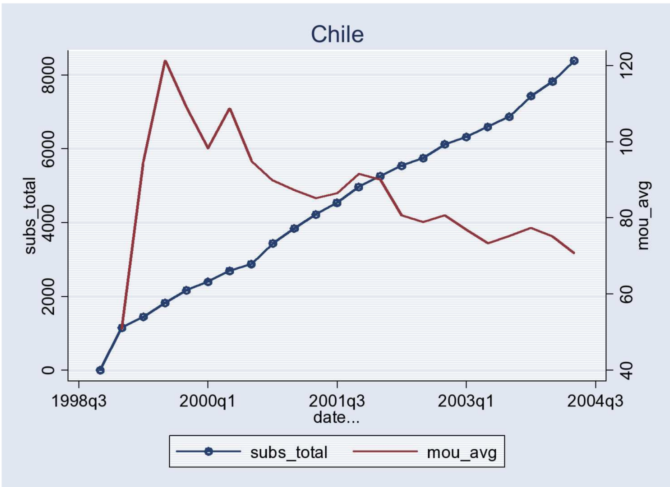


Figure 1: Chile, Diffusion and average usage, 9/98 – 9/04

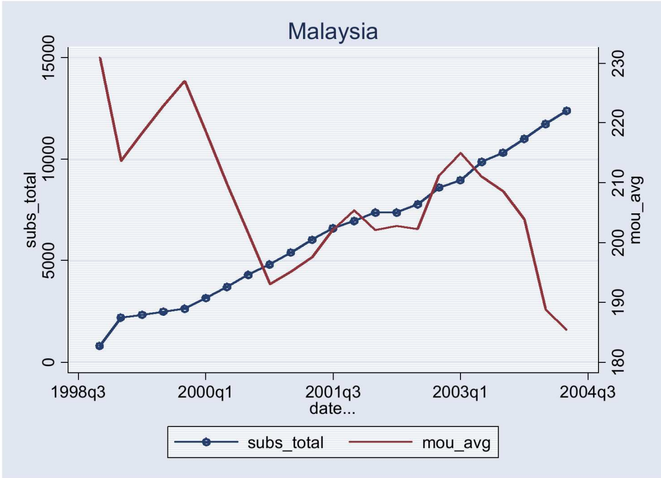


Figure 2: Malaysia, Diffusion and average usage, 9/98 – 9/04

From figure 1 we can see that Chile seems to follow the pattern of first increasing, then decreasing average usage. Figure 2 shows that average usage in Malaysia is decreasing fairly steadily over time. Simple regressions confirm that a linear time trend yields a poor fit for Chile (Slope: -.201, R^2 = .008), while it generates a relatively good fit for Malaysia (Slope: -1.048, R^2 = .374). Including a nonlinear term improves results for Chile to an R^2 of .253, while results remain similar for Malaysia (R^2 = .423).

The above descriptive statistics suggest that usage patterns can vary significantly across countries, despite the fact that diffusion is s-shaped in both countries. Clearly, these statistics should be interpreted with caution since we do not control for important country- and firm-level variables. For example, it may be significant that Malaysia had an approximately 50% higher penetration rate throughout the sample and that the Malay economy grew by about 55% in the same time period versus 20% growth in Chile. The different patterns in average usage suggest that cross-country usage is worth studying in more detail. In the following section therefore, we will first run some fairly standard diffusion regression before looking at determinants of average usage by cellular operator.

V Empirical specification and results

Global diffusion of mobile telephony

We start our empirical analysis by looking at the diffusion of mobile penetration in countries from our sample. The wide coverage of the sample implies that we consider countries that already reached near full penetration along with countries in which mobile telephony just takes off. To provide some descriptive summary of the diffusion process we estimate country-wise logistic diffusion equation of the form:

$$SUBS_t = \frac{SUBS^*}{1+\exp(-\beta(t-\tau))}, \qquad (1)$$

where $SUBS^* = \gamma POP$.

$SUBS_t$ denotes the number of subscribers at time t, and POP measures the population of a country. The potential number of adopters $SUBS^*$, i.e. the saturation level to which $SUBS_t$ converges, is assumed to be equal to a fraction γ of the total country's population. The other two parameters describing the diffusion process, τ and β, stand for timing and speed of the diffusion respectively. τ indicates the inflection point of the logistic s-curve, thereby shifting the curve forwards and backwards. β gives the growth rate of $SUBS_t$ relative to its distance to the saturation level: $\frac{dSUBS_t}{dt}\frac{1}{SUBS_t} = \beta\frac{SUBS^* - SUBS_t}{SUBS^*}$. In particular, the growth rate attains its maximum, $\frac{\beta}{2}$, at the inflection point $t = \tau$.

Table 3 presents Nonlinear Least Squares (NLS) estimates of the country-wise regressions. τ is measured in quarters: the average τ is approximately 163, which corresponds to the 4th quarter of 2000. In order to stress the different diffusion stage across countries in our sample, we draw the actual and fitted penetration levels for three groups of countries: leaders, followers, and laggards, based on the estimates of the timing coefficient τ in figures 3 – 5.

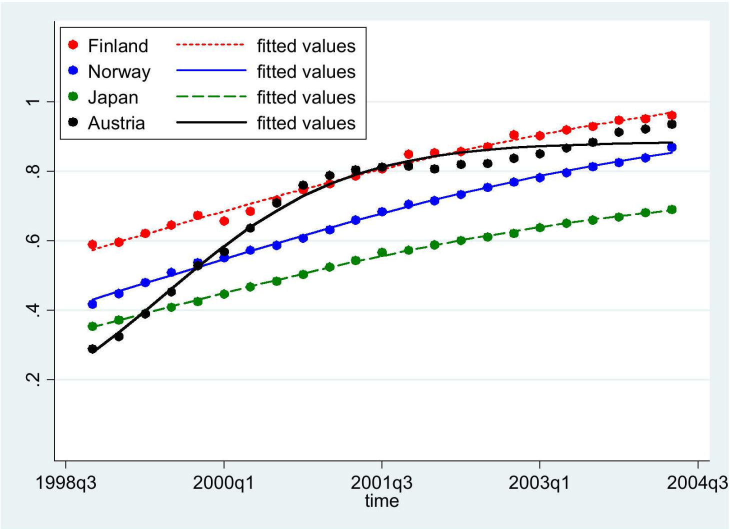


Figure 3. Leaders' mobile penetration diffusion

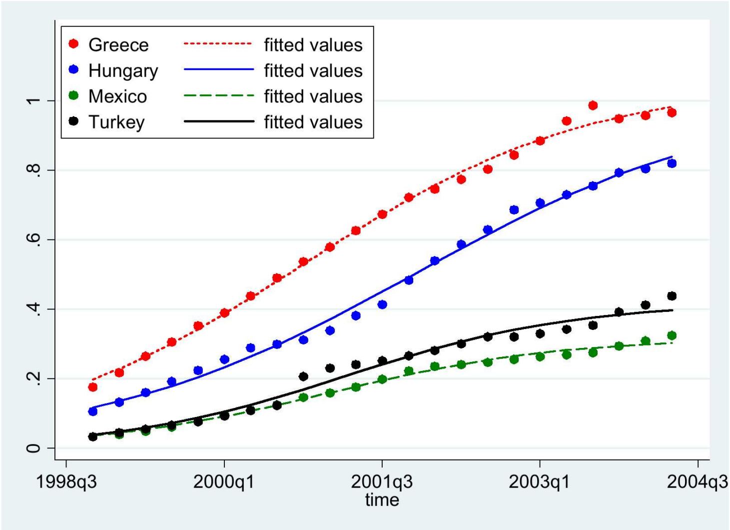


Figure 4. Followers' mobile penetration diffusion

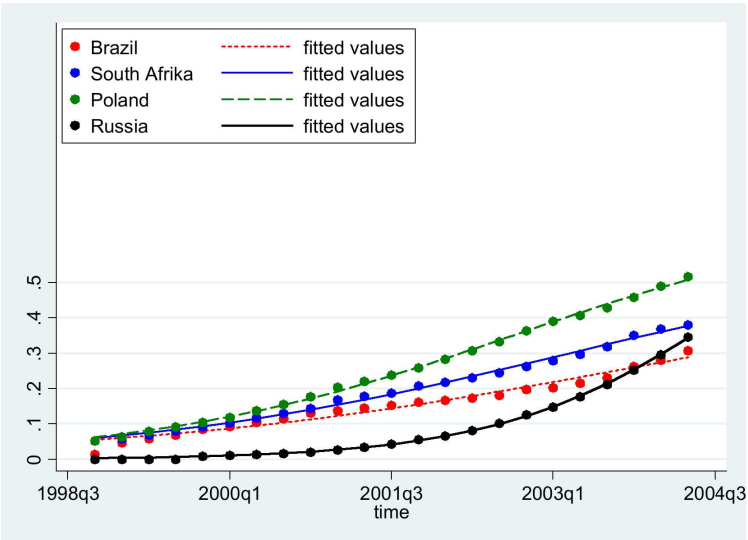


Figure 5. Laggards' mobile penetration diffusion

Usage regressions

In a second step, we run regressions on the average usage of adopters in our sample countries. Usage intensity has so far been largely ignored in the literature on cellular telephony and many other communication technologies, which is surprising given the implications of different usage patterns for firms and policymakers. For example, firms will most likely consider a technology that is widely adopted, but not used intensively, a failure because it does not generate a steady revenue stream for the firm. Similarly, patterns of replacement of an existing technology will be different depending on the usage intensity of the new technology – adoption subsidies that make it easy for users to own multiple devices but continue using the incumbent technology are not particularly beneficial, both from a firm's and society's point of view. After having established in the previous section that diffusion of cellular telephony follows the expected pattern and having identified countries at different diffusion stages, we therefore take average usage intensity as the dependent variable and try to identify the determinants of usage intensity in our sample of countries. Note that our dataset allows us to run usage regressions on an operator level. This is useful since operators in the same country may have different characteristics, for example the proportion of prepaid users or the installed base of subscribers, which both are expected to have an impact on the average usage intensity of this particular operator. Of course, including average prices by operator lets us uncover own-price and cross-price effects on communication demands.

Our baseline specification of the cellular phone usage reads as follows:

$$MoU_{ijt} = \alpha_{ij} + \beta_1 * CellP_{ijt} + \beta_2 * CellP_{i(-j)t} + \beta_3 * FixedP_{it} + \beta_4 * CellSubs_{ijt} + \beta_5 * CellSubs_{i(-j)t} +$$
$$+ \beta_6 * FixedSubs_{it} + \gamma * \boldsymbol{X}_{ijt} + e_{ijt}, \qquad (2)$$

where subscripts i, j, and t stand for country, cellphone operator, and time, respectively. The dependent variable is the average monthly minutes of use. We consider both own and cross price effects on cellphone usage by including the operator's (j) own price, the average price of other cellphone operators in the country ($-j$), as well as the price of local fixed-line connection, in the regressions. Similarly, we distinguish between the operator's own network of subscribers, the subscribers to other cellphone operators, and the fixed-line subscribers. To facilitate international comparability, all price variables are in US cents and the subscribers variables are normalized by the total country's population.

The vector **X** contains a number of other factors that correlate with usage: GDP per capita, share of prepaid-card users in the own subscriber base, and the time on air. GDP controls for income effects on cellphone usage. The share of prepaid-card users in the own subscriber base captures the impact of nonlinear pricing and self selection of consumers into the light users' group. The time-on-air variable that indicates the time since the launch of the service by an operator captures the increasing usability of cellular telephone due to network roll-out and possibly consumer habit formation. Finally, the α_{ij}s capture the unobserved heterogeneity across countries and operators driven by different pricing regimes (Receiving Party Pays vs. Calling Party Pays), different tastes for communications services (Italians tends to talk more than Swedes), incumbent's first-mover advantage, and other country and operator-specific effects that are time invariant.

The standard textbook argument suggests positive cross-price elasticity for substitutes and negative one for complements. We then expect a positive coefficient on the price of other cellphone operators in equation (2), as the cellular services offered by competing operators are clearly substitutes. The relationship between fixed-line and cellular phones is less clear. The empirical literature finds evidence of both substitution and complementarity between them by looking at subscription rates (Ahn and Lee, 1999; Barros, Cadima, 2000; Sung, Kim, and Lee, 2000; Rodini, Ward, and Woroch, 2003). In our setting, the substitution/complementarity between fixed-line and cellular service is primarily captured by the cross-price effect. The coefficient on fixed-line subscriber network, however, is also indicative of the relationship between these two communications platforms. For instance, a negative effect of fixed-line network size on cellular usage in equation (2) would suggest shifting usage between the platforms induced by a subscription decision, i.e. a substitution relationship.

The subscriber network variables in equation (2)—the main focus of this study—capture the changing cellphone usage patterns along with the diffusion of the cellular service among consumers. By including these variables we intend to shed some light on the mechanisms behind the diffusion process. As already mentioned, the economic literature delivers some predictions concerning the usage intensity of a diffusing technology. If the diffusion is solely driven by consumer heterogeneity—accompanied by falling price or increasing quality over time—,we expect a negative coefficient on the network size variable, as subscribers that are added to the network later on decrease the average usage intensity. This effect can be linked to the Stoneman's (1999) *rank effect*, which says that earlier (later) adopters have a higher (lower) preference for the product. If, however, strong *network effects* are present, we expect that increasing communications opportunities due to growing network will offset the rank effect leading to an increasing usage intensity of a network (Cabral, 2002). Depending on which of the effects dominates—*rank/consumer-heterogeneity effect* or *network effect*—the own cellular network variable will carry negative or positive sign, respectively.[5]

Unlike the *rank effect*, the *network effect* will also work for subscribers to the other cellular operators (*-j*) within a country.[6] This is because by adding additional subscribers to the competing cellular networks the overall network and thereby communications opportunities grow while leaving the composition of the own subscriber base unchanged. On the other

[5] The word-of-mouth diffusion models do not deliver any prediction concerning the usage intensity of a diffusing technology.
[6] This is true to the extent that the network effects operate on the cellular industry level and not solely on the operator level.

hand, the subscribers-to-the-other-cellular-operators variable, unlike own-subscribers variable, captures the *substitution effect* that we already mentioned in the case of fixed-mobile substitution. Although, holding fixed-line and cellular connection at the same time—a prerequisite for the usage substitution between platforms—is much more common than holding two cellular connections with different operators; the latter is also observed, in particular in the matured cellular telecom markets (Wireless Intelligence, 2006). We then expect the sign on the network of the other cellular operators in equation (2) to be determined be the relative extent of the *network* and the *substitution effects*.[7]

Finally, we expect cellphone usage to exhibit positive income effect, captured by the positive coefficient on GDP in equation (2). The other control variables: the share of prepaid users in the operator's network and the time on air are expected to have negative and positive effect on cellphone usage, respectively.

Our regression results are reported in tables 4 and 5. To strip out operator-specific effects α_{ij}s, we apply first-differenced (FD) estimation.[8] Further, to test robustness of our results, we consider two alternatives to the linear specification in (2): the log-linear and the log-log specifications. Besides the robustness check, the log-log specification is useful, as its coefficients can be interpreted as elasticities. Finally, we control for the possible endogeneity of prices and adoption decision in the usage equation by applying instrumental variable (IV) technique. We take the number of cellular operators and country-specific time dummies as instruments for prices and for subscriber figures we take their values lagged by 2 periods.

We start interpretation of our empirical results by looking at the FD regressions (table 4). In general, our control variables—share of prepay users, time on air, and GDP—are significant

[7] By the same token, we might expect the fixed-line-subscribers variable to also capture some network effects; although, the extent of the network effects between fixed and cellular networks is most likely to be smaller than between cellular networks of different operators.

[8] An alternative would be to use fixed-effects (FE) estimation. The high autocorrelation of the error term (close to 1) that we found under FE suggested that FD provides more efficient estimates (Wooldridge, 2002).

and have expected signs across specifications. The average usage on an operator network decreases with the share of prepay users and increase with the time on air as well as GDP. Further, we find a consistent negative own-price effect and a consistent positive cross-price effect on cellphone usage. These effects are strongly significant across all specifications. From the linear specification we read that an increase of the revenue per minute from cellular service by 1 US cent is accompanied by a decrease in the average monthly usage of a customer by 1.8 minutes. In terms of elasticity, the corresponding coefficient in the log-log specification equals -0.36. The magnitude of the cross-price effect is smaller, as we expect from imperfect substitutes.

The price of fixed-line telephony, however, is not significant when entered in simple form into the estimation equation. To further investigate the relationship between the old and the new telecommunications technologies we interact the fixed-line price with a diffusion stage indicator. The indicator is constructed from the estimates of the country-wise diffusion regressions (table 3) and equals one in periods after a country reached the inflection point of the cellular telephony diffusion (τ) and zero otherwise. The motivation behind the varying fixed-price coefficient is that the relationship between the old and the new telecommunications technology might be quite different depending on the diffusion stage of the latter. Initially, when the market penetration of cellphones is low, we might expect a lot of traffic between the fixed-line and the cellular networks, as only very few contacts have a cellular connection. On the other hand, when the cellular market penetration reaches full market size, all communications needs can in principle be satisfied on the cellular network alone. Therefore, we might expect that the two technologies change form initial complements to substitutes in the later stages of the diffusion process. Our results provide some support for this hypothesis. Our new composite variable is positive and significant in the linear and the log-log specifications, which means that the two technologies become more substitutable in the mature diffusion stage. This rising substitutability is accompanied with a significant negative coefficient on the fixed-line price suggesting initial complementarity. At the same time, we find the fixed-line penetration to have a significant negative effect on the cellphone usage. The magnitude of this effect seems to diminish in the later diffusion stages. As we already mentioned, the negative coefficient on the fixed-line network indicates that the two telecommunications platforms are substitutes. The coexistence of the complementary and the substitution elements we find is in line with the results in Sung et al. (2000). They report that the number of Korean mobile subscribers is positively (negatively) correlated with the number of fixed-line disconnects (new connections), which suggests substitution. At the same time, they find the stock of fixed lines being positively correlated with the number of mobile subscribers, providing evidence of complementarity.

Further, turning to the subscriber network variables we find strong evidence of the consumer heterogeneity effect dominating the network effect. The coefficient on own market penetration is significantly negative in all specifications, which means that adding subscribers to own cellular network decreases the average cellphone usage. The magnitude of this effect is not marginal. As we read from the linear specification for instance, the average minutes of use decrease by 1.64 with an increase of the penetration by 1 percentage point. We do not find indication of the network effect in the cellular competitors' network variable either. On the contrary, the coefficient on market penetration of cellular competitors is either insignificant or negative depending on functional specification. This negative coefficient gives some evidence of the effect of multiple cellular connections held by a consumer.[9] The inability to identify the network effects in cellphone usage is somewhat surprising given that the previous literature found network effects to significantly contribute to the speed of the cellular diffusion process (Koski and Kretschmer, 2005a; Liikanen et al., 2004). While our results do not imply that

[9] This effect is also referred to as multihoming (Doganoglu, Wright, 2006).

there are no network effects, we can at least conclude that they do not outweigh the consumer heterogeneity effect, and that no significant network effects seem to operate across different cellular networks.

Finally, to account for possible endogeneity of our regressors, we report also the results of instrumental variable (IV) regressions in table 5. Most of our previous results are confirmed, with the important exception that now the cross-price effects are no longer significant. In particular, the price of fixed-line connection is not significant, even after the introduction of intersection with the diffusion stage indicator. This suggests that our fixed-mobile substitution results should be treated with caution.

VI Interpretation and Discussion

The relationship between mobile usage and the network size is determined by two countervailing forces: network effects and consumer heterogeneity effects. Network effects arise as the growing installed base of subscribers allows them to satisfy more communications needs, hence the average number of calls increases with network size. Consumer heterogeneity effects imply that usage of telecommunications services decreases with the installed base of subscribers, as less and less eager users subscribe to the service over time and "dilute" the intensity of use as installed base grows.

One of the problems in estimating the relative strength of these effects is that adding subscribers has the dual effect of enlarging an operator's network and adding lower-preference users to the network. Our regressions suggest that in all specifications the heterogeneity effect strongly dominates the network effect since the coefficient on own network penetration is consistently negative and significant. A potential strategy to isolate the network from the composition effect was to consider the subscribers of competing networks since the composition of an operator's own network does not change while overall network size grows. In our regressions, however, we find that competing network size does not have a significant positive effect on own usage intensity. While this does not imply that there are no network effects, we can at least conclude that they do not outweigh the heterogeneity effect, and that no significant network effects seem to operate across different cellular networks.

As we already mentioned, there is small but growing literature on substitutability of fixed-line and cellular telephony. Our regression results suggest that the degree of substitutability depends strongly on the stage of the diffusion process. There seems to be a significant degree of complementarity between fixed-line and cellular usage in the earlier stages of the diffusion curve. These effects however grow significantly weaker as countries advance in their diffusion stage, suggesting that an incumbent technology like fixed line telephony may foster diffusion at the start of cellular diffusion. This result should be treated with caution though, as it is not supported after controlling for the endogeneity of fixed-line price in the cellphone usage equation. On the other hand, the two telecommunications platforms seem to be substitutes in terms of subscriptions, as the size of fixed-line network is negatively correlated with the cellphone usage. This substitution effect becomes weaker, however, as diffusion progresses, possibly due to consumer heterogeneity.

We also find (weak) support for increasing usage intensity over time after other determinants are controlled for. This could be due to learning effects, habit formation, or even indirect network effects which we do not pick up in our network size variables. We aim to build on this tentative result.

VII Conclusions and further research

We study the usage patterns of cellular telephony over time using data from 41 countries over the 1998-2004 time period. Our reduced-form regressions have uncovered a number of interesting findings. First, it seems that consumer heterogeneity is considerable and network effects are moderate in comparison. Second, we find some evidence of fixed-mobile usage complementarity in the early stages of diffusion. At the same time we observe substitution of fixed-line with cellular minutes driven by the changes in fixed-line subscriber base. This effect also seems to wear off later as cellular telephony becomes more established.

Although the coefficients are consistent across most specifications, they are not always significant, and we will refine the model further to obtain more robust results. We are also planning to explore a number of different avenues in future work:

Functional form of network effects and heterogeneity: Our current reduced-form approach does not allow us to make precise inferences about the shape of the preference distribution or the functional form of network effects individually, but rather the net effect of both. One way to further separate out composition and network effects would be to assume a sufficiently general distribution of preferences (taken, e.g., from Rogers, 2003), a functional form for network effects (e.g., Swann, 2002), and a degree of compatibility between the networks of different operators (e.g. Grajek, 2004). This would enable us to interpret the coefficients we obtain quantitatively, obviously at the cost of having to make some, possibly quite restrictive assumptions.

Role of prepaid consumers: We find unequivocally that the proportion of prepaid consumers has a negative effect on average usage intensity, as expected. It would be interesting to see if the overall effect of prepaid users is to increase network size and thus usage incentives for contract consumers. This would show up in our regressions as a positive effect of competitors' prepaid consumers on own average usage. Further, we hope to be able to take advantage of differences in the share of prepaid consumers for different operators in the same country.

Getting some further insight on the shape of consumer preferences would have significant implications for firm and policymaker behavior. Strong consumer heterogeneity would suggest that early adopters are much more profitable than later ones. This would then make introductory pricing a double-edged sword: On the one hand, securing these early customers is likely to have long-term benefits, while on the other hand these early consumers are likely to represent a large proportion of a firm's profits.[10] Similarly, diffusion policies will be assessed on their expected impact, which depends on the distribution of consumer preferences and the intensity of network effects.

This study is the first to our knowledge that empirically tries to disentangle the consumer-heterogeneity and the network effect on technological diffusion. We believe that while there have been a number of recent studies on the diffusion of mobile telephony (including our own) recovering some information on the underlying parameters and the subsequent causes of diffusion is a crucial next step in the study of new technologies. Our study is also the first to allow for time-varying effects of an incumbent technology, which has important implications for policymaker and firms in their incentives to phase out existing technologies.

[10] There is an extensive literature in marketing science that concerns itself with the optimal pricing path of new products based on assumptions about the s-shaped diffusion curve, but not on the origins of the s-curve (see, e.g., Krishnan et al., 1999).

Bibliography

Ahn, H., Lee, M.-H. (1999). An econometric analysis of the demand for access to mobile telephone networks, *Information Economics and Policy* 11, 297-305.

Barros, P.P. and Cadima, N. (2000): The impact of mobile phone diffusion on the fixed-line network. *CEPR discussion paper* 2598.

Battisti, G., P. Stoneman (2006). The intra-firm diffusion of new process technologies, *International Journal of Industrial Organization*, forthcoming.

Cabral, L. (2002). Equilibrium, Epidemic and Catastrophe: Diffusion of Innovations with Network Effects. Mimeo, *New York University.*

Cabral, L., T. Kretschmer (forthcoming). Standards Battles and Public Policy. Forthcoming in S. Greenstein and V. Stango (eds.), *Standards and Public Policy*, Cambridge University Press.

Dekimpe, M., P. Parker and M. Sarvary (2000), Global Diffusion of Technological Innovations: A Coupled-Hazard Approach. *Journal of Marketing Research*, 47-59.

Doganoglu, T., J. Wright (2006). Multihoming and compatibility, *International Journal of Industrial Organization*, forthcoming.

Doganoglu, T., L. Grzybowski (2004). Estimating Network Effects in the Mobile Telecommunications Industry in Germany. Mimeo, *University of Munich.*

Farrell, J., P. Klemperer (2005). Coordination and Lock-In: Competition with Switching Costs and Network Effects. Forthcoming, *Handbook of Industrial Organization*, Vol 3.

Geroski, P. (2000): Models of technology diffusion. *Research Policy* 29, 603-625.

Grajek, M. (2004). Estimating Network Effects and Compatibility in Mobile Telecommunications, *WZB* Working Paper SP II 2003-26.

Hoernig, S. (2005). Equilibrium Confusion? Mimeo, *Universidad Nova de Lisboa.*

Koski, H., T. Kretschmer (2005a). Entry, Standards and Competition: Firm Strategies and the Diffusion of Mobile Telephony. *Review of Industrial Organization*, 1/26 (2005), 89-113.

Koski, H., T. Kretschmer (2005b). Emergence of Dominant Designs in Mobile Handsets. Mimeo, *Helsinki School of Economics* and *London School of Economics*.

Krishnan, T., F. Bass and D. Jain (1999). Optimal Pricing Strategy for New Products. *Management Science*, 1650-1663.

Liikanen, J., Stoneman, P., Toivanen, O. (2004). Intergenerational Effects in the Diffusion of New Technology: The Case of Mobile Phones. *International Journal of Industrial Organization* 22, 1137-1154.

Miravete, E., L. Röller (2004). Estimating Markups under Nonlinear Pricing Competition. *Journal of the European Economic Association*, 526-535.

Naranayan, S., P. Chintagunta, and E. Miravete (2005). The Role of Self Selection and Usage Uncertainty in the Demand for Local Telephone Service. Mimeo, *University of Pennsylvania*.

Rodini, M., M. Ward, G. Woroch (2003). Going Mobile: Substitution between Fixed and Mobile Access, *Telecommunications Policy* 27, 457-476.

Rogers, E. (2003). *Diffusion of Innovations*. Fifth Edition, Free Press, New York.

Sung, N., Kim, Ch.-G., and Lee, Y.-H. (2000). Is a POTS Dispensable? Substitution Effects Between Mobile and Fixed Telephones in Korea. *Available at SSRN*: http://ssrn.com/abstract=222288 or DOI: 10.2139/ssrn.222288

Swann, P. (2002). The Functional Form of Network Effects. *Information Economics and Policy* 14, 417-429.

Wireless Inteligence (2006). Multiple devices: real penetration in Western Europe.

Wooldridge, J., M. (2002). Econometric Analysis of Cross Section and Panel Data, *Cambridge and London: the MIT Press*.

Table 1. Variable definitions and descriptive statistics

Variable	Definition	Mean	Std. Dev.	Min	Max
MoU	Average monthly minutes of use	177.49	117.97	56	960
CellP	Average revenue per minute (US cents)	21.41	9.41	3.26	79.03
FixedP	Price of a local fixed-line connection (US cents)	8.46	5.58	0	23
CellSubs(j)	Own subscribers as population's share (%)	15.92	12.12	0.03	54.42
CellSubs(-j)	Subscribers to competing operators as population's share (%)	34.12	19.53	0.61	85.49
FixedSubs	Fixed-line subscribers as population's share (%)	47.03	18.36	10.06	75.76
Prepay	Share of prepay users among own subscribers (%)	44.22	29.31	0	97.74
OnAir	Time since the launch of service (quarters)	21.99	12.99	0	50
GDP	GDP per capita (000's US dollars)	21.06	11.90	0.76	51.98
Stage	Diffusion stage indicator (1 after a country reached the inflection point of the cellular telephony diffusion; 0 otherwise)	0.77	0.42	0	1

Table 2. Descriptive statistics by year (variable definitions as in Table 1)

	1998	1999	2000	2001	2002	2003	2004
MoU	162.42	157.67	170.82	176.68	177.15	185.26	198.44
CellP	35.70	31.88	23.60	19.36	18.30	19.64	19.31
FixedP	8.79	9.63	8.27	7.86	7.75	8.56	9.61
CellSubs(j)	6.64	9.62	12.57	15.02	17.06	18.99	20.38
CellSubs(-j)	12.55	17.00	24.40	32.55	37.55	42.89	45.94
FixedSubs	49.28	48.39	47.37	47.39	46.99	45.22	45.68
Prepay	24.16	30.11	37.76	43.81	47.65	50.71	49.63
OnAir	14.32	15.67	16.80	19.14	22.93	26.43	29.43
GDP	20.45	21.31	20.09	18.84	19.76	22.02	25.39
Stage	0.00	0.17	0.63	0.82	0.88	0.96	0.99

Table 3. Country-wise logistic diffusion coefficients

	gamma	beta	tau	observations	R^2
Argentina	0.187	0.280	157.1	23	0.991
Australia	0.933	0.118	162.3	23	0.999
Austria	0.887	0.290	157.7	23	0.999
Belgium	0.824	0.292	160.4	23	0.998
Brazil	0.534	0.105	175.5	23	0.993
Canada	0.498	0.111	162.1	23	1.000
Chile	0.561	0.177	166.2	23	0.996
China	0.291	0.183	169.7	23	0.999
Colombia*	.	.	.	.	.
Czech Republic	1.028	0.247	164.5	23	0.999
Denmark	1.106	0.136	161.2	23	0.999
Egypt	0.104	0.198	168.1	23	0.996
Finland	1.130	0.080	154.6	23	1.000
France	0.698	0.239	159.5	23	1.000
Germany	0.765	0.311	160.2	23	0.997
Greece	1.057	0.185	163.0	23	0.999
Hungary	1.008	0.165	167.3	23	0.999
India*	.	.	.	.	.
Ireland	0.830	0.286	160.2	23	0.996
Israel	1.004	0.184	160.4	23	0.999
Italy	1.029	0.180	158.8	23	1.000
Japan	0.777	0.102	156.9	23	1.000
Korea	0.828	0.124	160.0	23	0.998
Malaysia	0.639	0.143	169.4	23	0.997
Mexico	0.319	0.227	164.0	23	0.998
Netherlands	0.820	0.299	158.7	23	0.998
New Zealand	0.710	0.230	160.0	23	0.997
Norway	0.997	0.094	157.9	23	1.000
Poland	0.698	0.150	170.5	23	1.000
Portugal	1.086	0.194	160.6	23	0.999
Russia	0.873	0.231	178.9	23	1.000
Singapore	0.889	0.231	160.0	23	0.998
South Africa	0.579	0.126	172.1	23	0.999
Spain	0.929	0.221	161.0	23	0.999
Sweden	1.252	0.088	160.7	23	1.000
Switzerland	0.880	0.235	159.2	23	0.999
Thailand	0.461	0.264	169.9	23	0.996
Turkey	0.420	0.233	164.7	23	0.994
United Kingdom	0.924	0.262	159.9	23	0.999
United States	0.674	0.101	161.3	23	1.000
Venezuela	0.269	0.218	159.7	23	0.999
Cross-country average	0.756	0.193	162.9		

* Missing coefficient indicate that the NLS estimation procedure did not converged.

Table 4. First-Differenced (FD) Estimation Results

Dependent variable: Average minutes of use (MoU)	Linear specification		Log-linear specification		Log-log specification	
	(1)	(2)	(3)	(4)	(5)	(6)
Own price	-1.821***	-1.833***	-0.012***	-0.012***	-0.361***	-0.360***
	(0.229)	(0.230)	(0.001)	(0.001)	(0.046)	(0.046)
Average price of mobile competitors	0.588***	0.575***	0.004***	0.004***	0.195***	0.195***
	(0.148)	(0.149)	(0.001)	(0.001)	(0.041)	(0.041)
Price of local fixed-line connection	-0.016	-3.081**	0.002	-0.008	-0.044	-0.146***
	(0.520)	(1.306)	(0.003)	(0.009)	(0.031)	(0.039)
Fixed-line price * diffusion stage indicator		3.404**		0.012		0.129***
		(1.395)		(0.009)		(0.047)
Own market penetration	-1.640***	-1.632***	-0.007***	-0.007***	-0.039**	-0.039**
	(0.590)	(0.595)	(0.002)	(0.002)	(0.016)	(0.017)
Market penetration of mobile competitors	-0.008	0.015	-0.002*	-0.002	-0.053**	-0.052*
	(0.266)	(0.262)	(0.001)	(0.001)	(0.026)	(0.026)
Fixed-line market penetration	-2.573*	-6.050***	-0.009*	-0.026***	-0.041	0.194
	(1.385)	(1.785)	(0.005)	(0.009)	(0.181)	(0.394)
Fixed-line penetration * diffusion stage indicator		4.020*		0.019*		-0.261
		(2.082)		(0.011)		(0.448)
Share of own prepay users	-0.150	-0.131	-0.001**	-0.001**	-0.020**	-0.020**
	(0.104)	(0.104)	(0.001)	(0.001)	(0.010)	(0.010)
Time on air	1.776**	1.855**	0.005	0.005*	0.039	0.039
	(0.849)	(0.846)	(0.003)	(0.003)	(0.025)	(0.026)
GDP	1.327***	1.335***	0.010***	0.010***	0.178***	0.179***
	(0.241)	(0.243)	(0.001)	(0.001)	(0.048)	(0.048)
Adjusted R-squared	0.117	0.118	0.217	0.218	0.217	0.219
Number of observations	1634	1634	1634	1634	1356	1356

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are robust to heteroscedasticity and autocorrelation in the error term.

Table 5. First-Differenced (FD) Instrumental Variable (IV) Estimation Results

Dependent variable: Average minutes of use (MoU)	Linear specification		Log-linear specification		Log-log specification	
	(1)	(2)	(3)	(4)	(5)	(6)
Own price	-1.898***	-1.910***	-0.008***	-0.008***	0.029	0.024
	(0.647)	(0.643)	(0.003)	(0.003)	(0.176)	(0.176)
Average price of mobile competitors	0.675	0.670	0.000	0.000	-0.179	-0.174
	(0.523)	(0.519)	(0.002)	(0.002)	(0.184)	(0.184)
Price of local fixed-line connection	2.171	6.406	-0.004	0.044	0.009	0.089
	(1.804)	(6.812)	(0.009)	(0.037)	(0.031)	(0.184)
Fixed-line price * diffusion stage indicator		-4.764		-0.054		-0.092
		(6.963)		(0.037)		(0.188)
Own market penetration	-1.867**	-1.871**	-0.006**	-0.006**	-0.022*	-0.023*
	(0.742)	(0.744)	(0.002)	(0.002)	(0.012)	(0.012)
Market penetration of mobile competitors	0.617	0.630	-0.003	-0.003	-0.052*	-0.049
	(0.574)	(0.572)	(0.002)	(0.002)	(0.030)	(0.031)
Fixed-line market penetration	-1.676	-4.197*	-0.007	-0.034**	-0.112	0.059
	(1.173)	(2.334)	(0.005)	(0.013)	(0.162)	(0.491)
Fixed-line penetration * diffusion stage indicator		2.736		0.029**		-0.169
		(2.573)		(0.013)		(0.408)
Share of own prepay users	-0.479***	-0.449***	-0.002***	-0.002***	-0.011*	-0.010
	(0.113)	(0.117)	(0.001)	(0.001)	(0.006)	(0.006)
Time on air	1.540	1.642	0.009**	0.010***	0.037	0.038
	(1.064)	(1.035)	(0.004)	(0.004)	(0.026)	(0.026)
GDP	1.191***	1.212***	0.009***	0.009***	0.160***	0.164***
	(0.268)	(0.271)	(0.001)	(0.001)	(0.046)	(0.041)
Adjusted R-squared	0.064	0.064	0.086	0.090	0.071	0.071
Number of observations	1511	1511	1511	1511	1325	1325

* p<0.1, ** p<0.05, *** p<0.01; standard errors in parentheses are robust to heteroscedasticity and autocorrelation in the error term.

endogenous	instruments
mobile prices and prepay share	#mobile operators, country-specific time dummies
fixed-line prices, diffusion-stage indicator	#mobile operators, time dummies
mobile and fixed-line penetration	lagged by 2 periods

Simulation of Merger in Mobile Telephony*

Lukasz Grzybowski† Pedro Pereira‡

Katholieke Universiteit Leuven Autoridade da Concorrência

June 2006

PLEASE, DO NOT QUOTE NOR CITE, VERY INCOMPLETE VERSION

Abstract

The purpose of this paper is to assess unilateral effects of merger in mobile telephony in Portugal. We estimate system of demands for subscriptions to mobile services using a standard aggregate nested logit model for quarterly data in the period between January 1998 and December 2004. We find that network effects significantly determine consumers decision about subscription to mobile telephony. Demand estimates support the assumption of static Nash-equilibrium in prices. Mobile services provided by particular operators are close substitutes. However, the notified merger may result in significant price increases, even in the presence of substantial cost efficiencies.

Key Words: *mobile telephony, merger simulation, network effects*

JEL Classification: L13, L43, L93

*The opinions expressed in this article reflect only the authors' views, and in no way bind the institutions to which they are affiliated.

†Faculty of Economics and Applied Economics, Naamsestraat 69, 3000 Leuven, Belgium, Phone: +32 496531489, E-mail: Lukasz.Grzybowski@econ.kuleuven.be

‡AdC, Rua Laura Alves, 1050-188 Lisboa, Portugal, e-mail: jppereira@oninetspeed.pt

Contents

1 Introduction **3**

1.1 Literature review 5

1.2 Development of Mobile Telephony in Portugal 8

2 Empirical Model **9**

2.1 Network Effects in Mobile Telephony 11

2.2 The Data 12

2.3 Supply Side 12

3 Estimation Results **15**

3.1 Demand Side 15

3.2 Merger Simulation 16

4 Conclusion **18**

1 Introduction

The purpose of this paper is to assess unilateral effects of merger in mobile telephony in Portugal. This is a very relevant policy question because of pending merger notification among two out of three network operators in Portugal. The complicity of consumer behavior and firm strategies in mobile telecommunications industry poses many challenges for economic modelling. However, in this study we use only limited aggregate data which constraints the sophistication of the estimated model. Also, the competition authorities themselves have limited time to collect data and evaluate merger effects. Personnel constraints are critical, which makes it difficult to conduct sophisticated analysis. These are very important practical arguments for usage of simple methods. Therefore, in this study we employ rather standard methods and standard functional forms which have been successfully employed in earlier literature. This framework allows relatively quickly to get a feeling about the unilateral effects of a merger between network operators. It is also important to note that, so far in Europe, only very few merger simulations were conducted and no study at all for mobile telecommunications industry. Therefore, this study proposes some simple framework to start with.

In the first stage we estimate demand functions using a standard aggregate nested logit model á la Berry (1994). We assume that consumers first decide whether to subscribe to fixed telephony services only or to mobile and fixed services together. By normalizing with respect to the utility of fixed telephony services, it is possible to impute the mean utility levels of subscribing to mobile telephony services via a simple transformation of observed market shares. We then posit a relatively straightforward linear utility for subscribing to mobile services, and search for parameters that allow our linear model to best explain the observed mean utility levels.

Similarly to Doganoglu and Grzybowski (2006) we investigate whether network effects have an impact on the decision about subscription to mobile telephony services in Portugal. It is an important aspect of this analysis because by ignoring network effects we could overestimate price elasticities of demand and underestimate markups. In modelling network effects, we use the lagged total share of subscribers in the population to proxy for network size. This assumes perfect compatibility between services of the providers and the lack of price mediated network effects due to different on-net and off-net prices.

Our analysis is based on quarterly data on subscriptions and average revenues per minute as a proxy for prices in the period between January 1998 and December 2004. This kind of data should be easily accessible for competition authorities.[1] An important modelling issue is the fact that consumers often sign long term deals with their service providers and therefore do not engage in decision making every month. We have information about the number of consumers using pre-paid and postpaid services which allows us to approximate the number of locked in consumers and hence the number of consumers which make decision about subscription to mobile services each month.[2] A simple first difference of the observed subscription levels and the number of locked in consumers should yield a good approximation for the number of renewed or new contracts sold. The other important variables we need are firm specific prices but these are hard to find in a prepared form. We use average per minute revenues as a proxy for prices, that is, consumers make subscription decision based on observed per minute prices.

We use these variables, i.e. total installed base, proxies for sales and prices along with quarterly dummies to estimate a system of demand equations. Our results suggest that network effects played a significant role in the diffusion of mobile services in Portugal. Afterwards, using the estimates from the demand side we derive the estimates of marginal costs. Demand estimates support the assumption of static Nash-equilibrium in prices, while joint profit maximization turns out to be unrealistic assumption resulting in negative computed marginal costs. Thus, we assume that observed prices result from pure strategy Nash equilibrium. Furthermore, we regress the estimates of marginal costs on the data on operating costs, which was reported by the firms. Our computed marginal costs are well explained by firm specific dummies and the average operating costs per subscriber with the coefficient of one. Thus, we use the estimates

[1] It is even available in some business reports, for instance by Merill Lynch.

[2] We do not have precise information about the number of people with expiring contracts and with nil switching costs.

of demand parameters and marginal costs to simulate unilateral effects of a merger between two network operators: TMN and Optimus. Finally, we simulate price changes in the presence of cost efficiencies as well as the estimates of changes in consumer surplus, profits and total welfare. Because of the assumption, that relatively few consumers are locked in, we get very high estimates of price elasticities. Thus, we may perceive our case as the most extreme situation, in which the potential price increase is the lowest.[3] Clearly, incentives to raise prices are stronger when price elasticities are smaller. However, even in our case, notified merger will lead to significant price increases, also in the presence of substantial cost efficiencies.

The next section provides a short overview of empirical literature on demand estimation in the telecommunications industry, network effects and merger simulations. In the further section we present a brief overview of the mobile industry in Portugal. The model which we use for econometric analysis is presented in section 2 along with data description, and the estimation results follow in section 3. Finally, we conclude our analysis in section 4.

1.1 Literature review

There is a large body of literature on estimation of demand for telecommunications services in fixed line and mobile industry. Ben-Akiva, McFadden and Train (1987) conduct an analysis of the choice of local telephone service plans using a nested logit model. A number of other studies provides estimates of price elasticities of demand for mobile subscriptions and demand for mobile originated calls. Ahn and Lee (1999) use cross-country data on 64 countries to analyze determinants of demand for access to mobile telephony. Tishler, Ventura and Watters (2001) estimate demand for mobile subscriptions and for usage of mobile services in Israel. Rodini, Ward and Woroch (2002) estimate the substitutability of fixed and mobile services for telecommunications access using data on U.S. households. Okada and Hatta (1999) analyze dependence between usage of mobile and fixed line services using data for Japan. Cadima and Barros (2000) estimate simultaneously logistic diffusion functions for mobile and fixed line telephony in Portugal.

[3] In case of perfect substitutes there would be no price increase at all.

They find a negative impact of mobile penetration on fixed line density. Finally, Grzybowski and Doganoglu (2006) use nested logit model to estimate demand for subscriptions to mobile services in Germany. In this study we follow their approach and extend it by merger simulation.

There is also a growing body of literature on estimation of demand systems when consumers make both discrete and continues choices starting with Hanemann (1984) and Dubin and McFadden (1984). Telecommunications industry is one of the industries, in which consumers make interrelated choices. They choose network operator, tariff plan and how much to use the phone. There are a few recent empirical studies, which use discrete/continuous models in application to telecommunications industry. Chintagunta, Narayanan and Miravete (2005) develop a structural discrete/continues model to account for both plan choice and usage decisions of consumers in local telephony. Economides, Seim and Viard (2004) adopt a similar framework to estimate a model for choice and usage of local calling plans and investigate the welfare effects of entry of competitors into a monopoly market. Doganoglu and Lange (2004) estimate a discrete/continouse choice model for long distance telecommunications services in Germany. They account for unobserved heterogeneity in the valuation of calls and in the cost of using alternative call-by-call providers and estimate a mixed logit model of provider choice jointly with the duration demand.

In this study we use aggregate data to estimate demand for subscriptions to mobile services. We refrain from estimating demand for usage due to data limitations but we will try to develop such model in the further research. As suggested by Grzybowski and Doganoglu (2006) and some other before, in estimating demand for subscriptions to mobile services it is important to account for the presence of network effects. Otherwise the estimates of price elasticities may be overestimated. There is a growing body of literature that tries to measure indirect and direct network effects in a variety of network industries. A number of earlier papers focuses on estimating a hedonic price function for products that exhibit network effects. Brynjolfsson and Kemerer (1996) use the hedonic pricing model to determine the impact of network effects, defined as compatibility with the dominant standard, on the prices of microcomputer spreadsheets.

Similar approaches are employed by Hartmann and Teece (1990), Gandal (1994), Economides and Himmelberg (1995), Moch (1995) and Gröhn (1999).

More recent papers develop structural models, for instance, Gandal, Kende and Rob (2000) study the diffusion of CD technology and find that the number of CD titles available has an impact on the consumer's willingness to adopt to the CD player. Park (2003) analyzes the role of network effects in the standard war between VHS and Betamax video recording systems. Similarly, Ohashi (2001) estimates a random utility model and measures the role of network externalities in the diffusion of VCR in the U.S. between 1978 and 1986. Clements and Ohashi (2004) estimate indirect network effects in the U.S. video game market between 1994 and 2002 using a nested logit model. For an excellent review of the theoretical and empirical literature on network effects and switching costs, we refer the reader to Farrell and Klemperer (2004).

The empirical studies that account for network effects in the telecommunications industry are relatively scarce. Most studies focus on the diffusion of telecommunications services and use reduced form regressions and diffusion models. The presence of network effects has usually not been taken into account. For instance, Gruber and Verboven (2001) estimate a logistic diffusion model for the EU countries and find that regulation and technological progress are important for the growth of mobile industry. Wallsten (2002) uses data on the telecommunications industry worldwide to analyze whether the sequence of reforms, such as establishing a regulatory authority and privatization of the incumbent, matters. Koski and Kretschmer (2005) analyze the effects of regulation and competition on the development of mobile telephony.

There are only a few recent studies which explicitly acknowledge network effects in the telecommunications industry. Bousquet and Ivaldi (1997) estimate network effects in fixed-line telephony in terms of usage. Kim and Kwon (2003) use a consumer survey to analyze Korean mobile telephony and conclude that consumers prefer carriers that have larger consumer bases. Birke and Swann (2004) use household survey data to identify price-mediated network effects in mobile telephony in the UK. Grajek (2003) estimates the magnitude of network effects in the Polish mobile telephone industry during the period 1996-2001. In this study we estimate demand for subscriptions to mobile services using the same empirical framework as suggested by Grzybowski and Doganoglu (2006).

After estimating a system of demands we proceed with simulation of unilateral effects of merger between two network operators. There is an emerging literature on predicting merger effects. For instance, Nevo (2000) estimates a general random coefficient model to study merger effects in the U.S. ready-to-eat cereal industry. Pinkse and Slade (2004) follow a distance metric approach to study mergers in the brewing industry. Ivaldi and Verboven (2005) analyze unilateral effects between Volvo and Scania in the European Union. In this study we employ a similar approach. First, we use nested logit model to estimate demand system. Afterwards we test for static Bertrand-Nash equilibrium using data on operating costs and simulate potential price changes due to a merger under this assumption.

1.2 Development of Mobile Telephony in Portugal

The firm associated with the telecommunications incumbent in Portugal, *Tmn*, started its activity in 1989 with the analogue technology *C-450*. In 1991, the sectorial regulator, *ICP-ANACOM* , assigned two licenses to operate the digital technology *GSM 900*. One of the licenses was assigned to *Tmn* and the other one to the entrant *Vodafone*. In 1997, the regulator assigned three licenses to operate the digital technology *GSM 1800*. Two licenses were assigned to *Tmn* and *Vodafone* and a third one to the entrant *Optimus*, which was also granted another *GSM 900*.[4] After its inception in 1989, the Portuguese mobile telephony industry enjoyed an exponential diffusion, analyzed in Pereira and Pernias (2004). After entering the market in 1992, *Vodafone* rapidly gained market share. During the duopoly period, i.e., from 1992 to 1997, *Tmn* and *Vodafone* essentially shared the market. The entry of *Optimus* led to an asymmetric split of the market.

[4]Both of the licenses for *GSM 900* and for *GSM 1800* were assigned through public tenders, following the *EU* Directives 91/287 and 96/2, respectively. The first Directive instructed Member States to adopt the *GSM* standard, and the second to grant at least 2 *GSM 900* licenses and to allow additional firms to use *GSM 1800*. System *GSM 900* operates on the 900 MHz frequency. System *GSM 1800* operates on the 1800 MHz frequency.

2 Empirical Model

We model demand for mobile subscriptions by a discrete-choice model, as discussed in Anderson, de Palma and Thisse (1992) and Berry (1994). We follow the estimation strategy proposed by Berry (1994) and invert market-share equations to find the implied mean levels of utility for each alternative. We then posit a functional form for this utility in terms of observed and unobserved variables. The unobserved variable serves as our econometric error term and is interpreted as the mean value of consumers' valuations for unobserved product characteristics, such as product quality, for instance.

We assume that all consumers have access to a fixed line at home, work or public places. In the first stage they decide whether to continue using a fixed telephone alone or to buy a mobile and as well. In the second stage the consumers choose a network operator. This is a standard nested logit structure, where one branch is degenerated and no further choices are made. The utility of an outside option for consumer i at time t is denoted by U_{i0t} and may vary in time due its dependence on prices of fixed line services, for instance. The utility derived by consumer i from using a fixed-line together with mobile services of network operator j can be written as

$$U_{ijt} = U_{i0t} + r_j - \alpha p_{jt} + V(z_t^e) + \zeta_{gt} + \xi_{jt} + (1-\sigma)\epsilon_{ijt} = \delta_{jt} + \zeta_{gt} + (1-\sigma)\epsilon_{ijt} \qquad (1)$$

where r_j is the stand alone value, p_{jt} represents service price and $V(z_t^e)$ is the expected network benefit, which we discuss in detail in the next subsection. The variable ζ_{gt} is a common value of all products in group $g = \{0, 1\}$ and has a distribution dependent on σ. The nest $g = 0$ stands for fixed line alone and $g = 1$ represents the choice of mobile telephony together with a fixed line. By normalizing with respect to the utility of the outside option, the choice of alternatives becomes independent of the determinants of the fixed line utility. The consumer's tastes for products within the nest may be correlated. When the choice of alternatives in the nest is independent, which implies that $\sigma = 0$, nested logit is reduced to a simple logit. Finally, ξ_{jt} accounts for the population average of the unobserved utility of operator j and ϵ_{ijt} is the idiosyncratic taste variable, which has a double exponential distribution.[5] The probability of

[5] The only firm characteristics in the model are prices, stand alone values and unobserved qualities.

consumer i to subscribe to mobile network operator j is equal to:

$$P_{ijt} = P_{ijt}(g = 1)P_{ijt}(U_{ijt} \geq max_{k\in\{1,....,N\},k\neq j}U_{ikt}) \tag{2}$$

and may be written in a closed form formula as:

$$P_{ijt} = \frac{exp(\delta_j/(1-\sigma))}{D_g}\frac{D_g^{1-\sigma}}{1+\sum_{g=1}^{G} D_g^{1-\sigma}}$$

where D_g is defined by:

$$D_g = \sum_{k\in G_g} exp(\delta_k/(1-\sigma))$$

Probability (2) is equivalent to the share of network operator j in the total number of consumers which are free to make a subscription decision at instant t, i.e. $s_{jt} = P_{ijt}$.

Following Berry (1994), we invert observed market shares to compute mean utility levels for each product and treat them as observed. Using the observed utility level and our specification in (1), we arrive at the following estimation equation

$$\log(s_{jt}) - \log(1 - s_t) = r_j - \alpha p_{jt} + V(z_t^e) + \sigma log(\overline{s}_{jt|g=1}) + \xi_{jt} \tag{3}$$

where s_{jt} represents the share of operator j in the total number of consumers that make decisions about the subscription and $s_t = \sum_j s_{jt}$. The share of operator j in the total sales of mobile services is denoted by $\overline{s}_{jt|g}$. The unobserved utility, ξ_{jt}, serves as the econometric error term.

The specification of utility function (1) is representative of consumers with sufficiently low (zero) switching costs and of new consumers. Otherwise, the utility function may depend on the previous choice due to switching costs, for instance. Because we miss precise data on the number of switching consumers and their choices of network operators, we have to make simplifying assumptions.[6] We assume that there are three types of consumers. Consumers with sufficiently low switching costs and new consumers choose network operators, while consumers with high switching costs are locked-in and continue using the same mobile services. Only consumers

[6] It is even more difficult if not impossible to determine the number of consumers which have no switching costs but decide not to switch.

which signed long term contracts are assumed to be locked-in. We use data on accumulated total subscriptions and prepaid subscriptions to approximate sales. Sales of operator j in period t can be computed by the difference $y_{jt} = Z_{jt} - Z^c_{jt-1}$ where Z_{jt} stands for the total number of subscribers and Z^c_{jt-1} is the number of contract subscribers in the last period.[7] The total number of consumers who can make subscription decisions is given by $m_t = M_t - \sum_{m=1}^{N} Z^c_{mt-1}$ and represents our market size. Here, M_t is the whole population of Portugal. Thus, the share of subscribers of network operator j in the total number of consumers that can make subscription decisions is given by $s_{jt} = y_{jt}/m_t$. The share of the outside good is computed as $s_{0t} = 1 - s_t = 1 - \sum_{i=1}^{N} s_{jt}$.

2.1 Network Effects in Mobile Telephony

So far we have not specified how consumers form expectations about network size and how the network benefit function is formulated. Most of the empirical and theoretical literature on network effects assumes linear network benefits. Swann (2002) examines the assumptions on communications needs which are necessary for the utility function to be either linear or s-shaped in the network size. He argues that an s-shaped utility function in the network size is more realistic for an average consumer and that the shape may differ for pioneers, medium adopters and late adopters. In the time period considered in this study, the mobile telephony market in Portugal was in its fastest growth phase. Thus, the network benefits should be well approximated by a simple linear function $V(z^e_t) = \beta z^e_t$.

We employ a very simple rule for the formation of expectations. We assume that consumers think that the network penetration in the last period will be realized in the current period. When the market reaches a steady state, such formation of expectations will be fulfilled. Networks are fully compatible and their users may freely communicate with each other. Thus the expected penetration is represented by the sum of lagged installed bases divided by the size of population, $z^e_t = \sum_j Z_{jt-1}/M_{t-1} \equiv z_{t-1}$. A new subscriber to any of the networks brings the same marginal

[7] Alternatively, as the extreme case, we may assume that all consumers of mobile services in the last period are locked-in. We test robustness of estimates assuming that 50% of prepaid consumers are locked in as well.

utility.

Clearly consumers derive network benefits from a fixed line network as well. In the last years, however, changes in the number of subscribers to fixed line telephony were negligible. Given our assumption regarding the linear network benefit function and the linear form of utility function in (1), any network benefits from the fixed line are cancelled when we normalize with respect to the outside option.

2.2 The Data

The data used in this study was provided by the Portuguese Competition Authority on quarterly basis for the time period from January 1998 to December 2004. We use the information on total subscriptions and on the number of users of pre-paid cards to construct sales. Average prices per minute are constructed by dividing the total firm revenue from calling by the total number of minutes.[8]

To account for the endogeneity of prices and the within group shares we use instrumental variables. We have to find instruments which are correlated with prices and within group shares, but uncorrelated with the unobservable demand shocks. The error terms may be autocorrelated due to the character of data. Thus the usage of lagged endogenous variables, such as lagged consumer base, could be problematic. Other standard candidates for instruments could be input factors, see for instance, Evans and Heckman (1983). We have firm specific information on operating costs and the number of employees, which we use as instruments together with the 4th quarter dummy, which is also variable in the model.

2.3 Supply Side

We assume that firms are aware of the consumer preferences that is they can calculate their expected subscriber bases as a function of market prices. We start the analysis of merger effects

[8]Data on subscriptions and revenues per minute are also available in some business reports, for instance by Merill Lynch. We used one source of data for consistency.

by deriving the elements of substitution matrix for nested logit noted by:

$$S_t = \begin{bmatrix} \frac{\partial s_{1t}}{\partial p_{1t}} & \frac{\partial s_{2t}}{\partial p_{1t}} & \frac{\partial s_{3t}}{\partial p_{1t}} \\ \frac{\partial s_{1t}}{\partial p_{2t}} & \frac{\partial s_{2t}}{\partial p_{2t}} & \frac{\partial s_{3t}}{\partial p_{2t}} \\ \frac{\partial s_{1t}}{\partial p_{3t}} & \frac{\partial s_{2t}}{\partial p_{3t}} & \frac{\partial s_{3t}}{\partial p_{3t}} \end{bmatrix}$$

where

$$s_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} = \begin{cases} -\frac{\alpha}{1-\sigma} s_{jt}\left[1 - \sigma \overline{s}_{jt|g=1} - (1-\sigma)s_{jt}\right] & \text{if } k = j; \\ \frac{\alpha}{1-\sigma} s_{jt}\left[\sigma \overline{s}_{kt|g=1} + (1-\sigma)s_{kt}\right] & \text{if } k \neq j. \end{cases}$$

where s_{jt} is the share of network operator j in sales at time t and $\overline{s}_{jt|g=1}$ is the within group share. The cross price elasticity of demand for outside option is given by

$$\frac{\partial s_{0t}}{\partial p_{kt}} = \alpha s_{kt} s_{0t}$$

where $s_{0t} = 1 - s_t$ is the share of the outside option. We may calculate all elements of matrix S_t using the estimates of σ and α from the demand side. Assuming that firms are myopic and that they set prices only in respect to consumers which are not locked in, we may write the profit function for firm f at time t as:[9]

$$\Pi_{ft} = \sum_{j \in F_f} \left[(p_{jt} - mc_{jt})s_{jt}m_t - FC_{jt}\right] \tag{4}$$

where F_f is the subset of networks owned by operator f, c_{jt} represents marginal cost, FC_{jt} is the fixed cost and s_{jt} is the share of operator j in total sales at time t. Consumers with contracts are excluded as we assume that firms cannot change the prices in their contracts, that is $s_{jt}m_t = Z_{jt} - Z^c_{jt-1}$, where Z_{jt} is the installed base of firm j at time t.[10] In the pre-merger situation each network is owned by different operator. When post-merger firm controls more then one network it internalizes the diversion rate between owned networks. We may define an

[9]Clearly, given the industry dynamics the assumption of static equilibrium is very restrictive but pre-paid users constitute about 80% of all consumers.

[10]In the following we assume that a Nash equilibrium exists. See Anderson et al. (1992) for a proof of existence for the nested logit model with multiproduct firms, assuming symmetry.

ownership matrix $\Delta_{3\times3}$ with values determined by the following rule:

$$\Delta_{jk} = \begin{cases} 1 & \text{if the same operator owns networks j and k;} \\ 0 & \text{otherwise} \end{cases}$$

Thus, the first order conditions may be written as:

$$\frac{\partial \Pi_{ft}}{\partial p_{kt}} = s_{kt} m_t + m_t \sum_{j=1}^{N} \Delta_{jk} \frac{\partial s_{jt}}{\partial p_{kt}} (p_{jt} - mc_{jt}) = 0 \tag{5}$$

and in matrix notation:

$$q_t + \Delta \bullet S_t (p_t - mc_t) = 0 \tag{6}$$

where q_t is a 3×1 vector of elements s_{jt} and similarly p_t and mc_t are 3×1 vectors. $\Delta \bullet S_t$ is known as the Hadamard or element by element matrix product. From this expression, assuming separate ownership ($\Delta = I$) and static Nash equilibrium in prices, we may get the estimates of marginal costs

$$mc_t^n = p_t + (I \bullet S_t)^{-1} q_t \tag{7}$$

Furthermore, we assume that firms collude and jointly maximize profits in which case the ownership matrix is full of ones. The vector of estimates of marginal costs at time t in this case is noted as mc_t^c.

Assuming that operating costs are a good proxy for marginal costs we may test under which equilibrium assumption our cost estimates are at best explained by the data on operating costs. We consider the following dependence between marginal and operating costs:

$$mc_{jt} = \eta_j + \gamma X_{jt} + \omega_{jt} \tag{8}$$

where η_j is firm specific constant, X_{jt} are operating costs and ω_{jt} is the measurement error. Afterwards we use demand and cost estimates to simulate price effects of a merger between TMN and Optimus.

3 Estimation Results

3.1 Demand Side

The demand for mobile subscriptions is dependent on average per-minute call prices, the lagged total penetration of mobiles and a dummy for the fourth quarter to account for the seasonal effect. All coefficients are common across three demand functions. First, demands are estimated using ordinary least squares (OLS) and two stage least squares (2SLS). As instruments we use firm specific operating costs, number of employees and the 4th quarter dummy. The Breusch-Godfrey test indicates autocorrelation of the error terms in all three demand equations. We account for the problem of endogeneity and autocorrelation by estimating the covariance matrix of parameters by general method of moments (GMM) using the Newey-West estimator.

The estimates of all parameters are significant, as presented in Table (1). In particular, σ is estimated to be 0.986, which implies that mobile services provided by different network operators are very close substitutes but the hypothesis of $\sigma = 1$ may be rejected. There is a very high correlation of choices within the nest. Large t statistics maybe due to collinearity. Therefore, we re-estimate the model in differences, in which case the constant r_j cannot be identified. Estimation results are presented in Table (2) under the assumption that prepaid consumers are not locked in and in Table (3) assuming that 50% of prepaid consumers are locked in. In the first case, the estimation results are very similar to the initial estimates. The estimate of price coefficient α increases to -0.365 and the estimate of σ is equal to 0.987. Under the assumption that more consumers are locked in, the estimates of σ and network effects decrease, while price coefficient remains almost unchanged. However, in this situation computed marginal costs using formula (7) are negative in the case of TMN. We use the estimates from Table (2) in further calculations.

The own and cross price elasticities of demand in the nested logit model are specified as:

$$E_{p_{kt}}^{s_{jt}} = \begin{cases} -\frac{\alpha}{1-\sigma} p_{jt} \left[1 - \sigma \overline{s}_{jt|g=1} - (1-\sigma) s_{jt}\right] & \text{if } k = j; \\ \frac{\alpha}{1-\sigma} p_{kt} \left[\sigma \overline{s}_{kt|g=1} + (1-\sigma) s_{kt}\right] & \text{if } k \neq j. \end{cases}$$

where s_{jt} is the share of network operator j in sales at time t and $\overline{s}_{jt|g=1}$ is the within group share. Table (4) presents the average elasticities for GMM estimates for period January 1998 – December 2004. We also calculate the elasticity of demand for mobile services in total, which are given by

$$E_{p_{jt}}^{s_t} = -\alpha s_{jt}(1 - s_t)\frac{p_{jt}}{s_t}$$

The values in Table (4) are interpreted as follows: on average in the period January 1998 – December 2004, a 1% price increase by Optimus resulted in 0.058% decrease in total sales of mobiles. Similarly, a 1% price increase by TMN and Vodafone led to a decrease in total sales by 0.191% and 0.134%, respectively. These elasticities suggest that mobile and fixed line subscriptions are rather weak substitutes. The elasticity of demand for mobile services in respect to the past installed base is specified as

$$E_{z_{t-1}}^{s_t} = \beta z_{t-1}(1 - s_t).$$

If the previous period total installed base increased by 1%, current period sales would surge on average by 0.821%. This indicates strong network effects. If there were no network effects, the industry growth would be stimulated only by price changes and the penetration level in the absence of network effects could be much lower. In addition network effects also have a negative impact on the equilibrium prices. There is also a significant 4th quarter effect.

3.2 Merger Simulation

In this section we follow the procedure adopted by Nevo (2000), Ivaldi and Verboven (2005) and other studies employing discrete choice model for the simulation of unilateral effects of mergers. We use the estimates of demand system presented in Table (2) to compute the marginal costs under the assumption of Nash equilibrium and joint profit maximization. We find that for the joint profit maximization to hold, the marginal costs would have to be negative. This is due to the fact that according to our estimates mobile services are very close substitutes while the substitution with the outside good – fixed line telephony seems to be rather small. Furthermore we regress marginal costs resulting from static Nash-Bertrand equilibrium on operating costs data provided by the firms, as given by equation (8). Because of autocorrelation we estimate the model in differences with separate ρ for each operator. The estimation results are presented in Table ().[11]

It seems that the Nash-Bertrand marginal costs are well explained by the data on operating costs received from the firms and firm dummies, as presented on Figure (). Visually the best fit we get for Optimus, followed by Vodafone. Even thought the assumptions we made are very restrictive, this result suggests that there may be some rationale for using static Nash-Bertrand equilibrium in the further analysis of the effects of merger between Optimus and TMN. We account for the ownership of both Optimus and TMN by a single operator by adjusting the elements of the matrix Δ. The post-merger prices solve the following equation

$$p^* = \widehat{mc}_t + (\Delta^{post} \bullet S_t(p^*))^{-1} s_t(p^*) \tag{9}$$

where $\widehat{mc}_t$ are the marginal costs implied by the demand estimates and the pre-merger ownership structure. This equation imposes static Nash-Bertrand pre-merger conduct and cost structure which remains the same before and after the merger. Afterwards we will simulate price changes in the presence of cost efficiencies.

Based on our model of demand and supply we find that in result of a merger between Optimus and TMN, the equilibrium prices may increase by values presented in Table ().[12] In the first simulation we assume that there are no cost efficiencies. The further two simulations consider cost efficiencies of 5% and 10%, respectively. In the last simulation we assume that after merger less efficient Optimus will be able to provide services with the marginal cost level of TMN.[13] Thus, keeping the consumption constant, we observe significant price increases in all situations considered. Even in the presence of large cost efficiencies of 10% there are significant price increases. There are incentives to redistribute sales from the less efficient Optimus to the more efficient TMN which benefits the firm staying outside the merger. The merging firms redistribute profits while Vodafone has incentives for a small price increase but enjoys huge

[11] This Table was temporarily taken out from this version due to confidentiality reasons

[12] This Table was temporarily taken out from this version due to confidentiality reasons

[13] Obviously, these simulations do not consider adjustments in consumption resulting from price changes, that is in the number of calls and minutes. But the price elasticities of demand for calls and duration are rather small.

increase in profits. The smallest price increase takes place when Optimus is able to provide services at the level of marginal costs of TMN.

Following Anderson et al. (1992), and Ivaldi and Verboven (2004), using the assumptions of the nested logit model, the net consumer surplus CS equals:

$$CS = \frac{1}{\alpha}\left(1 + \sum_{g=1}^{G} D_g^{1-\sigma}\right) \tag{10}$$

It measures the attractiveness of the set of J+1 products in monetary terms, after subtracting the price consumers have to pay. [To be continued...]

4 Conclusion

In this paper we simulate unilateral effects of merger in mobile telephony in Portugal. In the first stage we estimate demand functions using a standard aggregate nested logit model and investigate whether network effects have an impact on the decision about subscription to mobile telephony services in Portugal. In the estimation we use limited aggregate data which constraints the sophistication of the estimated model, that is, quarterly data on subscriptions and revenues per minute between January 1998 and December 2004. An important modelling issue is the lock-in of consumers, since consumers often sign long term deals with their service providers and therefore do not engage in decision making every month. We use information about the number of consumers using pre-paid and postpaid services to approximate the number of consumers which make decision about subscription to mobile services each month.

Mobile services provided by particular operators are close substitutes. Our results suggest also that network effects played a significant role in the diffusion of mobile services in Portugal. Afterwards, using the estimates from the demand side we derive the estimates of marginal costs. Demand estimates support the assumption of static Nash-Bertrand equilibrium, while joint profit maximization turns out to be unrealistic assumption resulting in negative computed marginal costs. Furthermore, we regress the estimates of marginal costs on the data on operating costs, which was reported by the firms. Our computed marginal costs are well explained by firm specific dummies and the average operating costs per subscriber with the coefficient of one. We use the estimates of demand parameters and marginal costs to simulate unilateral effects of a merger between two network operators: TMN and Optimus. Finally, we simulate price changes in the presence of cost efficiencies as well as changes in consumer surplus, profits and total welfare. Because of the assumption, that relatively few consumers are locked in, and resulting high estimates of price elasticities, we may perceive our case as the extreme situation, in which the potential price increase is the lowest. However, even in this case, notified merger may lead to significant price increases, even in the presence of substantial cost efficiencies.

Bibliography

Anderson, S., De Palma, A. and J-F. Thisse. 1992. "Discrete Choice Theory of Product Differentiation." *MIT Press.*

Barros, P. and N. Cadima, 2000. "The impact of mobile phone diffusion on the fixed - link network," *CEPR Discussion Paper*, no.2598.

Berry, S.T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, 25(2), pp. 242-262.

Birke, D. and G.M.P. Swann. 2004. "Network effects in mobile telecommunications: An empirical analysis." mimeo.

Bousquet A. and M. Ivaldi. 1997. "Optimal pricing of telephone usage: An econometric implementation." *Information Economics and Policy*, 9(3), pp.219-239.

Brynjolfsson, E. and C. Kemerer. 1996. "Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market." *Management Science*, 42, pp.1627-1647.

Chintagunta, P.K., Narayanan, S. and E.J. Miravete, 2005. "The Role of Self Selection and Usage Uncertainty in the Demand for Local Telephone Service," University of Chicago, unpublished.

Clements, M.T. and H. Ohashi. 2004. "Indirect Network Effects and the Product Cycle: Video Games in the U.S., 1994-2002." *Journal of Industrial Economics*, forthcoming.

Doganoglu, T. and L. Grzybowski, 2005. "Estimating Network Effects in Mobile Telephony in Germany," Ludwig Maximilians Universitat Munchen, unpublished.

Doganoglu T. and K. Lange, 2004. "Estimating Models of Long-Distance Carrier Choice and Cost of Using Alternatives," Ludwig Maximilians Universitat Munchen, unpublished.

Dubin J.A. and D.L. McFadden, 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, pp.345-362.

Economides, N. and Ch. Himmelberg. 1995. "Critical Mass and Network Size with Application to the US Fax Market." Discussion Paper no. EC-95-11, Stern School of Business, N.Y.U.

Economides, N., K. Seim, and V.B. Viard, 2005. "Quantifying the Benefits of Entry into Local Phone Service," working paper, Stanford Graduate School of Business, Stanford, CA.

Evans, D.S. and J.J. Heckman. 1983. "Multiproduct Cost function Estimates and Natural Monopoly Tests for the Bell System." in: Breaking up the Bell, North-Holland, New York, pp.253-287.

Farrell, J. and P. Klemperer. 2004. "Coordination and Lock-In: Competition with Switching Costs and Network Effects." preliminary draft chapter for Handbook of Industrial Organization, Vol. 3.

Gandal, N. 1994. "Hedonic Price Indexes for Spread Sheets and a Test of Network Externalities." em Rand Journal of Economics, 25(1), pp.160-70.

Gandal, N., Kende, M. and R. Rob. 2000. "The Dynamics of Technological Adoption in Hardware/Software Systems: The Case of Compact Disc Players." *RAND Journal of Economics*, 31, pp.43-61.

Goolsbee, A. and P.J. Klenow. 2002. "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." *Journal of Law and Economics*, Vol. XLV(2), October, pp.317-344.

Grajek, M. 2003. "Estimating Network Effects and Compatibility in Mobile Telecommunications", CIC working paper.

Greene, W.H. 1997. Econometric Analysis: *New Jersey: Prentice-Hall Inc.*.

Gröhn, A. 1999. "Network Effects in PC Software: An Empirical Analysis." Mimeo, National Bureau of Economic Research, Cambridge, MA.

Gruber H. and F. Verboven. 2001. "The Diffusion of Mobile Telecommunication Services in the European Union." *European Economic Review*, 45, pp.577- 588.

Hanemann, W.M., 1984. "Discrete/Continuous Models of Consumer Demands," *Econometrica*, 52, pp.541-561.

Hartman, R., and D.J. Teece. 1990. "Product Emulation Strategies in the Presence of Reputation Effects and Network Externalities: Some Evidence from the Microcomputer Industry." *Economics of Innovation and New Technology*, pp.157-182.

Ivaldi, M. and F. Verboven. 2004. "Quantifying the effects from horizontal mergers in European competition policy," *International Journal of Industrial Organization*, 23, pp.669-691.

Katz, M. and Shapiro, C. 1985. "Network Externalities, Competition and Compatibility." *American Economic Review*, 75(3), pp. 424-440.

Kim, H.S. and N. Kwon. 2003. "The advantage of network size in acquiring new subscribers: a conditional logit analysis of the Korean mobile telephony market." *Information Economics and Policy*, 15(1), pp.17-33.

Koski, H. and T. Kretschmer. 2005. "Entry, Standards and Competition: Firm Strategies and the Diffusion of Mobile Telephony." *Review of Industrial Organization*, 26, pp.89-113.

Moch, D. 1995. "Ein hedonischer Preisindex für PC-Datenbanksoftware: Eine empirische Untersuchung", in: Harhoff, D./Müller, M. (Hrsg.): Preismessung und technischer Fortschritt, Baden Baden.

Nevo, A. 2000. "Mergers with differentiated products: the case of the ready-to-eat cereal industry." *Rand Journal of Economics*, pp.395 421.

Ohashi, H. 2003. "The Role of Network Externalities in the U.S. VCR Market, 1978-86." *Journal of Economics and Management Strategy*, 12(4), pp.447-494.

Park, S. 2003. "Quantitative Analysis of Network Externalities in Competing Technologies: The VCR Case." *The Review of Economics and Statistics*, 86, pp.937 945.

Pinkse, J. and M. Slade, 2004. "Mergers, brand competition, and the price of a pint," *European Economic Review*, 48(3), pp.617643.

Rodini, M., Ward M. and G. Woroch, 2003. "Going Mobile: Substitution Between Fixed and Mobile Access," *Telecommunications Policy*, 27, pp.457-476.

Statistisches Bundesamt Wiesbaden, 1999. "Neuberechnung des Verbraucherpreisindex für Telekommunikationsdienstleistungen auf Basis 1995", Sonderdruck aus Wirtschaft und Statistik.

Swann, G.M.P. 2002. "The functional form of network effects", *Information Economics and Policy*, 14(3), pp. 417-429.

Train, K.E., McFadden, D.L. and M. Ben-Akiva, 1987. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," *Rand Journal of Economics*, 18(1), pp.109-23.

Wallsten, S.J. 2003. "Of Carts and Horses: Regulation and Privatization in Telecommunications." *Journal of Policy Reform*, 6(4).

Appendix

Table 1: Nested logit estimates

Variable	OLS Estimates	std	2SLS Estimates	std	GMM Estimates	std
r_j	-2.448	0.124	-2.438	0.126	-2.371	0.055
α	-0.063	0.319	-0.184	0.328	-0.118	0.037
σ	0.987	0.055	0.978	0.055	0.986	0.004
β	4.755	0.108	4.778	0.109	4.574	0.109
Christmas	0.181	0.050	0.179	0.050	0.216	0.054
mse Opt.	0.0373		0.0369		0.0414	
mse Tmn	0.0377		0.0378		0.0417	
mse Vod.	0.0380		0.0387		0.0427	
N*Obj.	2.6381		0.9019		13.2785	

Table 2: Nested logit estimates in differences

Variable	OLS Estimates	std	2SLS Estimates	std	GMM Estimates	std
α	-0.138	0.576	-0.397	0.997	-0.365	0.0485
σ	1.185	0.362	0.986	0.561	0.987	0.0168
β	3.813	0.431	4.401	0.485	4.399	0.0880
Christmas	0.098	0.022	0.059	0.028	0.072	0.0119
mse Opt.	0.0169		0.0178		0.0175	
mse Tmn	0.0173		0.0182		0.0177	
mse Vod.	0.0162		0.0185		0.0180	
N*Obj.	1.1419		0.3559		17.92	

Table 3: Nested logit estimates in differences (50% of prepaid users are locked in)

Variable	OLS Estimates	std	2SLS Estimates	std	GMM Estimates	std
α	-0.046	0.417	-0.365	0.652	-0.347	0.047
σ	1.163	0.191	0.882	0.455	0.889	0.015
β	2.116	0.351	2.758	0.415	2.759	0.167
Christmas	0.076	0.018	0.024	0.024	0.024	0.005
mse Opt.	0.0112		0.0131		0.0131	
mse Tmn	0.0122		0.0131		0.0132	
mse Vod.	0.0106		0.0149		0.0148	
N*Obj.	0.7699		0.2302		19.64	

Table 4: Average price elasticities – prices and past consumer base

	Optimus	TMN	Vodafone	Mobiles s_t
Optimus	-7.61	1.74	1.74	-0.058
TMN	3.38	-3.27	3.38	-0.191
Vodafone	3.66	3.66	-7.59	-0.134
Net.Effect				0.821

The Virtual Location of E-Tailers
Evidence from a B2C E-Commerce Market

Julia Häring
ZEW (Centre for European Economic Research)
April 2005

Abstract
An Internet retailer's (e-tailer's) outstanding virtual location enhances the probability of being noticed by potential customers. The notion of a virtual location for e-tailers refers to the analogy to the physical location. In the empirical analysis, an e-tailer's Internet search engine rank as well as its advertising activities in search engines serve as proxies for the virtual location. The results suggest that it is optimal for e-tailers to complement a high search engine rank with investments in online advertising. Moreover, banner ads seem to serve as price advertising mechanism, whereas sponsored links rather seem to be used in order to signal outstanding customer service.

JEL–Classification: L81, L29, M37, C25
Keywords: virtual location, online advertising, search engines

Address: Centre for European Economic Research (ZEW)
Research Group Information and Communication Technologies
P.O. Box 10 34 43
D–68034 Mannheim
Germany
Phone: +49/621/1235–134
Fax: +49/621/1235–333
E-Mail: `haering@zew.de`

Acknowledgements: Financial support by the Landesstiftung Baden-Württemberg foundation is gratefully acknowledged. I would like to thank Irene Bertschek, Dirk Czarnitzki, Avi Goldfarb, Tobias Hagen, Ulrich Kaiser, Friedhelm Pfeiffer, Manuel Trajtenberg, Thorsten Wichmann, and the seminar participants at the Centre for Economic and Business Research (Copenhagen) for their suggestions and comments, and to Margit Vanberg for drawing my attention to purchasing contact lenses online. I am indebted to Tobias Köhne for his research assistance in collecting and putting together the primary data set.

1 Introduction

Today, it is widely accepted that the Internet is not the "great equalizer" it was expected to be for competition and retailer prices (see for example Smith and Brynjolfsson, 2001; Clay, Krishnan, Wolff and Fernandes, 2002). Empirical analyses have mainly focused on differentiation in retailer service as a reason for price premiums, which induce price dispersion and impede the observation of the law of one price. Smith, Bailey and Brynjolfsson (2000, p. 110) argue that there are certain general interest websites which make some online retailers easier to be found than their competitors and refer to the huge amounts of money invested in portals and "content sites". This phenomenon is interpreted by the authors as "neural real estate" (see Smith et al., 2000, p. 110).

It should be obvious from everyday observation that not all online retailers can be found by uninformed consumers in an equally easy way. Instead, some online shops are easier to be found than others. This seems to be an analogy to the location in the physical world where shop owners invest considerable amounts of money in superior locations, for example in highly frequented shopping malls or pedestrian areas. On the Internet, search engines or news portals are the highly frequented locations where hyperlinks leading to other sites on the Web are noticed by Internet surfers with a larger probability. This Internet analogy to the concept of location in the physical is termed *virtual location*. This follows the view by Hunter (2003) who argues from a juridical point of view that cyberspace is perceived as a place with spatial characteristics, and discusses the implications for legal regulation.

This paper analyses empirically the determinants of the virtual location of e-tailers[1]. As most consumers seem to use relatively little sophisticated techniques when searching the Web (this is one of the results of the study by Machill, Neuberger, Schweiger and Wirth, 2003), an outstanding virtual location is a crucial factor for the attraction of potential customers to e-commerce sites. Thus, optimising their virtual location becomes an integral part of e-tailers' overall strategy. In the empirical analysis, the virtual location will be proxied by an online retailer's position in the Google search results list and by online advertising activities, since these may be the most important factors making it more likely for consumers to get to know about a specific online shop.[2] The observed activities of online advertising are context dependent banner ads and sponsored links in the 10 most-widely used Internet search engines.[3]

A recent study investigating the market for Internet search engines in German language by Machill et al. (2003) suggests that having a prominent location on the Internet is crucial for attracting consumers to websites.[4] The study consists of two parts: a survey among operators of search engines in German language, and an

[1]Throughout the paper, the terms 'e-tailer' and 'online retailer' are used interchangeably and refer to firms selling products online via e-commerce websites.

[2]There will of course be repeat purchasers or potential customers who directly access a specific e-tailer which they are aware of due to retailer branding or word of mouth. Hence, similarly to the physical location, the virtual location is crucial for the attraction of new customers.

[3]Examples for banner ads and sponsored links are depicted in Figures 2 and 3 in the Appendix.

[4]The author of this paper is not aware of any comparable study for a further language area.

experiment, in which the steps of Internet users when solving specified search tasks are observed and analysed. Machill et al. (2003, p. 92) report extensive linkages between different search engines, as many of them share the same search technology. Furthermore, it is well-known that several highly-frequented search engines rely on the Google technology, and that the Google website itself has a dominant market share. Both factors combined lead to consumers being highly dependent on the information provided by just a few search sites on the Web.

Most consumers are neither aware of the economic dominance of the Google technology, nor aware of the existence of sponsored links on results pages (see Machill et al., 2003, p. 94).[5] Furthermore, search engine providers state in the survey, that users click-through to just a few links of the results pages and that nearly 70% of the users do not examine more than the first two pages of results, corresponding to the first 20 hits. This is confirmed by the experimental results of the study (see Machill et al., 2003, p. 255): 81% of the participants evaluated only the first page of results, further 13% the first and second pages, implying that only 6% considered more than the first 20 entries. The results suggest that being visible on the first or second page of the results list should improve the ability to attract consumers quite a lot. A means to reach this goal is the use of advertising in Internet search engines, thus serving as a virtual location. The virtual location being proxied by both the rank in the Google list of results and contextual advertising in different Internet search engines is in line with these arguments.

The contribution of this paper is twofold: This is – to the best of the author's knowledge – the first attempt to relate the virtual location to e-tailer strategy resulting from profit-maximising behaviour. Second, the analysis is based on a unique set of primary data from the online market for contact lenses relevant for consumers living in Germany. The data set contains observations for 146 different online retailers collected between March and September 2002 on a monthly basis. It is not a selected sample but it represents the whole population of online shops, which are relevant for consumers living in Germany.

The primary data set contains the prices and range of products offered by retailers in approximately the whole market for contact lenses. The data are merged with retailer characteristics as well as information on the virtual location of online retailers. Neither is the observed product range restricted to a predetermined subset of products, nor are the retailers selected. The observations are collected directly at the e-tailers' websites instead of using shopbot data or data from price comparison websites.

The results presented in this paper suggest that it is optimal for profit maximising e-tailers to complement a high search engine rank by investments in online advertising. Moreover, banner ads seem to serve as price advertising mechanism, whereas sponsored links rather seem to be used by e-tailers in order to signal out-

[5] This is supported by the fact that in contrast to journalism, where advertising and editorial contents are usually clearly separated according to a code of conduct, this is not common in the field of Internet search engines (see Machill et al., 2003, p. 92). Also Silk, Klein and Berndt (2001, p. 140) report a "blurring of the traditional distinction between advertising and editorial content on the Internet."

standing customer service. The search engine rank appears to remain relatively stable over time suggesting that the virtual locations for the whole market are in an equilibrium during the period under observation.

The paper is organised as follows: Section 2 describes the related literature, and Section 3 provides background information on the market for online advertising. Section 4 describes the data set, which is analysed in Section 5. Section 6 concludes.

2 Related Literature

So far, the existing literature on aspects of the virtual location has been sparse. The growing importance for online retailers to be prominently placed on the Internet has been brought to the discussion by Smith et al. (2000). Hunter (2003) discusses the metaphor of "cyberspace as place". Based on a theoretical model, Smith (2002) distinguishes between online retailers with a high awareness on the consumer side and those with a low awareness and focuses on implications on the pricing strategies of the two groups. That location merges with brand in the virtual space, is an observation by Tang and Lu (2001) who analyse price dispersion among online retailers.

There is a growing strand of literature dealing with advertising on the Internet, most of which tackles the effect of price advertising on competition (see for example Stahl (2000) or Baye and Morgan (2001)). Stahl (2000) focuses on the relation between online price advertising and pricing in e-commerce and the welfare implications. Baye and Morgan (2001) analyse e-tailers' incentives to advertise prices on a gatekeeper's site (such as a price comparison site where firms pay in order to have their prices listed) and the competitive effects of such price advertising. The predictions of this model are tested in an experimental setting by Morgan, Orzen and Sefton (2003). From a theoretical point of view, Baye and Morgan (2004) consider the effects of informational and promotional advertising in a common model.[6] Similarly to the other studies, the effects of both types of advertising on price competition are investigated.

There are two main points differing between the aforementioned empirical papers and this paper. First, other studies mostly rely on data from a specific gatekeeper's website (for example Baye, Morgan and Scholten (2004)). An important drawback for such approaches is the neglect of retailers never listing on the specific price comparison site considered which may imply a highly selective sample. In contrast to this, also retailers never advertising are observed in the empirical analysis presented in Section 5. A second difference between the papers cited above and this paper is their focus on price advertising and price competition, whereas this paper deals with e-tailer *brand* advertising as a proxy for the virtual location of e-tailers.

The role of location on the Internet has received attention in a recent paper from Baye, Gatti, Kattuman and Morgan (2004), where a different aspect of virtual

[6] Informational advertising refers to the listing of prices on a price comparison site attracting price sensitive consumers, whereas promotional advertising is understood as creating loyal customers and thus enticing website traffic away from price comparison sites.

location is investigated: effects of the position of e-tailers in the price quotation list of an Internet price comparison site on the decision to click-through to a specific e-tailer are analysed. The estimates suggest that customers favour products that are listed higher on the screen. It is important to note that this effect is modelled and estimated independently of the relative price compared to the other firms in the list.[7] Also Ellison and Fisher Ellison (2004) include the rank of the retailer in their estimation explaining the demand for computer memory modules, but retailers are automatically sorted according to price on the shopbot site underlying their sample. Therefore, the effect of the order cannot be distinguished from the price effect. Smith and Brynjolfsson (2001) find in an empirical analysis of the click-through behaviour at an Internet shopbot for books that a considerable fraction of consumers does not decide in favour of the cheapest retailer. The influence of an e-tailer's position on the screen can however not be analysed, as retailers are automatically sorted according to price.

The question of how to identify the ideal location for paid online advertising is addressed by Bhatnagar and Papatla (2001). They directly address the issue of narrowly targeting potential customers on the Internet. Their discussion starts from the consideration that a firm's customer segment would be best found on the websites of its competitors which however would deny to sell advertising space. The discussion focuses on ways to identify adequate websites to advertise at and considers consumer search behaviour in the analysis. Consumer response to banner ads is discussed and empirically analysed in Chatterjee, Hoffman and Novak (2003) or Manchanda, Dubé, Goh and Chintagunta (2004), for example.

In this paper, it is argued that the virtual location is tied to e-tailers' rankings and context dependent advertising efforts in Internet search engines. This view is supported by Machill et al. (2003) where a systematic evaluation of the role and power of Internet search engines can be found. The study discusses the role and market power of search engines in the German language area of the Internet but the results should be transferable to the English part of the Internet without major obstacles. The authors focus on the market structure in the search engine market and additionally present an extensive experimental study of user behaviour when using Internet search engines (selected results of this study have already been reported in the Introduction).

3 The Market for Online Advertising

Online advertising faces increasing problems of lacking acceptance by Internet users. A recent study for Germany revealed that the proportion of Internet users tolerating advertising on websites as a necessary instrument for financing websites has shrunk to 41 percent from 53 percent at the beginning of 2001 (see Fittkau & Maaß (2003), cited in ECIN (2003a)). Simultaneously, the proportion of Internet users claiming to ignore online advertising has risen from 34 to 41 percent. Both Kenny and Marshall

[7] ? are able to assess the roles of the price and the position on the screen separately, as firms are not listed according to price on the price comparison site they observe.

(2000) and Hoffman and Novak (2000) report average click rates as low as 0.5 percent for banner ads. Hoffman and Novak (2000) argue in this context that the optimal placement of online advertising activities is crucial in order to achieve higher click rates.

One way to an exact targeting of the relevant consumer group is contextual marketing (see Luo and Seyedian, 2004). Contextual marketing is defined by Kenny and Marshall (2000, p. 120) as using the Internet "to deliver tailored messages and information to customers at the point of need". One instrument of contextual marketing are banner ads and sponsored links appearing together with the results list after search engine queries for specific keywords, which have been chosen by the marketer. *Banner ads* refer to ads which are graphically emphasised, for example by using coloured boxed, graphics or pictures to draw attention on them. In contrast, *sponsored links* have an impact by their unconspicious placement on top of the results list. The optical appearance of these prominently placed links is meant to differ as least as possible from the ordinary list of results. Examples for banner ads and sponsored links are depicted in Figures 2 and 3 in the Appendix.

According to Jupiter Media Metrix (as cited in ECIN (2003b)), 42 percent of online purchases are initiated via a preceding search which makes context dependent advertising in Internet search engines an ideal advertising channel for online retailers. If advertising is linked to specified search terms, people interested in the product qualify as possible customers by the word(s) they actively search for. In addition to the better targeting of the audience, the advertiser incurs costs only if the user actually clicks on the banner ad or sponsored link.

The evolution of online advertising in the market for glasses and contact lenses during the period covered by the data set underlying this paper is depicted in Figure 1.[8] According to W3SCAN.COM, the spending for online advertising rose from 29,879 Euro during the first wave of data collection in March 2002 to 48,855 Euro during the weeks following the last collection of data in September 2002. In total, 614,063 Euro were spent for online advertising for optics in 2002. These figures are published online by the company W3SCAN.COM on their website (www.w3scan.com).

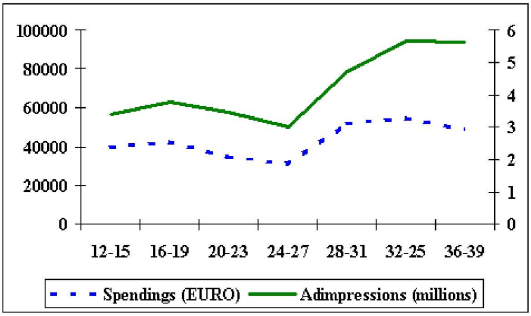
4 Data

The analysis of markets for optometric devices has a long tradition in economics (see for example Benham, 1972; Kwoka, 1984). The online market for contact lenses was chosen due to several criteria which make the products suitable for both being sold via e-commerce and being analysed empirically.[9] The data set used for the analysis is condensed from a data set with monthly observations of e-tailers for contact lenses which were observed between March and September 2002.

[8]Unfortunately, there is no separate information on the segment for contact lenses excluding glasses available.

[9]These criteria are discussed in Häring (2003) where also a more detailed description of the data set can be found.

Figure 1: Market for online advertising in the weeks of data collection



Source: www.w3scan.com

Before collecting the primary data set, the online shops for contact lenses had to be identified. This was done by searching for the German word for contact lenses in its two possible spellings (*Kontaktlinsen* and *Contactlinsen*) in the ten most-widely used Internet search engines at the beginning of March 2002. The list of search engines can be found in Table 5 in the Appendix. From each of these search queries, the first 250 results were evaluated in order to identify sellers of contact lenses.

The primary data set contains monthly information on both the range of products and the prices offered by the e-tailers which results in more than 20,000 price observations. Relevant product attributes of the contact lenses and the service characteristics of the online shops were evaluated once during the period of data collection and then merged with the price data. This was appropriate since none of the online shops underwent major changes, and also no product was relaunched with retention of its original name.

Furthermore, information on the e-tailers' virtual locations was included. The virtual location is proxied by context dependent banner ads and sponsored links in the ten most-widely used search engines[10] and by the rank in the Google search engine at each time of data collection when a search for the German word for contact lenses in its two possible spellings was conducted. Both forms of online advertising are linked to the searches for contact lenses in the search engines as described above. The ranks in the Google list of results were only considered for the first 10 pages of results (i.e. ranks 1 to 100). Price comparison sites played no role in the market for contact lenses at the time of data collection.

For the analysis of the link between e-tailer strategies and their virtual location, the original data set was condensed. The characteristics of the online shops underwent a factor analysis in order to obtain factors describing the service of the online retailers. Names were assigned to the factors according to the underlying variables they represent. The result are five factors indicating: Convenient navi-

[10] A list of these can be found in Table 5 in the Appendix.

gation, superior customer service, a favourable return policy & supply of lens care products, security and trustworthiness features of the websites, and special services for customers using contact lenses for the first time.[11] The width of the product range offered by the e-tailers is measured by the number of different products offered. Clearly, the probability of the online shop covering different product segments of the contact lense market increases with the number of distinct products offered. Each e-tailer's overall price level in comparison to its competitors is proxied by the average of its standardised prices. For this standardisation, each product price is divided by the average price over all e-tailers for this product.

The result is an unbalanced panel data set with 929 observations on a monthly basis for 146 different e-tailers. On average, the e-tailers are observed in 6.4 of the seven months of data collection. Descriptive statistics of the variables used in the empirical analysis can be found in the Appendix.

5 Empirical Analysis

5.1 Framework

The determinants of virtual location are analysed using data from the online market for contact lenses. In the case of online advertising, it is obvious that firms deliberately choose whether or not they advertise and which kind(s) of advertising to invest in. But also the rank in the Google results list can be interpreted as an outcome of the profit maximising behaviour of e-tailers. Thus, both dimensions reflect the underlying latent profit maximisation of the e-tailers, and both dimensions of the virtual location can be investigated using the latent profit index framework which will briefly be described in the following.

The latent profit is an analogue to the latent utility in the consumer choice literature (see McFadden, 1974). The latent profit index of an e-tailer when choosing alternative j out of J possible choices is not observable and assumed to consist of a systematic and a stochastic component:

$$y^* = \mathbf{x}\beta + \epsilon \tag{1}$$

where the latent profit y^* is not observable, $\mathbf{x}$ is a vector of observable characteristics, β is a parameter vector, and ϵ is a stochastic error term. The observed outcome y takes on one of the values 1, ..., J indicating the chosen alternative. Thereby, some information about the latent index is revealed but the underlying profit level cannot be fully recovered. In order to use the observed information on the virtual location of e-tailers, assumptions about the decision process are made in order to estimate the relationship between various e-tailer attributes and their virtual location. These are explained in the context of the estimation problem in the following two subsections.

[11] The construction of these factors is explained in Häring (2003).

5.2 An Empirical Analysis of The Virtual Location

5.2.1 Correlation between Search Engine Rank and E-Tailer Strategy

It has been argued in the previous sections that a superior position in the results list of search engines is of considerable importance. Since most of the popular search engines nowadays rely on the Google technology, the achieved position in the Google list of results is a valid proxy for the virtual location of online retailers.

Regarding the result of Machill et al. (2003) according to which most consumers evaluate only the first or the first two pages of results when using Internet search engines, the observed Google ranks are grouped into categories for the first (position 1-10) and the second (position 11-20) page of results in the list. The rest of the e-tailers is sorted into the third category (position 21 and above). The frequency distribution over these ordered categories is denoted in Table 6 in the Appendix.

In order to identify the strategies pursued by e-tailers which are correlated with a high Google rank, the rank is explained by the width of the product range the e-tailer offers, its relative price level, the amount of online advertising and the five factors describing retailer services. Since the rank categories are ordered (a higher rank is better), an ordered discrete choice model is the appropriate tool for analysis. The latent index y^* in Equation 1 can be used to estimate the correlations between the e-tailers' strategies and their Google rank using an ordered probit model[12]. In the case of three categories it is assumed that we observe:

$$\begin{array}{ll} y=0 & \text{if } y^* \leq \mu_1 \\ y=1 & \text{if } \mu_1 < y^* \leq \mu_2 \\ y=2 & \text{if } \mu_2 < y^* \end{array}$$

where $\mu_1 < \mu_2$ are unknown threshold parameters. By using the ordered probit model, a standard normal distribution for the error term $\epsilon \mid \mathbf{x}$ is assumed, and the unknown parameters β, μ_1 and μ_2 are estimated by maximum likelihood. The estimated parameter value $\hat{\beta}_k$ for regressor x_k does not correspond to the marginal effect of a change in x_k on the conditional probability to observe outcome j, denoted as $Pr(y = j \mid \mathbf{x})$. These marginal effects $\partial p_j(\mathbf{x})/\partial x_k$ have to be computed separately for each outcome category. This has been done in Table 1 where the estimation results can be found. The standard errors are estimated by nonparametric bootstrapping in order to account for the inclusion of generated regressors, which were obtained by the factor analysis.

The estimated marginal effects reveal that a wider product range and a higher price level are associated to a higher probability of being ranked on the first two pages of results in the Google search engine. Moreover, the probability of achieving a rank between 1 and 20 decreases with the service level offered by the e-tailers. But poor service can apparently be offset by investments in online advertising. The probabilities of being ranked on the first two pages of results are positively correlated with the numbers of banner ads and of sponsored links, respectively. On the contrary, the probability of being ranked lower than the first 20 entries increases

[12] For a textbook treatment of the ordered probit model, see Wooldridge (2002, ch. 15.10).

Table 1: Analysis of rank in Google list of results

	Dependent categorical variable: rank in Google list of results					
	Rank 1-10		Rank 11-20		Rank $\geq$ 21	
	m.eff.	s.e.	m.eff.	s.e.	m.eff.	s.e.
Width of product range $*10^{-2}$	0.079***	0.019	0.124***	0.027	-0.203***	0.040
Relative price level	0.058***	0.022	0.091***	0.027	-0.149***	0.047
Online advertising:						
Number of banner ads	0.010***	0.003	0.015***	0.005	-0.025***	0.007
Number of sponsored links	0.003**	0.001	0.004**	0.002	-0.007**	0.003
Retailer service:						
Convenient navigation	-0.010***	0.004	-0.015***	0.006	0.025***	0.008
Superior customer service	-0.012	0.011	-0.019	0.017	0.032	0.028
Return policy + lens care	-0.010	0.017	-0.016	0.027	0.025	0.043
Security + trustworthiness	-0.019	0.013	-0.030	0.020	0.049	0.032
New CL user service	-0.002	0.005	-0.004	0.008	0.006	0.013
Thresholds:						
μ_1 (s.e.)			3.767 (0.586)			
μ_2 (s.e.)			4.260 (0.568)			
Number of observations			929			
Log likelihood			-257.4			
Wald test (χ^2_9)			143.2			
Pseudo R^2			0.230			

Notes: Marginal effects of ordered probit estimation. The marginal effects are calculated at the mean values of continuous variables. ***, **, * denotes significance at the 1%, 5% and 10% level.

with offering more favourable service. Also cheaper and more specialised e-tailers (offering a narrower product range) are significantly more often found below the first 20 entries in the list of results. E-tailer's investment in online advertising is however negatively correlated with an adverse rank.

The causality of the observed relationship between an e-tailer's rank in the Google list of results and its online advertising activities could, however, go in both directions. On the one hand, e-tailers could use online advertising as a complement to a high search engine rank. On the other hand, it is coceivable that online advertising is used as a means to overcome an unfavourable search engine rank. In which way the causality goes, can not be disentangled using the data at hand. A higher rank being associated with a higher price level as compared to the competitors could stem from a reverse causality in the sense that online retailers with a superior search engine rank are possibly able to exploit this prominent position through price mark-ups. This possibility is suggested by Smith et al. (2000).

5.2.2 Determinants of the Advertising Decision

In order to capture an additional dimension of the concept of virtual location, the use of different types of online advertising is explained by the profit-maximising strategy of e-tailers and the search engine rank which can be expected to be exogenous to the advertising decision. Thus, the estimated effects are interpreted in a causal sense. The fundamental decision with respect to online advertising is analysed, implying that the observed outcome is a categorical variable, indicating the type of online advertising strategy.[13]

The dependent variable distinguishes between online retailers investing only in banner ads, those investing only in sponsored links, online retailers investing in both forms of online advertising, and those without any online advertising.[14] The frequency distribution of the online advertising categories can be found in Table 7 in the Appendix. The advertising decision is explained by the width of the product range, the relative price level, the Google rank and the factors describing the e-tailer services.

On the contrary to the search engine ranks, the categories describing the online advertising decision are not ordered but only mutually exclusive. The latent profit index idea of Equation 1 can be specified leading to the multinomial logit model (see for example Wooldridge, 2002, ch. 15.9). The multinomial logit model is based on the assumption that the observed outcome y is the one which the individual (or firm) attaches the largest latent utility (or here: profit) to. The probability of observing alternative j can in our case with three possible forms of advertising be described as:

$$Pr(y = j \mid \mathbf{x}) = \frac{exp(\mathbf{x}\beta_j)}{1 + \sum_{h=1}^{3} exp(\mathbf{x}\beta_h)}$$

where $j = 1, 2, 3$. Once the probabilities for $j = 1, 2, 3$ are specified, the probability for no online advertising $Pr(y = 0 \mid \mathbf{x})$ is known, because the probabilities must sum to unity. The parameter vector β is estimated by maximum likelihood. The estimated coefficients and the marginal effects $\partial p_j(\mathbf{x})/\partial x_k$ computed separately for each outcome category are denoted together with their estimated standard errors in Table 2.

[13]The actual numbers of banner ads or sponsored links are thus neglected.

[14]Note that only key word-related advertising in Internet search engines is observed in the data. Other types of activities such as banner ads on portals or general interest websites, for example, are neglected.

Table 2: Analysis of online advertising activities

	Dependent categorical variable: type of online advertising activities							
	None		Banner ads		Sponsored links		Both	
	m.eff.	s.e.	m.eff.	s.e.	m.eff.	s.e.	m.eff.	s.e.
Width of product range $*10^{-2}$	-0.019	2.112	0.004	0.906	0.018	1.663	-0.002	0.435
Relative price level	0.270***	0.085	-0.168**	0.075	-0.035	0.025	-0.067**	0.027
Google rank (ref.: $\geq$ 21):								
Google rank 1-10	-0.231***	0.084	0.093*	0.053	0.026	0.030	0.112*	0.059
Google rank 11-20	-0.329***	0.104	0.042	0.033	0.091	0.065	0.196**	0.094
Retailer service:								
Convenient navigation	-0.030***	0.008	0.003	0.004	0.023***	0.007	0.005**	0.002
Superior customer service	0.009	0.008	-0.015*	0.008	0.005*	0.003	0.000	0.002
Return policy + lens care	-0.038	0.030	0.015	0.013	0.011	0.010	0.012	0.010
Security + trustworthiness	-0.010**	0.005	0.003	0.003	0.003	0.003	0.004*	0.002
New CL user service	0.009	0.034	-0.002	0.014	-0.007	0.013	-0.001	0.009
Number of observations	929							
Log likelihood	-372.29							
Wald test (χ^2_{27})	306.46							
Pseudo R^2	0.3237							

Notes: Marginal effects of multinomial logit estimation. The marginal effects are calculated at the mean values of continuous variables and the base categories of dummies. The marginal effects of dummy variables are calculated as discrete changes in the expected value of the dependent variable. ***, **, * denotes significance at the 1%, 5% and 10% level.

The width of the product range offered by online retailers has no significant influence on their decision to promote their web shop online. By contrast to this, the relative price level of e-tailers significantly influences the online advertising strategy: The more expensive an e-tailer is, the more probably it does not invest in online advertising. Moreover, e-tailers being ranked higher in the Google list of results invest more probably in both types of online advertising, whereas e-tailers with less favourable ranks do not invest in online advertising with a larger probability. This suggests that e-tailers tend to complement a superior Google rank with online advertising, further enhancing their visibility on the web and thus increasing the probability of being found by potential customers.

Being relatively cheaper in comparison to the competitors significantly increases the probability of investing at least in banner ads. This implies that banner ads seem to be used as a means of signalling low prices.[15] On the contrary, the positive marginal effects of the variables describing retailer services which are estimated for the probability of using sponsored links indicate that sponsored links seem to serve as a vehicle for signalling superior customer service. Both results appear plausible, when the different optical designs of banner ads and sponsored links are considered. Banner ads are the more aggressive type of advertising because they are graphically highlighted and easily distinguishable from the list of search results. If relatively cheaper e-tailers use banner ads with a higher probability, this result can be interpreted as evidence for banner ads targeting fairly price-sensitive consumers which respond to this type of advertising design. Besides, sponsored links which are designed to hide on top of the list of search results and which differ only slightly from ordinary search results, can be thought of as appealing to consumers searching for a reliable supplier or for superior e-tailer quality instead of hunting for a bargain.

In order to check the robustness of the results, various further specifications of the multinomial logit model have been estimated. The estimated coefficients of these specifications are denoted in Table 3 together with the coefficients of the original specification which is equivalent to specification (2). The standard errors are estimated by nonparametric bootstrapping in order to account for the inclusion of generated regressors, which were obtained by the factor analysis. The first specification explains the online advertising decision by the same explanatory variables except the 5 factors describing e-tailer services. Specification (3) uses the one-month lag of the search engine rank instead of the current rank to avoid endogeneity of the Google rank. In the fourth specification, an additional variable is included, describing if the e-tailer is a "pure e-tailer" – meaning an e-tailer being active only in the online business and not running an additional physical store. These results are fairly robust to these different specifications, and it must be noted that the estimated effects of the search engine rank are robust to different ways of accounting for e-tailer quality attributes.

[15] Banners were only sporadically used for price advertising during the period of data collection.

Table 3: Robustness check

	Dependent categorical variable: type of online advertising activities							
	(1)		(2)		(3)		(4)	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Only banner ads								
Width of product range $*10^{-2}$	2.076***	0.639	0.461	0.926	0.202	1.181	0.303	1.185
Relative price level	-13.117***	2.123	-20.368***	4.607	-20.595***	5.517	-20.811***	5.766
Pure e-tailer							0.248	0.371
Google rank (ref.: $\geq$ 21):								
Google rank$_t$ 1-10	1.960	1.508	2.877*	1.514				
Google rank$_t$ 11-20	1.640	6.138	2.299	6.077				
Google rank$_{t-1}$ 1-10					2.716***	0.762	2.670***	0.789
Google rank$_{t-1}$ 11-20					2.334	8.066	2.301	7.773
Retailer service:								
Convenient navigation			0.376*	0.213	0.399*	0.220	0.402*	0.223
Superior customer service			-1.770***	0.668	-1.773***	0.603	-1.903***	0.662
Return policy + lens care			1.837*	1.099	1.897**	0.830	1.908**	0.869
Security + trustworthiness			0.364	0.238	0.341	0.250	0.349	0.257
New CL user service			-0.195	1.032	-0.338	0.509	-0.303	0.511
Constant	9.468***	1.908	15.724***	3.926	15.946***	4.701	16.053***	4.895
Only sponsored links								
Width of product range $*10^{-2}$	2.476**	1.102	1.400	1.502	1.567	1.588	1.320	1.612
Relative price level	-5.757***	1.358	-2.971	1.946	-2.830	2.011	-2.644	2.315
Pure e-tailer							-0.383	0.767
Google rank (ref.: $\geq$ 21):								
Google rank$_t$ 1-10	1.426	8.768	1.398	8.870				
Google rank$_t$ 11-20	2.141	1.976	2.585	2.064				
Google rank$_{t-1}$ 1-10					1.657	8.618	1.569	8.788
Google rank$_{t-1}$ 11-20					2.955	1.952	2.911	1.818
Retailer service:								
Convenient navigation			1.777***	0.430	1.846***	0.409	1.868***	0.466
Superior customer service			0.378	0.252	0.327	0.251	0.385	0.319
Return policy + lens care			0.864	0.533	0.915**	0.435	0.915*	0.469
Security + trustworthiness			0.272	0.253	0.323	0.229	0.312	0.292
New CL user service			-0.522	0.620	-0.553**	0.249	-0.603**	0.259
Constant	1.715	1.423	-1.800	2.186	-2.039	2.228	-2.071	2.540
Both								
Width of product range $*10^{-2}$	1.730**	0.814	-0.491	0.934	0.188	1.082	2.356**	1.051
Relative price level	-10.349***	1.487	-14.002***	2.465	-14.829***	3.177	-14.599***	3.696
Pure e-tailer							2.264***	0.607
Google rank (ref.: $\geq$ 21):								
Google rank$_t$ 1-10	2.771***	0.537	3.600***	0.697				
Google rank$_t$ 11-20	2.933***	0.653	4.314***	0.809				
Google rank$_{t-1}$ 1-10					3.683***	0.849	3.118***	0.849
Google rank$_{t-1}$ 11-20					4.715***	0.931	3.913***	0.902
Retailer service:								
Convenient navigation			1.003**	0.432	1.228***	0.440	1.426***	0.495
Superior customer service			0.043	0.353	0.032	0.411	-0.237	0.321
Return policy + lens care			2.440*	1.371	2.688***	0.869	2.997***	0.827
Security + trustworthiness			0.842**	0.390	1.003***	0.361	0.825**	0.397
New CL user service			-0.258	1.287	-0.371	0.518	-0.154	0.631
Constant	6.260***	1.519	8.722***	2.490	8.970***	3.173	6.952**	3.516
Number of observations	929		929		779		779	
Log likelihood	-457.19		-372.29		-306.11		-296.23	
Wald test (χ^2_{27})	173.78***		306.46***		261.82***		256.70***	
Pseudo R^2	0.1695		0.3237		0.3385		0.3599	

Notes: Estimated coefficients of multinomial logit estimation with "no online advertising" as comparison group. The marginal effects are calculated at the mean values of continuous variables and the base categories of dummies. The marginal effects of dummy variables are calculated as discrete changes in the expected value of the dependent variable. ***, **, * denotes significance at the 1%, 5% and 10% level.

Further checks, such as splitting the sample into a high- and a low-quality subsample and then estimating separately on both samples, could not be performed due to the sample size. The validity of the independence of irrelevant alternatives (IIA) assumption has been checked by a Small/Hsiao test which does not reject the IIA assumption. The results are available from the author on request.

Some effort has been put into the question whether the search engine rank is endogenous with respect to the advertising strategy, that is, whether the Google rank can really be treated as an exogenous variable in Table 2. In order to gain some insight into the problem two test strategies[16] both being based on an instrumental variable approach have been performed. The width of the product range offered by the e-tailers appears to be usable as an instrument for the search engine rank, as it turned out to have no effect on the advertising strategy in Table 2 but to affect the search engine rank even after conditioning on other covariates, as can be seen in Table 1.

A plausible interpretation for this result stems from the unknown mechanism by which Google ranks its results: online shops offering a wider product range tend to consist of a larger bundle of websites than online shops offering less products. This possibly affects the number of links which is used by Google as one criterion among others to rank the search results. Besides, there is no clear theoretical relationship between the number of products offered by an e-tailer and its advertising strategy, as specialized suppliers offering a limited product variety as well as e-tailers offering the full range of available products may both invest in online advertising or not.

As the width of the product range is the only available potential instrumental variable the Google rank is (incorrectly) treated as a continuous variable being censored at 101 (using dummy variables describing the Google rank, one instrumental variable would have been necessary for each dummy variable). First, the exogeneity test by Smith and Blundell (1986) for the probit model is applied with the null hypothesis of exogeneity. For performing this test all advertising strategies are tested separately, which is possible due to the independence of irrelevant alternatives assumption. Second, a simple plug-in test (analogous to a 2SLS approach) is applied where in a first stage the Google rank is regressed on the covariates and the instrumental variable (see Wooldridge, 2002, p. 474). In the second stage, a multinomial logit augmented by the residual from the first stage regression is estimated. The null hypothesis is then tested by a t-test on the coefficient of the residual. Using both test strategies, the null hypothesis of endogeneity can be rejected and the search engine rank can be used as an exogenous variable in the analysis of the determinants of the online advertising strategy. The results are available from the author on request.

5.2.3 Changes in the Virtual Location

In the next step it would be interesting to investigate changes in the virtual location, particularly in the search engine rank. The frequency distribution of changes in the

[16] Both tests were originally developed for different limited dependent variable models but are also applicable to the multinomial logit case.

Google rank is described in Table 4. It can be seen that substantial changes in the rank do not occur very often. Most e-tailers stay in the same rank category in which they were ranked in the previous month. Of course, changes occur in the exact list of results, but who would notice if a seller with position 56 in May would climb to position 53 in June? Unfortunately, the number of substantial changes in the Google results is too small to be analysed econometrically. It can only be concluded that search engine ranks appear to remain relatively stable over time. This relatively low turnover in search engine ranks suggests that the market was in equilibrium during the period of observation.

Table 4: Changes in the search engine rank

rank in month t	rank in month $t+1$		
	1-10	11-20	≥ 21
1-10	27	5	2
11-20	7	16	12
≥ 21	1	14	697

6 Conclusions

In this paper, a first empirical analysis of the profit maximising strategies of e-tailers with respect to their virtual location has been provided. The important role of an outstanding virtual location in the attraction of new customers has been discussed.

The empirical results suggest that for e-tailers optimising their strategy with respect to their virtual location, a high Google rank and online advertising appear to be complements, as e-tailers being ranked among the first 20 results decide to use banner ads as well as sponsored links with a higher probability than e-tailers being adversely ranked. Moreover, banner ads seem to serve as price advertising mechanism targeting price-sensitive customer groups, whereas sponsored links rather seem to be used in order to signal outstanding customer service to more quality-oriented consumers. A test strategy using the width of product range as instrumental variable has revealed that the search engine rank can indeed be assumed to be exogenous in the analysis of the online advertising strategy.

To complete the descriptive picture on e-tailers and their virtual location, there must be kept in mind that there is a significant correlation between the combination of relatively high price level/wider product range/inferior customer service, and a high search engine rank. Considering the development over time, the search engine rank appears to remain relatively stable. This suggests that the market was in equilibrium during the period of observation.

Up to now, the literature on the virtual location of firms has been sparse. The role of promotional (or brand) advertising online has been illustrated by Baye and Morgan (2004), and this is the only empirical study examining promotional advertising in the online environment which is known to the author of this paper. Future research should focus on the competing roles of promotional and informational (or price) advertising: What exactly is the trade-off for e-tailers when allocating their budget for online advertising? In this context – but also for the isolated analysis of virtual location – information on consumer response to online advertising would be quite useful. This lack of information could be alleviated using clickstream data (like for example in Baye, Gatti, Kattuman and Morgan (2004), Smith and Brynjolfsson (2001) or Goldfarb (2002)). Using such data would alleviate the problem but probably not solve it completely, as only click-throughs to an e-tailer's website can be observed, but not actual purchasing decisions or consumer characteristics.

References

Baye, M. R., Gatti, J. R. J., Kattuman, P. and Morgan, J. (2004). Estimating Firm-Level Demand at a Price Comparison Site: Accounting for Shoppers and the Number of Competitors, *mimeo*, download: http://www.econ.cam.ac.uk/faculty/gatti/wip.html.

Baye, M. R. and Morgan, J. (2001). Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets, *American Economic Review* **91**(3): 454–474.

Baye, M. R. and Morgan, J. (2004). Brand and Price Advertising in Online Markets, *mimeo*, download: http://www.nash-equilibrium.com/baye/Promotion.pdf.

Baye, M. R., Morgan, J. and Scholten, P. (2004). Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site, *Journal of Industrial Economics* **52**(4): 463–496.

Benham, L. (1972). The Effect of Advertising on the Price of Eyeglasses, *Journal of Law & Economics* **15**(2): 337–352.

Bhatnagar, A. and Papatla, P. (2001). Identifying Locations for Targeted Advertising on the Internet, *International Journal of Electronic Commerce* **5**(3): 23–44.

Chatterjee, P., Hoffman, D. L. and Novak, T. P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts, *Marketing Science* **22**(4): 520–541.

Clay, K., Krishnan, R., Wolff, E. and Fernandes, D. (2002). Retail Strategies on the Web: Price and Non-Price Competition in the Online Book Industry, *Journal of Industrial Economics* **50**(3): 351–367.

ECIN (2003a). Online-Werbung in der Krise? Download: www.ecin.de/news/2003/11/25/06466/.

ECIN (2003b). Suchmaschinen-Marketing: ROI in Echtzeit. Download: www.ecin.de/marketing/suchmaschinenreoi/.

Ellison, G. and Fisher Ellison, S. (2004). Search, Obfuscation, and Price Elasticities on the Internet, *Discussion Paper 10570*, NBER, Cambridge, Massachusetts.

Fittkau & Maaß (2003). *17. Nutzeranalyse W3B*, Fittkau & MaaßConsulting GmbH, Hamburg.

Goldfarb, A. (2002). Analyzing Website Choice Using Clickstream Data, *in* M. R. Baye (ed.), *The Economics of the Internet and E-Commerce*, number 11 in *Advances in Applied Microeconomics*, JAI.

Häring, J. (2003). Different Prices for Identical Products? Market Efficiency and the Virtual Location in B2C E-Commerce, *Discussion Paper 03-68*, Centre for European Economic Research, Mannheim.

Hoffman, D. L. and Novak, T. P. (2000). Advertising Pricing Models for the World Wide Web, *in* B. Kahin and H. R. Varian (eds), *Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*, MIT Press, Cambridge, Massachusetts, pp. 45–61.

Hunter, D. (2003). Cyberspace as Place and the Tragedy of the Digital Anticommons, *California Law Review* **91**(2): 439–519.

Kenny, D. and Marshall, J. F. (2000). Contextual Marketing: The Real Business of the Internet, *Harvard Business Review* **78**(6): 119–125.

Kwoka, J. E. (1984). Advertising and the Price and Quality of Optometric Services, *American Economic Review* **74**(1): 211–216.

Luo, X. and Seyedian, M. (2004). Contextual Marketing and Customer-Orientation Strategy for E-Commerce: An Empirical Analysis, *International Journal of Electronic Commerce* **8**(2): 95–118.

Machill, M., Neuberger, C., Schweiger, W. and Wirth, W. (2003). Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen, *in* M. Machill and C. Welp (eds), *Wegweiser im Netz: Qualität und Nutzung von Suchmaschinen*, Verlag Bertelsmann Stiftung, Bielefeld, pp. 13–490.

Manchanda, P., Dubé, J.-P., Goh, K. Y. and Chintagunta, P. K. (2004). The Effect of Banner Advertising on Internet Purchasing, *mimeo*, University of Chicago, Graduate School of Business, Chicago.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior, *in* P. Zarembka (ed.), *Frontiers of Econometrics*, Academic Press, New York.

Morgan, J., Orzen, H. and Sefton, M. (2003). A Laboratory Study of Advertising and Price Competition, *Technical Report 2003-04*, University of Nottingham, Centre for Decision Research and Experimental Economics, Nottingham.

Silk, A. J., Klein, L. R. and Berndt, E. R. (2001). The Emerging Position of the Internet as an Advertising Medium, *Netnomics* **3**(2): 129–148.

Smith, M. D. (2002). The Law of One Price? The Impact of IT-enabled Markets on Consumer Search and Retailer Pricing, *Working Paper 2002-35*, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh. http://www.heinz.cmu.edu/wpapers/retrievePDF?id=2002-35.

Smith, M. D., Bailey, J. and Brynjolfsson, E. (2000). Understanding Digital Markets: Review and Assessment, *in* E. Brynjolfsson and B. Kahin (eds), *Understanding the Digital Economy*, MIT Press, Cambridge, Massachusetts, pp. 99–136.

Smith, M. D. and Brynjolfsson, E. (2001). Consumer Decision-Making at an Internet Shopbot: Brand Still Matters, *Journal of Industrial Economics* **49**(4): 541–558.

Smith, R. J. and Blundell, R. W. (1986). An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply, *Econometrica* **54**(3): 679–686.

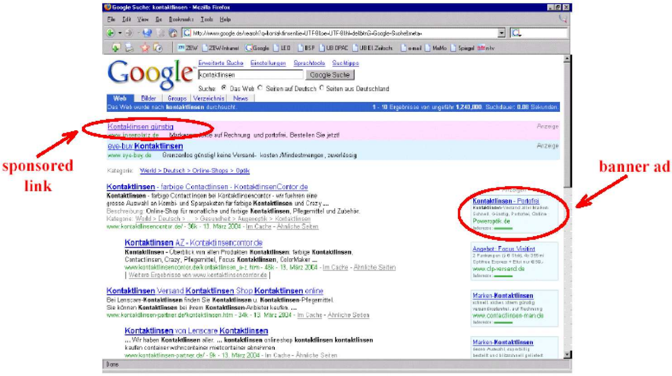
Stahl, D. O. (2000). Strategic Advertising and Pricing in E-Commerce, *in* M. R. Baye (ed.), *Industrial Organization*, Vol. 9 of *Advances in Applied Microeconomics*, JAI, Amsterdam, pp. 69–100.

Tang, F.-F. and Lu, D. (2001). Pricing Patterns in the Online CD Market: An Empirical Study, *Electronic Markets* **11**(3): 171–185.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

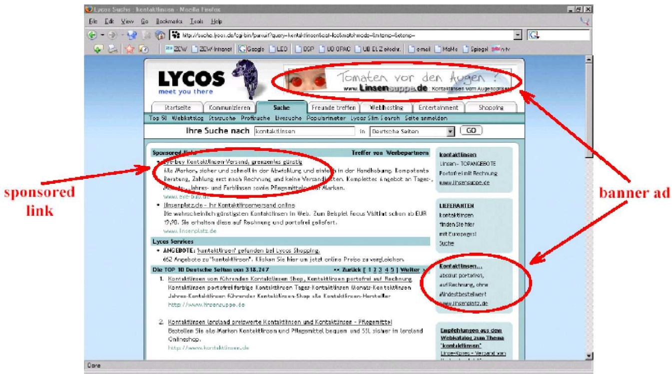
Appendix

Figure 2: Banner ads and sponsored links in a Google list of results



Note: Screenshot from 14.03.2004

Figure 3: Banner ads and sponsored links in a Lycos list of results



Note: Screenshot from 14.03.2004

Table 5: 10 most-widely used search engines

www.google.de
www.yahoo.de
search.msn.de
www.lycos.de
www.t-online.de
www.altavista.de
www.web.de
www.metager.de (meta search engine)
www.fireball.de
suche.aol.com

Source: www.webhits.de, 04.03.2002.

Table 6: Frequency distribution of Google rank categories

Google rank	Frequency
Google rank 1-10	41
Google rank 11-20	41
Google rank $\geq$ 21	847
Number of observations	929

Table 7: Frequency distribution of online advertising categories

Online advertising activities	Frequency
Banner ad	66
Sponsored link	34
Both	43
None	786
Number of observations	929

Table 8: Descriptive Statistics of explanatory variables

	Mean	Std. Dev.
Width of product range	21.743	17.661
Relative price level	1.024	0.135
Number of banner ads	0.300	1.062
Number of sponsored links	0.670	2.917
Google rank (ref.: not ranked 1-100):		
Google rank 1-10	0.044	
Google rank 11-20	0.044	
Google rank $\geq$ 21	0.912	
Retailer service:		
Convenient navigation	0.015	0.844
Superior customer service	0.002	0.813
Return policy + lens care	0.018	0.810
Security + trustworthiness	0.008	0.813
New CL user service	0.011	0.794
Number of observations	929	

Multi-Sided Markets: Competing with Network Externalities and Price Discrimination*

Bruno Jullien†

April 2006
first draft: March 2000

Abstract

The paper examines competition between multi-sided platforms, allowing for more than two sides, and analyzes the profits and the pricing strategies that emerge, assuming that externalities are not too large. It focuses on a Stackelberg pricing game, assuming a particular resolution of the coordination game played by platforms' users, that favours the leader. Price discrimination helps a platform coordinating the choices of consumers, and the pattern of prices that a platform should use to conquer market shares is characterized. Equilibria involve low profits, and even under the most advantageous conditions, a platform's profit may be adversely affected by asymmetries in multi-sided externalities. As a result, platforms may be better of interconnecting their multi-sided services. Results are then used to characterize the equilibria of a game of perfect price-discrimination by competiting networks, where the largest network is shown to be too small from a welfare perspective. Finally two-sided market equilibria are characterized (with sequential and simultaneous prices), and implications for inefficiencies and quality choices are discussed.

1 Introduction

The recent literature on two-sided markets has developed tools for the analysis of platform industries.[1] Two-sided platforms offer services used by two

*The author is strongly indebted to Bernard Caillaud. The author thanks Jacques Crémer, Jean Tirole and Jean Charles Rochet for their detailed comments, and Patrick Rey for useful discussions, as well as seminar participants at IDEI, University of Essex, University College London, University of East Anglia, Ente Einaudi, London Business School, ZEW and the European Summer Symposium in Economic Theory. Part of this work was conducted while visiting the European University Institute.

†University of Toulouse (IDEI and GREMAQ), bjullien@cict.fr.

[1]See Caillaud and Jullien (2001 and 2003), Rochet and Tirole (2003) and Armstrong (2005). See also Rochet and Tirole (2006) and Jullien (2005) for general presentations.

types of agents to interact. Because the interaction requires the cooperation of the parties on the platform, the service involves externalities. This combined with complex price structures[2], leads to pricing strategies that depart from the traditional trade-off between the volume of sales and the margin, and reallocate the benefits of the platform-intermediated interaction between members.[3]

So far the literature has been confined to models with two sides. This paper studies competition between platforms with two or more sides interacting on the platforms. It defines "multi-sided externalities" as externalities between distinct and well identified groups of agents, for which it is possible to charge different prices, and show how to extend the insights from two-sided markets to this case. Extending the model to an arbitrary number of sides allows to obtain new insights on business strategies used by platforms to conquer or preserve market shares, to clarify the role of price discrimination, and to analyze the extent of market power that one can expect in such markets.

The "multi-sided market" set-up emerges when a platform intermediates activities involving two or more distinct types of agents, as an internet portal providing a search engine along with a B2C auction platform. It also emerges when platforms, and more generally networks, discriminate by setting different prices for different users of the same service.[4] For instance, software suppliers discriminate between residential and professional users, as well as geographic areas. As it will be clear below, two groups of users that are subject to price-discrimination by a platform or a network can be be treated as two distinct sides in a multi-sided market context. The set-up thus covers a wide range of activities, such as intermediation activities, financial market places, telecommunication services, postal services, medias, operating systems and software applications, research centers,....

The paper examines a situation where various types of users have different valuations for the goods and services available on the platforms, as well as for the participation of other sides to the platform. It is assumed that each side is homogeneous, that users register with only one platform - there is single-homing[5]-, and that multi-sided externalities are not too large

[2]Platforms can charge different prices to the two sides. They can also discriminate between usage and registration.

[3]The two-sided approach has provided valuable insights on the functioning of markets such as credit cards (Wright (2003), Rochet and Tirole (2001)), intermediation (Caillaud and Jullien (2001, 2003)), mobile telephony termination charges (Wright (2003), Armstrong (2003)), media content (Anderson and Coates (2005), Ferando et al (2003)).

[4]The case of second degree price discrimination is not considered here. Damiano and Li (2005a) propose an analysis of endogenous sorting with a self-selection mechanism for a matching monopoly.

[5]The literature on two-sided markets distinguishes between single-homing and multi-

compared to the value attached by a user to the direct consumption of the goods offered by the platform. To this respect, this work is complementary to Caillaud and Jullien (2003), Ambrus and Argenziano (2004), or Damiano and Li (2005b) who consider pure intermediation services.

Due to multi-sided externalities, users of platforms' services face a coordination problem in their purchasing decision that may generate multiple equilibria (Katz and Shapiro (1985)). The paper assumes a particular equilibrium selection, obtained by choosing a (strong) platform and imposing that a user joins it whenever this is his choice in at least one equilibrium of the subgame where users choose between platforms. This allows to establish the maximal level of market power that a platform can obtain when coordination failure biases the game in its favor, by extending the concept of favorable expectations used by Hagiu (2004) to analyze the emergence of dominant platforms in the software or the video game markets, and by Cailaud and Jullien (2003).[6]

The core of the paper is devoted to characterizing the profit that a strong platform can obtain in a sequential pricing game where it is a Stackelberg leader. This allows to focus on the market power of the leader, and to derive an upper bound on the profit of a platform in the simultaneous move game, which is presented for the case of two sides in a second part.

As emphasized by the literature on two-sided markets, the design of the price structure must account for the coordination issue. Thus the price charged to a user reflects not only the cost of service but also the externalities associated with his/her participation. One intuition for this specificity of a platform's pricing strategies is that each individual is not only a user but also an "input", since his/her participation creates value for other user. As an input, the user is a scarce resource. Firms then compete both to sell the outputs (interactions) and to buy the inputs (participation). Competition may then lead to aggressive strategies offering very advantageous terms, even a subsidy, to some sides and exploiting the network externalities to generate profits on other sides: a divide-and-conquer strategy.[7] Such a strategy allows a platform overcoming the coordination problem by transferring part of the surplus to targeted customers and creating a bandwagon effect.[8] While for two-sided markets, the side generating the highest externality is favoured,

homing, where the latter refers to the possibility of joining two platforms. (See Caillaud and Jullien (2003) and Rochet and Tirole (2003)).

[6]To this extent, this paper is also complementary to contributions on two-sided markets that focus on selection criteria limiting to some extent the level of coordination failure (Ambrus and Argenziano (2004) for instance).

[7]See Caillaud and Jullien (2001, 2003), Innes and Sexton (1993).

[8]This bears similarity with the dynamic mechanism that leads a network to subsidize early customers. See Farell and Saloner (1986), Katz and Shapiro (1986,1992). See also Bensaid and Lesne (1996) and Cabral, Salant and Woroch (1999) for applications to dynamic monopoly pricing.

there is no such ranking with multiple sides and the paper shows how to identify the sides targeted with low prices.

It is shown that competition eliminates the ability of a platform to capture a share of the surplus generated by the multi-sided externalities. The effect is stronger when externalities are asymmetric: Sides valuing less the participation of others become the object of an intense competition as they are more value-enhancing relatively to others. The platform that succeeds in attracting them gains a competitive advantage, and as a result, competition for these sides dissipates the profits obtained with bandwagon effects. This may even prevent the existence of a pure strategy equilibrium in a simultaneous pricing game. This strong intensity of competition is one of the key difference with the more conventional analysis of network competition with uniform prices.

One consequence of multi-sided externalities is that the market allocation will be inefficient. Unsurprisingly, there may be excessive sales by the strong platform. More surprisingly, there may be excessive sales by the weak platform. Even a strong platform offering a uniformly superior quality may fail to cover the market and may leave some profitable niche for the rival, as a mean to soften competition on other sides.

Another consequence is that platforms may be better off interconnecting the multi-sided services, i.e. allowing their members to interact with the other platform's members. Without interconnection, platforms are extremely aggressive in building market shares, exploiting multi-sided externalities through cross-subsidies. Interconnection suppresses the motives for cross-subsidization and restores the benefits of differentiation.[9]

An alternative strategy for a platform aiming at reducing the intensity of competition, that emerges in the simultaneous pricing game, is to degrade the quality it offers to some targeted side, thereby increasing horizontal differentiation and inducing market sharing. This allows a platform to commit not to compete for this side.

As mentioned above, there is a close connection between the economics of multi-sided markets, and the economics of price-discrimination in network industries. While there is already a substantive literature on competition with network effects under uniform prices (see Katz and Shapiro (1994) and Economides (1996)), little attention has been devoted to price-discrimination, and this despite the fact that the practice is widespread for network services. Discriminating networks are special instances of multi-sided markets. The paper illustrates this by applying the analysis to perfect price-discrimination by competing networks with size related network effects. Increasing the value of network externalities raises the equilibrium

[9] This reasoning assumes away exclusionary motives and externalities within sides.

size of the largest network, but not as much as welfare maximization would require. Thus in equilibrium the largest network is too small. This result is similar to the result obtained by Argenziano (2005) for network competing without price-discrimination.

The paper is organized as follows. Section 2 presents the model and assumptions. Section 3 presents general results concerning the analysis of competitive strategies and of equilibria. Section 4 then presents the multi-sided applications while section 5 focuses on the two-sided market case. In particular, section 4 examines the case on perfect price-discrimination with one-sided network effects, and a few illustrative examples. Section 5 derives the equilibria under both the sequential and the simultaneous timing, and discusses the implications of the results. Section 6 discusses interconnection. Section 7 concludes. All the proofs are in Appendix.

2 The model

2.1 A monopoly platform

Consider a sole platform with a production cost normalized to 0. The platform bundles a consumption good and an intermediation service subject to externalities. For instance, for a software, the good would correspond to the basic service and while the multi-sided service would correspond to the ability to run independent applications. There are J different types of users, each represented by a "side" composed of a mass m_j of identical agents. The set of sides is denoted $\mathcal{J}$.

The perceived quality of the good varies across types of users, denoted u_j for side j (intrinsic value). The intermediation service value depends on the composition of the population that the platform allows to reach, and each side values differently the participation of every other sides to the platform. The valuation of a user on side j for the participation of n_l members of side $l \neq j$ is $\beta_{jl} n_l$, where the coefficients β_{jl} are nonnegative. Overall the gross utility of a side j user joining the platform is $u_j + \sum_{l \neq j} \beta_{jl} n_l$. Notice that the model doesn't consider the possibility of within-side network effects. With homogeneous sides, the extension to the case where users care also about their own side, $\beta_{jj} \geq 0$, raises no difficulty, and is presented at the end of the section, where it will be shown that the positive analysis would be similar to a redefinition of utilities. The choice of not considering it here is made first to focus on the multi-sided aspect and avoid confusion, and second because welfare conclusions would be more ambiguous.

Remark 1 *Given homogeneity within sides and the fact that sides will not split between platforms, the model can be also interpreted as a model of competition with network effects and perfect price discrimination, with J*

individuals and arbitrary preferences. Indeed, setting $m_j = 1$ for all sides, there will be no distinction between the behavior of a side and the behavior of a single individual with the same preference. Thus all the conclusions derived apply as well to competing network under perfect price discrimination (see section 4.1).

The platform charges a different price for each side, the vector of prices being $P = \{p_1, ..., p_J\}$. Given these prices, each user decides whether to join or not. Consumers coordinate on a rational expectation equilibrium of this allocation game.[10] Due to network effects, there may be a multiplicity of such equilibria.

The maximal prices that a monopoly platform can set are $u_j + \sum_{l \neq j} \beta_{jl} m_l$. At these prices, each user is willing to join provided that all others do. From now on and for conciseness, I consider only prices that are below this maximal monopoly prices.[11]

Whether the monopolist is able to sell at these maximal prices depends on the coordination process of users. The reason is that there is also another equilibrium allocation of users where none of them join.

More generally, for any vector of prices such that $p_j \geq u_j$ for all j, there is the possibility that users fail to coordinate and that the platform doesn't sell at all. Thus when the users' coordination process leads to the least favorable allocation, the monopolist must set at least one price p_j below u_j. Suppose that $p_1 < u_1$, then side 1 joins irrespective of what the others are doing. Given that, the platform can set a price above u_2 for side 2: any price $p_2 < u_2 + \beta_{21} m_1$ ensures that side 2 joins as well, since given that side 1 joins, a member of side 2 joins irrespective of what the others are doing. Using this reasoning recursively, we see that the platform can cover the market by setting prices $p_j = u_j + \sum_{l<j} \beta_{jl} m_l$, even faced with the least favorable market conditions. Obviously the reasoning doesn't depend on the order of the sides so that, if $\sigma(.)$ denotes an order (a permutation) on the set of sides, prices $p_j = u_j + \sum_{\sigma(l)<\sigma(j)} \beta_{jl} m_l$ (or slightly below) allow the platform to cover the market as well. Thus, even in the worst case, the monopoly can extract part of the value of network externalities.[12]

The general conclusion from this section is that prices serve two purposes for a monopoly platform. First they serve the standard purpose of extracting the rent from users. Second, it may help the platform to overcome coordination failure.

[10] Each individual's consumption decision maximizes his utility given the prices and the equilibrium allocation of other individuals, including others members of its side.

[11] Prices above these levels are irrelevant for the analyis of strategies and equilibria, and would be ruled out for instance by elimination of weakly dominated strategies.

[12] This is similar to the analysis of unique implementation with externalities in Segal (2003).

2.2 Two competing platforms

Suppose now that two platforms, denoted S (strong) and W (weak), compete for the provision of services. Denote u_j^k the utility derived by a member of side j from the good of platform k, and let $\delta_j = u_j^S - u_j^W$ be the "stand-alone" quality differential for side j. If n_l^S and n_l^W members of side l join S and W respectively, a side j user joining platform k benefits also from network effects $\beta_{jl} n_j^k$. Given that, side j utility at platform k writes as

$$U_j^k = u_j^k + \sum_{l \neq j} \beta_{jl} n_l^k - p_j^k.$$

I focus on the case where the network externalities at the platform level are not too large compared to the intrinsic value of the good. More specifically it is assumed that:

Assumption 1 (weak externalities): $\forall j : u_j^W \geq \sum_{l \neq j} \beta_{jl} m_l$.

To give an example, the benefits of using the same text editor are confined within small communities, and presumably smaller than the value of using a text editor. Third generation mobile base services include voice services or E-mail access under universal connectivity, while the provision of more sophisticated services (TV programs for instance) may have a multi-sided nature.[13] The assumption doesn't hold for credit cards or matching agencies for instance, for which network effects swamp the intrinsic value.[14]

It must be stressed that results differ if assumption 1 is not verified.[15] Hence, this assumption is not innocuous.

A full characterization of the equilibria of the general game with multi-sided markets is not tractable.[16] The core of the paper focuses on the Stackelberg game with S leader. Equilibria for the simultaneous move equilibria, as well as the reverse sequential timing, are discussed afterward. The competitive game is thus composed of three stages:

Stage 1.1 : Firm S sets prices P^S.

Stage 1.2 : Firm W sets prices P^W.

[13] Notice that this case involves also pricing of interactions (calls), that is assumed away in the paper.

[14] One can also reinterpret this assumption in terms of compatibility between services or standards. For instance let θ measure the quality of interconnection between two communication networks. If side j receives a utility $U_j^k = \theta \sum_{l \neq j} \alpha_{jl} m_l + (1-\theta) \sum_{l \neq j} \alpha_{jl} n_l^k$, we can define $u_j^k = \theta \sum_{l \neq j} \alpha_{jl} m_l$ and $\beta_{jl} = (1-\theta)\alpha_{jl}$. Weak externalities then means that θ is large enough.

[15] See section 5.3.5.

[16] Moreover a pure strategy equilibrium may not exist in the simultaneous pricing game.

Stage 2 : Users simultaneously decide which platform to join if any.

Faced to prices $P^S = \{p_1^S, .., p_J^S\}$ and $P^W = \{p_1^W, ..., p_J^W\}$, users coordinate in stage 2 on a rational expectation equilibrium allocation (REA). A REA is such that the choice of each user is individually optimal given the behavior of the other users. An allocation rule, denoted $\mathcal{A}(P^S, P^W)$, is a mapping from the set of prices to the set of users allocations, that assigns to every price vector a REA. An equilibrium consists in an allocation rule $\mathcal{A}$, and of equilibrium prices of the pricing game where the profit of platform k at prices P^S and P^W is the profit obtained when sides are allocated according to $\mathcal{A}$. The choice of the allocation rule is thus key to the determination of equilibrium prices.

This paper focuses on the extreme case where platform S benefits from an advantage because users' tend to coordinate on it. While this notion may be ambiguous in a general context, it takes the form of a simple selection criteria when there are positive network externalities, due to the following result:

Lemma 1 (maximal inclusiveness) *Fix the prices P^S and P^W, and consider all REAs for these prices. Let $\mathcal{K}$ be the set of sides with a positive participation to platform k in* ***at least one*** *REA and $\mathcal{S}$ be the set of sides with full participation to platform s in* ***all*** *REAs. Then there exists a REA, denoted $\mathcal{D}^k(P^k, P^s)$, such that all sides in $\mathcal{K}$ join platform k, and only sides in $\mathcal{S}$ join platform s.*

The key feature is that the value of a platform uniformly increases when new users are added to its customer base. According to a bandwagon effect, moving sides from s to k raises the incentives to join k for all individuals and reduces the incentives to join s. Thus there exists a unique REA that at the same time maximizes the market share of platform k and minimizes the market share of platform s. This follows from the fact that there are strategic complementarity in the users allocation process (see Topkis (1979), Vives (1990)).

It is assumed that platform S dominates the coordination process in the sense that users coordinate on $\mathcal{D}^S(P^S, P^W)$. To avoid inexistence problems that may be created by discontinuities at indifference points of users, the allocation is allowed to differ from $\mathcal{D}^S$ at such points. Defining $\bar{\mathcal{D}}^S$ as the closure of $\mathcal{D}^S$:[17]

Assumption 2 (favorable expectations): $\forall (P^S, P^W), \mathcal{A}(P^S, P^W) \in \bar{\mathcal{D}}^S(P^S, P^W)$.

[17]For almost all prices, $\bar{\mathcal{D}}^S$ coincides with $\mathcal{D}^S$. At points of discontinuity, agents of some side joining S are indifferent between S and W. A slight reduction of the price of W for this side would induce them to join W, and the bandwagon effect may induce other sides to follow. $\bar{\mathcal{D}}^S(P^S, P^W)$ thus includes also $\lim_{\varepsilon \to 0} \mathcal{D}^S(P^S, P^W - \varepsilon)$.

Admittedly a strong criteria, the assumption allows capturing in a simple way the role of users coordination in the emergence of market power in platform industries.[18] It postulates that coordination failure in the market is always resolved "in favor" of platform S, which to some extent is focal. Which firm is focal at a particular point in time may then depend on history or some exogenous factor. The assumption then confers to platform S the ability to charge prices above cost, and may thus be viewed as a way to focus on maximal market power. The term favorable expectations is borrowed from Hagiu (2004) who, building on Caillaud and Jullien (2003), uses a similar concept to analyze the emergence of dominant platforms in a two-sided model where sides choose the platform sequentially.

Alternative selection criteria that have been used in the literature tend to limit the lack of coordination through various means. For instance, one alternative is to assume that consumers coordinate on a Pareto undominated equilibrium of the allocation subgame. Ambrus and Argenziano (2005) use a different concept, coalitional rationalizability, with a similar property. Limiting coordination failure then limits the potential exercise of market power.

To see the difference, consider the case of two networks competing with uniform prices for identical agents who care only about the size of the network: a consumer joining network k receives a utility $u^k + \beta N^k - p^k$, where N^k is the size of network k and p^k is its price. Suppose that $u^S \geq u^W$. Then the Pareto criteria or coalitional rationalizability imply that all consumers join network k whenever this generates the highest utility, so that in equilibrium, platform S serves the market at price $u^S - u^W$. By contrast, under favorable expectations for S, all consumers coordinate on platform S whenever $u^S + \beta M - p^S > u^W - p^W$. Thus in equilibrium the price of platform S is now $u^S - u^W + \beta M$, augmented by the value of network effects.

Understanding the best-response in terms of prices when expectations favor one platform is also useful for the more general case where $\mathcal{A}$ is not restricted. Indeed the easiest way to support any equilibrium is to assume that, following a deviation from equilibrium strategies by platform k, users coordinate on $\mathcal{D}^{-k}(P^{-k}, P^k)$.[19] This methodology is used for instance in Caillaud and Jullien (2003) to characterize the set of equilibrium profits and allocations of a pure matching model under simultaneous pricing and no restriction except the monotonicity of the equilibrium selection.

[18] The concept of expectations is viewed here in a game theoretic manner, as it refers to the beliefs of players on the other players' choice of platforms.

[19] This minimizes the deviation profits, without restricting the allocation of consumers on the equilibrium path.

2.3 Within-side cross-effects

In the paper, it is assumed that there is no externality between members of the same side. One can allow for such an externality by setting $\beta_{jj} \geq 0$. The utility writes as

$$U_j^k = u_j^k + \sum_{l=1}^{J} \beta_{jl} n_l^k - p_j^k.$$

It is straightforward to see that lemma 1 still holds, so that the concept of favorable expectations is well defined. Then the value of the within-side externalities β_{jj} can be attributed to S :

Lemma 2 *Suppose that $\beta_{jj} \geq 0$ for some j. Consider an alternative model that differs only by the facts that for all j, $\beta_{jj} = 0$ while the value of J's good is $u_j^S + \beta_{jj} m_j$. Then the two models generate the same maximal allocation rule $\mathcal{D}^S(P^S, P^W)$.*

Thus the market allocation at given prices can be derived "as if" the intrinsic value differential were $\hat{\delta}_j = \delta_j + \beta_{jj} m_j$ with no network effect within sides. Apart for the welfare analysis, all the results of the paper extend to this case replacing δ_j by $\hat{\delta}_j$. Thus positive network effects between members of the same side would strengthen the position of S.

3 The strong leader: general results

In the sequential game considered, intuition suggests that platform S benefits from the presence of externalities. This is not the case because platform S must account for the possibility that platform W uses "divide-and-conquer" strategies to overcome its disadvantage. To do so, W has to favor some sides over the others: W can charge a low or even negative price for one side,[20] if selling to this side allows to sell to another side at a high price. To follow insights from Innes and Sexton (1993), one may view the situation as one in which W faces users who have the possibility to form a coalition (to join S). To prevent the formation of the coalition, W needs to "bribe" some users. However it needs not bribe all users but only enough of them to ensure that the value of the sub-coalition composed of the remaining users is reduced to a point where it becomes unattractive. This strategy solves the coordination problem and allows a weak platform to capture part of the value of network externalities.

[20] Given that costs are normalised to 0, prices should be interpreted as margins and a negative price as a price below marginal cost. In the model, the real price may (but need not) be not only below marginal cost but indeed negative. One interpretation in this case is that the customer receives free access to the platform, and is subsidized through additional free goods and services.

We will show that due to these strategies, the platform S selling to a subset $\mathcal{K}$ of sides cannot obtain a profit larger than $\sum_{j\in\mathcal{K}} \delta_j m_j$, which represents simply the quality differential associated with the good offered by the platform.

To illustrate this, suppose that there are only two sides and that S decides to sell to both sides in equilibrium. Assuming that $\beta_{12} \leq \beta_{21}$, let us build the best strategy that would enable W to attract both sides. For any prices such that $p_j^W \geq p_j^S - \delta_j - \beta_{jl} m_l$, assumption 2 implies that side j joins S if the other side does. Thus to attract some users, W needs to set one price p_j^W below $p_j^S - \left(\delta_j + \beta_{jl} m_l\right)$, say p_1^W. Then side 1 joins platform W irrespective of what the other side does. This is known by side 2, so that W attract side 2 at a price p_2^W such that $u_2^W + \beta_{21} m_1 - p_2^W \geq u_2^S - p_2^S$. Indeed the choice for a member of side 2 is now between joining side 1 or staying with S. Under the above condition, the former dominates unambiguously. The maximal price that W can set on side 2 is thus $p_2^S - \delta_2 + \beta_{21} m_1$. By attracting side 1, W has reduced the attractiveness of S for side 2 by an amount $\beta_{21} m_1$, and increased its own attractiveness by the same amount. The overall profit from this strategy is $\left(p_1^S - \delta_1 - \beta_{12} m_2\right) m_1 + \left(p_2^S - \delta_2 + \beta_{21} m_1\right) m_2$. If S sells to both sides in equilibrium, this must be non-positive:

$$p_1^S m_1 + p_2^S m_2 \leq \sum_{j=1}^{2} \delta_j m_j - (\beta_{21} - \beta_{12}) m_1 m_2 \tag{1}$$

We see that the maximal profit is bounded above by $\sum_{j=1}^{2} \delta_j m_j$, and the difference depends on the level of asymmetry in externalities. The remaining analysis extends this result to an arbitrary number of sides.

3.1 Market covering by the leader

We start by considering the maximal profit that platform S can derive when it chooses to corner the market, i.e. to attract all sides. Given our assumption on the users coordination process, S will cover the market for prices P^S and P^W such that the maximal utility a user can obtain with S is larger than the minimal utility a user may obtain with W :

$$\forall j : u_j^S + \sum_{l \neq j} \beta_{jl} m_l - p_j^S \geq u_j^W - p_j^W,$$

where the LHS is the utility of joining platform S when all other users do, while the RHS is the utility of joining platform W if nobody else does.

To cover the market, S must set prices such that this inequality holds for all positive prices p_j^W, since otherwise platform W could attract side j with a positive price.

Using $\delta_j = u_j^S - u_j^W$, platform S must necessarily set prices such that

$$\forall j : p_j^S \leq \delta_j + \sum_{l \neq j} \beta_{jl} m_l. \tag{2}$$

Notice that assumption 1, weak externalities, implies that $\delta_j + \sum_{l \neq j} \beta_{jl} m_l \leq u_j^S$. Thus condition (2) implies that the utility level that side j would obtain joining alone S is nonnegative: $u_j^S - p_j^S \geq 0$. This ensures that, under condition (2), in any subgame following $S's$ pricing decision, and in particular for any price P^W, side j joins S if not W. Thus when discussing potential deviation for W, we need not worry about the possibility that some users decide to stay out of the market, which simplifies greatly the analysis.

The two-sided reasoning developed above extends to the multi-sided context, in the sense that, when designing a pricing strategy to conquer markets shares, platform W decides for each pair of sides j and l, whether it is preferable to set a low price for side j and to charge side l for the value it attaches to joining j, or the reverse. The former possibility would call for undercutting S on side j (by $\beta_{jl} m_l$) and to add a mark-up on side l (of value $\beta_{lj} m_l$) and it will be the preferred option if side l values more the participation of side j, than side j values the participation of side l, $\beta_{lj} > \beta_{jl}$.

The difficulty with multi-sided externalities is that there may be conflicts between the preferred strategies for different pairs of sides, and the necessity to overcome the coordination failure at the multi-sided level. In order to "start" the process that allows to generate a bandwagon effect, W will need to subsidy heavily some sides.[21] This leads to an implicit ranking of sides from the most favoured to the most heavily taxed ones. This notion of ranking already appeared in the analysis of the monopoly platform and is captured by a permutation on the set of sides.

This leads to define

$$\Omega(\mathcal{J}) = \max_{\sigma(.)} \left\{ \sum_{j \in \mathcal{J}} \sum_{\substack{l \in \mathcal{J} \\ \sigma(l) > \sigma(j)}} (\beta_{lj} - \beta_{jl}) m_l m_j \right\} \tag{3}$$

$\Omega(\mathcal{J})$ captures the profit that W can generate with cross-subsidization, the maximization coming from the optimal choice of targeting. This term has a very simple interpretation: when side j is attracted "before" side l, W must give a subsidy to the members of side j equal to their opportunity cost $\beta_{jl} m_l$ of leaving side l, but it can charge an extra amount $\beta_{lj} m_j$ to the

[21] I use the term "subsidy" to mean here that the side is charged a price below $p_j^S - \delta_j$. This is slightly abusive as there may not be any real subsidy and prices may remain above cost. Cross-subsidy means in this context that the mark-up over $p_j^S - \delta_j$ is negative for some sides and positive for others.

members of side l corresponding to the value of joining side j. The net effect is then $(\beta_{lj} - \beta_{jl})m_l m_j$.

The maximal profit that S can generate when covering the market is then equal to the total quality differential on the good minus the profit that W could get by exploiting multi-sided externalities.

Proposition 2 *Suppose that S sells to all sides in equilibrium, then its profit is* $\bar{\Pi}^S = \sum_{j=1}^{J} \delta_j m_j - \Omega(\mathcal{J})$.

The formal proof is in appendix, but a sketch goes as follows. Assume that prices P^S verify (2) and suppose that W tries to sell to a subset $\mathcal{K}$ of sides. For conciseness take $\mathcal{K} = \{1, 2, ..., K\}$. Firm W has to set at least one price small enough so that $u_j^S + \sum_{l \neq j} \beta_{jl} m_l - p_j^S < u_j^W - p_j^W$, which ensures the willingness of a member of side j to join alone. Say that the price p_1^W is slightly below $p_1^S - \delta_1 - \sum_{l>1} \beta_{1l} m_l$. The resulting price is such that it is a dominant strategy for a member of side 1 to join W. Given that it is commonly known that side 1 joins W, any price p_2^W such that $u_2^S + \sum_{l>2} \beta_{2l} m_l - p_2^S < u_2^W + \beta_{21} m_1 - p_2^W$ induces a member of side 2 to join as well. W can thus convince side 2 to join with a price p_2^W slightly below $p_2^S - \delta_2 - \sum_{l>2} \beta_{2l} m_l + \beta_{21} m_1$. The process can then continue: given p_1^W and p_2^W, members of sides 1 and 2 join W, so that a member of side 3 joins at a price that makes it more attractive to join sides 1 and 2, as opposed to staying with sides 3 and above in $S's$ platform. More generally, side j is charged by W the largest price that ensures that its members join given that all sides $l < j$ join. The pricing strategy is thus built in such a way that after j rounds of elimination of dominated strategies in stage 2, there are j sides for which the only remaining strategy is to join W. The resulting prices are:[22]

$$p_j^W = p_j^S - \delta_j + \sum_{l<j} \beta_{jl} m_l - \sum_{l>j} \beta_{jl} m_l$$

Summing up over the sides we obtain $W's$ profit:

$$\sum_{j=1}^{K} p_j^W m_j = \sum_{j=1}^{K} p_j^S m_j - \sum_{j=1}^{K} \delta_j m_j + \sum_{j=1}^{K} \sum_{l=j+1}^{K} (\beta_{lj} - \beta_{jl}) m_l m_j - \sum_{j=1}^{K} \sum_{l=K+1}^{J} \beta_{jl} m_l m_j$$

More generally, platform W has the choice of the set $\mathcal{K}$ of sides targeted, and of the order in which sides are subsidized. Indeed W benefits from subsidizing the sides valuing less the network effects, since the required subsidy is smaller than the value of the network externalities that W can extract from the other sides. To compute equilibrium conditions, we must thus account for

[22] Here equality generates the maximal profit but one should think of W as setting prices slightly below equality.

any possible order between sides. Denote $\sigma(.)$ a permutation on the set of sides, where $\sigma(l) > \sigma(j)$ means that j is ranked before l. The interpretation is that W sets a price $p_j^W = p_j^S - \delta_j + \sum_{\sigma(l)<\sigma(j)} \beta_{jl} m_l - \sum_{\sigma(l)>\sigma(j)} \beta_{jl} m_l$ for side j, such that a member of this side is willing to join provided that it is sure that all sides ranked below join as well. The maximal profit that W can get on a subset $\mathcal{K}$ of sides is obtained by summing the corresponding prices over the sides and optimizing on the order σ. Define

$$\Omega(\mathcal{K}) = \max_{\sigma(.)} \left\{ \sum_{j\in\mathcal{K}} \sum_{\substack{l\in\mathcal{K} \\ \sigma(l)>\sigma(j)}} (\beta_{lj} - \beta_{jl}) m_l m_j \right\} \tag{4}$$

$$B(\mathcal{K}) = \sum_{j\in\mathcal{K}} \sum_{l\in\mathcal{J}\setminus\mathcal{K}} \beta_{jl} m_l m_j. \tag{5}$$

Then the maximal profit that W can obtain by selling to the subset $\mathcal{K}$ is

$$\sum_{j\in\mathcal{K}} p_j^S m_j - \sum_{j\in\mathcal{K}} \delta_j m_j + \Omega(\mathcal{K}) - B(\mathcal{K}).$$

$B(\mathcal{K})$ is the total value of the network effects between sides within $\mathcal{K}$ and sides outside $\mathcal{K}$. Since W sells only to $\mathcal{K}$, it must compensate its customers for the foregone value of network externalities. $\Omega(\mathcal{K})$ has the same interpretation as $\Omega(\mathcal{J})$. Because platform W can choose the order of targeting, $\Omega(\mathcal{K})$ is non-negative

Lemma 3 *For all* $\mathcal{K}$, $\Omega(\mathcal{K}) \geq 0$, *with equality if and only if for all* j, l *within* $\mathcal{K}$, $\beta_{jl} = \beta_{lj}$.

Remind that for the moment, it is assumed that platform S covers the market. This is the case if prices P^S are such that there is no subset $\mathcal{K}$ of sides that W could attract with positive profits. This amounts to the following condition:

$$\forall\mathcal{K} : \sum_{j\in\mathcal{K}} p_j^S m_j \leq \sum_{j\in\mathcal{K}} \delta_j m_j - \Omega(\mathcal{K}) + B(\mathcal{K}). \tag{6}$$

This provides us with bounds on the total profit derived by S from any particular subset of sides. These bounds have three components. The first component includes the absolute advantage of S on the sides targeted, captured by the " stand alone" quality differentials δ_j. The second term $\Omega(\mathcal{K})$ captures the impact on profit of $W's$ discrimination between sides. The last term captures the fact that when W chooses not to sell to some sides, it must compensate its clients for the value of networks effects they would receive if they stayed with S.

Clearly an upper bound on the profit that S can obtain by selling to all sides is the bound obtained for the set of all sides $\mathcal{K} = \mathcal{J}$. It turns out that S can set prices that generate this maximal profit and verify all the other constraints, hence the result of proposition (2).

3.2 Market sharing

A bound on profits can also be obtained when S doesn't cover the market and sell only to a subset $\mathcal{K}$, by considering the extra profits that W could earn by changing its prices for sides within $\mathcal{K}$, while maintaining the prices of sides outside $\mathcal{K}$ at their equilibrium values.

Corollary 1 *Suppose that S sells to sides within $\mathcal{K}$ only in equilibrium, it's profit is smaller than or equal to $\sum_{j\in\mathcal{K}} \delta_j m_j - \Omega(\mathcal{K}) - B(\mathcal{K})$.*

The bound is the same as before accounting for the fact that since sides outside $\mathcal{K}$ join W, the value of W is augmented by the corresponding value $B(\mathcal{K})$ of multi-sided effects. While the same logic applies, we will see in the section on two-sided markets that in the case where S doesn't cover the market this bound may not always be attainable. The reason is that the bound is obtained by fixing the prices set by W for the sides it serves in equilibrium and varying the other prices only. It thus ignore the possibility that W uses more complex strategies involving all the prices.

These more complex strategies don't matter when multi-sided effect are pairwise symmetric, that is when for all $j, l : \beta_{jl} = \beta_{lj}$. Indeed, in this case the implicit ranking of sides is irrelevant so that corollary (1) gives the profit.

Proposition 3 *When multi-sided effect are pairwise symmetric ($\forall j, l : \beta_{jl} = \beta_{lj}$), platform S can obtain the profit $\sum_{j\in\mathcal{K}} \delta_j m_j - B(\mathcal{K})$ by selling to sides within $\mathcal{K}$.*

In this case platform S chooses the set $\mathcal{K}$ so as to maximize this profit.

A key implication of the results is that S can't benefit from the presence of multi-sided externalities, despite assumption 2, since its profit is bounded by the profit $\sum_{j\in\mathcal{K}} \delta_j m_j$ it would obtain if no users attached a value to the participation of other sides ($\beta_{jl} \equiv 0$, for all j and l).[23] When multi-sided externalities are symmetric ($\beta_{jl} = \beta_{lj}$), their effects in the divide-and-conquer strategy cancel out and $\Omega(\mathcal{K})$ vanishes. Multi-sided externalities with a platform are neutral for S in this case, and what matters is the values missing externalities between platforms. Then introducing some asymmetry can

[23] This contrasts with the case of networks effects between members of the same side, as S would benefit from the presence of positive network effects within sides (see the discussion section).

only worsen the case for S. In particular if $\delta_j \equiv 0$ for all j, the Stackelberg leader leaves the whole market to its rival.

The results also shows that

Corollary 2 *If $\bar{\Pi}^S \geq 0$ and for all subsets $\mathcal{K}$*

$$\sum_{j \notin \mathcal{K}} \delta_j m_j + B(\mathcal{K}) \geq \Omega(\mathcal{J}) - \Omega(\mathcal{K}).$$

then platform S covers the market at equilibrium.

One point worth noticing when comparing (6) and corollary (1) is that the difference between the two formulas is $2B(\mathcal{K})$. Indeed the opportunity cost for S of not serving side j is equal to twice the value of the externality between this side and $S's$ clientele. The reason is that the value of the externality is transferred to W, thus the total value of S is reduced by $B(\mathcal{K})$ while the value W can offer increases by $B(\mathcal{K})$.

3.3 The weak leader

To complete this section let us just mention what occurs if instead of S, platform W is the price leader. The difference with before is that under assumption 2, S can capture all the surplus from network effects without relying on cross-subsidization. By attracting a side client of W, S could then capture the value of the externalities between this sides and its own clients, but also the value created to its clients when the new side joins the platform. This generates a total potential value differential in favor of S equal to the sum of the quality differential and the total increase in the value of network effects on $S's$ platform when it attracts the whole population instead of sides within the equilibrium set $\mathcal{K}$ only.

Proposition 4 *Suppose that W is a Stackelberg leader. The profit that W obtains when it sells to sides within $\mathcal{L}$ and S sells to sides within $\mathcal{K} = \mathcal{J} \backslash \mathcal{L}$ is $\sum_{j \in \mathcal{L}} (u_j^W - u_j^S) m_j - B(\mathcal{L}) - B(\mathcal{K})$.*

In particular, if W covers the market it obtains profits

$$\bar{\Pi}^W = -\sum_{j \in \mathcal{J}} \delta_j m_j - \sum_{(j,l) \in \mathcal{J} \times \mathcal{J}} \beta_{jl} m_j m_l.$$

4 Multi-sided markets

4.1 Perfect price-discrimination with uniform network effects

As pointed above, the analysis is the same if a side is constituted by a single individual. Setting $m_j = 1$, then J is the number of individuals and

p_j is interpreted as an individualized price. Thus one application of the multi-sided framework developed is the case of perfect price-discrimination with network effects. Let us apply this framework to the conventional case where consumers care only about the size of the network, by assuming in this section:

Property N : $\forall j, l : \beta_{jl} = \beta, \ m_j = 1, \ \delta_j \geq \delta_{j+1};$

where individuals are ranked by decreasing order of preference for network S.

Assume for the moment that all individuals have identical taste for the goods, so that for all j : $\delta_j = \delta$. From proposition (3), when able to set different prices to different sides, platform S covers the whole market with profit δ if positive, while S cannot profitably attract a set of consumers if $\delta < 0$. In the case where $\delta = 0$, either of the two platforms may sell to all individuals in equilibrium.

Thus, with uniform preferences, the equilibrium allocation is efficient and platform S sells if and only if it offers a higher quality than platform W. The weakest platform is able to overcome the coordination problem and to pass the full value of the surplus to its customers, which eliminates inefficiencies.[24]

Allowing the intrinsic utility to differ across consumers, the profit for the network S serving sides in $\mathcal{K}$ is $\sum_{j \in \mathcal{K}} \delta_j - \beta K(J-K)$ where K is the cardinal of $\mathcal{K}$, which implies that platform S serves a set $\mathcal{K} = \{1, ..., K\}$ and choose the size

$$K^S(\beta) = \arg\max_K \sum_{j=1}^{K} \delta_j - \beta K(J-K)$$

We see that the benefits of adding one more side is

$$\delta_{K+1} - \beta(K+1)(J-K-1) + \beta K(J-K) = \delta_{K+1} - \beta(J-2K),$$

larger than δ_j if $K > \frac{J}{2}$. When it serves more than half of the market, the leader would serve all sides for which it has a larger quality u_j^k and more, but if it is smaller than its rival, it may refrain from including some side for which it is more efficient. The reason is that the leader benefits from maximizing asymmetries between the size of the network as this minimizes the value of the externalities between users of different platform.

Proposition 5 *Assume property N holds. For* $\beta_2 > \beta_1$: $\left|K^S(\beta_2) - \frac{J}{2}\right| \geq \left|K^S(\beta_1) - \frac{J}{2}\right|$. *If the profit is quasi-concave in* K *when* $\beta = \beta_2$, *then eiher* $K^S(\beta_2) \geq K^S(\beta_1) \geq J/2$ *or* $K^S(\beta_2) \leq K^S(\beta_1) \leq J/2$.

[24] One consequence of this is that if we drop assumption 2, and consider a simultaneous pricing game, there is a unique efficient equilibrium.

The proposition says that under quasi-concavity, increasing β increases the size of the largest network. Quasi-concavity is required as for large β, some tipping may occur where a small platform S decides to cover the market. For instance suppose that $\sum_{j=1}^{K}\delta_j$ is maximal at $K^S(0) < 1/2$, but that $\sum_{j=1}^{J}\delta_j > 0$. For a low but positive β, platform S reduces its market share to $K^S < K^S(0)$. But at some point, its profit will fall below $\sum_{j=1}^{J}\delta_j$, so that it will prefer to serve the whole population and $K^S(\beta)$ may jump to J.

Thus competition with perfect price discrimination tends to generate less balanced allocations than competition based solely on intrinsic values. However the allocation is too balanced from a social welfare perspective. Indeed welfare writes as

$$\begin{aligned}&\sum_{j=K+1}^{J}u_j^W+\sum_{j=1}^{K}u_j^S+\beta K^2+\beta(J-K)^2\\ =\;&\left(\sum_{j=1}^{J}u_j^W+\beta J^2\right)+\left(\sum_{j=1}^{K}\delta_j-2\beta K(J-K)\right)\end{aligned}$$

Comparing with the profit of S, we see that, up to the first constant term, the difference between $S's$ profit and welfare is $\mathcal{B}(\mathcal{K}) = \beta K(J-K)$, the value of the externalities between members of different platforms, maximal at $K = J/2$, when the platforms are of equal sizes. Divide-and conquer strategies by W prevents S from capturing all the value of the externalities generated when attracting more sides. Here S obtains only half of them. This implies that the allocation is biased toward the allocation generating the least value of network effects, namely the balanced allocation.

Proposition 6 *Under property N, let total welfare be maximal at K^*. Then eirther $K^* \le \min\{1/2, K^S(\beta)\}$ or $K^* \ge \max\{1/2, K^S(\beta)\}$. If in addition welfare is strictly quasi-concave in K, then either $K^* \le K^S(\beta) \le 1/2$, or $K^* \ge K^S(\beta) \ge 1/2$.*

Thus the market share of the network that should be the largest from a welfare perspective is too small in equilibrium. If we add the requirement that welfare is quasi-concave in K, still the largest network is the right one. Therefore welfare maximization would require increasing the size of the largest network. The conclusion is similar to the result of Argenziano (2005) on one-sided externalities. She analyzes a model of competition between networks with uniform prices and concludes that the largest network size is suboptimal in equilibrium. The same result holds under perfect price-discrimination but for different reasons. In the case of uniform prices, the result follows from the relation between mark-ups and network effects. In the present case, this is due to an attempt of the leader to protect its market share, faced to "excessive" competition due to price-discrimination.

4.2 A multi-service platform

Consider the case where consumers on one side ($j = 1$) find partners for $J-1$ potential trades not related with each other. One may think for instance of a mall used for one-point shopping, side $j = 1$ includes the consumers, while each side $l > 1$ consists in stores selling a specific good searched by the consumers.[25] The mall sets prices to stores and to consumers. In this case we have: $\beta_{1l} > 0$, $\beta_{l1} > 0$, and $\beta_{jl} = 0$ otherwise. To simplify assume that

$$\forall l : \beta_{1l} = \beta_b < \beta_{l1} = \beta_s$$

Thus the value β_b attached by a consumer to the presence of one more store is smaller than the profit per consumer expected by a store. Then we have

$$\begin{aligned}\bar{\Pi}^S &= \sum_{j=1}^{J} \delta_j m_j - (\beta_s - \beta_b)\left(\sum_{j=2}^{J} m_j\right) m_1, \\ &= \sum_{j=1}^{J} \delta_j m_j + \beta_b \left(\sum_{j=2}^{J} m_j\right) m_1 - \sum_{j=2}^{J} \beta_s m_1 m_j.\end{aligned}$$

The divide and conquer strategy calls for subsidizing consumers by an amount equal to the total value they derive from the presence of stores and charge stores for the access to consumers.

Now suppose that one store is more attractive than others, and can obtain an individualized price. As already pointed we can consider this store as a single side, side 2 with $m_2 = 1$ and $\beta_{12} > \beta_b$. If it is the case that

$$\beta_b < \beta_{l1} = \beta_s < \beta_{12},$$

we have

$$\begin{aligned}\bar{\Pi}^S &= \sum_{j=1}^{J} \delta_j m_j - (\beta_{12} - \beta_s) m_1 - (\beta_s - \beta_b)\left(\sum_{j=3}^{J} m_j\right) m_1. \\ &= \sum_{j=1}^{J} \delta_j m_j + \beta_b \left(\sum_{j=3}^{J} m_j\right) m_1 - \beta_{12} m_1 - \left(\sum_{j=3}^{J} \beta_s m_1 m_j - \beta_s m_1\right).\end{aligned}$$

[25]This case may also involve negative externalities within sides due to competition between sellers of the same good at the mall, but this can be dealt along the line explained in the next section.

Then in this case we see that a divide-and-conquer strategy should subsidy store 2, and set a smaller subsidy then before for consumers. Compared to above, the profit increases by $2\beta_s - \beta_b - \beta_{12}$, which may be positive if β_{12} is close to β_s, but negative if β_{12} is large. Thus the fact that one store can be used as a leading store attracting consumers may not be an advantage if platforms compete by subsidizing the participation of this store.

4.3 Price discrimination and horizontal differentiation in a two-sided market

Consider a two-sided market with sides b and s, and a population of 1 on each side. Side b values the participation of the other side at β_b and the other at β_s. Suppose that on each side the population is decomposed in two groups of equal size. The only difference between members of each group on side j is the valuation of goods. One group attaches a value u_j to platform S and 0 to the other, while the other attaches a value of 0 platform S and u_j to platform W.

Then we have

$$\begin{aligned}\bar{\Pi}^S &= \frac{u_b}{2} - \frac{u_b}{2} + \frac{u_s}{2} - \frac{u_s}{2} - |\beta_b - \beta_s| \\ &= -|\beta_b - \beta_s|.\end{aligned}$$

The strong platform cannot profitably sell to all customers. Suppose that it serves only half of the market on each side, then the bound on profit is

$$\frac{u_b + u_s}{2} - \frac{|\beta_b - \beta_s|}{4} - \frac{|\beta_b + \beta_s|}{4} = \frac{u_b + u_s - \max(\beta_b, \beta_s)}{2}.$$

Since by assumption $u_j > \beta_j$, this profit is positive. One can easily verify in this case that this profit is achievable. In this case thus the market will not tip, but rather segment.

5 Two-sided markets

In order to obtain more specific results on market shares and profits, let us focus on a two-sided market: $J = 2$. We maintain the sequential timing but we will extend the results to a simultaneous pricing in the next section and show that the conclusions are robust.

5.1 Sequential timing

If platform S covers the market, its profit is:

$$\bar{\Pi}^S = \delta_1 m_2 + \delta_2 m_2 - |\beta_{21} - \beta_{12}| \, m_1 m_2.$$

The question here is whether S prefers to cover the market, or to let W sell to some side.

To fix ideas, suppose that S gives up on side 2 and serves only side 1. From corollary (1), S cannot expect more than $\delta_1 m_1 - \beta_{12} m_1 m_2$. But this bound may not be attainable. The reason is that corollary (1) didn't consider all possible strategies that W can use. Indeed, to sell to side 1, platform W could just undercut (by an amount $\delta_1 - \beta_{12} m_2$) the price p_1^S charged by platform S to this side keeping the same price for side 2, which yields the above bound on S's profit. But an alternative strategy for platform W is to exploit the bandwagon effect associated to the success of attracting side 1. To do so, platform W increases the price it charges to side 2 by an amount equal to the value ($\beta_{21} m_1$) attached by side 2 to the participation of side 1. Obviously this raises again the issue of coordination failure, as at the new price side 2 could prefer to join platform S. Platform W must then secure side 1 by undercutting the price set by platform S by a much larger amount than before, indeed by an amount $\delta_1 + \beta_{12} m_2$. If side 2 values the externality much more than side 1, this strategy is the most profitable for W, and prevent it requires that S reduced its price below the level of corollary (1). As a result we obtain:

Lemma 4 *When $J = 2$ and S sells to side j only, its profit is $\Pi_j^S = \delta_j m_j - \max\{\beta_{jl}, \beta_{lj} - \beta_{jl}\} m_l m_j$.*

Thus, if $\beta_{21} > 2\beta_{12}$, irrespective of whether S covers the whole market or sell to 1 only, the main limitation on its strategy is always the indirect profit that W can obtain on side 2 by attracting side 1 with a subsidy.

Intuition suggests that, when sharing the market and selling only to side j, S should leave to W the opportunity to gain a high profit on side l alone and thus set a high price p_l^S. The proof of the lemma shows that the optimal price on side l is then $p_l^S = u_l^S + \beta_{lj} m_j$: it lets W capturing its maximal profit $u_l^W m_l$, with a price $p_l^W = u_l^W$, while maintaining the willingness of side 2 to join S along with side 1. An interpretation is that S induces cooperation by W with some type of "stick and carrot" strategy which can be stated as: "you can serve one side with high profit, but don't try to be aggressive on the other side, or I will take it back".

Let us now turn to the equilibrium analysis, which amounts to the comparison of the three levels of profit, $\bar{\Pi}^S$, Π_1^S and Π_2^S. The equilibrium is then as follows:

Proposition 7 *Assume that there are two sides and $\beta_{12} \leq \beta_{21}$, then:*
Both sides join S if

$$\delta_1 m_1 + \delta_2 m_2 \geq (\beta_{21} - \beta_{12})\, m_1 m_2,$$
$$\delta_1 \geq -\beta_{12} m_2 \text{ and } \delta_2 \geq \inf\{0, \beta_{21} - 2\beta_{12}\} m_1;$$

Side 1 *joins* S *and side* 2 *joins* W *if:*

$$\begin{aligned} \delta_1 &\geq \max\{\beta_{12}, \beta_{21} - \beta_{12}\}m_2, \\ \delta_2 &< \inf\{0, \beta_{21} - 2\beta_{12}\}m_1; \end{aligned}$$

Side 2 *joins* S *and side* 1 *joins* W *if:*

$$\begin{aligned} \delta_1 &< -\beta_{12}m_2 \\ \delta_2 &\geq \beta_{21}m_1. \end{aligned}$$

In the remaining cases, all sides joins W.

The market share of platform S is represented in the space (δ_1, δ_2), for the case $\beta_{12} < \beta_{21} < 2\beta_{12}$ in Figure 1. Straight lines delineate the ranges where platform S covers the market, or sells only to one side, or to none. The dotted lines delineate the same range but for the allocation that maximizes total surplus.

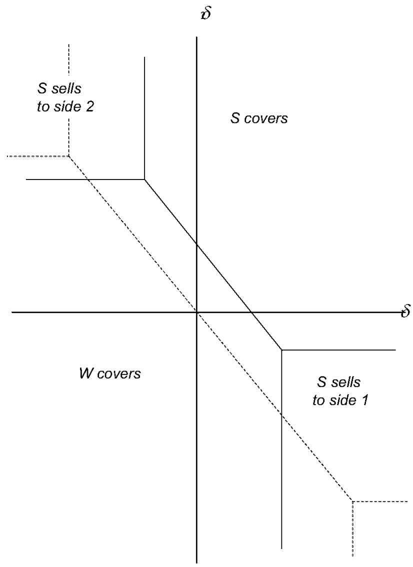


Figure 1

The first point is that S may be unable to sell, even in cases where both δ_1 and δ_2 are positive. Even a platform that offer better qualities and dominates the coordination process in stage 2 may not be able to generate a positive profit on the market. While this is exacerbated by the sequential timing and the second mover advantage of W, we will see that this is deeper and the same conclusion holds under a simultaneous timing.

The second point is that the profit loss due to multi-sided externalities is always higher when S serves only one side than when it serves two sides. As a consequence, if S sells, it will not give up on a side for which it offers a higher intrinsic utility:

$$\text{for } l \neq j,\ \delta_l \geq 0 \Rightarrow \bar{\Pi}^S \geq \Pi_j^S.$$

The motive for abandoning one side is thus grounded into a quality advantage of W. In particular market sharing can only occur when there is a sufficient degree of horizontal differentiation ($\delta_j > 0 > \delta_l$).

5.2 Simultaneous pricing

The Stackelberg profits derived above are upper bounds on the profits that the platforms can achieve in the simultaneous pricing game. With network effects however, these profits may not be reached in a simultaneous pricing game. This section assumes that the population is composed of two sides. We need also to strengthen assumption 1 to

Property $2SM$: $J = 2,\ \beta_{12} < \beta_{21}$,[26] and for all k and $j,\ l : u_j^k > \beta_{jl} m_l$.

Consider an equilibrium where S covers the market. From the analysis of $W's$ best response in the sequential game, $S's$ prices must verify:

$$\begin{aligned} p_j^S &\leq \delta_j + \beta_{jl} m_l, & (7)\\ 0 &\leq p_1^S m_1 + p_2^S m_2 \leq \bar{\Pi}^S; & (8) \end{aligned}$$

along with the condition for the market allocation:

$$p_j^W \geq p_j^S - \delta_j - \beta_{jl} m_l. \tag{9}$$

The question is now whether one can set prices for W in such a way that S doesn't deviate. This requires binding condition 9 on both sides. When S's prices are nonnegative, this is sufficient to obtain an equilibrium. However, some price may need to below cost, and in this case S may be tempted to serve only the other side. One way to look at this issue is to analyze the opportunity cost of not selling to side j served at a loss. S looses the (negative) income $p_j m_j = (\delta_j + \beta_{jl} m_l)\, m_j$, and since side j joins W instead of S, the valuation of platform S by the other side decreases by an amount

[26]Notice that a strict inequality holds. With $\beta_{12} = \beta_{21}$, the results applies by labeling side 1 the side that maximizes $\delta_j m_j$.

equal to the value of the externality $\beta_{lj}m_j$, while the valuation of platform W increases by the same amount. This forces S to decrease the price for side l by twice this amount. The net opportunity cost for S of not selling to side j is thus $(\delta_j + \beta_{jl}m_l + 2\beta_{lj}m_l)\, m_j$. S prefers to cover the market rather than to sell to l only when this opportunity cost is positive.

A similar reasoning allows to define for the case where W covers the market the opportunity cost for W of not selling to side j. The next proposition shows that an equilibrium with one platform covering the market exists provided that the profit is nonnegative and that the opportunity cost of not selling to any of the two sides is nonnegative:

Proposition 8 *Under property $2SM$, there exists an equilibrium with simultaneous pricing where platform k covers the market (with profit $\Pi^k \in [0, \bar{\Pi}^k]$) if and only if $\bar{\Pi}^k \geq 0$ and:*

$$\delta_1 m_1 + (\beta_{12} + 2\beta_{21})\, m_1 m_2 \geq 0 \text{ and } \delta_2 m_2 + (\beta_{21} + 2\beta_{12})\, m_1 m_2 \geq 0 \text{ when } k = S;$$
$$\delta_1 m_1 - (\beta_{12} + 2\beta_{21})\, m_1 m_2 \leq 0 \text{ and } \delta_2 m_2 - \beta_{21} m_1 m_2 \leq 0 \text{ when } k = W.$$

The analysis of market sharing situations follows the same lines. Suppose that S serves side j and W serves side l. Then it must be profitable for S not to sell to side l. The relevant criteria is again the opportunity cost for S of not selling to side l which must be nonpositive.

For W, a differences arises in the opportunity cost of not selling to side 1, due to differences in strategies when W serves side 2 and when W covers the market.

Proposition 9 *Under property $2SM$, there exists an equilibrium with simultaneous pricing where platform S sells to side j with profit $\Pi^S \geq 0$ and platform W sells to side l with profit $\Pi^W \geq 0$ if and only if:*

$$\delta_1 m_1 + (\beta_{12} - 2\beta_{21})\, m_2 m_1 \geq \Pi^S \text{ and } -\delta_2 m_2 - (\beta_{21} + 2\beta_{12})\, m_1 m_2 \geq \Pi^W, \text{ if } j = 1,\ l = 2;$$
$$-\delta_1 m_1 - (\beta_{12} + 2\beta_{21})\, m_1 m_2 \geq \Pi^W \text{ and } \delta_2 m_2 - \beta_{21} m_1 m_2 \geq \Pi^S, \text{ if } j = 2,\ l = 1.$$

The global equilibrium configuration is always the same and is depicted on Figure 2. Dotted lines delineate the range of market sharing in the sequential game.

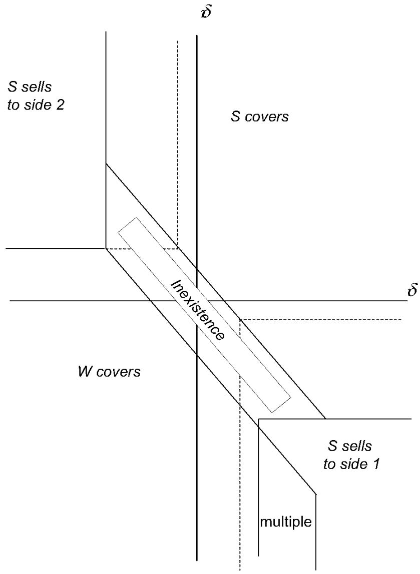


Figure 2

Compared to the sequential game, there is less market sharing. In particular S covers more often. Notice also that S obtains less than its Stackelberg profit when it sells to side 1 only, and when it covers the market while it were sharing the market as a leader.

5.3 Discussion

5.3.1 Welfare

Let us compare the equilibrium configuration with the efficient allocation in the two-sided case. Market sharing maximizes total surplus whenever $\delta_1\delta_2 < 0$ and $(\beta_{12} + \beta_{21})m_1m_2 < \min\{|\delta_1|\, m_1, |\delta_2|\, m_2\}$, which corresponds to a pattern of strong horizontal differentiation with small network effects. Otherwise platform S should cover the market if $\delta_1 m_1 + \delta_2 m_2 > 0$, while platform 2 should cover the market if $\delta_1 m_1 + \delta_2 m_2 < 0$. The dotted lines in Figures 1 show the delimitations of the various range of quality differentials. There can be excessive sales by S which can take the form of inefficient market covering by S or inefficient market sharing.

More surprisingly, given the fact that S benefits from favorable expectations, there is also the possibility of excessive sales by the weak platform W both in the sequential game and in the simultaneous game. In the sequential case, it occurs when it would be optimal that platform S covers the market, and can take two forms. First, S may not sell at all, when both quality differentials are small and network effects are asymmetric. Price discrimination may allow W to generate a profit that outweigh the quality differentials in favor of an efficient leader. Second S may prefer to let its competitor sell to one side, that it would be efficient to add to $S's$ platform, when this weakens the competition on the other side.

To illustrate this effect, set $m_1 = m_2 = 1$, and suppose that side 1 values only good S but not network effects or good W: $u_1^W = 0$, $\beta_{12} = 0$. Suppose that side 2 has no value for good S: $u_2^S = 0$. Assuming that $u_1^S > \beta_{21} > u_2^W$, the efficient allocation has both sides at platform S. However the optimal strategy for S is to give up on side 2 by setting a price β_{21} on side 2. This allows to charge $u_1^S - \beta_{21}$ to side 1, while W charges u_2^W to side 2. This conclusion holds despite the fact that side 2 would be willing to pay a larger amount (β_{21}) to join S than the price charged by W.

To see why this is the case, consider what occurs if platform S lower its price p_2^S to zero. W can not sell to side 2 alone at a positive price and has no other option than to attract both sides or none. But with a divide-and-conquer strategy subsidizing side 1, W can generate profit: $\left(p_1^S - u_1^S\right) + \left(u_2^W + \beta_{21}\right)$. To avoid that S would need to reduce p_1^S below $u_1^S - \beta_{21} - u_2^W$. This means reducing the price on side 1 by an amount u_2^W. Thus even at a price that would generate almost no profit on side 2, S would have to reduce its price p_1^S by a finite amount compared to the level it can charge with market sharing.

In both cases, market sharing or market covering by S, selling to side 1 allows W to create and capture some extra value on side 2. Thus, despite $u_1^W = 0$, side 1 has some value for W, who is willing to compete for it. But in the market sharing case, this value is just the externality for side 2, while when S covers the market, it is the sum of the externality and the intrinsic value, since both come together. By giving away side 2, S allows its competitor to extract the intrinsic value without having to fight for the other side, reducing the value attached by W to the participation of side 1 to its platform. Thus S weakens competition.[27]

In the simultaneous price game, there is also the possibility that W covers the market while it should be shared. This occurs in the range where there are multiple equilibria.

---

[27] As it will appear, this also arises in the context of the simultaneous pricing game.

5.3.2 Multi-sided network effects and market shares

Compared with the case where there is no externality (where S serves side j if $\delta_j \geq 0$), the range of parameters for which the market is shared between the two platforms is reduced. Without surprise, network effects tends to generate some tipping, extending the range of market covering by one platform. But this can translate into an increase or a decrease of the strong platform's market share. When S has a large quality differential on some side (δ_j high) and a rather small disadvantage on the other (δ_l negative but small in absolute value), S will cover the market, in order to secure side j. On the other hand, with strong horizontal differentiation, S is not able to obtain positive profits and leaves the whole market to W.

It is also fairly easy to see in the two-sided market case that when the value attached to the externalities by one side moves closer to the value attached by the other side, the strong leader's profit raises (weakly) and S is more likely to serve both sides. Thus, network effects are the more detrimental to S's profits and market shares, the more asymmetric they are.

5.3.3 Strategic degradation of quality for targeted customers

One of the general principle that emerges is that head-to-head competition dissipates profit. One way to escape from such situation of intense competition is to achieve enough horizontal differentiation. In the present model, this means shifting from a market covering equilibrium to a more peaceful market sharing situation. When a platform controls the quality of the good at the individual level, it can reach this objective by degrading quality for some customers.[28]

The idea relies on the fact that with simultaneous pricing, S may sell more than when it is a leader. This means that there are cases where S would prefer to share the market, but cannot because at any price that allows W to sell, S has the ability to attract the whole market. Degrading the quality on one side is one way to induce the more profitable market sharing situation.

To illustrate this phenomenon, suppose that S covers the market in equilibrium with maximal profit $\bar{\Pi}^S = (\delta_1 + \beta_{12} m_2) m_1 + (\delta_2 - \beta_{21} m_1)\, m_2 > 0$. When $\delta_1 + \beta_{12} m_2 < 0$ (which is compatible with equilibrium conditions), S sells to side 1 at a loss to protect its market. Consider what happens if S can costlessly and publicly reduce u_1^S before the price game (holding u_2^S constant), and does so up to a point where the new quality differential falls short of $-(\beta_{12} + 2\beta_{21}) m_2$. Then the new equilibrium involves market sharing: S sells to side 2 only with profit $(\delta_2 - \beta_{21} m_1)\, m_2 > \bar{\Pi}^S$. It is thus

[28] Both the motive for quality reduction and the way it is achieved differ substantially from models of damaged goods and screening as Denekere and McAfee (1996) and Hahn (2000).

profitable to do so.

The point here is that S would like to commit not to compete on side 1, as an alternative to being 'forced' to include side 1 in its platform despite a competitive hedge in favor of W on this side. A targeted degradation of quality is one way to achieve such a commitment (a "puppy dog" strategy). The same phenomenon may hold for W as well.

More generally when a platform can choose the technology and affect perceived qualities, and when it can't gain a large quality advantage on both sides, it will have incentives to shift its technological choices toward the preferred technology of one side and the least preferred technology of the other side. This may generate inefficiencies in technological choices and even result in the choice of a dominated technology.

5.3.4 Inexistence of a pure strategy equilibrium

With simultaneous pricing, a pure strategy equilibrium fails to exist when the platforms offer goods of similar characteristics. The natural question is whether this due to the assumption 2 (favorable expectations).

Clearly, relaxing assumption 2 allows to extend the set of equilibria.[29] But, although the range of existence increases when all market allocations are considered (at the cost of more multiplicity), the inexistence problem remains if platforms are not differentiated enough and there is some asymmetry in externalities. To see that, notice that in any pure strategy equilibrium of a simultaneous pricing game, the profit of a platform is bounded above by the Stackelberg profit it can derive when it is "strong". Otherwise there would be a strategy of the rival that would raises its profits by extending it market share. But if the differences between the intrinsic values of the goods are small, no platform can profitably sell as a Stackelberg leader. In particular, with two sides, a pure strategy equilibrium fails to exist if $\sum_j \left|u_j^S - u_j^W\right| m_j < \min\{\beta_{12}, \beta_{21} - \beta_{12}\} m_1 m_2$. The inexistence issue thus appears to be a robust phenomenon.

5.3.5 Large network effects

In the paper, externalities are assumed to be small compared to intrinsic values (assumption 1). For pure multi-sided services, the only source of value is the presence of other users, and assumption 1 doesn't hold. A two-sided market where $u_j^W = 0$ is considered in Caillaud and Jullien (2001 and 2003). The key difference with strong externalities is that S may sell to side j at a price p_j^S above the value of the good u_j^S, because the value of externalities can outweigh the intrinsic values. Then S can generate an extra profit $p_j^S - u_j^S > 0$ on side j that platform W will be unable to compete

[29] To generate equilibria in this case, it is sufficient to assume that when platform k deviates, consumers coordinate on $\mathcal{D}^{-k}(P^{-k}, P^k)$.

away with a divide-and-conquer strategy. The reason is that if W convinces side l to join with a low price, side j stops to join S. The largest price that W can set for side j is then $u_j^W + \beta_{jl}m_l$, independent of p_j^S and smaller than $p_j^S - \delta_j + \beta_{jl}m_l$. Platform W's ability to compete is thus reduced. This effect allows S to extract part of the value of multi-sided externalities without increasing the profit that W can generate by stealing $S's$ clients. The dominant platform's profit may then exceed $\sum_j \delta_j m_j$.

6 Platform interconnection

A natural question that the analysis allows to address is whether platforms would prefer to sell "compatible" services. This can be achieved when, while selling differentiated goods, platforms interconnect and allow clients to interact with members of the both platforms. An example of this strategy is provided by the alliances between European financial market places. Let us compare the equilibrium profits with the profits of platforms interconnected under a bill and keep agreement, where no payment between platforms is involved.

Users of interconnected platforms benefit from externalities with all users, irrespective of their choice. For any individual, the comparison between the two platforms thus reduces to the comparison between the goods', thus of $u_j^S - p_j^S$ and $u_j^W - p_j^W$. The competition game reduces to a standard Bertrand type game, where each side constitutes a specific market. Firm S sells to side j if and only if $u_j^S \geq u_j^W$. The profit that the interconnected platform S is then $\sum_{j=1}^{J} \max\{\delta_j, 0\} m_j$.

When platform S is not interconnected, multi-sided effects force it to reduce its prices to prevent divide-and-conquer strategies by W. Moreover, S may have to sell to some sides despite the fact that $\delta_j < 0$, because there is an opportunity cost of letting them joining the competitor. In other words the traditional incentive to reach a large population when there are network externalities is preserved, although platform S is not able to appropriate the efficiency gains associated with these externalities. Choosing to interconnect may then be one way to avoid this problem and to keep only profitable customers.

Proposition 10 *The profit of S is larger when platforms are interconnected.*

Given that the bounds on the $S's$ profits are derived under the most favorable conditions for S, this conclusion holds for any choice of the market allocation of users and doesn't depend on assumption 2. When platforms are not interconnected, the ability to exploit multi-sided effects strengthens competition and harms profits. The platform has to bear the cost of this exacerbation of competition, and prefers to be cooperative.

Although profits in the case where platforms set prices simultaneously don't always coincide with profits of the Stackelberg game, it is fairly easy to see from the proofs that the bounds on profits derived in corollary (1) are valid for this case as well. It follows that the result extends to simultaneous pricing.

Corollary 3 *Assume that platforms set prices simultaneously. The profit of each platform in a pure strategy equilibrium is larger when they interconnect.*

This conclusion doesn't rely on assumption 2 (favorable expectations) since the bounds on profits are valid for any equilibrium of the simultaneous game. It does rely however on the absence of within-sides network effects. Indeed positive network effects between members of the same side may refrain S from interconnecting with W, as it can obtain an extra profit on each side served, that vanishes under interconnection (see section 2.3). For instance the profit that S can obtain by cornering the market is now augmented by the total value of within sides network effects:

$$\bar{\Pi}^S = \sum_{j=1}^{J} \delta_j m_j + \sum_{j=1}^{J} \beta_{jj} m_j^2 - \Omega(\mathcal{J}),$$

which may be above the interconnected profit.

Installed bases may also change the conclusion. Having a large installed base confers an advantage that vanishes if platforms are interconnected. Thus the platform must weight the benefits it can derive from the installed base, and the cost due to more competitive pressure.

7 Conclusion

The paper derives striking results that deserve to be extended in several directions.

For one thing, sides were supposed to be homogeneous, which favours tipping. Extension of multi-sided markets to heterogenous sides may help to understand market sharing configurations and differentiation strategies.

The divide and conquer strategies may require selling below cost, which may be very risky in environments with demand uncertainty, since the platform may fail to recover the subsidy on others sides. The strategies may also require negative prices and these may be hard to target. They may attract clients who just grab the subsidy and are not interested by the platform's services. Moreover these clients may generate negative externalities on other members of the platform. The strategy would then be to use in-kind subsidies targeted at attracting only active users of the platform, but these are usually costly while the model assumes no social cost associated with subsidies.

The analysis should be extended to account for dynamic consideration. Dynamics may allow to identify more precisely the pattern of cross-subsidy since sides may have to join a platform at different dates. A dynamic divide and conquer strategy may require to run negative cash-flows for sometime, and thus need financial resources. With an imperfect capital market, this means that platforms having access to a deep-pocket should have a strong advantage over financially constrained platform. The analysis of platform competition should then devote special care to financial aspects.

References

[1] Ambrus A. and R. Argenziano (2005): "Asymmetric Networks in Two-Sided Markets", mimeo, Harvard University.

[2] Anderson S. and Coate S. (2005): "Market Provision of Broadcasting: A Welfare Analysis", *Review of Economic Studies,* 72, 947-972.

[3] Argenziano R. (2005): "Differentiated Networks: Equilibrium and Efficiency", mimeo, university of Essex.

[4] Armstrong M. (2002): "The Theory of Access Pricing and Interconnection", in Cave, Majumdar and Volgelsang eds. *Handbook of Telecommunication Economics,* North-Holland.

[5] Armstrong M. (2005): "Competition in Two-Sided Markets", forthcoming *Rand Journal of Economics.*

[6] Arthur B. (1989): "Competing Technologies, Increasing Returns and Lock-in by Historical Events", *Economic Journal* 99, 199-214.

[7] Bensaid B. and J.P. Lesne (1996): "Dynamic Monopoly Pricing with Network Externalities", *International Journal of Industrial Organization.* 14, 837–855.

[8] Cabral L., Salant D. and Woroch G. (1999): "Monopoly Pricing with Network Externalities", *International Journal of Industrial Organization.* 17, 199-214.

[9] Caillaud B. and B. Jullien (2001): "Competing Cybermediaries", *European Economic Review* (Papers & Proceedings) 45, 797-808.

[10] Caillaud B. and B. Jullien (2003): "Chicken & Egg: competition among intermediation service providers", *Rand Journal of Economics,* 34, 309-328.

[11] Crémer J., Rey P. and J. Tirole (2000): "Connectivity in the Commercial Internet", *Journal of Industrial Economics*, 48, 433-472.

[12] Damiano E. and H. Li (2005a): "Price Discrimination and Efficient Matching", mimeo, university of Toronto.

[13] Damiano E. and H. Li (2005b): "Competing Matchmakers", mimeo, university of Toronto.

[14] Denekere R. and P. McAfee (1996): "Damaged Goods", *Journal of Economics and Management Strategy* 5(2), 149-174.

[15] David P. (1985): "CLIO and the Economics of QWERTY", *American Economic Review Papers and Proceedings* 75, 322-337.

[16] Economides N. (1996): "The Economics of Network", *International Journal of Industrial Organization.* 14, 673-700.

[17] Farrell J. and G. Saloner (1985): "Standardization, Compatibility and Innovation", *Rand Journal of Economics* **16,** 70-83.

[18] Ferrando J., Gabsewitcz J., Laussel D. and Sonnac N. (2003): "Two-Sided Network Effect and Competition: an Application to Media Industries", mimeo, CREST-LEI.

[19] Fudenberg D. and J. Tirole (2000): "Pricing a Network Good to Deter Entry", *Journal of Industrial Economics,* 48(4):373-390.

[20] Hagiu A (2004): "Optimal Pricing and Commitment in Two-Sided Markets", mimeo, Princeton University.

[21] Innes R. and R. Sexton (1993): "Customer Coalitions, Monopoly Price Discrimination, and Generic Entry Deterrence", *European Economic Review* **37,** 1569-1597.

[22] Jullien B. (2000): "Competing in Network Industries: Divide and Conquer", IDEI Working Paper.

[23] Katz M. and K.Shapiro (1985): "Network Externalities, Competition and Compatibility", *American Economic Review* **75,** 424-440.

[24] Katz M. and K. Shapiro (1986): "Technology Adoption in the Presence of Network Externalities", *Journal of Political Economy* 94, 822-841.

[25] Katz M. and K. Shapiro (1992): "Product Innovation with Network Externalities", *Journal of Industrial Economics.*

[26] Katz M. and K. Shapiro (1994): "System Competition and Network Effects", *Journal of Economic Perspective* **8,** 93-115.

[27] Rochet J.C. and J. Tirole, (2002): "Cooperation Among Competitors: Some Economics of Payment Card Associations." *Rand Journal of Economics*, Vol. 33 , 1–22.

[28] Rochet J.C. and J. Tirole (2003): "Platform Competition in Two-Sided Markets", *Journal of the European Economic Association 1,* 990-1029.

[29] Rochet J.C. and Tirole, J. (2006): "Two-Sided Market: a Progress Report", forthcoming *Rand Journal of Economics.*

[30] Segal I. (2003): "Coordination and Discrimination in Contracting with Externalities: Divide and Conquer?" *Journal of Economic Theory,* 113(2), 147-181

[31] Topkis D.M. (1979): "Equilibrium Points of Nonzero-Sum n-Person Submodular Games", *Siam Journal of Control and Optimization* **17**(6), 773-787.

[32] Vives X. (1990): "Nash Equilibrium with Strategic Complementarities", *Journal of Mathematical Economics* **19(**3**),** 305-321.

[33] Wright J. (2002): "Access Pricing under Competition: an Application to Cellular Networks", *Journal of Industrial Economics,* 250, 289-315.

[34] Wright J. (2003): "Optimal Card Payment System", *European Economic Review* 47, 587-617.

A Appendix to section 2

Proof of lemma 1. Consider two $REA : i = 1, 2$, where sides allocate according to ${}^i n_j^k \in [0, m_j]$ and let $\mathcal{K}_i$ be the set of side with ${}^i n_j^k > 0$, and $\mathcal{S}_i$ the set of sides with ${}^i n_j^s = m_j$. Then

$$\begin{aligned}
j &\in \mathcal{K}_i \Rightarrow u_j^k + \sum_{l\in\mathcal{K}_i} \beta_{jl} \left({}^i n_l^k\right) - p_j^k \geq \max\{u_j^s + \sum_{l\notin\mathcal{S}_i} \beta_{jl} \left({}^i n_l^k\right) + \sum_{l\in\mathcal{S}_i} \beta_{jl} m_l - p_j^s, 0\} \\
&\Rightarrow u_j^k + \sum_{l\in\mathcal{K}_i} \beta_{jl} m_l - p_j^k \geq \max\{u_j^s + \sum_{l\in\mathcal{S}_i} \beta_{jl} m_l - p_j^s, 0\} \\
j &\in \mathcal{J}\backslash(\mathcal{S}_i \cup \mathcal{K}_i) \Rightarrow 0 \geq \max\left\{u_j^k + \sum_{l\in\mathcal{K}_i} \beta_{jl} \left({}^i n_l^k\right) - p_j^k, u_j^s + \sum_{l\notin\mathcal{S}_i} \beta_{jl} \left({}^i n_l^k\right) + \sum_{l\in\mathcal{S}_i} \beta_{jl} m_l - p_j^s\right\} \\
&\Rightarrow \max\left\{u_j^k + \sum_{l\in\mathcal{K}_i} \beta_{jl}^i m_l - p_j^k, 0\right\} \geq u_j^s + \sum_{l\in\mathcal{S}_i} \beta_{jl} m_l - p_j^s
\end{aligned}$$

Now suppose that we impose that no member of in $\mathcal{J}\backslash(\mathcal{S}_1 \cap \mathcal{S}_2)$ join s, and that all members of sides in $\mathcal{K}_1 \cup \mathcal{K}_2$ join k. Since for the sides in $\mathcal{J}\backslash(\mathcal{S}_1 \cap \mathcal{S}_2)$:

$$\max\{u_j^k + \sum_{l\in\mathcal{K}_1\cup\mathcal{K}_2} \beta_{jl} m_l - p_j^k, 0\} \geq u_j^s + \sum_{l\in\mathcal{S}_1\cap\mathcal{S}_2} \beta_{jl} m_l - p_j^s,$$

the minimal benefit that a member of a side in $\mathcal{J}\backslash\mathcal{S}_1 \cap \mathcal{S}_2$ can obtain when not joining s is larger than the maximal benefit it can gain when joining s. Therefore the optimal strategy is either to join k or not to join at all. Moreover for a user in $\mathcal{K}_1 \cup \mathcal{K}_2$, $u_j^k + \sum_{l\in\mathcal{K}_1\cup\mathcal{K}_2} \beta_{jl} m_l - p_j^k \geq 0$, implying that the optimal strategy for this user is indeed to join k. Now take an equilibrium of the game with strategy spaces restricted as described. This is an equilibrium of the allocation game with the property that $n_l^k \in \{0, 1\}$, $\mathcal{K}_1 \cup \mathcal{K}_2 \subset \mathcal{K}$ and $\mathcal{S} \subset \mathcal{S}_1 \cap \mathcal{S}_2$.

Taking a maximal element completes the proof. ■

Proof of lemma 2. Notice that for maximal inclusiveness implies that no side split between the two platforms. Then, maximal inclusiveness implies that $\mathcal{D}^S(P^S, P^W)$ necessarily verifies the following conditions

$$\begin{aligned}
j &\in \mathcal{K}\ iff: \delta_j + \beta_{jj} m_j + \sum_{l\in\mathcal{K}\backslash j} \beta_{jl} m_l - p_j^S \geq \max\left\{\sum_{l\in\mathcal{S}\backslash j} \beta_{jl} m_l - p_j^W, 0\right\}, \\
j &\in \mathcal{S}\ iff: \sum_{l\in\mathcal{S}\backslash j} \beta_{jl} m_l - p_j^W > \max\left\{\delta_j + \beta_{jj} m_j + \sum_{l\in\mathcal{K}\backslash j} \beta_{jl} m_l - p_j^S, 0\right\}, \\
j &\in \mathcal{J}\backslash(\mathcal{S}_i \cup \mathcal{K}_i)\ otherwise.
\end{aligned}$$

where the second inequality comes from the fact that joining platform j cannot generate more profit for an individual than when the whole side joins S or no platform. Notice that any allocation that verifies these conditions is a REA. We can thus define $\mathcal{D}^S(P^S, P^W)$ by taking a maximal element in the set of $REAs$ that verifies these conditions. ■

B Appendix to section 3

Lemma 5 *Given prices P^S such that $p_j^S \leq u_j^S$, the maximal profit that W can obtain is smaller or equal to Π iff:*

$$\forall \mathcal{K} \subset \mathcal{J}, \sum_{j \in \mathcal{K}} p_j^S m_j \leq \sum_{j \in \mathcal{K}} \delta_j m_j - \Omega(\mathcal{K}) + B(\mathcal{K}) + \Pi. \tag{10}$$

Proof. It is true if $J = 1$. Suppose the proposition holds up to $\#\mathcal{J} = J - 1$ and consider $\#\mathcal{J} = J$. Suppose first that W decides to sell to only $J - 1$ sides, leaving outside one side (say $j = J$) by setting p_{2J} high. Given that side J stay with S, the problem of attracting $K < J$ sides within J sides is the same as with only $J - 1$ sides but with a utility for side j equal to $u_j^S + \beta_{jJ} m_J + \sum_{l=1}^{J-1} \beta_{jl} n_l^S - p_j^S$ when joining S. Condition 10 for $K < J$ then ensures that attracting less than J side can't yield more than Π.

Consider now attracting all the sides. For at least one of the side the price p_j^W must be smaller than $p_j^S - (\delta_1 + \sum_{l \neq j} \beta_{jl} m_l)$. Let us say it is side 1. We then set $p_1^W = p_1^S - \left(\delta_1 + \sum_{l>1} \beta_{1l} m_l\right)$ and side 1 joins W for sure. Here the problem is reduced to the $J - 1$ sides $j \geq 2$, but now with utility $u_j^S + \sum_{l=2}^J \beta_{jl} n_l^S - p_j^S$ at platform S, and utility $u_j^W + \sum_{l=2}^J \beta_{jl} n_l^W + \beta_{j1} m_1 - p_j^W$ at W. The total profit on the $J - 1$ sides is less than $\Pi - (p_1^S - (\delta_1 + \sum \beta_{1l} m_l)) m_1$, which is precisely condition (10) for $\sigma(1) = 1$. Therefore W can't get more than Π.

From lemma 5, S serves all the market if and only if

$$for\ all\ \mathcal{K} : \sum_{j \in \mathcal{K}} p_j^S m_j \leq \sum_{j \in \mathcal{K}} \delta_j m_j - \Omega(\mathcal{K}) + B(\mathcal{K}). \tag{11}$$

W.l.o.g. let us assume that $\Omega(\mathcal{J})$ is obtained for the order $1, 2, ...J$. Suppose that S sets prices

$$p_j^S = \delta_j - \sum_{l<j} \beta_{jl} m_l + \sum_{l>j} \beta_{jl} m_l. \tag{12}$$

Then $\sum_{j \in \mathcal{J}} p_j^S m_j = \sum_{j \in \mathcal{J}} \delta_j m_j - \Omega(\mathcal{J})$. It remains to show that at these prices, the constraints (11) are verified.

Fix a subset $\mathcal{K}$ and let σ be such that $\Omega(\mathcal{K}) = \sum_{j\in\mathcal{K}} \sum_{\substack{l\in\mathcal{K} \\ \sigma(l)>\sigma(j)}} (\beta_{lj} - \beta_{jl})m_l m_j$. Suppose that $\sum_{j\in\mathcal{K}} p_j^S m_j > \sum_{j\in\mathcal{K}} \delta_j m_j - \Omega(\mathcal{K}) + B(\mathcal{K})$ and let us show that it leads to contradiction. The condition writes

$$\sum_{j\in\mathcal{K}} \left(\sum_{l<j} \beta_{jl} m_l - \sum_{l>j} \beta_{jl} m_l \right) m_j < \sum_{j\in\mathcal{K}} \sum_{\substack{l\in\mathcal{K} \\ \sigma(l)>\sigma(j)}} (\beta_{lj} - \beta_{jl}) m_l m_j - \sum_{j\in\mathcal{K}} \sum_{l\notin\mathcal{K}} \beta_{jl} m_l m_j$$

Define $\hat{\sigma}$ as the order $\mathcal{K}$:

$$\begin{aligned} \hat{\sigma}(j) &< \hat{\sigma}(l)\ Sf\ j \in \mathcal{K}\ anW\ l \notin \mathcal{K}, \\ \hat{\sigma}(j) &< \hat{\sigma}(l) \Longleftrightarrow \sigma(j) < \sigma(l)\ Sf\ j \in \mathcal{K}\ anW\ l \in \mathcal{K}, \\ \hat{\sigma}(j) &< \hat{\sigma}(l) \Longleftrightarrow j < l\ Sf\ j \notin \mathcal{K}\ anW\ l \notin \mathcal{K}. \end{aligned}$$

Then tedious computation shows that

$$\sum_{j\in\mathcal{J}} \sum_{l>j} (\beta_{lj} - \beta_{jl}) m_l m_j < \sum_{j\in\mathcal{J}} \sum_{\hat{\sigma}(l)>\hat{\sigma}(j)} \left(\beta_{lj} - \beta_{jl}\right) m_l m_j - 2 \sum_{j\in\mathcal{K}} \sum_{\substack{l\notin\mathcal{K} \\ l<j}} \beta_{lj} m_j m_l$$

But this implies that $\Omega(\mathcal{J}) < \sum_{j\in\mathcal{J}} \sum_{\hat{\sigma}(l)>\hat{\sigma}(j)} \left(\beta_{lj} - \beta_{jl}\right) m_l m_j$, a contradiction. Therefore, it must be the case that (11) holds. ∎

Proof of lemma 3. Denote $\Omega^{\sigma}(\mathcal{J}) = \sum_{j\in\mathcal{J}} \left(\sum_{\sigma(l)>\sigma(j)} (\beta_{lj} - \beta_{jl}) m_l m_j \right)$. For any order σ on $\mathcal{J}$ there is an exact reverse ordering $\hat{\sigma}$ and it verifies $\Omega^{\sigma}(\mathcal{J}) = -\Omega^{\hat{\sigma}}(\mathcal{J})$. Hence the maximum over all permutations is non-negative. The maximum is zero iff $\Omega^{\sigma}(\mathcal{J}) = 0$ for all σ. Suppose this is the case. Fix j, l. Consider a permutation σ such that $\sigma(j) = J - 1$ and $\sigma(l) = J$, and $\bar{\sigma}$ that coincides with σ except that $\bar{\sigma}(j) = J$ and $\bar{\sigma}(l) = J - 1$. Then $0 = \Omega^{\sigma}(\mathcal{J}) = \Omega^{\hat{\sigma}}(\mathcal{J}) + 2(\beta_{lj} - \beta_{jl}) m_j m_l$ which implies that $\beta_{jl} = \beta_{lj}$. The same proof holds for $\mathcal{K} \subset \mathcal{J}$. ∎

Proof of corollary 1 . Let us fix the prices of S at their equilibrium values. Suppose that W set prices for sides $j \notin \mathcal{K}$ at their equilibrium values minus an arbitrarily small amount, and attempts to design prices for sides $j \in \mathcal{K}$ so as to attract some of them. In doing so W can't loose its equilibrium customers (sides $j \notin \mathcal{K}$) because our criterion for the allocation of users ensures that bringing new customers to W (which reduces the attractiveness of S) can't induce W' customers to change their behavior and join S (otherwise they would have done so in the equilibrium configuration). We can then analyze $W's$ strategy as above, taking into account the fact that sides outside $\mathcal{K}$ join W. The resulting reduced game is thus the same game as above played on the sides $j \in \mathcal{K}$ only, but where the intrinsic value u_j^W is replaced by $u_j^W + \sum_{l\notin\mathcal{K}} \beta_{jl} m_l$. An upper bound on $S's$ profit is thus its maximal profit when its covers a market consisting of only sides within $\mathcal{K}$, with the new "stand-alone" value differential defined as $\delta_j - \sum_{l\notin\mathcal{K}} \beta_{jl} m_l$ instead of δ_j. The result of corollary 1 then follows. ∎

Proof of proposition 3. Suppose that S refuse to sell to sides outside $\mathcal{K}$ and sets prices within $\mathcal{K}$:

$$p_j^S = \delta_j - \sum_{l \notin \mathcal{K}} \beta_{jl} m_l.$$

Its profits is $\sum_{j \in \mathcal{K}} \delta_j m_j - B(\mathcal{K})$. Then W obtains (the equilibrium profit selling to sides outside $\mathcal{K}$:

$$\Pi^W = \sum_{j \notin \mathcal{K}} u_j^W m_j + \max_\sigma \sum_{\substack{j,l \notin \mathcal{K} \\ \sigma(l)<\sigma(j)}} \beta_{jl} m_j m_l.$$

In any deviation, W would sell to all sides outside $\mathcal{K}$ because S refuses to sell to them. The analysis of $W's$ strategy is the same as before but replacing $p_j^S - \delta_j$ by u_j^W for sides outside $\mathcal{K}$, accounting for the fact that they would not buy from S. Consider a set $\mathcal{S}$ targeted by W with $\mathcal{K} \subset \mathcal{S}$, and a permutation σ on $\mathcal{S}$. Then W's profit is

$$\begin{aligned}
& \sum_{l \notin \mathcal{K}} u_j^W m_j + \sum_{\substack{j,l \notin \mathcal{K} \\ \sigma(l)<\sigma(j)}} \beta_{jl} m_j m_l + \sum_{j \notin \mathcal{K},} \sum_{\substack{l \in \mathcal{S}\cap\mathcal{K} \\ \sigma(l)<\sigma(j)}} \beta_{jl} m_j m_l + \sum_{j \in \mathcal{S}\cap\mathcal{K}} \left(p_j^S - \delta_j\right) m_j \\
& + \sum_{\substack{j,l \in \mathcal{S}\cap\mathcal{K} \\ \sigma(l)<\sigma(j)}} \left(\beta_{jl} - \beta_{lj}\right) m_j m_l + \sum_{j \in \mathcal{S}\cap\mathcal{K}} \sum_{\substack{l \notin \mathcal{K} \\ \sigma(l)<\sigma(j)}} \beta_{jl} m_j m_l - \sum_{j \in \mathcal{S}\cap\mathcal{K}} \sum_{l \notin \mathcal{S}} \beta_{jl} m_j m_l \\
\leq\ & \Pi^W + \sum_{j \notin \mathcal{K},} \sum_{l \in \mathcal{S}\cap\mathcal{K}} \beta_{jl} m_j m_l - \sum_{j \in \mathcal{S}\cap\mathcal{K}} \sum_{l \notin \mathcal{K}} \beta_{jl} m_j m_l - \sum_{j \in \mathcal{S}\cap\mathcal{K}} \sum_{l \notin \mathcal{S}} \beta_{jl} m_j m_l \\
\leq\ & \Pi^W - \sum_{j \in \mathcal{S}\cap\mathcal{K}} \sum_{l \notin \mathcal{S}} \beta_{jl} m_j m_l.
\end{aligned}$$

where we make use of the fact that pairwise symmetry implies:

$$\sum_{\substack{j \notin \mathcal{K}, l \in \mathcal{S}\cap\mathcal{K} \\ \sigma(l)<\sigma(j)}} \beta_{jl} m_j m_l + \sum_{\substack{j \in \mathcal{S}\cap\mathcal{K}, l \notin \mathcal{K} \\ \sigma(l)<\sigma(j)}} \beta_{jl} m_j m_l = \sum_{j \notin \mathcal{K}, l \in \mathcal{S}\cap\mathcal{K}} \beta_{jl} m_j m_l$$

and

$$\sum_{\substack{j,l \in \mathcal{S}\cap\mathcal{K} \\ \sigma(l)<\sigma(j)}} \left(\beta_{jl} - \beta_{lj}\right) m_j m_l = 0$$

Thus W would not deviate. ■

Proof of corollary 2. The result follows from the comparison of the value of the profit when S covers the market is covered and the bounds on market sharing profits. ■

Proof of proposition 4. In equilibrium S sets the price for $j \in \mathcal{K}$ at

$$p_j^S = u_j^S - \max\{u_j^W + \sum_{l \in \mathcal{L}} \beta_{jl} m_l - p_j^W, 0\} + \sum_{l \in \mathcal{K}} \beta_{jl} m_l$$

which is the maximal price at which a member of side j is willing to join S when sides $l \in \mathcal{K}$ do. If S decides to attract sides in $\mathcal{H} \subset \mathcal{L}$, it can do so by setting prices for all sides in $\mathcal{K} \cup \mathcal{H}$, including those sides it already serves:

$$\hat{p}_j^S = u_j^S - \max\{u_j^W + \sum_{l \in \mathcal{L} \backslash \mathcal{H}} \beta_{jl} m_l - p_j^W, 0\} + \sum_{l \in \mathcal{K} \cup \mathcal{H}} \beta_{jl} m_l$$

The gain in profit is then $\sum_{j \in \mathcal{H}} \hat{p}_j^S m_j + \sum_{j \in \mathcal{K}} (\hat{p}_j^S - p_j^S) m_j$. Notice first that for $j \in \mathcal{K}$, the price differential $\hat{p}_j^S - p_j^S$ is larger than $\sum_{l \in \mathcal{H}} \beta_{jl} m_l m_j$, with equality when $p_j^W \geq u_j^W + \sum_{l \in \mathcal{L}} \beta_{jl} m_l$.. The gain in $S's$ profit is thus minimal when W sets very high prices for these sides, with a lower bound $\sum_{j \in \mathcal{H}} \hat{p}_j^S m_j + \sum_{j \in \mathcal{K}} \sum_{l \in \mathcal{H}} \beta_{jl} m_l m_j$.

If W sells to all sides in $\mathcal{L}$, it sets prices such that for all subset $\mathcal{H}$:

$$\sum_{j \in \mathcal{H}} p_j^W m_j \leq \sum_{j \in \mathcal{H}} (u_j^W - u_j^S) m_j + \sum_{j \in \mathcal{H}} \sum_{l \in \mathcal{L} \backslash \mathcal{H}} \beta_{jl} m_l m_j - \sum_{j \in \mathcal{H}} \sum_{l \in \mathcal{K} \cup \mathcal{H}} \beta_{jl} m_l m_j - \sum_{j \in \mathcal{K}} \sum_{l \in \mathcal{H}} \beta_{jl} m_l m_j$$

An upper bound on the profit is then obtained for $\mathcal{H} = \mathcal{L}$:

$$\sum_{j \in \mathcal{L}} p_j^W m_j \leq \sum_{j \in \mathcal{L}} (u_j^W - u_j^S) m_j - B(\mathcal{L}) - B(\mathcal{K}).$$

If it is a Stackelberg leader, W can obtain this profit by setting a price $p_j^W = u_j^W - u_j^S - \left(\sum_{l=1}^J \beta_{jl} m_l - \sum_{l \in \mathcal{K}} \beta_{lj} m_l \right)$ for all sides within $\mathcal{L}$. ■

C Appendix to section 4

Proof of proposition 5.

$$\begin{aligned}
& \sum_{j=1}^{K} \delta_j - \beta_1 K(J-K) - \left(\sum_{j=1}^{K^S(\beta_2)} \delta_j - \beta_2 K^S(\beta_2)(J - K^S(\beta_2)) \right) \\
= & \sum_{j=1}^{K} \delta_j - \beta_2 K(J-K) - \left(\sum_{j=1}^{K^S(\beta_2)} \delta_j - \beta_2 K^S(\beta_2)(J - K^S(\beta_2)) \right) \\
& + (\beta_2 - \beta_1) \left(K(J-K) - K^S(\beta_1)(J - K^S(\beta_1)) \right) \\
< & \ 0 \ if \ K^S(\beta_1)(J - K^S(\beta_1)) < K(J-K).
\end{aligned}$$

Thus $K^S(\beta_1)(J - K^S(\beta_1)) \geq K^S(\beta_2)(J - K^S(\beta_2))$ which reduces to $\left| K^S(\beta_2) - \frac{J}{2} \right| \geq \left| K^S(\beta_1) - \frac{J}{2} \right|$.

Suppose that in addition the profit is quasi-concave in K at β_2. Then if $K^S(\beta_2) \geq \frac{1}{2}$, the profit at $\beta = \beta_1$ is increasing on $K < 1/2$ and thus $K^S(\beta_1) \geq \frac{J}{2}$ which implies $K^S(\beta_2) \geq K^S(\beta_1)$. The symmetric argument holds for $K^S(\beta_2) \leq \frac{1}{2}$. ■

Proof of proposition 6. The equilibrium size of S maximizes $w(K)+\beta K(J-K)$, where $w(K)=\sum_{j=1}^{K}\delta_j - 2\beta K(J-K)$ denote the welfare gain. Suppose that $K^* \geq 1/2$. Then for $K > K^*$, $\beta K^*(J-K^*) > \beta K(J-K)$ so that $K^S(\beta) \leq K^*$. If in addition $w(K)$ is quasi-concav ∎

D Appendix to section 5

Proof of lemma 4 and proposition 7. Suppose that S sells only to side 1 (the same proof holds for side 2). Imposing w.l.o.g. $p_2^S \leq u_2^S + \beta_{21}m_1$, the equilibrium price of W is $p_2^W = p_2^S - \delta_2 - \beta_{21}m_1$.

W will conform to selling to side 2 only if

$$\begin{aligned} p_2^W m_2 &\geq p_1^S m_1 + p_2^S m_2 - \delta_1 m_1 - \delta_2 m_2 + (\beta_{12} - \beta_{21}) m_1 m_2 \\ p_2^W m_2 &\geq p_1^S m_1 - \delta_1 m_1 + u_2^W m_2 - \max\{u_2^S - p_2^S, 0\} m_2 + (\beta_{21} - \beta_{12}) m_1 m_2 \end{aligned}$$

This reduces to

$$\begin{aligned} \delta_1 m_1 - \beta_{12} m_1 m_2 &\geq p_1^S m_1 \\ \delta_1 m_1 + \max\{u_2^S - p_2^S, 0\} m_2 - u_2^S m_2 + p_2^S m_2 - \beta_{21} m_1 - (\beta_{21} - \beta_{12}) m_1 m_2 &\geq p_1^S m_1 \end{aligned}$$

The LHS of the second condition is maximal at the price $p_2^S = u_2^S + \beta_{22} m_2 + \beta_{21} m_1$, which yields:

$$\begin{aligned} \Pi_1^S &= \min\{\delta_1 m_1 - \beta_{12} m_1 m_2, \delta_1 m_1 - (\beta_{21} - \beta_{12}) m_1 m_2\} \geq p_1^S m_1. \\ p_2^W &= u_2^W. \end{aligned}$$

This completes the proof the lemma. Proposition 7 follows by comparing the profit levels. ∎

Proof of proposition 8. First a platform can't cover the market in equilibrium unless $\bar{\Pi}^k \geq 0$.

Suppose that $\bar{\Pi}^S \geq 0$. Set the prices p_j^S verifying conditions 7 and 8. Then $W's$ best response is not to sell at all provided that $\bar{\Pi}^S \geq \Pi^S$. Equilibrium prices of W must verify $p_j^W = p_j^S - \delta_j - \beta_{jl} m_l$, since otherwise either W would sell or S could raise its price. Thus this is an equilibrium if S prefers to cover the market than to sell to one side only. To sell to side j alone, S must set a price $\hat{p}_j^S$ such that $u_j^S + \beta_{jj} m_j - \hat{p}_j^S \geq u_j^W + \beta_{jl} m_l - p_j^W$, which amounts to $\hat{p}_j^S = p_j^S - 2\beta_{jl} m_l$. Thus an equilibrium must verify $p_1^S m_1 + p_2^S m_2 \geq p_j^S m_j - 2\beta_{jl} m_l m_j$ or : $p_l^S \geq -2\beta_{jl} m_j$. Equilibrium prices then exist if $\delta_j + \beta_{jl} m_l \geq -2\beta_{lj} m_l$ for both sides .

Suppose now that $\bar{\Pi}^W \geq \Pi^W \geq 0$. Choose prices p_j^W verifying conditions **??** and **??**. S can't obtain a positive profit because its profits is the maximum between $\sum p_j^W m_j - \bar{\Pi}^W$, $\left(p_1^W + \delta_1 - \beta_{12} m_2\right) m_1$ and $\left(p_2^W + \delta_2 - \beta_{21} m_1\right) m_2$, and they are all nonpositive. It must then be the case that for one side $p_j^W \leq p_j^S - \delta_j - \beta_{jl} m_l$ and for the other side $p_l^W \leq p_l^S - \delta_l + \beta_{lj} m_j$. Moreover this must hold at equality, since otherwise W could raise its prices. W can't obtain more than Π^W by covering the market if $\sum p_j^S m_j - \bar{\Pi}^S \leq \Pi^W$. The unique prices that are compatible with the fact that W covers the market and obtains at most Π^W are thus $p_1^S = p_1^W + \delta_1 + \beta_{12} m_2$, $p_2^S = p_2^W + \delta_2 - \beta_{21} m_1$. Notice that at price $p_1^S - \varepsilon$, side 1 joins W even alone. So $W's$ options if it doesn't cover the market are to sell to side 1 at p_1^W or to side 2 at $p_2^S - \delta_2 - \beta_{21} m_1 = p_2^W - 2\beta_{21} m_1$. It thus chooses to cover only if

$$\Pi^W \geq \max\left\{p_1^W m_1, p_2^W m_2 - 2\beta_{21} m_1 m_2\right\}$$

or

$$\begin{array}{rcl} p_2^W & \geq & 0 \\ p_1^W & \geq & -2\beta_{21} m_2 \end{array}$$

We can find such prices whenever $-2\beta_{21} m_2 \leq -\delta_1 + \beta_{12} m_2$, $0 \leq -\delta_2 + \beta_{21} m_1$, and $-2\beta_{21} m_1 m_2 \leq \bar{\Pi}^W$. ■

Proof of proposition 9. Assume that S sells to side 1. Each users should prefer to stay with its side rather than to join the other side. This yields

$$\begin{array}{rcl} u_1^S + \beta_{11} m_1 - \max\left\{u_1^W + \beta_{12} m_2 - p_1^W, 0\right\} & \geq & p_1^S \\ u_2^W - \max\left\{u_2^S + \beta_{22} m_2 + \beta_{21} m_1 - p_2^S, 0\right\} & \geq & p_2^W \end{array}$$

Clearly to sustain the equilibrium, the best is to put prices p_1^W and p_2^S at their minimal value given profits, which yields:

$$p_1^S = p_1^W + \delta_1 - \beta_{12} m_2 \text{ and } p_2^W = p_2^S - \delta_2 - \beta_{21} m_1$$

Equilibrium conditions then write for W:

$$\begin{array}{rcl} p_2^W m_2 & \geq & p_1^S m_1 - \delta_1 m_1 - \beta_{12} m_1 m_2 \\ & & +u_2^W m_2 - \max\left\{u_2^S m_2 + \beta_{22} m_2^2 - p_2^S m_2, 0\right\} + \beta_{21} m_1 m_2 \\ p_2^W m_2 & \geq & p_1^S m_1 + p_2^S m_2 - \delta_1 m_1 - \delta_2 m_2 + \beta_{12} m_1 m_2 - \beta_{21} m_1 m_2 \end{array}$$

For S we then get

$$p_1^S m_1 \geq p_2^W m_2 + p_1^W m_1 + \delta_1 m_1 + \delta_2 m_2 + (\beta_{12} + \beta_{21}) m_1 m_2.$$

Overall this yields equilibrium conditions

$$\begin{array}{rcl} \delta_1 + (\beta_{12} - 2\beta_{21}) m_2 & \geq & p_1^S \\ \delta_1 - \beta_{12} m_2 & \geq & p_1^S \\ -\delta_2 - (2\beta_{12} + \beta_{21}) m_1 & \geq & p_2^W \end{array}$$

Given that $\beta_{12} \leq \beta_{21}$, the second constraint is implied by the first which gives the bounds on profits.

If S sells to side 2 the three conditions are the same, reverting the role of the two sides. Given that $\beta_{12} \leq \beta_{21}$, the relevant constraints are now $\delta_2 - \beta_{21} m_1 \geq p_2^S$ and $-\delta_1 - (\beta_{12} + 2\beta_{21}) m_2 \geq p_1^W$. ■

E Appendix to section 6

Proof of proposition 10 and corollary 3. Without interconnection S obtains less than $\max_{\mathcal{K}} \left\{ \sum_{j \in \mathcal{K}} \delta_j m_j \right\} = \sum_{j \in \mathcal{J}} \max \{\delta_j, 0\} m_j$. ■

Do Consumers Buy Less of a Taxed Good?*

Hans Jarle Kind†, Marko Koethenbuerger‡ and Guttorm Schjelderup§

June, 2006

Abstract

This paper shows that consumers may buy more of a taxed good if it is sold by a two-sided platform firm. Two-sided platform industries serve distinct customer groups that are connected through interdependent demand, and include major businesses such as the media industry (newspapers/magazines and advertisers), banking (cardholder and merchant), and the software industry (users and application developers). The paper compares ad valorem and specific taxes and shows that they may have opposite effects on quantities sold, and that the ad valorem tax - the most commonly used tax throughout the OECD - has effects on prices and quantities not previously recognized.

Keywords: Two-sided markets, ad-valorem taxes, specific taxes

JEL Codes: D4; D43; H21;H22; L13

*We are indebted to Simon Anderson, Tore Ellingsen, Kåre P. Hagen, Andreas Haufler, Eckhard Janeba, Bruno Jullien, Agnar Sandmo, Bertil Tungodden, and Frank Staehler for valuable comments. The usual disclaimer applies.

†Norwegian School of Economics and Business Administration, Bergen, Norway. E-mail: hans.kind@nhh.no.

‡CES, University of Munich, Germany. E-mail: marko.koethenbuerger@ces.vwl.uni-muenchen.de.

§Norwegian School of Economics and Business Administration, Bergen, Norway. E-mail: guttorm.schjelderup@nhh.no.

1 Introduction

A benchmark result in economics is that consumers buy less of a taxed good.[1] In this paper we show a new result. Consumers may actually buy more of a taxed good if it is sold by a two-sided platform firm. In particular, a higher ad-valorem tax may lower the end-user price and increase sales.

By definition, a two-sided platform firm serves two different groups of customers that are connected through interdependent demand.[2] Our analysis shows that in such markets, an increase in the ad valorem tax in one side of the market affects the relative profitability between the two markets, making the firm want to shift its earnings to the market where the tax rate is unchanged. To see the logic involved consider a firm that sells good A in market A and good B in market B. Suppose there is a positive externality from market A to market B, say, in the sense that sales in market B are positively correlated with sales in market A. In such a case the firm may generate more income in market B if it reduces the price and increases output in market A.

An example of the incentive mechanism at work above can be illustrated by a media firm, which derives income from selling a newspaper and advertisements, and where the income from advertisements depends positively on newspaper sales. An increase in the ad valorem tax rate on the newspaper may induce the media firm to rely more on income from advertisements because it can reduce the burden of the tax by lowering the price on the newspaper and attract more readers. This will increase the profit margin of the media firm in the market for advertisements. In the extreme, a very high tax on newspapers can lead the media firm to provide the newspaper free of charge and rely on income from advertising only.

In order to bring forth our results we set up a general model of a two-sided market, and show the exact conditions for when a tax increase causes the end-user price to fall and demand to rise. For the sake of convenience, we use the media

[1] An overview of the tax incidence literature is given by Fullerton and Metcalf (2002).

[2] Evans (2003a,b) provides examples and classifications of two-sided markets.

example above to explain our results, but our model is general in nature.

Two-sided platform firms are found in major businesses such as the media industry, the financial sector, real-estate brokerage, and the computing industry. Many two-sided platform firms receive favorable tax treatment. Newspapers, for example, are taxed at a reduced rate or completely exempted from value-added taxation in most countries, since governments consider such publications to be an essential channel for disseminating vital information about e.g. culture, politics, and international affairs. Similarly, many countries exempt credit card services from value added taxation, partly in order to disseminate the use of these services among all income groups.

An important feature of a two-sided platform firm is that its pricing strategies must account for interactions between the demands of multiple customer groups and the externalities that arise in these relationships (Rochet and Tirole, 2003). In the media industry, advertising may be perceived as a nuisance (a negative externality) or a benefit (a positive externality) by readers.[3] A media firm can internalize this externality by charging advertisers and readers/viewers appropriately. We show that the sign and size of such externalities are decisive for the effects of changes in ad valorem tax rates.

Our analysis is related to a growing literature on Industrial Organization that analyzes the price-setting behavior of firms in two-sided markets. However, this literature does not consider taxation issues.[4] The literature on indirect taxation, on the other hand, does not consider two-sided markets.[5]

Closest to the spirit of our analysis are an early paper by Edgeworth (1925) and follow-up contributions by Hotelling (1932), Wicksell (1934), and Bailey (1954). Edgeworth argued that higher commodity taxation under certain conditions may

[3] The nature of these externalities is further discussed in the formal model below.

[4] See for instance Rochet and Tirole (2003, 2004), Crampes, Haritchabalet and Jullien (2005), and Armstrong (2005).

[5] E.g., Keen and Delipalla (1992), Dierickx, Matutes and Neven (1998) and Anderson et. al. (2001a, 2001b). For a survey, see Fullerton and Metcalf (2002).

reduce end-user prices if demand for two different goods is directly interrelated. This possibility has later been labelled Edgeworth's Taxation Paradox. As an illustration, Edgeworth considered demand for first-class and third-class railway tickets. His assessment was that a tax imposed on first-class tickets may give the railway company an incentive to reduce the price of the untaxed good - third-class tickets - in order to sell more of it. Indeed, under certain conditions the price of both types of tickets will fall subsequent to the tax increase.[6]

There are probably many reasons why no link has been made between Edgeworth's Taxation Paradox and indirect taxation in two-sided markets. First, the specific example used by Edgeworth is admittedly peculiar, and may explain why it has been almost forgotten in the literature.[7] Second, the example given by Edgeworth relates to a one-sided market with substitutes.[8] Third, Edgeworth focused solely on specific taxes, and not on the more widely used ad valorem tax. As shown in our analysis, the most interesting policy recommendations arise when we compare the effects of specific and ad valorem taxes, as indeed more recent contributions have emphasized (e.g., Delipalla and Keen 1992).

The rest of the paper is organized as follows: Section 2 sets up a model of a two-sided platform and analyzes the effects of an ad valorem tax on quantities and prices. Section 3 discusses policy implications and carries out an analysis with respect to specific taxes, while Section 4 presents conclusions and discusses some extensions to the model.

---

[6] See Creedy (1988) for a good overview and discussion of this literature.

[7] For example, it is not mentioned in the Handbook of Public Economics (see Fullerton and Metcalf 2002).

[8] Bailey (1954) shows the precise conditions under which the tax paradox may arise, and points out a mistake in Hotelling's analysis.

2 The Model

Consider a two-sided monopoly platform firm selling good C to consumers, say, at price q, and good X to producers, say, at price p. We assume that both the consumers and the producers are price takers. Let c and x denote the respective quantities of the two goods. The inverse demand function for each good is downward-sloping, so that the own-price effects are negative; $q_c \equiv \partial q/\partial c < 0$ and $p_x \equiv \partial p/\partial x < 0$ (subscripts henceforth denote partial derivatives). The willingness to pay for each good, however, may also depend on how much is sold of the other good. The sale of good X imposes a positive externality on the willingness to pay for good C if $q_x > 0$ and a negative externality if $q_x < 0$.[9] In the same manner, good C may impose a positive or negative externality on demand for good X. The inverse demand functions can thus be written as $q = q(c, x, \omega)$ and $p = p(c, x, \omega)$, where ω is a vector of other factors that may affect demand, including the general VAT rate (T) in society.

Examples that fit the model structure above can be found in many sectors of the economy, such as the media industry (or banking) where C is a newspaper/broadcasting (banking; credit cards) and good X is advertising space (banking; shops that accept cards). For the sake of convenience and to emphasize the economic intuition and policy relevance of our results, we shall in what follows relate our model to a media firm (the platform) selling a newspaper to readers and advertising space to firms.

An ad valorem tax (t) is levied on good C, which implies that the platform receives the price $q/(1+t)$ from the consumers. The tax rate t may deviate from the general VAT rate T. Our focal point here is to examine the effects of a change in the tax rate t, holding T and other elements in ω fixed. For this reason we do not model ad valorem taxes on ads, and in what follows we shall suppress ω in the

[9] This is an externality since producers and consumers are price takers. Thus, they do not take into account the effect of their actions on the demand in either side of the market.

inverse demand functions.[10]

The platform has the following profit level:

$$\pi = \max_{c,x} \left[xp(c,x) + \frac{cq(c,x)}{1+t} - k(c,x) \right], \tag{1}$$

where $k(c,x)$ is the cost function, with $k_i \geq 0$ $(i = c, x)$.

The first-order condition for good X $(\pi_x = 0)$ implies

$$[p + xp_x] + \left[cq_x (1+t)^{-1}\right] = k_x. \tag{2}$$

The first squared bracket in equation (2) is marginal revenue on the advertising side of the market of selling more ads. In optimum, this term would be equal to marginal cost (k_x) in a standard one-sided market. The second squared bracket captures the fact that sales of advertising (good X) may influence sales of newspapers (good C). This term is negative if demand for newspapers is decreasing in the level of advertising (that is, $q_x < 0$), while it is greater than zero if advertising imposes a positive externality on demand for newspapers. In the former case, the level of advertising should be set lower than the level that maximizes profit in the advertising market in isolation, while the opposite is true if a larger advertising volume increases demand for newspapers.

Setting $\pi_c = 0$ we further find that

$$\left[(q + cq_c)(1+t)^{-1}\right] + xp_c = k_c. \tag{3}$$

The first squared bracket is marginal revenue from selling the newspaper (good C) to consumers, and should be equal to k_c if $p_c = 0$. However, if demand for ads is higher the larger the number of readers $(p_c > 0)$, profit is maximized by raising the sale of newspapers beyond the volume that maximizes profit in the consumer market (and vice versa for $p_c < 0$).

[10] We hold T fixed, reflecting the view that the general VAT level is determined by more overriding concerns than targeted tax policy in one single market. Note also that interfirm value-added taxes (e.g. on ad revenues) are fully rebated by the government under VAT.

From the first-order conditions we see that equilibrium prices and quantities on both sides of the market depend on the tax rate. Since $p = p(c, x)$ and $q = q(c, x)$, we therefore find that the price changes subsequent to a tax increase are given by

$$\frac{dp}{dt} = p_x \frac{dx}{dt} + p_c \frac{dc}{dt}, \quad \text{and} \quad \frac{dq}{dt} = q_c \frac{dc}{dt} + q_x \frac{dx}{dt}. \tag{4}$$

The second-order conditions for profit maximum require that $\pi_{xx} < 0$, $\pi_{cc} < 0$, and $H \equiv \pi_{xx}\pi_{cc} - \pi_{xc}^2 > 0$.

In order to have a two-sided market, there must be positive externalities from at least one side of the market to the other.[11] The implication is that $p_c > 0$ and/or $q_x > 0$, but whether both terms are positive depends on the particularities of the industry in question. Related to our media example, we cannot predetermine the sign of q_x, since empirical evidence does not give a clear answer as to whether consumers consider advertising to be a good or a bad. However, other things equal, it is reasonable to assume that the willingness to pay for advertising (p) is increasing in the number of readers. We shall consequently assume that $p_c > 0$:

Assumption 1: $p_c > 0$.

It should be emphasized that the model is applicable to two-sided markets in general and that our mathematical derivations and results also hold for $p_c \leq 0$ (in which case two-sidedness requires $q_x > 0$).

For the analysis to follow, the sign of π_{xc} is of particular relevance. Differentiating equation (2) or (3) we find

$$\pi_{xc} = p_c [1 + \varepsilon_p] + q_x (1 + t)^{-1} [1 + \varepsilon_q] - k_{xc}, \tag{5}$$

where $\varepsilon_p \equiv \frac{x}{p_c}\frac{\partial p_c}{\partial x}$ and $\varepsilon_q \equiv \frac{c}{q_x}\frac{\partial q_x}{\partial c}$.

[11] Evans (2003b) defines a two-sided market as one where we have (a) two distinct groups of customers, (b) positive network externalities (at least from one of the customer groups to the other), and (c) an intermediary that internalizes the externalities between the groups. See Rochet and Tirole (2004) for a more formal definition.

The variable π_{xc} measures how the marginal profitability of selling advertising space, π_x, changes if the number of readers increases. One might think that π_{xc} is positive, given the assumption that the willingness to pay for advertising is increasing in the number of readers; $p_c > 0$. However, this is not necessarily true. To see why, note that $\partial p_c/\partial x < 0$ if the marginal value of a larger readership for the advertisers is decreasing in the advertising volume. Thus, the first term in (5) may be negative; this is the case when the elasticity of p_c with respect to x is smaller than minus one ($\varepsilon_p < -1$). The interpretation of the second term in (5) is similar; this term is negative if consumers are ad-lovers ($q_x > 0$) and $\varepsilon_q < -1$, or if consumers dislike ads ($q_x < 0$) and $\varepsilon_q > -1$. Summing up, it is thus clear that the sign of π_{xc} is ambiguous.[12] In order to simplify the discussion in the main text, we shall assume that $\pi_{xc} > 0$:

Assumption 2: $\pi_{xc} > 0$.

In the Appendix we discuss how to interpret our results if $\pi_{xc} < 0$.

2.1 Non-positive externalities from the producer side

In what follows, we examine in detail how different assumptions regarding the externalities between the two groups affect output and prices.

2.1.1 Zero externalities from the producer side ($q_x = 0$)

If readers are indifferent to the advertising level, we have $q_x \equiv 0$. To find how a higher value-added tax affects sales in the two sides of the market, we totally differentiate first order conditions (2) and (3) to find[13]

$$\left.\frac{dc}{dt}\right|_{q_x=0} = \frac{-\pi_{xx}\left(xp_c - k_c\right)}{H\left(1+t\right)} > 0 \text{ if } \left(xp_c - k_c\right) > 0. \tag{6}$$

[12] Note also that with a sufficiently high value of k_{xc}, π_{xc} may be negative even if the first two terms in (5) are positive.

[13] The full derivation is stated in the Appendix.

With $q_x = 0$ we further see from equation (4) that the effect on prices is

$$\left.\frac{dq}{dt}\right|_{q_x=0} = q_c \left.\frac{dc}{dt}\right|_{q_x=0} < 0 \text{ if } (xp_c - k_c) > 0. \tag{7}$$

The sign of dc/dt and dq/dt is determined by the sign of $(xp_c - k_c)$, which has a straightforward economic explanation: The willingness to pay for advertising increases by p_c units if the newspaper attracts one more reader. With a total advertising volume equal to x, the value for the newspaper of attracting one extra reader is thus equal to xp_c. If this value is greater than the marginal cost k_c of an extra copy of the newspaper (i.e., $xp_c - k_c > 0$), we see from equations (6) and (7) that $dc/dt > 0$ and $dq/dt < 0$. From equations (2) and (3) we further find that a larger number of readers allows the newspaper to sell more advertising:

$$\left.\frac{dx}{dt}\right|_{q_x=0} = \frac{\pi_{xc}(xp_c - k_c)}{H(1+t)} > 0 \text{ if } xp_c - k_c > 0. \tag{8}$$

To put the result sin eqs. (6) and (7) into perspective, we show that, in sharp contrast to results obtained in one-sided markets (see e.g., Delipalla and Keen, 1992), a higher VAT may increase sales ($dc/dt > 0$) and reduce the price ($dq/dt < 0$) of the good subject to higher taxes.

The reason for this rather paradoxical result is that in the market where the tax is increased (newspaper) the profitability falls relative to the profitability in the other market (advertising). The firm therefore wants to shift income from the consumer side of the market to the producer side. In order to do so, it must increase the sale of newspapers, since this leads to a higher demand for ads (as is evident from (8)). To obtain higher sales of the newspaper ($dc/dt > 0$), the price of the newspaper must be reduced ($dq/dt < 0$).[14]

Since $p(c, x)$ is downward-sloping in own quantity, an increase in the advertising volume tends to reduce p ($p_x < 0$). At the same time, the firm can charge a higher

[14]To see the intuition for this result as clearly as possible, assume that t approaches infinity. Obviously, the newspaper would then have no reason to charge a positive consumer price. However, it can still raise revenue through the advertising market and give the newspaper away for free.

advertising price if the size of the readership increases (since $p_c > 0$). Consequently, it is uncertain whether the price of advertising will go up or down:

$$\left.\frac{dp}{dt}\right|_{q_x=0} = p_x \frac{dx}{dt} + p_c \frac{dc}{dt} \gtreqless 0 \text{ if } \pi_{xc} > 0.$$

2.1.2 Negative externalities from the producer side ($q_x < 0$)

When demand for newspapers depends negatively on the advertising level, we have $q_x < 0$. One might think that higher value-added taxes are more likely to reduce sales of newspapers the more consumers dislike ads (as an increase in x, motivated by profit shifting to the ad market, now lowers c). However, total differentiation of equations (2) and (3) makes it clear that the opposite is true:

$$\left.\frac{dx}{dt}\right|_{q_x<0} = \left.\frac{dx}{dt}\right|_{q_x=0} + \left(\frac{1}{1+t}\right)^2 \frac{\overbrace{\pi_{cc} c q_x}^{+}}{H} \tag{9}$$

$$\left.\frac{dc}{dt}\right|_{q_x<0} = \left.\frac{dc}{dt}\right|_{q_x=0} + \left(\frac{1}{1+t}\right)^2 \frac{\overbrace{(-\pi_{xc} c q_x)}^{+}}{H}. \tag{10}$$

The first term in (9) and (10) shows how advertising and newspaper sales respond to a tax increase if consumers are indifferent about ads ($q_x = 0$). As argued above, this term may be positive or negative. The second term, though, is unambiguously positive and increasing in the consumers' disutility of ads. The reason is that if sales in the newspaper market are adversely affected by advertising ($q_x < 0$) the media firm has a smaller advertising level than the volume which maximizes profit in the advertising market (c.f. equation (2)). With a heavier taxation of newspaper sales, this effect becomes less important for the media firm. Other things equal, it is optimal to increase the volume of ads, but in order to facilitate a rise in demand for advertising the size of the readership must increase. The latter requires a reduction in the price charged by the media firm, and more so the stronger the consumers' distaste for advertising. In particular, this implies that the tendency for the consumer price to fall subsequent to a tax increase is even more pronounced when $q_x < 0$ than when $q_x = 0$.[15] It should be noted, though, that we still cannot sign the change in the price of advertising if both the advertising level and the size of the readership increase.

We can now state:

Proposition 1: *Suppose that $q_x \leq 0$. A sufficient condition for a higher value-added tax on good C to increase equilibrium quantities of both goods is that $xp_c > k_c$. The price of good C (inclusive of VAT) is lowered, while the sign of the change in the price of the untaxed good (X) is ambiguous.*

2.2 Positive externalities from the producer side $(q_x > 0)$

If demand for good C depends positively on output of good X, we have $q_x > 0$. Such a positive externality is characteristic for the banking industry, where consumers presumably have a higher willingness to pay for holding a credit card the larger the number of merchants that accept it. It may also be characteristic for specialized magazines, where $q_x > 0$ reflects a taste for commercials (ad-lovers). Car ads in automobile magazines or perfume ads in beauty magazines may well be appreciated by readers (Depken II and Wilson, 2004). In what follows, we continue to relate the model to the media market.

Equations (9) and (10) still hold when consumers are ad lovers, but with the important difference that the last terms in both equations turn from positive to negative, that is,

$$\left.\frac{dx}{dt}\right|_{q_x>0} = \left.\frac{dx}{dt}\right|_{q_x=0} + \left(\frac{1}{1+t}\right)^2 \frac{\overbrace{\pi_{cc} c q_x}^{-}}{H} \tag{11}$$

[15] With $q_c < 0$ and $q_x < 0$ it follows immediately from equation (4) that $dq/dt < 0$ if $dx/dt > 0$ and $dc/dt > 0$, and that the price reduction is larger the more consumers dislike ads.

$$\left.\frac{dc}{dt}\right|_{q_x>0} = \left.\frac{dc}{dt}\right|_{q_x=0} + \left(\frac{1}{1+t}\right)^2 \frac{\overbrace{(-\pi_{xc}cq_x)}^{-}}{H}. \tag{12}$$

If $q_x > 0$ is small, the last term is insignificant relative to the first one, so that $sign(dx/dt) = sign(dx/dt|_{q_x=0})$ and $sign(dc/dt) = sign(dc/dt|_{q_x=0})$. As shown above, we then have that quantities of both goods increase ($dc/dt > 0$ and $dx/dt > 0$) if $xp_c > k_c$. However, if q_x is sufficiently high, it follows from equations (11) and (12) that the sales of newspapers and advertising are decreasing in taxes. To see why, notice that the newspaper has more commercials than the quantity which maximizes profit on the advertising side when consumers love ads (c.f. equation (2)). An increase in VAT, though, implies that it becomes less profitable for the media firm to attract readers by having a large advertising volume. Instead, the media firm will have incentives to reduce the level of advertising, and approach the volume that maximizes profit on the advertising side. If q_x is sufficiently high, both the level of advertising and the demand for the media product will therefore fall.

Finally, note from equation (4) that the signs of both dp/dt and dq/dt are ambiguous when output on both sides of the market is decreasing in t.

We can now state:

Proposition 2: *If $q_x > 0$, but is relatively small, a higher value-added tax on good C may increase output on both sides of the market. If $q_x > 0$ is sufficiently high, a higher VAT reduces output on both sides of the market. The effects on prices of higher taxes are ambiguous.*

3 Policy Implications

In most countries newspapers are subject to a reduced value-added tax rate. In Germany they are subject to a rate of 7% (16% is the regular rate) while in e.g. the UK and Denmark they are exempted from value-added taxation all together (European Commission, 2004). Newspapers are also either fully or partially exempted from sales taxes in a number of U.S. states. The reason for the preferential tax treatment is that governments consider newspapers to be essential for the dissemination of vital information regarding for instance culture, politics and international affairs. A lower tax rate is thought to reduce the newspaper price and, more importantly, to increase the circulation of the media product. The reasoning is in line with tax incidence analysis in a one-sided market, but the analysis above shows that this need not hold for the newspaper industry, which typically operates in a two-sided market. On the contrary, a lower VAT may reduce the sales of newspapers.

The analysis provides a framework for thinking about taxation also in other industries. Credit card services, which are VAT-exempted in the European Union, is one example.[16] As the value of holding a credit card is increasing in the number of merchants accepting it, and vice versa, we have $p_c > 0$ and $q_x > 0$. If, for historical reasons, governments have wanted to increase the usage of credit cards, the policy of exempting the use from VAT may have been effective. Presently, though, the use of credit cards has become so widespread that the network effects presumably are more or less exhausted (which in particular means that q_x is small) in most European countries. Since the marginal costs of a transaction for the platform (the credit card company) are close to zero ($k_c \approx 0$) the analysis suggests that abolition of preferred VAT treatment of credit card services need not have a large negative impact on its use. In fact, we cannot disregard the possibility that the opposite may happen; imposing a VAT may further expand the network size on both the credit card holder and the merchant side of the market.[17] This illustrates that the effectiveness of reducing the VAT-rate may depend crucially on whether or not we

[16] Auerbach and Gordon (2002) discuss the desirability of taxation of financial services. Challenging the current European practice, they recommend taxation of financial transactions. Their analysis resorts to the standard one-sided market view.

[17] This certainly does not mean that credit card companies would welcome a VAT. On the contrary, differentiating the equilibrium value of equation (1) with respect to t, and using the envelope theorem, we find $d\pi/dt = -q(x,c)(1+t)^{-2} < 0$. The profit level is thus strictly decreasing in the tax rate.

consider a mature industry.

The discussion above makes it clear that it is difficult to ascertain the effects of VAT changes in two-sided markets without detailed knowledge about marginal costs and externalities. The difficulties are particularly large in the media industry. Depken II and Wilson (2004), for instance, find that advertising is considered to be a good in some paper magazines and a bad in others. Presumably, we find a similar variety in the public's attitude to advertising in the newspaper industry. So what can the government do if it wants to spur output of newspapers? A more accurate policy than changing the VAT rate, would be to subsidize newspapers. To see this, let the profit level of the media firm be given by

$$\pi = \max_{c,x} \left[xp(c,x) + \left(\frac{q(c,x)}{1+t} - \tau \right) c - k\left(c,x\right) \right],$$

where τ is a specific tax on newspapers. The first-order conditions for the platform are the same as before, except that the specific tax now appears as an additional cost term in selling newspapers (c.f. equations (2) and (3)):

$$\pi_x = 0 \qquad => \qquad [p + xp_x] + \left[q_x c \left(1+t\right)^{-1} \right] = k_x \tag{13}$$

and

$$\pi_c = 0 \qquad => \qquad \left[\left(1+t\right)^{-1} \left(q_c c + q\right) \right] + xp_c = k_c + \tau. \tag{14}$$

Totally differentiating (13) and (14), holding t fixed, we find

$$\frac{dc}{d\tau} = \frac{\pi_{xx}}{H} < 0 \text{ and } \frac{dx}{d\tau} = -\frac{\pi_{cx}}{H} < 0. \tag{15}$$

Equation (15) makes it clear that specific taxes unambiguously have a negative impact on output in both markets, independently of consumer preferences for ads. The reason is that higher specific taxes are equivalent to increased unit costs, as shown by equation (14). Since higher unit costs lower the marginal profitability for any given output, it is optimal to reduce sales of newspapers ($dc/d\tau < 0$). As a result, the advertising level falls ($dx/d\tau < 0$).

As discussed in the introduction, Edgeworth's tax paradox asserts that a higher specific tax on one of two goods under certain conditions may reduce the price of both.[18] To see that this holds in our context of a two-sided market, note first that for the consumer price we have

$$\frac{dq}{d\tau} = q_c \overbrace{\frac{dc}{d\tau}}^{+} + \overbrace{q_x}^{?} \overbrace{\frac{dx}{d\tau}}^{-} . \tag{16}$$

Equation (16) is unambiguously positive if consumers dislike ads ($q_x < 0$). However, with ad-lovers ($q_x > 0$) the second term is negative, reflecting that the consumers' willingness to pay for the newspaper falls when the level of advertising decreases. If this effect is sufficiently strong, we obtain $dq/d\tau < 0$.

We likewise find that

$$\frac{dp}{d\tau} = p_x \overbrace{\frac{dx}{d\tau}}^{+} + p_c \overbrace{\frac{dc}{d\tau}}^{-} \tag{17}$$

is negative if the fall in readership, $p_c\,(dc/d\tau)$, dominates the increase in ads, that is $p_x\,(dx/d\tau)$. Equations (15) - (17) thus show that an increase in τ may reduce output and prices of both goods.[19]

To summarize:

Proposition 3: *A higher specific tax on good C reduces output of both goods. If p_c and q_x are positive and sufficiently large, prices fall (Edgeworth's tax paradox).*

The analysis in Sections 2 and 3 makes it clear that raising ad valorem taxes and specific taxes may have opposite quantity effects. The reason for this is that with

[18] See also Bailey (1954) for a discussion.

[19] An example that yields this result is the following. Let $p = -x/10 + c$, $q = z - c/10 + x$ and $\pi = xp + (q - \tau)\,c - x^2 - c^2$. Then we have that $\partial^2\pi_1/\partial c\partial x = 2 > 0$. It is easily verified that all second-order conditions are satisfied. Solving $\partial\pi/\partial c = \partial\pi/\partial x = 0$ we find $p = x = 50\,(z - \tau)/21$, $q = 131z/42 - 89\tau/42$ and $c = 55\,(z - \tau)/21$, from which it is immediately clear that a higher tax rate reduces all prices and quantities. Related to the media market, we may intuitively regard the reduction in readership (resp. advertising) as a quality reduction of the newspaper from the advertisers' (resp. readers) point of view. Other things equal, this leads to a lower willingness to pay for the newspaper on both sides of the market.

specific taxes, there is a one-to-one relationship between tax payments and quantity, while there is no direct link between output and the burden of taxation under ad valorem taxation. Indeed, subsequent to a higher ad valorem tax the firm can in principle both reduce tax payments and increase the quantity by lowering the price. Put differently, in this case the firm has two instruments at its disposal, compared to only one under specific taxes. In a multi-sided market this difference has profound implications for firm behavior, and thus for public policy in a context where output per se is considered to be important.

4 Conclusion

Traditional analysis of tax incidence has focused on conventional (one-sided) markets, where the sale of one good does not directly affect the sale of other goods. In such markets a general insight is that indirect taxes are partly shifted (or even overshifted) onto consumers, resulting in lower sales of the taxed good. Our analysis has shown that this result is challenged in a two-sided market. If demand for the taxed good matters for sales of a product in another market, the incidence of taxation changes. In a two-sided market an increase in an ad valorem tax may, under certain conditions, lead to lower prices for both goods as well as to higher sales. The results obtained under ad valorem taxation are in sharp contrast to our findings under specific taxation, where a higher tax unambiguously has a negative effect on output.

Our study has been carried out in a monopoly setting. An interesting path for future research would be to check the robustness of our results under different market structures. However, we believe that the main results in this paper would survive under oligopoly as well. As long as firms have some market power, a tax increase on one side of the market implies that the firms will have incentives to shift profit to the other side of the market. In an appendix, available from the authors upon request, we show that this conjecture holds in a simple duopoly model with linear demand functions.

Even though our discussion is related to the media market, we have not incorporated any of the particularities of the media market or the advertising market into the model. The reason is that we have used a model sufficiently general in structure to highlight the most common mechanisms in two-sided markets. We have also abstained from welfare analysis. Such analysis would hinge on specific characteristics of the industry in question. In the media market, this would for instance require that we make assumptions about whether advertising is persuasive or informative. This said, we believe that there is also a need for industry-specific analysis, both theoretically and empirically.

5 Appendix

Derivation of the relationship between quantities and ad valorem taxes

We assume that the second order conditions hold with non-negative prices and quantities, so that the equilibrium is characterized by first order conditions (2) and (3). To find how a higher value-added tax affects prices on the two sides of the market, we totally differentiate (2) and (3). We then find

$$\begin{aligned}\pi_{xx}\frac{dx}{dt}+\pi_{xc}\frac{dc}{dt} &= \left(\frac{1}{1+t}\right)^2 cq_x \\ \pi_{xc}\frac{dx}{dt}+\pi_{cc}\frac{dc}{dt} &= \left(\frac{1}{1+t}\right)^2 (q+cq_c)\,.\end{aligned}$$

Making use of the first-order condition (3), the effect of the tax on quantities is now given by

$$\frac{dx}{dt}=\left(\frac{1}{1+t}\right)^2\frac{\pi_{xc}\left(1+t\right)\left(xp_c-k_c\right)+\pi_{cc}cq_x}{H} \tag{18}$$

and

$$\frac{dc}{dt} = -\left(\frac{1}{1+t}\right)^2 \frac{\pi_{xx}(1+t)(xp_c - k_c) + \pi_{xc}cq_x}{H}. \tag{19}$$

Consequences of relaxing the assumption that $\pi_{cx} > 0$

Suppose that $\pi_{xc} < 0$ and $q_x = 0$. From equations (6) and (7) we see that a higher ad valorem tax still increases sales of the newspaper and reduces its price if $xp_c - k_c > 0$: thus the media firm's incentive to sell a larger number of newspapers in order to shift revenue to the advertising side is unaltered. However, from equation (8) we find that $dx/dt < 0$ if $\pi_{xc} < 0$.

If $q_x < 0$, we know that there will be less advertising than the volume which maximizes profit on the advertising side of the market. If the ad valorem tax rate on sales of newspapers increases, the media firm will care less about the revenue it captures directly from the readers. This is true independent of whether $\pi_{xc} > 0$ or $\pi_{xc} < 0$. The second term in equation (9) shows that this effect makes the media firm sell more advertising space if t increases. However, the second term in equation (10) makes it clear that this tends to reduce the sales of newspapers.

To grasp the intuition for this result, assume that $\pi_{xc} < 0$ because k_{xc} is large. In order to save costs, the media firm will then have incentives to reduce the circulation of the newspaper when the advertising volume increases.[20]

The case where $q_x > 0$ has a similar interpretation. If the consumers are ad lovers, the newspaper has more ads than the level that maximizes profit on the advertising side of the market. Independent of the sign on π_{xc}, the newspaper will therefore reduce the advertising level if t increases ($dx/dt < 0$). However, a lower advertising level means that the marginal profit of selling newspapers increases if $\pi_{xc} < 0$, which induces the newspaper to sell more newspapers ($dc/dt > 0$).

The effects of assuming $\pi_{xc} < 0$ when we consider specific taxes are analogous,

[20] For the same reason, we see from equation (15) that a higher specific tax on newspapers - which always reduces sales of newspapes - increases the advertising volume if $\pi_{xc} < 0$.

and seen from equations (15) - (17).

References

[1] Anderson, P.S., A. de Palma, and B. Kreider (2001a).The efficiency of indirect taxes under imperfect competition, *Journal of Public Economics* 81, 231-251.

[2] Anderson, P.S., A. de Palma, and B. Kreider (2001b). Tax incidence in differentiated product oligopoly, *Journal of Public Economics* 81, 173-192

[3] Armstrong, M. (2005), Competition in Two-Sided Markets, Mimeo, UCL.

[4] Auerbach, A.J. and R. Gordon (2002). Taxation of Financial Services under a VAT, *American Economic Review (Papers and Proceedings)*, 92, 411 - 416.

[5] Bailey, M. J. (1954). Edgeworth's Taxation Paradox and the Nature of Demand Functions, *Econometrica,* 22, 72-76.

[6] Crampes, C., C. Haritchabalet, and B. Jullien (2005), Advertising, Competition and Entry in Media Industries, CESifo Working Paper No. 1591, Munich

[7] Creedy, J. (1988). Wicksell on Edgeworth's Tax Paradox, *Scandinavian Journal of Economics*, 90, 101-112.

[8] Depken II, C. A. and D. P. Wilson (2004). Is Advertising Good or Bad? Evidence from U.S. Magazine Subscriptions, *Journal of Business*, 77, S61-S80.

[9] Dierickx, I., C. Matutes and D. Neven (1998). Indirect Taxation and Cournot Equilibrium, *International Journal of Industrial Organization,* 6, 385-399.

[10] European Commission (2004). VAT Rates Applied in the Member States of the European Community, DOC/2008/2004, Brussels.

[11] Edgeworth, F.Y. (1925). Papers Relating to Political Economy, vol. 1, Macmillan, London.

[12] Evans, D. S. (2003a). The Antitrust Economics of Two-Sided Markets, *Yale Journal of Regulation*, 20 , 325 - 381.

[13] Evans, D. S. (2003b). Some Empirical Aspects of Multi-Sided Platform Industries, *Review of Network Economics*, 2 , 191 - 209.

[14] Fullerton, D. and G.E. Metcalf, (2002). Tax Incidence, in: A. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, vol. 4, 1787-1872 (North-Holland, Amsterdam).

[15] Hotelling, H. (1932). Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions, *Journal of Political Economy*, 40, 577-616.

[16] Keen M, And S. Delipalla (1992): The comparison between ad valorem taxes and specific taxation under imperfect competition, *Journal of Public Economics* 49, 351-367.

[17] Pindyck and Rubinfeld (2005), Microeconomics, 6^{th} ed, Prentice Hall, ch 11.

[18] Rochet, J.C. and J. Tirole (2003), Platform Competition in Two-Sided Markets, *Journal of the European Economic Association*, 1, 990-1029.

[19] Rochet, J.C. and J. Tirole (2004), Two-Sided Markets: An Overview, IDEI Working Paper, n. 275.

[20] Wicksell, K. (1934), Lectures on Political Economy, Routledge, London.

% ( ) * - .)))) (/)0
1 %2 3(4 5 3 *% 6 . 7 %- *. / ((

6. 5 8 9: =

: 1 8 8)*.
@ 5) A* :) B

&( )* , - . &/ 001

9 -) (3

This article aims at analysing the factors than can explain the motivations of firms for investing in a website and for the degree of use of this site in an informative and/or in a commercial way. We also study the importance of reorganisations of the firm in the success of e-commerce. Firstly, we seek to find in the theoretical and the empirical literature the factors underlined as influencing the creation of a website, the choice of the richness in terms of contents, and the use of the site as a new business channel. Secondly, it permits us to formulate different hypothesis we test using a Luxembourg database. The results show that there are large differences between factors influencing creation behaviours and degree of involvement in the general use of the website. But, concerning the two uses the firm can make with its site, there are many similarities in the determinants of intensity of website uses both as an informational tool and as an e-commerce one. Regarding the links between reorganisations thanks to management ICTs and the setting up of an e-commerce website, our study displays a significant complementarity.

C AD*)0 Internet strategies, e-commerce, firm reorganisation

E1 6 4()) 3(* 0 L14, L21, L86

* CREM – Centre de Recherche en Economie et Management – UMR CNRS 6211
Faculté des Sciences Economiques, 7 place Hoche, 35065 Rennes cedex
E-mail : ludivine.martin@univ-rennes1.fr, Tel : +33 (0)2 23 23 33 35, Fax: +33 (0)2 23 23 35 09

BF * . 3 *

From March 2000 and the Lisbon European Council, the European Union has set the ambitious objective of modernising the European economy[1] harnessing the benefits of the opportunities given notably by Information and Communication Technologies (ICTs). The long term strategy defined by this Council and included in the policy agenda of the Union is well-known as the "Lisbon Strategy".

The first stage of this strategy was the endorsement of the eEurope 2002 Action Plan at the Feira European Council in June 2000. This plan aimed at co-ordinating regional, national and European actions to increase the number of individuals and enterprises connected to the Internet and to stimulate their uses. Because Small and Medium-sized Enterprises (SMEs[2]) are the backbone of the European economy[3], they received the highest priority and the plan included a specific action to help such enterprises to use Internet in order to find new business opportunities. The GoDigital initiative was implemented by the Commission in order to: *2 34. 5 67 8 9 *4 : 4 5 * =* 4. 5 34 4 (*(* 4 A 5 3*B * 9)4 * C3 5 4) A 4 (5 4 - 9* D 3*3 9E* (Commission of the European Communities, 2001, p.4).

The second stage was the eEurope 2005 Action Plan adopted at the Sevilla European Council in June 2002. Since 2000, the use of basic ICTs and access to the Internet had grown and in 2002 was closed to saturation levels. In fact, in Europe-15, 92% of SMEs[4] and 99% of large firms used computers. Concerning Internet, in the same time and area, 79% of SMEs and 97% of large enterprises had an access to the Web. Because eEurope 2002 permitted to connect a large part of SMEs, the new priority was given to the development of e-business[5] and viable business models *2 * * 4) 3 5 * 34 D **B 99 4) 8. 4D * D 9 9 (9 5 D 4(. 3*B *& (5 4 * * 4. 5 B 9* * FGHE* (Commission of the European Communities, 2003a, p.3).

---

[1] *2H I 4 9 *4(&9 * *9) 9* * 5 3 54)4 * C*(3 (J *4 - 34 * 49* 34 D **B ((& 3 A 4 (5 - 9 (34 4 & * 4 (3 D - 4) 9. 9* - 34 4 3 5 4 * * 4 (- ** I4- 9 (5 * 9 43 34 9 4 E* (Lisbon European Council, 2000, p.2).

[2] According to the European Union definition (Commission of the European Communities, 2003b), SMEs are private enterprises with less than 250 employees. More precisely, are considered as "small", enterprises with a number of employees between 10 and 49 and as "medium-sized", enterprises with a number of employees between 50 and 249.

[3] In 2003, there were 18.7 million private enterprises in Europe-15. More than 99% of these firms employed less than 250 persons (Commission of the European Communities, 2004). Moreover *2 67 8 9 5 * 9. - 9* * 9 4) (A &9 4. 3 4) I4- 9 9 9 - (5 5 4. ( )4 * D . 9 D (- . 9 99 (9E* (Commission of the European Communities, 2001, p.5).

[4] The data come from the Community enterprise survey on ICT usage conducted in 2002 (quoted by Commission of the European Communities, 2003a, p.25).

[5] *2H * : - . 9 99= 34B 9 - 4* 34 3 M. & 5 (9 5 4 N (* 9*. 3*: 5 4) - . 9 99 D 43 99 9 *4 A * - 9*. 9 4) (5 * * 3 4 45 9E* (Commission of the European Communities, 2002, p.13).

The Commission conscious of the opportunities given by e-commerce and of SMEs' difficulties to set up an e-commerce website and also to restructure their production and distribution processes to make best use of Internet technologies decided to establish an European e-Business Support Network[6]. The aim of this network is to co-ordinate and strengthen actions in support of firms to do e-business in a favourable environment[7]. It was also conceived for firms to share practical experience and information on funding support concerning ICTs adoption and uses. Because the end of the eEurope 2005 is impending, a new initiative, the i2010 programme, is going to perpetuate the actions carried on up to now. Moreover, it is time to draw up an assessment of the progresses in terms of websites and e-commerce practice diffusion made during the eEurope 2005 Action Plan. Thus, since 2002 in Europe-15, there has been in average an increase in the number of enterprises present on the Web with a single page or a real website, but since 2003 there has been, especially in Luxembourg, a stagnation or even a decrease in the diffusion rate of firms' sites. According to Eurostat[8] in 2002 the EU-15 average rate[9] of website ownership was around 55% of firms employing 10 persons and more, 62% in 2004 and 67% in 2005. If a focus is made on Luxembourg, we can notice that there has been a fast growth between 2002 and 2003 (with a rate of 61% of firms employing 10 persons and more in 2003 compared to 51% one year before), and then a stagnation with a rate around 62% in 2004 and 2005.

In spite of this stagnation, Luxembourg exhibits a very higher diffusion rate of websites than countries with the smallest ones, i.e. Portugal and France with rates around 30%. But this country has an important possibility in terms of growth, knowing that first, Nordic countries reached rates around 80%, and secondly that the development of online commerce is important in European households[10] and especially in Luxembourg, which is on the top of the EU-15 in terms of household online commerce, with 32% of the population (half of the Internet users living in Luxembourg) which have used Internet to buy or order online in the second term of 2004, against 18% of the population (34% of Internet users) in the first term of 2003.

---

[6] Further information is provided on the portal of this network: www.e-bsn.org.

[7] The improvement of firms environment was an important objective in eEurope 2005. Regional and national actions were encouraged by the Commission with a particular emphasis on trust and confidence of consumers in order to help the development of e-commerce.

[8] Source: Eurostat, Information society statistics, available on: http://epp.europa.cec.eu.int/ for European data and STATEC/CEPS-Instead for Luxembourg data.

[9] The average was calculated on a limited number of sectors (manufacturing, trade, hotels and restaurants, transport and communication, renting and business activities, and personal service activities) in order to have a comparability for all countries.

[10] In EU-15, in the second term of 2004, 21% of the population have bought or ordered online against 15% in the first term of 2003, that is to say 37% of Internet users in 2004 and 31% in 2003.

Are the factors that contribute to website creation behaviours different from those that influence the degree of involvement in the site use? As the site may be considered as a two sides tool which permits both to inform customers and to create a business channel, do the explanatory factors influence differently the intensity of website used as an information tool or as an e-commerce one? Do reorganisations undertaken by firms thanks notably to adoption of management ICTs (allowing to accelerate production and distribution processes) involve a higher setting up of e-commerce website?

That are the questions we seek to provide pieces of response in the remainder of the article. We successively analyse the determinants of website creation and of the degree of involvement in the website. Furthermore, we distinguish between the uses of the website and study the differences and similarities in factors influencing use behaviours. Generally, because of a lack of information concerning reorganisations in surveys, there exists few studies which study the e-business phenomenon. We try to offset this shortage thanks to the richness of items available in our database. Moreover this richness allow us to insert in our various investigations a large number of explanatory variables we can find in existing empirical studies.

On the one hand, we find that there are large differences between factors influencing creation behaviours and those influencing the degree of involvement in the general use of the website. On the other hand, concerning the two uses the firm can make with its website, there are many similarities in the determinants of intensity of website uses both as an informational tool and as an e-commerce one. Regarding the links between reorganisations thanks to management ICTs and the setting up of an e-commerce website, our study displays a significant complementarity. Thus, management technologies that improve the distribution process of the firm (such as ICTs linked with the logistic system or with customers systems) have a positive influence on e-commerce practice.

The remainder of this article is organised as follows. In the next section we present both theoretical and empirical literatures concerning the different factors influencing website creation and intensity of involvement in the site they highlight. After a detailed description of the database, we present both the dependent and explanatory variables introduced in our econometric analyses. Then we present the results and discuss the main empirical findings.

FH(3 *) 4 3 / D -)) / . 2 (.)

The aim of this paper is to analyse the behaviours of firms facing the choice of investing in a particular use of Internet: the website. The website meets different needs for the firms: it is both a tool to inform and present the activities of the firm, and a tool to build a new business channel. Therefore, to ascertain the motivations of the firms towards its website, we will analyse the factors influencing the setting up of the website and the use of this website especially in an informative way or in an e-commerce perspective, and the relative importance of each factors in the different levels of the firm involvement.

In order to address this task, we begin with a review of both the theoretical literature and the empirical studies concerning the adoption of ICTs and in particular the setting up of a website and the use of this site. Generally agreed opinion on technologies diffusion is that the diffusion is better understood as a series of stages (Rogers, 1962; Gillepsie, Richardson & Cornford, 1995). That's why we can distinguish three subjects of study, in the empirical literature, that cover three stages of the website creation and use.

First, we find the website creation, that is generally modelled as a binary choice, i.e. to have a website or not (Sadowski, Maitland & Van Dogen[11], 2002; Dholakia & Kshetri[12], 2004 and Martin & Pénard[13], 2005). Secondly, some articles study the nature of the site (Lucchetti & Sterlacchini[14], 2004; Martin & Pénard, 2005). The focus is made on the understanding of the richness of the site in terms of contents. This richness can be grasped in different ways. As in Lucchetti & Sterlacchini (2004), a composite indicator can be constructed from the functions provided by the site, or as in Martin & Pénard (2005), a distinction can be made between elaborate websites and others according to the functions offered[15]. Thirdly, the last stage studied

[11] This study concerns 264 enterprises located in the Netherlands and employing between 1 and 499 persons (definition of SMEs by the OECD) interviewed during June and July 1998.
[12] This analysis uses 45 SMEs (OECD definition) located in New England region (U.S.A.) asked before October 2002.
[13] This study concerns 850 legal unit of firms located in Brittany (France) interviewed in early 2004.
[14] The sample is composed of 168 SMEs (EU definition) located in Central Italy and questioned in the end of 2000.
[15] The functions included in these two studies are somewhat different. In the first one, the functions included are: catalogue of products, detailed product description, on-line ordering facility, feedback forms other than e-mail address, on-line job offer, multilingual site and site referenced on common Internet engines. In the second one, in addition to the presence of the catalogue of products, on-line ordering facility, on-line job offer, multilingual site, we find firm presentation, intranet, collection of customer information, after sales support, traceability of products, order processing, on-line payment security and a regular updating.

concerning the use of a website is the practice of e-commerce (Chappell & Feindt[16], 2000; Dholakia & Kshetri, 2004 and Martin & Pénard, 2005).
In each of these stages, the factors displayed by the literatures can be classified into two types: internal factors and external ones (Dholakia & Kshetri, 2004).

2.1. Internal factors

The literature on ICTs adoption has highlighted different internal factors we choose to classify into three groups: resources, business characteristics and past experiences with ICTs.

/ ΦΦQ 94. 3 9

From a theoretical point of view, the enterprises' resources are essential to diversify firms' activities. First, in the resource-based theory, the firm as a collection of resources and competencies seeks to exploit its resources with the aim of finding new opportunities of diversification and growth in order to reinforce its position on the market (Penrose, 1959; Barney, 1991). By the means of a website, the firm can reach this objective. Actually, firms can develop rent-yielding activities such as improvement of trade relations with customers and a greater ability to distribute their products thanks to the possibility of selling on-line. It seems reasonable to assume that largest firms with a lot of resources (notably financial and human) would be in the best position to exploit the business opportunities created by Internet technologies. Secondly, we can consider models of technology diffusion such as epidemic models, information cascade models and probit models (Geroski, 2000). In epidemic models and information cascade models, it is supposed that the decision of adopting a new technology is influenced by the diffusion of information concerning this technology given by primary-adopter, the "more able"[17] firms. Because the adoption implies large sunk costs in terms of investment in the technology and in the organisational changes induced, the firms need to be adventurous and less risk adverse to adopt a technology with hypothetical benefit. But these models give no more information concerning the characteristics of those "more able" firms. The probit models or rank effects models (Karshenas & Stoneman, 1993) enhance these approaches and stress on the importance of individual characteristics in the adoption process, particularly resources measured generally by the size of the firm. Such models conjecture that firms have different gross returns from the use of new

[16] Chappell & Feindt (2000) realised a best practice analysis of e-commerce for 89 SMEs located in 17 countries and studied in July 1999.
[17] *2H *4* D4D *4 4) D4* * .9 9 9 34 D49 (4) : 4 - = (: 99 - =) 9 (*)4 9* * (( )). 9)4 *4) 9* E* (Geroski, 2000, p.608).

technologies. Consequently, firms with large resources are in a better position to adopt and are less affected by the uncertainty induced by ICTs adoptions and the reorganisations implied. Thus, such enterprises are less risk adverse and don't need much information on the technology to adopt. This approach also permits to understand the inertia observed when we looking at the diffusion of ICTs and called usually the "wait-and-see" attitude. Many firms wait to adopt a new technology until profitability is demonstrated. Limited capital is generally used as one of the best explanation to the relative lack of investment in ICTs by small firms (La Rovere, 1998; Sadowski * *Q*2002).

In empirical studies, the size of the firm has been held as a measure of its resources, but other variables introduced below can be considered as alternative proxies for this characteristic. Concerning the size, the effect on the website setting up, the site content and e-commerce practice is not well-defined. All the studies we are aware of introduced the size of the firm as an explaining variable but the results don't display the same results. Effectively, concerning the ownership of a website, while the analysis conducted by Sadowski * *O*(2002) finds no significant effect, others find a positive influence of the size (Dholakia & Kshetri, 2004; Martin & Pénard, 2005). Concerning both the choice of the functions the site provides and the e-commerce practice, none of the analyses find a significant effect of the size. But a best practice analysis of e-commerce by Chappell & Feindt (2000) highlights a possible negative impact of the size. The explanation proposed is that small firms have greater flexibility and ability to adjust to changes in the market and are this way more efficient in setting up a new business channel.

Other variables like the affiliation to a group, human resources, the age of the firm, and communication requirements (i.e. the need to communicate with different units of the firms, partners, and potential customers) are found in empirical studies and can also be proxies for the financial and human resources.

Regarding the effect of belonging to a group, it seems to be ambiguous. According to Lucchetti et Sterlacchini (2004), *2 (D (*) 9(3(4 . 3A& * *4 (4D* *3 4 45& 4 4*-. *MDDN) 9)) *(* 5 4. D CD 3 99) 3 34 9* *9 (*. 9* (*4 - 99 9A B 9 * (4D*4 (3 9 4 E* (p.166). In empirical studies, the setting up of a website seems to be positively correlated with the membership of a group (Martin & Pénard, 2005). In the investment in an advanced website, some studies highlight a negative influence of the group (Lucchetti et Sterlacchini, 2004) while others stress on a significant and positive influence of this characteristic (Martin & Pénard, 2005). Concerning the study of e-commerce practice, few studies carry on an analysis, and when it is the case no effect of the group is displayed (Martin & Pénard, 2005).

It seems reasonable to think that human qualifications are associated with faster adoption of ICTs, thanks to an easier adaptation of qualified employees to technological progress. Lucchetti et Sterlacchini (2004) insert this factor in their study of factors influencing the richness of the site and find a positive impact. According to Chappell & Feindt (2000), the qualification of managers seems to be important, because they need to be able to cope with growing opportunities associated with e-commerce.

Concerning the age of the firm, the effect is ambiguous: on the one hand, old firms have a priori more resources than young ones, but on the other hand, as the age of the firm is an indication of its date of creation, young firms may have been created to exploit the opportunities of Internet. Martin & Pénard (2005) shed light on this ambiguity: relatively old firms (compared with enterprises less than 10 years old) have a great probability to create a site, but the age of the firm seems to be a handicap both in the choice of creating an advanced website and in the choice of developing e-commerce.

Finally, the need to communicate with different units of the firms (Martin & Pénard, 2005), or with potential customers (Sadowski * Q2002), only introduced up to now in the first stage of the analysis of websites, seems to have a positive impact. It appears that multi-unit organisations have a huge potential to networking and have consequently a higher probability of ICTs adoption.

I A2*)) BFR9 * 94. 3 9 4 (-&*) (3 9 * . 3 * *&34 3 5 B 9* * * 9 * &9 4. (B D49 *B). 3 4 * -9* 9 ** 5 .D (.9 O

/ ΦØŒ. 9 99 3 3 9*39*

The creation and use of the website seem not only affected by the resources a firm possesses, but also by the characteristics of its business. As the business influence depends on the ICT studied, there is no theoretical work on the subject we focus on. According to Sadowski * Q(2002) or Chappell & Feindt (2000), some sectors like services and technology-intensive sectors (notably computing) are on the forefront of adopting and using Internet technologies compared to traditional sectors such as retailing. Thus, firms belonging to services and technology-intensive sectors have higher probabilities to create a website. Moreover, as the OECD highlights, 2T*U

*D4(. 3*9 4) 9 B3 9 A *4. 9 3 * 5- * (B 5)4 *4 34 * * 9. *( )4 D 3 9 4B * F* *E* (OECD, 2004, p.34). For Santarelli & D'Altri (2003) digitised or delivered on-line products facilitate e-commerce. According to Chappell & Feindt (2000), products that fit with lifestyles, aspirations of users have, at least for Business to Consumer, a high probability to be purchased on-line. As the characteristics of the product have an influence on the creation of a new business channel, it also influences the first stage, i.e. the creation of a site.

In empirical studies, the sectors are rarely introduced. When they are, no significance is most often displayed[18], as in Sadowski ** O.*(2002) for the creation of the site or in Lucchetti & Sterlacchini (2004) for the study of website content. To characterise the business of the firm, empirical studies sometimes introduce the presence of the firm on foreign market and the dependence with one customer. Lucchetti & Sterlacchini (2004) find, in their analysis of website content, that being an exporter and having commercial branches abroad influence both significantly and positively the richness of the site. In order to know whether or not preferential partnership with a customer is a brake on website and e-commerce deployment, Martin & Pénard (2005) test the influence of this dependence on the two last stages of website use. The regressions display no significant influence of such a dependency.

I A2*)) F6 *5 D4(. 3*9 * *)*) 9*& 9 4) . 9 9 4 * * *4 & (- 4 (9 4. ( ). 3 D49 *B &* D4- - *&4) 9 ** 5 . D -9* (4) . 9 5 * 99*)4 *B (V4 34 3 &O*

*/ P0O0R**:(9 (D 9* CD 3 9 * FGH9*

The cost of adopting a new technology has to include the costs associated with developing the new competencies needed to use correctly and to exploit for the best the new technology. These costs can be reduced if the firm has already some ICT competencies. Firms possessing specific competencies thanks to prior technologies use are more able to adopt new technologies. Moreover, thanks to different skills and experiences accumulated by the firm with prior ICTs adoption, it permits the firm to have an ICT culture and also a better perception of technologies benefits. Consequently, the firm is less risk adverse to new investment in ICTs. Finally, it

[18] Except in Martin & Pénard (2005), where the membership to industrial sectors compared to agro-food sector has a positive influence on website setting up, whereas retailing, transportation and professional services have non significant coefficient.

develops an in-house expertise, which avoids the firm to have to hand over responsibility for website creation, maintenance and updates to another firm. Thus, if the firm chooses to externalise, it resigns control over its website notably for updates, which can be a failure factor which contributes to website closes (Chappell & Feindt, 2000). Beyond the development of an ICT culture, the website setting up, and even more a commercial use of the site, needs to be associated with a business reorganisation to improve the performance of the firm. Thus the website need to be implemented as a part of a global strategy concerning the future development of the firm to ensure the success of the investment. In order to strengthen reactivity to customers' demands, the firm should have to adopt a flexible workplace organisation. While Milgrom & Roberts (1995) only consider two organisations of a firm, i.e. "mass production" and "modern manufacturing", Teece (1996) suggests that partial adoption of ICTs and new organisational practice is possible. So Bocquet and Brossard (2003) conclude that *2T*U ( )). 9 4 4) * 3 4 45&3 - 9. *4) * ( )). 9 4 4) 94 4 5 *4 D 3*3 9 (34 B 9 &E* (Bocquet & Brossard, 2003, p.3).

The empirical literature provides few tests of the impact of past ICTs use and of the introduction of tools that make more flexible the production and distribution processes. Dholakia & Kshetri (2004) find a positive impact of a composite indicator of prior technologies use[19] on both the stage of website setting up and the one of e-commerce practice. Martin & Pénard (2005) introduce some ICTs in the analysis of website creation and find that prior adoption of ICTs[20] has a positive influence on this stage. Furthermore, the probability of investing in an elaborate website increases with the quality of the Internet connection (Digital Subscriber Line (DSL) connection or leased lines), but the quality of the Web access has no influence on the investment in an e-commerce channel. Moreover, Martin & Pénard (2005) introduce in the two last stages some ICTs that can be interpreted as tools that make more flexible the organisation of the firm, such as Electronic Data Interchange (EDI) and Enterprise Resource Planning (ERP). The results stress on the positive impact of the existence of a technology that optimises the production and distribution processes such as EDI on the investment in an elaborate website. Furthermore, in the study of the factors that influence the setting up of a commercial website, the ERP, a technology that increases organisational flexibility and reactivity, has a positive influence.

[19] This composite indicator is composed of phone line, computer, fax machine, email account and toll free number.
[20] The ICTs introduced are local area network, pocket digital agenda, Internet access and mobile phones.

I A2*)) JFH 4 9 D4) 9D 3) 3 34 D* 3 9* A9*4 D 4 FGH9.9 (4) *3 4 45 9* * A * D4(.3*4 ((9*-.*4 D43 99 9 4) C- 9 4. ( ). 3 D49*B &* (4D*4 4) -9* (* 34 3 .9 4)* 9*O

2.2. External factors

The literature on ICTs adoption has underlined different external factors we classify into four groups.

*/ØØÞØG4 D **4*

According to Markides & Geroski (2004), the website opportunities in terms of commercial applications take the form of a "strategic investment". Actually, on the one hand, it implies a depreciation of existing advantages of established firms because it creates an alternative retailing channel that notably permits entry of "pure players" (like Amazon). But on the other hand, Internet has a minor effect on consumer habits, because it appears as a complementary channel for purchasing products (Martin & Pénard, 2005). The threat of entry encourages existing firms to adopt new technologies in order to keep their market power, to pre-empt as many rivals as possible thanks to the creation of an artificial entry barrier. So we can compare such a strategy to the "Top Dog"[21] of Fudenberg & Tirole (1984). In spite of a positive effect of competition on technology diffusion, it is possible that too much competition slows the diffusion process because it limits the possibility of extracting profits and high returns as the density dependence model conjectures (Geroski, 2000; Bocquet & Brossard, 2003).

In this connection, a market in which the competitive pressure is low (and so are the risks of being excluded from the market) the new technology has a high probability of being adopted and used. That's why, a niche market can be a good place to experiment with new technology and new business model (Chappell & Feindt, 2000).

In empirical studies as well as in the theoretical literature, it seems to be difficult to grasp the threshold that makes a distinction between the encouraging competition and the "too much" competition. As measures of competition like concentration ratio (i.e. the sum of the market share for the four biggest firms) are global measures of the intensity of competition, they don't capture this distinction (Martin & Pénard, 2005). It appears that the investment in an elaborated website

[21] This strategy implies an investment in order to eliminate existing or potential competitors.

or in a new business channel is negatively influenced by a high competition. This result implies that too much competition has a negative influence on the investment in the website. Finally, as far as we know, no empirical study has been conducted on the impact of evolving on a niche market on the probability of creating a website and using it.

I A2*)) KFH **4) *&9 4. ( ). 3 D49 *B &* B 9* * -9 * -. * *44 . 3 34 D **4 9 4. (B 5 *B D 3*O*

*/ ØØØ43 3 *B 9*

Before the decision of setting up a website and the choice of the degree of involvement, the firm needs information about the efficient website investment and the returns implied. Such information can be directly transmitted from current users who experienced the technology. This diffusion process by current users can perhaps be involuntary. Actually, information given by the success or failure of rivals adoption might influence potential users. Information cascades models show that it can be optimal for a firm, having observed the actions of others, to follow the behaviour of the preceding firms without regard to its own information on the efficiency of the technology (Geroski, 2000; Bocquet & Brossard, 2003).

In empirical studies, some papers seek to analyse the impact of the competitors' behaviours on two of the three stages we shed light on. Concerning the decision of a website creation, Sadowski * *Ø*(2002) introduce the actions of competitors but don't find a significant impact. Dholakia & Kshetri (2004) also analyse the actions of competitors as a pressure on the firms adoption motivations: they consider the perceptions a firm have about the uses of the Internet by competitors and they distinguish three uses concerning customer service, environmental monitoring and market research. The results display that the uses of Internet by rivals to improve customer services and for the supervision of the environment have a positive influence on the website creation. Dholakia & Kshetri (2004) introduce the same variables in their study of e-commerce setting up and find that a use of Internet to improve customer services by competitors has a positive influence. The characteristics of the territory where the firm is located can be introduced as a proxy for the intensity of communication between firms: for example, a location in a urban area can perhaps have a positive impact on website creation and use. This variable has been introduced in Martin & Pénard (2005), but the results show non significant effect of a urban location on the three stages of website setting up and investment.

I A2*)) LFFGH9 *(4D*4 (. 9 -& B 9 9 4. (- D49 *B & 9943 *(* *) (4D*4 4) -9* (. 9 O*

*/ ØØØØ7 A *D4*

Thanks to the Internet, firms have opportunities to reach a higher number of potential customers but the website investment still remain risky because the returns are not guarantied. These risks associated with this investment are a priori less important for well-known firms. According to OECD (2004), *24 3 *9B 345 *4 4) - (4 34 D & 9 (3 *4 4)) =9 3 (- *&L 9* 9* &(4 4)) E* (p.26). Thus, the brand image permits to avoid price competition or to maintain a value-for-money image[22] when the firm chooses to sell on-line (Brynjolfsson & Smith, 2000). As many customers still don't choose the lowest price and stay stick with a particular brand, well-known firms don't support a lot of risks. Studies on price competition on the web, such as Shankar, Rangaswamy & Pusateri (1998), show that the brand decreases consumer price sensitivity, which confirms the incitement for well-known firms to invest in a website and in an e-commerce channel.

As far as we know, no empirical study has introduced a measure of brand and renown of the firm to analyse their impact on the website setting up and the involvement in the use of the site.

I A2*)) MFH *4 9 D 4) - (9 4. (B D49 *B). 3 4 -9* 3 *4 (. 9 O*

/ ØOØ8 C 9. DD4 **

Before the adoption of a new technology, the firm needs information on the efficiency of the technology and on potential returns. These information can be fostered, for example, by business partners or by government support programs. Particularly, governmental policies can encourage the diffusion of website use and e-business practice thanks to the creation of *2) B4. - -. 9 99 B 4 * *) (D (3*- . 9E* (OECD, 2004, p.35). These actions undertaken by authorities can stand for the "central source" described in epidemic models (Geroski, 2000). This central source transmits information to potential adopters. These information reduce the search costs of the firm and, as a result, can induce a faster adoption of websites and beyond encourage the use of this Internet technology.

---

[22] *2G4 * &*4 * 34 4 B * ** [- 9 D4D *(* B4 * 3. 9*4 9 * [- 9 3*. & B & :9*3A&=9D 3 E* (Feindt, Jeffcoate & Chappell, 2002, p.58).

Sadowski * *Ɑ*(2002) assume that a support from partners can influence website setting up, but their empirical investigations don't highlight a significative impact. Martin & Pénard (2005) test the influence of a preferential partnership with a supplier on elaborated website and e-commerce deployment. The regressions display a negative influence of a dependence with a supplier in the analysis of the investment in an elaborated website. Poussing (2005) analyses the impact of the awareness of governmental initiatives undertaken in Luxembourg on the setting up of an e-commerce website. The empirical study introduces different initiatives like some actions instigated by European directives (i.e. the knowledge of electronic signatures, the code of electronic commerce, the "electronic commerce" committee, the Luxembourg office of accreditation that supervises the quality policy instigated by the code of electronic commerce, *3)O) *) and the most famous Luxembourg label that promotes e-commerce by developing customers trust and assuring the security of on-line transactions. The results display that only the quality label has a positive influence on the investment in an e-commerce website. This result confirms that confidence in an on-line retailer is essential to conclude transactions.

I A2*)) NF*8C* 9. DD4 *9 A 54B *D45 9 9 4. ( ). 3 D49*B & * ( )). 9 4 D43 994) -9*9 (34 3 O*

We want to test those hypotheses using data from firms located in Luxembourg. The next section describes the database.

JFO((

In the framework of eEurope action plans, EU countries, supervised by Eurostat, have conducted on an annual basis a survey on enterprises' access to ICTs and their use, and on e-commerce, since 2001. The aim of the project "ICT usage by enterprises" is to have comparable EU-level data to analyse the development of the Information Society. The data used in this study[23] relates to enterprises located in the Grand Duchy of Luxembourg. They were collected by the CEPS/INSTEAD[24] in collaboration with the Statec[25] and received financial support from the Community. The survey was sent by post, in the first half of the year 2005, to all enterprises

[23] Appendix 1 offers a full description of the survey data.
[24] CEPS/INSTEAD: Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques / International Networks for Studies in Technology, Environment, Alternatives, Development.
[25] The National Statistical Office of Luxembourg.

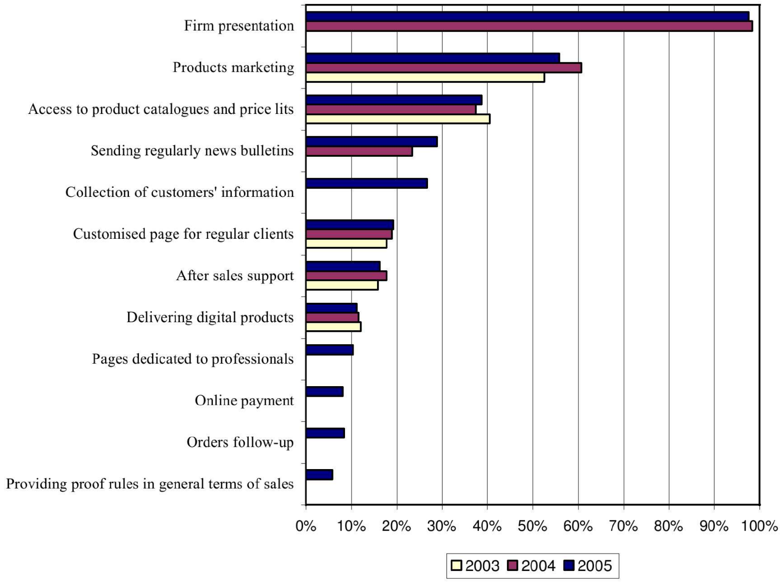
employing 10 persons and more, and exercising its activities in all the sectors of the economy except the financial one[26]. Therefore, 2472 enterprises were surveyed and 1161 answered (which entails a 47% response rate). The questionnaire takes over the Community one and inserts new items to analyse particularly the contents of firms' websites and the impact of government initiatives to promote ICTs adoption and e-commerce. The data gives some information about the characteristics of the firm surveyed and covers computers and communication technologies use, Internet access and use, e-security and sales, purchases via the Internet or other computer networks. As the questions concern the adoption and use of ICTs, the firms which are not equipped with computers (that is to say 29 firms) didn't need to respond these questions and consequently are excluded from the following of our study.

While for computerised firms located in Luxembourg the adoption of Internet reaches a rate of 95% in 2005 (96% in the EU-15), the rate of websites creation by firms employing 10 persons and more is smaller in the EU countries with a penetration rate of 67% in 2005 and in particular in the Grand Duchy where the rate is about 62%[27] (i.e. 698 of the respondents in our sample). Beyond the Internet presence of firms, the aim of our analyse is to focus on the nature of firms' websites. The data we use provide information on the content of websites. The survey gives details on most functionalities a website can offer to customers and suppliers.

[26] Since enterprises from the financial sector received a specific questionnaire different from other economic sectors, they were excluded from the data.

[27] Even though the Luxembourg is behind the EU-15 average in 2005, this phenomenon is recent. Effectively, in 2004, the Web presence of enterprises was around 62% both in the Grand Duchy and in EU-15.

H/. BFP -)). 3 * (4)



Note: firm presentation is provided in near 98% of the sites sampled.

As Figure 1 shows[28], the presentation of the firm is naturally the most provided function and is available in almost the whole sites. The website not only can offer the possibility of setting up a new retailing channel but can also be a way to canvass at a low cost new customers for the physical retailing channel. If we study in detail the second most frequent function (firms' products marketing), we notice that, for the 56% of firms which offer this function, about 73% received at least one order via Internet in 2004[29]. For the remaining percentage, the site represents a way to prospect customers, and to dynamise sales made in traditional stores. The availability of the product catalogue is also one of the most provided function since the introduction of the description of sites in the Luxemburger project "ICT usage by enterprises", but it appears more as a function to inform customers than a commercial one. Even if these functions show an interest for firms concerning e-commerce, on-line payments, after-sale support and a process of

[28] When prior questionnaires introduced the same functions, as the 2005 wave, in order to describe the nature of the site we provide information on the evolution of the content.
[29] Because the survey is conducted in the beginning of the year, the data on Internet sales and purchases concerns the year before.

following-up orders are not well-developed while they are important for the success of e-commerce.

Because the development of e-commerce is one of the most important priority and challenge for nationals and European authorities, we are going to look into on-line commerce in the Union and particularly in Luxembourg. The Grand Duchy is a country in which the on-line purchasing is important both for enterprises and individuals. Effectively, 41% of enterprises declared to order on-line products or services during 2004 and 34% in 2003 against a EU-15 average near 30% in 2003 and 2004. Concerning individuals, according to the Statec (2005), 32 % of the population used Internet to buy or order on-line in 2004 (18% one year earlier), which put the Luxembourg on top of the Union in terms of household on-line trade. Indeed, at the EU-15 level, we can notice that 21% of the population bought or ordered on-line in 2004 and 15% in 2003.

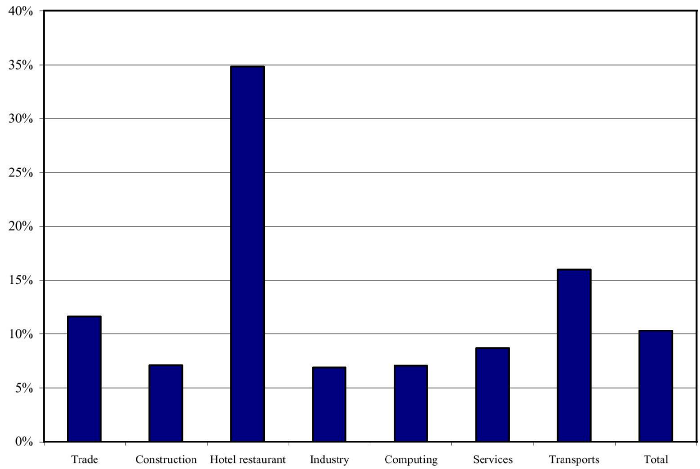
So, there is a real interest for on-line commerce in Luxembourg, but it does not result in a great development of direct on-line sales by enterprises. The Grand Duchy is still a very opened country concerning trade, and an important part of the purchases made on-line by enterprises and households are made abroad. Consequently, only 10.3% of firms received orders via Internet in 2004 and 11% in 2003, against 13% in the EU-15 in 2004 and 15% in 2003[30]. If we examine in details e-commerce, it exists sector-based differences, but because of a lack of data availability concerning purchases made by firms, we can only provide information on Business to Consumer (B to C) on-line trade. According to the Statec (2005) the products which are the most purchased on-line are books and magazines for which 59.2% of Internet users[31] declared at least a purchase in the first term of 2004. Trips come after with 47.7% and entertainment tickets with 33.5%. Pieces of clothing are purchased by 21% of Internet users. Then, CDs and DVDs of music and films are bought by 30.8% of Netsurfers. Software and electronic equipment are both bought by a little more than 20% of Internet users, and hardware equipment by 10.5%. Therefore, it seems that, at least concerning B to C, the sectors of commerce, spare time (like hotels and restaurants) and computing can exploit easily the dynamism of e-commerce. On the one hand, the sector of hotels and restaurants already exploits the opportunities given by e-commerce. Effectively, we can observe in our sample that nearly 35% of the firms exercising its activities on this sector

[30] If we limit the analysis of on-line sales to firms owning a website with the function of marketing the products, the percentage of firms which received on-line orders was about 23 in 2003 and 24.30 in 2004, but there was no comparison data available at the EU-15 level.

[31] The interest population is Internet users between 16 and 74 years.

received on-line orders during 2004. On the other hand, concerning firms exercising on the sectors of trade and computing, few of them received orders via Internet as Figure 2 displays, while the e-commerce seems to be important in these sectors.

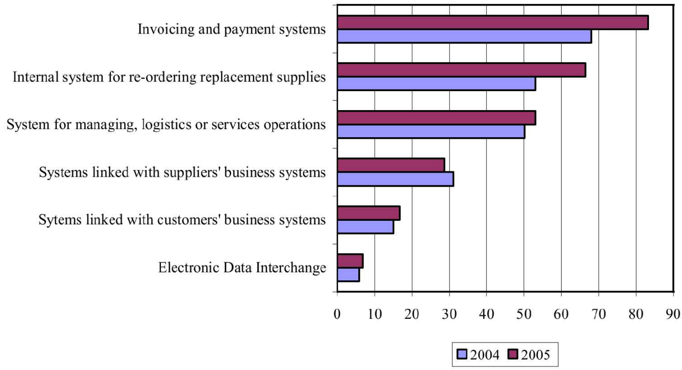
H/. F 4 * 3 2) - A) 3 *)



Note: on average, 10.3% of the firms sampled received at least an order via Internet in 2004.

Another priority for the Commission concerning e-commerce is the integration of this new distribution channel in the global strategy of firms, which implies the reorganisation of firms thanks notably to the adoption of management ICTs. The adoption of these ICTs effectively increases: 47.9% of firms in our sample have introduced management ICTs in 2005, against 43.4% one year before. As Figure 3 shows, these technologies are above all linked with the invoice system, the restocking one and logistic one. Associated with the website, such ICTs can enable the firm to reduce retailing costs. Moreover, because they permit to react rapidly to customers requirements, the firm can accelerate its distribution process in order to come up to customers' expectations and, consequently, can increase its markets shares and revenues.

H/. JF8 ((/ %)ES * %) (*2 /)A) % * %((/ *)T

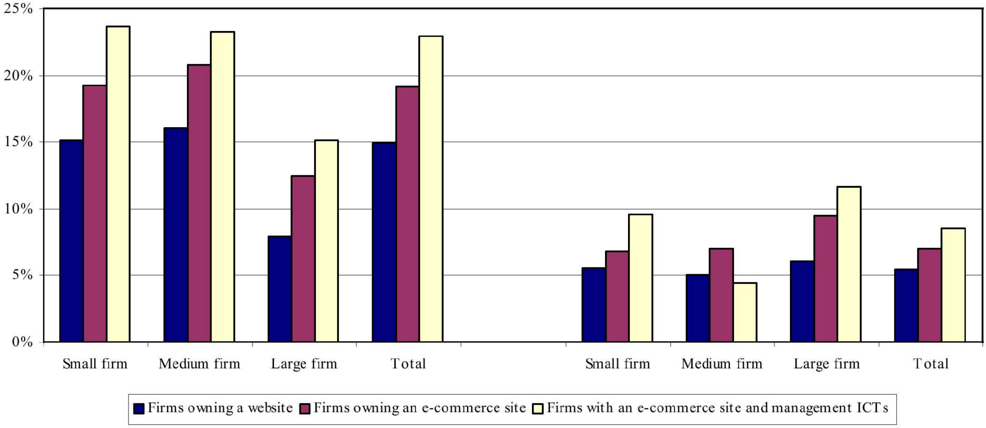


Note: in 2005, 83% of firms which introduced management ICTs chose to link these to invoicing and payment systems.

Items in the survey enable us to give some information concerning both market share increase and production and/or distribution costs reduction for firms with a website. For these firms, the advantages given by the website seem to locate more in the side of market opportunities than in the side of costs reduction. Effectively, for 15% of these firms the website permits to increase market share, while for only 5.5% it allows to decrease production and/or distribution costs. In order to analyse the importance of these two contributions of the website according to the implication of the firm both in e-commerce and in reorganisation, Figure 4 makes the link between integration of the website in the strategy of the firm and the consequences on market shares and costs reduction. Figure 4 shows this way that, more the firm gets involved in e-commerce and in the integration of this new distribution channel in its strategy via a reorganisation process, more the firm seems to benefit from the progress introduced by Internet. Moreover, concerning the increase of market share, it seems that both small and medium-sized firms take more advantages than large ones, and it seems to be the opposite in the side of costs reduction.

H/. KF 3* -. *)* D -) (A * (A4 * %)

8 (U) (3 () *) . 3 *



Note: 19% of firms owning an e-commerce website consider that their site contributes to increase market shares and near 7% consider that it permits to reduce costs.

KF * . 3 5((-4)

4.1. Dependent variables

Our first study is devoted to the analysis of the main determinants of the decision to invest in the creation of a website. Therefore, the dependent variable is a binary one taking the value 1 if the firm possesses a website and 0 otherwise. As said before, 698 firms in our sample own a website, that is to say 62% of the firms sampled.

In our second analysis, we focus on firms owning a website and study the characteristics of these firms depending on the use they make. Effectively, the website gives the firm possibilities to attain two main targets: inform customers and create a new business channel. Firms which use Internet to diffuse information on their activities do not have necessarily the same characteristics as the ones which create a new means to distribute their products via Internet. We distinguish the different functionalities the site provides between those used to inform customers and those introduced to create a on-line business channel, and we create three ordered variables capturing the intensity of the involvement in the use of a website.

The first one (information score) covers the uses of informing customers, i.e. six functions available on the site: a presentation of the firm, an access to product catalogues, customised page for regular clients, pages dedicated to professionals, the possibility of sending regularly bulletins on the current affairs of the firm, and finally the capacity of collecting information on Netsurfers visiting the site. Always in the perspective of understanding the involvement of the firm in the use of the website, we integrate the frequency of the update. The firm gets one point each time its website provides one function and we add no point if the update is made less than once a month, 1 if the update is made at least once a month and 2 if the update is made more than once a week.

The second variable (e-commerce score) groups together the uses linked to the creation of an e-commerce distribution channel. This score encompasses six functions: the marketing of the products, the delivering of digital products, the possibility of paying on-line, the diffusion of proof rules in general terms of sales, the access to the after sales support service department, and the possibility of following-up orders. We introduce the update in this variable in the same way than in the information score.

The last variable (total score) is created for a comparison with the others scores and makes no distinction between the needs met by the different functions provided in the site. So this variable makes the addition of the twelve functions a firm can provide on the website and introduces the frequency of the update.

In conclusion, the variables are constructed like this:

& *0* if no function available on the site and the update is made less than once a month;

& *P* if only one of the functions is provided on the site and the update is made less than once a month;

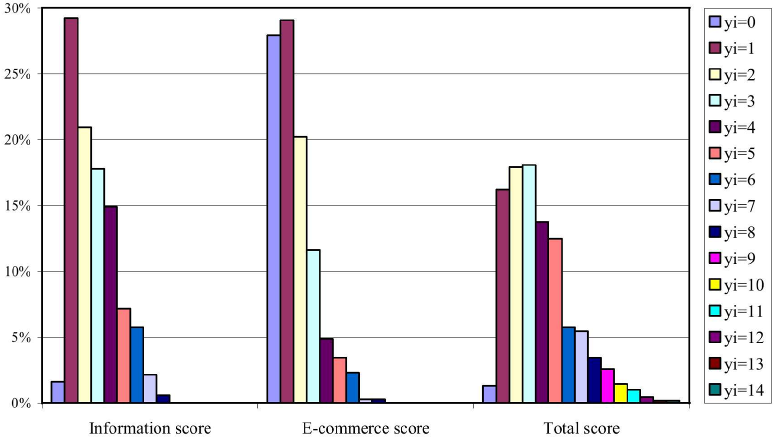
& *]* ;

& if all the functions are provided and the update is made more than once a week.

for the information score and the e-commerce score and *P* for the total score.

Figure 5 displays the distribution of these scores in the sample. As Figure 5 shows, there are a limited percentage of firms with & *0* in the information score and the total score compared with the e-commerce score, where more than 25% of firms owning a website do not propose at least one e-commerce function. The total score is on average 3.81. The information score is on average 2.74 while the e-commerce score is on average 1.59.

H/. LFO)3 2 * *)3*)



Note: 29% of firms owning a website obtain 1 point for both their information score and their e-commerce one. 16% obtain 1 point for their total score.

In a third analysis, we concentrate on an important concern for European authorities: the need of reorganisation for a firm which wants to succeed in e-commerce. So we analyse the link between the creation of an e-commerce website and the setting up of ICTs which allow the firm to accelerate its production and/or distribution processes. On the one hand, we define an e-commerce website as a site providing transactional functionalities, i.e. a site with at least one of the functions introduced in the e-commerce score[32]. On the other hand, as a proxy for the reorganisation a firm can engage, we use both the adoption of at least one of management ICTs[33] introduced in the survey and those ICTs taking one-by-one.

[32] For the record, transactional functions are: products marketing, digital products delivering, on-line payment, the diffusion of proof rules in general terms of sales, after sales support and/or order monitoring.

[33] For the record, management ICTs are: invoicing and payment systems, internal system for re-ordering replacement supplies, system for managing, logistics or services operations, systems linked with suppliers' business systems, systems linked with customers' business systems and electronic data interchange.

4.2. Explanatory variables

In order to determine which factors influence the website creation behaviours, involvement in the use of the site, and the decision to reorganise the firm via the adoption of management ICTs, we introduce different variables. As described previously, explanatory variables are split into two groups: internal factors and external ones.

*ØŒPF *) 3*4 9*

First, we want to analyse the influence of the resources possessed by the firm. So we introduce variables capturing the size of the firm, its organisational structure and a proxy for human resources capabilities[34]. In order to pick up the effect of size, we use the classification of the European Union concerning small, medium-sized and large firms. Therefore, we create three dichotomous variables concerning the number of employees (10-49, 50-249, 250+). To measure the influence of the organisational structure, we introduce two variables, the first taking the value 1 if the firm has more than one legal unit in its organisation, and the second taking the value 1 if the firm is a subsidiary of a group. Because we don't have information on the qualifications of managers and of the workforce, we introduce a proxy for human capabilities to use ICTs. This proxy is, in fact, a measure of the possibility for employees to attend ICTs training. These courses are notably devoted to a better use of operating systems, software, Internet, etc.

Secondly, we insert variables to characterise the business in which the firm operates: we introduce either sector of activities established across Luxembourg[35] or the nature of customers (the dichotomous variable takes the value 1 if customers are predominantly households), depending on the specification. Then, we introduce a variable to analyse the influence of being an exporter. Because of the country size and so of a high probability of making trade with foreign customers, we create a variable taking a value between 0 and 100 in accordance with the percentage of the firm's sales realised abroad. Finally, to test the effect of a dependence with a customer, another dummy variable indicates whether the firm concludes over 30% of its sales with a single customer.

Thirdly, we introduce variables for the level of adoption and use of two types of technologies: common communicant ICTs such as local computer network, Intranet, etc., and management

[34] As we have no information concerning the date of the setting up of the firm, the age of the firm is not introduced in our analysis.

[35] The different sectors are: industry; construction; trade; hotels, restaurants and cafés; transports; computing and services. As said before, the financial sector is excluded from the "ICT usage by enterprises" project.

ICTs presented above. Depending on the specification, we insert these ICTs one-to-one or as a composite variable. We use two composite indicators, one for common ICTs and one for management ICTs. The indicator takes the value 0 if the firm has no ICT proposed in the survey, and 1 or more depending on the number of technologies owned.

Another variable is introduced in the regressions concerning the sample of firms owning a website. This variable characterise the age of the site. More precisely, it is a category variable which is dependent on the date of the setting up of the website.

ØØŒ8C   ) 3*4 9*

First, we insert two groups of variables characterising the competition on the market. On the one hand, to capture the perception of the firm concerning the intensity of competition on its market, we create three dichotomous variables dependent of the intensity perceived: limited, intense and very intense. On the other hand, to have information on the importance of firm's position on a market, we insert four dichotomous variables (niche, follower, challenger and leader). Particularly, the variable "niche" takes the value 1 if the firm develops a niche on the market, i.e. if it exercises its activities on a good place to experiment with a new technology and a new business model.

Secondly, we introduce a variable to define the local incentives. As Luxembourg is a little country (the most distant town and the capital city are 70 kilometres away from each other) we don't think appropriate to introduce a variable characterising a location in a urban area. That's why, we insert a dichotomous variable capturing the setting up of an e-commerce website by main competitors. Thanks to this variable, we want to measure the influence of behavioural imitation.

Thirdly, in order to grasp the impact of the renown of the firm on creation and investment decisions, we create three binary variables (local, national and international), in accordance with the perception the firm declared concerning its brand and trade name.

Finally, we want to study the impact of external support the firm can have when it has to choose between investing or not in a website and in the use. So we introduce a dummy variable that takes the value 1 if the firm transacts over 30% of its purchases with a single supplier, and 0 otherwise. After, we include different European and Luxembourg initiatives designed to encourage firms to adopt and use ICTs and e-commerce and implemented within the scope of eEurope action plans. An item in the survey lists the main public authorities actions present in the

Grand Duchy. As for the analyse of prior ICTs use, we insert these initiatives one-to-one or as a composite variable depending on the specification. Eight public actions are included in the composite indicator which measures the intensity of knowledge concerning the different initiatives. These actions have different goals. Some public actions seek to promote use of ICTs thanks to the creation of communication platforms like "Handwriek-online" provided by the guild chamber or the portal "entreprises.lu"[36] provided by the government. These platforms aim to share out information, knowledge on best practise between firms, administrations, guild chamber and chamber of commerce, in order for firms to save time, money and efforts when they choose to adopt new technologies. There is another portal developped by the Ministry of Economy and Foreign Trade, focusing on information security. This portal named CASES[37] (Cyberworld Awareness Security Enhancement Structure) allows firms to get used to the theoretical side of network security and practical applications. Concerning the development of e-commerce there are many initiatives taken by the Luxembourg. Some actions are stimulated by the transposition of European directives. We find first the law related to Public Key Infrastructure (PKI) or electronic signatures adopted on September 8th 2000 and derived from an European directive[38]. This law was set up to assure security for trade, electronic administration and confidential data transfer. From this point of view, the government created in June 2001 the "electronic commerce" committee linked with the Ministry of Economy and Foreign Trade in order to verify the use of electronic signatures. Secondly, we find the code of electronic commerce. This code recovers both national and European legislation concerning e-commerce. Thanks to the law adopted on July 5th 2004, the Grand Duchy is in accordance with the European plan relating to electronic commerce. It allows notably to clarify protective rules concerning customers and to add a point to authentication organisms that allocate quality labels. The Luxembourg Office of Accreditation[39] (OLAS) was created to supervise the quality policy established by the code of electronic commerce. OLAS accredits notably the organisms that manage and deliver quality certificates and labels. The most important label concerning quality and security of e-commerce website is called "Luxembourg e-commerce certified". The attribution of this label is subordinated to an audit carried out by an organism accredited by OLAS. This label promotes e-commerce by developing customers trust and confidence, and

[36] For more information on the platform "Handwriek online", see: http://www.chambre-des-metiers.lu and on the portal "entreprises.lu", see: http://www.entreprises.public.lu.
[37] http://www.cases.public.lu
[38] 1999/93/EC (Official Journal of the European Communities, 2000).
[39] http://www.olas.public.lu

assuring the security of on-line transactions. Moreover, it offers guarantees concerning storage and treatment of information collected on customers.

LF1 3* *% 3). 4)

5.1. The determinants of a website setting up

Our first analysis is dedicated to the study of a website creation decision by a firm thanks to a binary Probit model presented in Box 1.

***7 BF8 * *4*/ A* - (AW*- %* 4**

The binary Probit model is used to study the factors that influence the decision of creating a website. The dependent variable (&) is dichotomous: it is ascribed the value 1 if the firm has a website and 0 otherwise. The estimate needs the definition of a latent variable in order to model the realisation of this event. This unobserved latent variable (&) corresponds to the net profit (or economic rent) obtained by the firm ($P(\!\!\!/D)$) when creating a website.

The latent variable is given by: $\&^{*} = \beta C' + \mu$

with C the vector containing the explanatory variables, β the vector of parameters relating to the variables C, and μ the random error term.

Prob(&=1) = P($\&^{*} \succ 0$) = $\mathrm{P}\left(\frac{\beta C'}{\sigma_\mu} + \frac{\mu}{\sigma_\mu} \succ 0\right) = \mathrm{P}\left(\frac{\mu}{\sigma_\mu} > \frac{-\beta C'}{\sigma_\mu}\right) = 1\text{-F}\left(\frac{-\beta C'}{\sigma_\mu}\right)$

Prob(&=0) = P($\&^{*} \le 0$) = $\mathrm{P}\left(\frac{\mu}{\sigma_\mu} \le \frac{-\beta C'}{\sigma_\mu}\right) = \mathrm{F}\left(\frac{\beta C'}{\sigma_\mu}\right)$

where , represents the cumulative distribution function.

$$, = \int_{-\infty}^{-\beta C'/\sigma} \frac{1}{\sigma_\mu \sqrt{2\pi}} . \exp\left[(-1/2)\cdot(\varepsilon)^2\right] \cdot (\varepsilon$$

In order to estimate the parameters β associated with $_a$, the solution usually retained consists in the normalisation of the standard error $_a$ to 1.

To obtain the estimated parameters, we maximise the log-likelihood function presuming that the error term is distributed normally with a mean equal to 0 and a variance equal to 1.

This log-likelihood function can then be written as:

$$Y45Y = \sum_{=1} \left[(\& \log(1 - , \ (-\beta C')) + (1 - \&) \log , \ (-\beta C') \right]$$

Results from the analysis of the creation of a website are presented in Table 1.

*bOPOPF *) 3*4 9*

Looking at the significance of the different variables concerning the firm resources, we find that financial resources have a positive influence on the decision of creating a website. Thus, small organisations (10-49 employees) have a lower probability of creation compared to mid-sized and large firms. Small firms seem to adopt a "wait-and-see" attitude. As expected, an organisation characterised by a high potential of networking, of knowledge sharing (i.e. a multi-unit organisation) has, ceteris paribus, a probability of possessing a website increased by more than 8% compared with organisations with only one unit. The results about our proxy for human competencies display a positive influence. So, when a firm enables its employees to attend ICTs training, the probability of owning a website is raised by about 8.5% whatever the estimated model (Model 1 or Model 2). Finally, being a part of a group (that can provide both financial and human resources) don't seem to have an impact in the choice of creating or not a website. Therefore, the results substantiate for the most part the hypothesis 1 formulated above.

Firms in the transport, computing, services and hotels/restaurants sectors have a probability of creating a website greater than firms in the industry, construction and trade sectors. Thus, technology-intensive sectors (computing) and sectors selling products that fit the lifestyle of consumers (tourism i.e. hotels/restaurants in our study) are different from others in term of their effect on the probability of possession. So the probability of having a website is increased by 28.6% when the firm evolves on the computing or tourism sectors and only 10-12% when the firm evolves on transports or services sectors. The fact of exporting products, the quality of the customers (i.e. household or firms), and a dependence with one customer reveal insignificant in all models. Thus, the results concerning the business characteristics don't really support the hypothesis 2.

Concerning attitudes and past experiences with ICTs, the results show the expected sign for the common communicant ICTs. The more a firm has adopted ICTs, the higher its probability of website creation is. Having a DSL connection to Internet, whatever the quality, has a positive influence. If we study in details individual impacts of ICTs introduced in the ICTs score (*3)O* Appendix 2, Table 4), we notice that the ICTs that have the greatest influence on website creation decision are the Local Area Network, the electronic forum, the presence of an electronic working group calendar and the use of an Intranet. About the adoption of management ICTs set up in order to make the organisation of the firm more flexible, the results show no significant parameters for both the score and the individual technologies (3)O Appendix 2, Table 4). Therefore, the regression results go in the direction of the hypothesis 3 only for common ICTs.

bOPOOC8 C   ) 3*4 9*

On the one hand, the intensity of competition on the market do not have an influence on the decision of creating or not a website. On the other hand, being the leader on the market has a positive impact on the setting up: the probability of having a site is increased by more than 11%, all else equal. In Model 2, in which we introduce the quality of the customers instead of the sectors in which the firm operates, it emerges that firms which develop a niche have a probability of creation raised by 7% ceteris paribus. As the intensity of competition has no significant coefficient, the hypothesis 4 is not substantiated in our study.

The possession of e-commerce websites by rivals has a positive influence on the website creation decision for an individual firm. This shows that following the behaviours of other firms is an important determinant for firms in establishing a presence on the Web. This results support the hypothesis 5 formulated above.

A firm that holds a brand or a trade name with a national renown has a probability of website creation increased by 10% compared with a firm with a local renown, ceteris paribus. An international renown has no influence on website creation. As at least one level of the renown is significant, the results support the hypothesis 6.

Neither a support given by a favoured partner (like a supplier) nor the fact for the firm to know government initiatives have a significant influence on the creation decision. If we watch in detail the regressions that introduce the different Luxembourg and European programs, it emerges that only the platforms developed by the guild chamber ("Guild chamber") and by the government ("Entreprises.lu") have a significant influence. But the coefficient of the guild chamber's platform "Hanwriek-online" (known by about 20% of the firms of the sample) is negative. The information available on this platform seems to reinforce the fear of failure and the fear concerning the investment returns associated with the creation of a website. Therefore, the results don't support the hypothesis 7.

(-4 BFH(3 *) 4 3 / 3 (* * (D -)

	8 * 4B		8 * 4	
	* H F	8 (/F F	* H F	8 (/F F
Constant	-0.6952** (0.3393)		-0.5248* (0.3175)	
(4 (3 *)				
Firm resources				
Small	-0.6287** (0.2833)	-0.2156	-0.5649** (0.2796)	-0.1972
Mid	ns		ns	
Large	Ref.		Ref.	
Multi-unit organisation	0.2501* (0.1414)	0.0890	0.2417* (0.1396)	0.0868
Group	ns		ns	
ICTs training	0.2286** (0.0960)	0.0837	0.2366** (0.0945)	0.0871
Business characteristics				
Industry	Ref.		-	
Construction	ns		-	
Trade	ns		-	
Horesca	1.0638*** (0.2789)	0.2860	-	
Transports	0.2879* (0.1734)	0.1015	-	
Computing	1.0545*** (0.3147)	0.2860	-	
Services	0.3643** (0.1641)	0.1277	-	
Household	-		ns	
Exports	ns		ns	
Dependence with a customer	ns		ns	
Attitudes and past experiences with ICTs				
ICTs score	0.1761*** (0.0345)	0.0654	0.1885*** (0.0329)	0.0704
Low connection	Ref.		Ref.	
Low DSL connection	0.4191*** (0.1024)	0.1503	0.4102*** (0.1012)	0.1483
High speed DSL	0.4341*** (0.1059)	0.1551	0.4564*** (0.1050)	0.1638
Management ICTs score	ns		ns	
17 (4 (3 *)				
Competition				
Limited competition	ns		ns	
Intense competition	Ref.		Ref.	
Very intense competition	ns		ns	
Leader	0.3335** (0.1361)	0.1183	0.3134** (0.1332)	0.1124
Challenger	ns		ns	
Follower	Ref.		Ref.	
Niche	ns		0.1990* (0.1177)	0.0725

		Local incentives		
Rivals have an e-commerce website	0.3132*** (0.0939)	0.1132	0.3494*** (0.0922)	0.1267
		Market power		
Local renown	Ref.		Ref.	
National renown	0.2896** (0.1210)	0.1072	0.2693** (0.1206)	0.1003
International renown	ns		ns	
		External support		
Supplier support	ns		ns	
Governmental initiatives score	ns		ns	
6 6	-621.4835		-635.0176	
W . * :	0.1753		0.1573	
*** 3 4A34())**	72.26%		70.85%	

*, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Standard errors are shown in brackets; ns: not significant; Ref.: reference group.
Note: in Model 1, the variable "multi-unit organisation" is significant at the 10% level. When a firm integrates different establishments, its probability of possessing a website is increased by 8.9%.

5.2. The determinants of the involvement in the use of a website

Our second analysis is dedicated to the study of the involvement of the firm in the use of its website in an informative way or in a commercial way, using count data models presented in Box 2.

***7 F8 * *4*/ A* 3*. ((%* 4** [K]

Models of count data (Poisson and Negative Binomial models) are used to determine the total obtained for the different scores we calculate, i.e. the number of functions and the quality of the update we can find on the site of a firm. The dependent variable is a count data which can take on the values: *0 P/]* with for the variables information score and e-commerce score, and *P* for the variable total score.

The Poisson regression model specifies that each $\&$ is drawn from a Poisson distribution with parameter λ which is related to the variables C.

$$\text{Prob}(c = \& / C) = \frac{^{-\lambda} . \lambda^{\&}}{\&!}$$

The most common formulation for λ is the log-linear model : $\ln \lambda = \beta' C$.

The expected score is given by: $8[c \ / C] = d \quad [c \ / C] = \lambda = \ ^{\beta' C}$ with $c = \ ^{\beta' C} + .$

To obtain the estimated parameters, we use maximum likelihood techniques.
The log-likelihood function can be written as:

$$Y45Y = \sum_{=1} [-\lambda + \&.\beta' C - \log \&!]$$

[40] For more information, see Greene (2003), Chapter 21.

Sometimes, we are confronted with a problem concerning the Poisson regression because of its implicit assumption that the variance of & equals its mean.
The negative binomial model (Hausman, Hall & Griliches, 1984; Cameron & Trivedi, 1986, 1990) relaxes the Poisson assumption: it's a natural formulation of cross section heterogeneity. It introduces an individual unobserved effect into the conditional mean, that captures an overdispersion existing in the data:

$$Y45(\lambda ..) = \beta' C + \varepsilon = Y45\,\lambda + Y45\ .$$

The distribution of & conditioned on C and . (i.e. e) remains Poisson with conditional mean and variance a is: $)(\&/C,.\ ) = \dfrac{{}^{-\lambda ..}\ .(\lambda ..\ )^{\&}}{\&!}$

The unconditional distribution of $)(\&/C)$ is: $\int_0^{+\infty})(\&/C,.\ ).5(.\ ).(.$

Assuming that the distribution of . is a gamma distribution with a mean normalised to 1, we obtain: $5(.\ ) = \dfrac{\theta^{\theta}}{\Gamma(\theta)}\ .\ ^{-\theta ..}\ ..\ ^{\theta-1}$

The density of & is then: $)(\&/C) = \dfrac{\Gamma(\theta+\&)}{\Gamma(\&+1).\Gamma(\theta)}\ .\ ^{\&}.(1-\ )^{\theta}$

where $= \lambda /(\lambda + \theta)$ with f the parameter of the gamma distribution. This is one form of the negative binomial distribution.
The distribution has conditional mean: λ and conditional variance: $\lambda\,(1+(1/\theta)\lambda\,)$.

The negative binomial model (NegBin) can be estimated by maximum likelihood.

To test the Poisson distribution against the negative binomial one, we test the hypothesis f=0 using a likelihood ratio test. The limiting distribution of the LR statistic is chi-squared with one degree of freedom. In accordance with this test, we present the non rejected model in Table 2.

Results from the analysis of the degree of involvement in a website use are presented in Table 2[41]. As opposed to the assumption we made when distinguishing between the different uses of the site, firms which use Internet to diffuse information on their activities have characteristics relatively close to firms which create a new commercial channel to distribute their products via the Web, except for few specific variables. In the remainder of our study, we don't comment specifically the results concerning the total score, in order to focus our attention on the two uses the site can provide.

[41] We work on the sub-sample of firms having a website. In order to verify that the choice of the content is independent from the choice of creation, we use three tests of independence between two binary decisions (LR Test). The first one is between the choice of creation and the choice of a number of functions superior or equal to the mean of the number of functions in the sites of the sample. The second one is between the choice of creation and the choice of a number of information functions superior or equal to the mean of the number of information functions provided by the sites of the sample. The third one is between the choice of creation and the setting up of an e-commerce website including at least one of the six functions considered in the e-commerce score. These tests (not displayed in the paper) show that the decisions are independent.

*bØØPOF *) 3*4 9*

The resources owned by a firm don't seem to have a great importance in the decisions of using intensively the site in an informative or in an commercial way. Only the fact of being a subsidiary of a group seem to have a (small) impact on the investment decision in informative functions. Therefore, the results don't really substantiate the hypothesis 1 formulated above.

Concerning the business characteristics, it appears that selling products that fit the lifestyle of consumers as the sector of tourism does have a positive influence on the degree of involvement in the website[42]. A little difference emerges between the informative use of the site and the commercial one. The positive effect seems, indeed, to be greater in the case of an investment dedicated to the creation of a new commercial channel. The expected number of commercial functions for firms evolving on tourism sector is about 45% higher than for firms of other sectors; and the expected number of informative functions for firms evolving on tourism sector is about 27% higher than for firms of other sectors. In the group of variables which characterise the business of the firm, the quality of the customers, the fact of selling abroad, and the dependence with a single customer highlight some differences between the two uses we distinguish. The fact that customers are predominantly households has a positive influence on the degree of involvement in the site for an informative aim, and no influence in e-commerce involvement. Selling products abroad and the dependence with one customer have only an influence on the commercial use of the site. The effects are both negative: this way, the number of commercial functions is reduced by about 0.03%[43] if the percentage of exports increases by 0.1, and by about 22% if an independent firm signs a contract with one customer for 30% or more of its sales. Therefore, the results substantiate for the most part the hypothesis 2 formulated above.

Concerning the attitudes and past experiences with ICTs, the ICTs score is significant only in the degree of involvement in informative use of the site. The setting up of an ICT culture and also a better perception of technologies benefits allow the firm to have the control of its site only for the competencies accumulated thanks to prior ICTs use. Thus, the firm has a greater facility to invest in an informative site instead of in commercial site for which management competencies are needed. If we observe in detail the impact of individual technologies introduced in the ICTs score

[42] The negative coefficient associated with the variable "Construction" shows that traditional sectors lag behind other sectors concerning the investment in a website.
[43] This negative effect seems contradictory with the hypothesis 2 but is very weak.

*(3)O*Appendix 3, Table 5), it appears that only the setting up of a group project scheduler has a positive[44] influence on the involvement in the site whatever the use considered. The prior use of management ICTs has, this way, a positive influence on the different degree of use of the website. Concerning the age of the site, the results highlight that the older the site is, the higher the expected number of both informative[45] and commercial functions are. In conclusion, the results substantiate the hypothesis 3 formulated above totally in the case of an informative use and partially in the case of a commercial use.

bØØŒ8C   ) 3*4 9*

Concerning the effect of competition on the degree of involvement in the website, the results highlight a difference between the two scores. On the one hand, as we discussed in the theoretical and empirical literatures, "too much" competition ("Very intensive competition") has a negative impact on the investment in an e-commerce website, and all positions on the market are non significant. On the other hand, if we observe the results of the information score, we find that the intensity of competition has non significant effects on the number of functions chosen, but being the leader on the market has a positive influence on the investment in a site that provides more information to potential customers. Such a position removes the risk of being excluded from the market if the investment failed. The regression results go in the direction of the hypothesis 4 at least partially for the commercial use of the site.

Concerning the effect of the behaviours of other firms on the decision of an individual firm to invest in an informative site and/or an e-commerce site the follow-up of rivals is important (as for the decision of creation). So, the results substantiate the hypothesis 5.

The fact of having a brand or a trade name known at an international level has a positive influence on the investment, especially, in an e-commerce website. The expected number of commercial functions for firms with an international renown is about 23% (or 30% according to the model studied) higher than for firms known locally or nationally. As at least one level of the renown is significant, the results support the hypothesis 6.

[44] The mail service seems to have a negative influence on the involvement in an informative use of the site, because the mail can be sufficient for the firm to give information to its customers.
[45] Note that the coefficient is only significant in the Model 2.

Both a support given by a favoured partner (like a supplier) and the fact for the firm to know a lot of government initiatives have a significant influence on the number of functions adopted by firms for their sites. The results display that a privileged relationship with a supplier has a positive influence that increases the degree of involvement in the website use. Knowing a lot of Luxembourg and European initiatives concerning ICTs and e-commerce deployment has a positive influence on the degree of use of the website. In details (*3)O*Appendix 3, Table 5), the guild chamber platform that diffuses information on best practices concerning firm adoption of technologies has a positive influence on the information score (but only in the Model 1). The electronic signature law, set up to ensure security for trade, e-government and confidential data transfer, has a positive influence on the use of the site in both an informative and/or a commercial aim. The results go in the direction presented in the hypothesis 7.

(-4 FH(3 *) 4 3 / %)) A* 5*45 % .) * (D -)

	* (4)3*		* %(*)3*		1 3*%% 3)3*	
	8 * 4B	8 * 4	8 * 4B	8 * 4	8 * 4B	8 * 4
4 3 %* 4	/	/	W*))*	W*))*	/	/
	* H F	* H F	* H F	* H F	* H F	* H F
Constant	0.6535*** (0.1473)	0.5256*** (0.1373)	0.4018** (0.1604)	ns	ns	-0.4552** (0.2154)
			(4 (3 *)			
			Firm resources			
Small	ns	ns	ns	ns	ns	ns
Mid	ns	ns	ns	ns	ns	ns
Large	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Multi-unit organisation	ns	ns	ns	ns	ns	ns
Group	ns	0.088* (0.0520)	ns	0.097* (0.0559)	ns	ns
ICTs training	ns	ns	ns	ns	ns	ns
			Business characteristics			
Industry	Ref.		Ref.		Ref.	
Construction	-0.1552* (0.0871)	-	-0.2059** (0.0961)	-	-0.2876** (0.1395)	-
Trade	ns	-	ns	-	ns	-
Horesca	0.3054*** (0.1170)	-	0.2720** (0.1272)	-	0.4448** (0.1782)	-
Transports	ns	-	ns	-	ns	-
Computing	ns	-	ns	-	ns	-
Services	ns	-	ns	-	ns	-
Household	-	0.1034* (0.0574)	-	0.1362** (0.0619)	-	ns
Exports	ns	ns	ns	ns	-0.0039*** (0.0013)	-0.0031** (0.0012)
Dependence with a consumer	-0.1356** (0.0603)	-0.1299** (0.0598)	ns	ns	-0.2268** (0.0971)	-0.2247** (0.0963)

			Attitudes and past experiences with ICTs			
ICTs score	0.0264*	ns	0.0408**	0.0385**	ns	ns
	(0.0155)		(0.0170)	(0.0164)		
Low connection	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Low DSL connection	ns	ns	ns	ns	ns	ns
High speed DSL	ns	ns	ns	ns	ns	ns
Management ICTs score	0.0613***	0.0594***	0.0416**	0.0423***	0.0780***	0.0713***
	(0.0148)	(0.0138)	(0.0163)	(0.0150)	(0.0230)	(0.0215)
			Characteristic of the website			
Age of the site	0.0232*	0.0329**	ns	0.0267*	0.0374*	0.0562***
	(0.0128)	(0.0128)		(0.0139)	(0.0202)	(0.0202)
			17 (4 (3 *)			
			Competition			
Limited competition	ns	ns	ns	ns	ns	ns
Intense competition	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Very intense competition	ns	ns	ns	ns	-0.1410*	-0.1411*
					(0.0720)	(0.0722)
Leader	0.1478**	0.1408**	0.1484**	0.1485**	ns	ns
	(0.0681)	(0.0680)	(0.0750)	(0.0742)		
Challenger	ns	ns	ns	ns	ns	ns
Follower	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Niche	ns	ns	ns	ns	ns	ns
			Market power			
Local renown	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
National renown	ns	ns	ns	ns	ns	ns
International renown	0.1366*	0.1850**	ns	0.1898**	0.2246*	0.3015**
	(0.0827)	(0.0839)		(0.0911)	(0.1315)	(0.1336)
			Local incentives			
Rivals have an e-commerce website	0.2102***	0.2364***	0.1501***	0.1750***	0.3291***	0.3637***
	(0.0448)	(0.0444)	(0.0492)	(0.0481)	(0.0699)	(0.0693)
			External support			
Supplier support	0.1079**	0.1202**	ns	0.1025*	0.1697**	0.1848**
	(0.0492)	(0.0483)		(0.0523)	(0.0764)	(0.0751)
Governmental initiatives score	0.0409***	0.0414***	0.0356***	0.0353***	0.0565***	0.0571***
	(0.0110)	(0.0110)	(0.0119)	(0.0119)	(0.0170)	(0.0171)
	0.0383**	0.0454***			0.0993**	0.1179***
	(0.0158)	(0.0162)			(0.0404)	(0.0420)
6 6	-1445.9956	-1454.141	-1226.3687	-1233.566	-1091.0442	-1100.3194
6 U 4 ** (*) * 0	chi2 (1) = 7.36 Prob≥chi2 = 0.003	chi2 (1) = 10.18 Prob≥chi2 = 0.001	chi2 (1) = 0.00 Prob≥chi2 = 0.500	chi2 (1) = 0.00 Prob≥chi2 = 0.500	chi2 (1) = 7.91 Prob≥chi2 = 0.002	chi2 (1) = 10.80 Prob≥chi2 = 0.001

*, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Standard errors are shown in brackets; ns: not significant; Ref.: reference group.

Note: for the information score (Model 2), the variable "household" is significant at the 5% level. The expected number of informative functions for firms which have predominantly households as customers is about 14% higher than for firms which have predominantly firms as customers.

5.3. The decision to reorganise the firm and the setting up of an e-commerce website

As one of the new priorities of European authorities is given to the development of e-business, we want to analyse briefly the connection that can exist between management ICTs and the setting up of a new business channel thanks to the opportunities given by the Web. If we observe the results displayed in Table 5 (*3)O*Appendix 3), we notice that the fact of possessing a management ICT linked with the business system of customers has a positive influence on the number of e-commerce functions provided on websites. In order to go further, we conduct a simultaneous study of the creation of an e-commerce website and the adoption of management technologies (ICTs taking in general and taking ICTs individually). For this purpose, we use a bivariate Probit model that permits to estimate jointly the probability of two simultaneous events and to evaluate the correlation degree between errors terms. The procedure consists, first to regress simultaneously the creation of an e-commerce website[46] and the adoption of management ICTs with the explanatory variables of the specification "Model 1" presented above (*3)O*Table 2), except the variable "Management ICT score", and second to calculate the correlation coefficient (*g*) between the errors of the two estimates. Then, we study the significance of the coefficient associated with *g*. We use a likelihood-ratio test which tests the null hypothesis $h_0 J g \; 0$. If *g* is different from 0 and positive, we can conclude, when the sign is positive, that there is a complementarity between the e-commerce website and management ICTs. The LR tests[47] are presented in the Table 3, and if a complementarity is found, we provide *g* to known its sign.

The results highlight a positive relationship between management ICTs whatever the ICT considered. So, one factor in the success of e-commerce is the adoption of ICTs that make more flexible the organisation in order to quickly come up to customers requirements. In details, as in the NegBin regression displayed in Appendix 3 (Table 5), this study confirms the fact that the possession of management ICTs linked with the business system of customers is complementary with the setting up of an e-commerce website. This specific analysis show that the adoption of ICTs for managing, logistics or services operations is also important when a firm choose to develop a new business channel.

[46] An e-commerce website provides at least one of the six functions introduced in the e-commerce score.
[47] The simultaneous regressions are not displayed in the paper.

(- 4 JF 4(*) 2 - D (3*%% 3 D -) (%((/ %)

Management ICTs in general	6 :) I	chi2(1)=	6.08733
		Prob>chi2=	0.0136
		0.1724** (0.0689)	
Internal system for reordering replacement supplies ("Reordering system")	6 :) I	chi2(1)=	1.21324
		Prob>chi2=	0.2707
		ns	
Invoicing and payment systems ("Invoicing system")	6 :) I	chi2(1)=	2.68342
		Prob>chi2=	0.1014
		ns	
ICTs for managing, logistics or services operations ("Logistic system")	6 :) I	chi2(1)=	7.68596
		Prob>chi2=	0.0056
		0.2028*** (0.0718)	
ICTs linked with suppliers' business systems ("Suppliers system")	6 :) I	chi2(1)=	2.47991
		Prob>chi2=	0.1153
		ns	
ICTs linked with customers' business systems ("Customers system")	6 :) I	chi2(1)=	10.3755
		Prob>chi2=	0.0013
		0.3164*** (0.0945)	
Electronic Data Interchange	6 :) I	chi2(1)=	2.236
		Prob>chi2=	0.1348
		ns	

MF * 34) *

In this article, we try to sum up the different factors than can explain the motivations for firms to invest in a website and for the degree of use this site in an informative and/or in a commercial way. At first, we try to find in the theoretical and the empirical literature the factors underlined as influencing the creation decision and the involvement of the firm when it chooses to set up a site. It permits us to formulate different hypothesis we test subsequently with the use of Probit models for the study of the creation decision, and with the use of count data models for the study of the degree of involvement in the uses of the site. The results of our empirical study show, on the one hand, that the factors influencing creation behaviours and degree of involvement in the use of the website are quite different. On the other hand, the results underline that there are many similarities in the determinants of intensity of website use both as an informational tool and as an e-commerce one. Website creation is, above all, influenced by the resources owned by the firms, i.e. financial, human, technological ones. However, these resources do not have a great importance in the choice of the number of functions provided by the site. The small size of the firm is a brake in website setting up, but has no influence on the site use once this site is created. A position of leader on the market influence positively both the site creation and the informative use of the site, but not the commercial use. The knowledge of the different initiatives instigated by both European and Luxembourg authorities has a positive influence on firms which already created a website when they choose the number of functions they want to include in their site. Sectors selling products that fit with the lifestyle of consumers and firms that have an international renown have, as expected, a higher probability than others to use intensely their site. After all, the setting up and use of a website is, for most part, an imitating behaviour consisting in following the actions, concerning e-commerce, taken by preceding firms. The adoption of management ICTs, the subject of a complementary study, has only an influence on the website use. Analysing the links between reorganisations thanks to management ICTs and the setting up of an e-commerce website, our analysis displays a significant complementarity. In details, the influence of certain ICTs (concerning the logistic of the firm and the relationship with customers) seem to be more important than others.

One limitation of this study lies in the way we measure some potential influencing factors such as human resources. Further researches have to address this limitation by finding better proxies. As the setting up of viable business models, thanks notably to firm reorganisations, is one of the priorities in the European diary, further researches should investigate more precisely the links existing between management ICTs adoption and e-commerce website creation.

: 3)

Barney J. (1991): "Firm Resources and Sustained Competitive Advantage", *i 4. 4) 7 5 *,* Vol.17, no.1, pp. 99-120.

Bocquet R. & Brossard O. (2003), "What are the Determinants of Information Technologies Adoption? A theoretical analysis and empirical evidence", *[4 A 5 D D F/ 8 8*, 22p.

Brynjolfsson E. & Smith M.D. (2000), "Frictionless Commerce ? A Comparison of Internet and Conventional Retailers", *7 5 *63 3* , Vol.26, N°4, pp.563-585.

Cameron A.C. & Trivedi P.K. (1986), "Econometric Models Based on Count Data: Comparisons and Applications of some Estimators and Tests", *i 4. 4) RDD (8 34 4 * 39*, Vol.1, N°1, pp.29-53.

Cameron A.C. & Trivedi P.K. (1990), "Regression-based tests for overdispersion in the Poisson model", *i 4. 4) 8 34 4 * 39*, Vol.46, Issue 3, pp.347-364.

Chappell C. & Feindt S. (2000), "Analysis of e-commerce pratice in SME", *G4 . 3 *4 9 k 6* * 5 9*, N° 37, 1st quarter, pp.47-69.

Commission of the European Communities (2001), "Helping SMEs to « Go Digital »", *G4 . 3 *4) 4 * G4 99 4 4) PW7 3 / 00P*, COM(2001) 136, 16p.
http://europa.eu.int/eur-lex/lex/LexUriServ/site/en/com/2001/com2001_0136en01.pdf

Commission of the European Communities (2002), "eEurope 2005: An information society far all", *G4 * -. *4 4) * 8. 4D G4 99 4 *4 * 6 B 8. 4D G4. 3* , COM(2002) 263, 22p.
http://europa.eu.int/information_society/eeurope/2002/news_library/documents/eeurope2005/eeurope2005_en.pdf

Commission of the European Communities (2003a), "Adapting e-business policies in changing environment: The lessons of the Go Digital initiative and the challenges ahead*E G4 . 3 *4) 4 * G4 99 4 4) / l 7 3 / 00W*, COM(2003) 148, 31p.
http://europa.eu.int/eur-lex/lex/LexUriServ/site/en/com/2003/com2003_0148en01.pdf

Commission of the European Communities (2003b), "Communities Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises", *m)) 3 i 4. 4) * 8. 4D I 4* , L 124 (2003), pp.36-41.
http://europa.eu.int/eur-lex/pri/en/oj/dat/2003/l_124/l_12420030520en00360041.pdf

Commission of the European Communities (2004), "SMEs in Europe 2003", *m- 9 B *4 & 4) 8. 4D 67 8 9j D4 *,* N°7, 94p.,
http://europa.eu.int/comm/enterprise/enterprise_policy/analysis/doc/smes_observatory_2003_report7_en.pdf

Dholakia R.R. & Kshetri N. (2004), "Factors Impacting the Adoption of the Internet among SMEs", *6 S. 9 99 8 34 4 39*, Vol. 23, pp. 311-322.

Feindt S., Jeffcoate J. & Chappell C. (2002), "Identifying success factors for rapid growth in SME e-commerce", *6 S. 9 99 8 34 4 39*, Vol. 19, pp.51-62.

Fudenberg D. & Tirole J. (1984), "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look", *H R 3 834 4 3j B D 9 (43 (59*, Vol.74, N°2, pp.361-366.

Geroski P.A. (2000), "Models of technology diffusion", *j 9 3 4 3&*, Vol. 29, N°4-5, April, pp.603-625.

Gillespie A., Richardson R. & Cornford J. (1995), "Information infrastructures and territorial development", *m8G L4 *FGG H 6[4 A9 4D4 F)4 *4 H3 4 45 9 (H *4 B 4D **, Paris, November.

Greene W.H. (2003), *834 4 * 3R &9 9*, 5th Edition, Prentice Hall, 1026p.

Hausman J., Hall B. H. & Griliches Z. (1984), "Economic Models for Count Data with an Application to the Patents-R&D Relationship", *834 4 * 3*, Vol. 52, N° 4, pp.909-938.

Karshenas M. & Stoneman P.L. (1993), "Rank, stock, order and epidemic effects in the diffusion of new process technologies : an empirical model", *j Rn i4. 4) 834 4 39*, Vol. 24, N°4, p.503-528.

La Rovere R.L. (1998), "Small- and Medium-Sized Enterprises and IT Diffusion Policies in Europe", *6 S. 9 99 834 4 39*, Vol. 11, pp.1-9.

Lisbon European Council (2000), *9(3&34 3. 94 9*, 23-24 March, DOC/00/8, 16p.
http://europa.eu.int/rapid/pressReleasesAction.do?reference=DOC/00/8&format=HTML&aged=0&language=FR&guiLanguage=en

Lucchetti R. & Sterlacchini A. (2004), "The adoption of ICT among SMEs: Evidence from an Italian survey", *6 S. 9 99 834 4 39*, Vol 23, pp.151-168.

Markides C. & Geroski P.A. (2004), "Racing to be second", *S. 9 99 6* *5&j B*, Vol.15, Issue 4, pp.25-31.

Martin L. & Pénard T. (2005) "Investing in a Website: a Top Dog or a Resource-Based Strategy for Firms?", *G4 . 3 *4 9k 6* *5 9* N° 59, 3rd quarter, pp.77-98.

Milgrom P. & Roberts J. (1995), "Complementarities and fit : Strategy, structure, and organizational change in manufacturing", *i4. 4) R334. * 5 (834 4 39*, Vol. 19, N°2-3, p.179-208.

OECD (2004), "ICT, E-business and SMEs", *6HFVFn V 7 8M00/ N V, Fn RY*, Organisation for Economic Co-operation and Development, Paris, 46p.

Official Journal of the European Communities (2000), "Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures", L13, 01/19/2000, pp.12-20.

Penrose E. (1959), *H * 4 &4) * 5 4 * 4) *)*, Oxford Basil Black Well, 272 p.

Poussing N. (2005), "Initiative des pouvoirs publics en faveur de la sécurité sur Internet. Quel effet sur le commerce électronique ?", *[4 A 5 D G8 6VF 9* (*, 16p.

Rogers E. (1962), *)). 9 4 4) 4B *4 9*, Free Press of Glencoe, New York, 367p.

Sadowski B.M., Maitland C. & Van Dogen J. (2002), "Strategic use of the Internet by small-medium-sized companies: an exploratory study", *F)4 *4 834 4 39 (4 3&*, Vol. 14, pp. 75-93.

Santarelli E. & D'Altri S. (2003), "The diffusion of e-commerce among SMEs : Theoretical implications and empirical evidence", *6 S. 9 99 834 4 39*, Vol. 21, pp.273-283.

Shankar V., Rangaswamy A. & Pusateri M. (1998), "The impact of Internet marketing on price sensitivity and price competition", *7 A * 5 63 3 (* F * *, INFORM College on Marketing Mini-Conference, Cambridge.

STATEC (2005), "Les technologies de l'information et de la communication en 2004", *S. * (. 6HRH8 G*, N°3-2005, 123p.

Teece D.J. (1996), "Firm organization, industrial structure and technological innovation", *i4. 4) 834 4 3 S B4. (m 5 *4* , Vol. 31, p.193-224.

9 22 7B0 . 5A (( )3 2 *

9-- 5(*	O)3 2 *	P *4)(%24 BBBJ %)T		H %) *D / (D -) RM %)T		H %) D (3*%% 3 D -) RKKL %)T	
		8 (	FO 5F	8 (	FO 5F	8 (	FO 5F
	(4 (3 *)						
	Firm resources						
Small	Size between 10 and 49 employees	0.7394	*0OW6/*	0.6819	*0O11P*	0.6787	*0O11b*
Mid	Size between 50 and 249 employees	0.2182	*0OPW*	0.2564	*0OW0*	0.2607	*0OW6b*
Large	250 employees and more	0.0424	*0O0P1*	0.0616	*0O 01*	0.0607	*0OW60*
Multi-unit organisation	Multi-unit organisation	0.1166	*0OWPP*	0.1433	*0OW601*	0.1438	*0OW6PW*
Group	Subsidiary of a group	0.2792	*0O*	0.3324	*0O1P*	0.3461	*0O11W*
ICTs training	Employees has possibility to attend ICTs training	0.3560	*0O1o0*	0.4355	*0Oo1/*	0.4652	*0OooW*
	Business characteristics						
Industry	Industrial sector	0.1343	*0OWPP*	0.1203	*0OWb1*	0.1191	*0OW W*
Construction	Construction sector	0.2597	*0OW1*	0.2264	*0OP*	0.2202	*0OP o*
Trade	Trade sector	0.2668	*0O /b*	0.2593	*0OW1*	0.2899	*0Ob /*
Horesca	Hotels, restaurants and cafés	0.0327	*0OP1 1o*	0.0415	*0OPool*	0.0584	*0OW*
Transports	Transport sector	0.1060	*0OW0 0*	0.0989	*0Oo 1*	0.0966	*0Oob*
Computing	Computing sector	0.0389	*0OPoW*	0.0573	*0OW1*	0.0607	*0OW60*
Services	Services	0.1617	*0OW W*	0.1963	*0OW61b*	0.1551	*0OW1/*
Household	1: customers are predominantly households ; 0: they are predominantly firms	0.1996	*0OW6oo*	0.1848	*0OW*	0.2090	*0O01 0*
Exports	% of the firm's sales realised abroad	19.76	*WOW*	19.6	*WPOW*	19.33	*W0O*
Dependence with a consumer	30% or more of the firm sales concluded with a single consumer	0.2041	*0O0W*	0.1891	*0OW6Po*	0.1618	*0OW1 1*
	Attitudes and past experiences with ICTs						
Intranet	Intranet	0.4682	*0Ooo/*	0.5616	*0Oo1b*	0.5640	*0Oo1*
Extranet	Extranet	0.2818	*0Ob0P*	0.3539	*0O1 b*	0.3573	*0O1ol*
Mail service	Electronic mail service	0.8542	*0OW6W0*	0.9126	*0O /1*	0.9011	*0Oo*
Videoconference	videoconference	0.0636	*0O /*	0.0860	*0O 0b*	0.0764	*0O1bo*
Electronic forum	Electronic forum	0.0813	*0O1W*	0.1175	*0OW//*	0.1213	*0OW1o*
Electronic working group calendar	Electronic working group calendar	0.2527	*0OW1*	0.3381	*0O1W*	0.3438	*0O1bb*
Project scheduler	Group project scheduler	0.1334	*0OW0P*	0.1777	*0OW/b*	0.1820	*0OW1W*
LAN	Local Area Network (wireless or not)	0.9700	*0OP1 0*	0.9914	*0Oo/*	0.9933	*0O Po*
ICTs score	Sum of the ICTs (without the Internet connection) (0-8)	3.1051	*PO 1P*	3.5387	*PO1b*	3.5393	*POW60*
Low connection	No or low connection to Internet	0.3489	*0O1 1*	0.2493	*0OWo*	0.236	*0O/bP*
Low DSL connection	DSL connection<2 Mb/sec	0.3322	*0O1P/*	0.3711	*0O W*	0.3798	*0O bo*
High speed DSL	DSL connection≥2Mb/sec	0.3189	*0O11W*	0.3797	*0O b1*	0.3843	*0O 1*
Management ICTs score	ICTs systems to manage placement and reception of orders	0.4876	*0O00P*	0.5487	*0Oo 0*	0.6022	*0Oo0*
Reordering system	Internal system for re-ordering replacement supplies	0.3260	*0O1 o*	0.3639	*0O Pb*	0.4022	*0Oo0o*
Invoicing system	Invoicing and payment systems	0.4099	*0Oo/0*	0.4742	*0Oool*	0.5146	*0O00W*
Logistic system	Linked to the system for managing, logistics or services operations	0.2677	*0O /o*	0.3209	*0O11/*	0.3640	*0O P1*
Suppliers system	Linked with suppliers' business systems	0.1387	*0OWb*	0.1590	*0OW110*	0.1888	*0OW6P*
Customers system	Linked with customers' business systems	0.0857	*0O 00*	0.1074	*0OW0oo*	0.1393	*0OW11*
EDI	Electronic Data Interchange	0.0698	*0Ob o*	0.0788	*0O1o1*	0.0966	*0Oob*
Management ICTs score	Sum of ICTs system of managing orders and EDI (0-6)	1.2977	*PO /*	1.5043	*POb/P*	1.7056	*POP11*

	Characteristics of the website						
Age of the site	Setting up of the website in: 1-2004-2005; 2-2003; 3-2002; 4-2001; 5-2000; 6-1998-1999; 7-before 1998	-	-	4.0544	*POWPb*	4.1506	*POW1*
	17 (4(3*)						
	Competition						
Limited competition	Limited competition	0.1325	*0OWWb/*	0.1232	*0OW o*	0.1258	*0OWW0*
Intense competition	Intense competition	0.4337	*0Oob*	0.4269	*0Oob*	0.436	*0Oo1*
Very intense competition	Very intense competition	0.4337	*0Oob*	0.4499	*0Ool*	0.4382	*0Oo1l*
Niche	The firm develops a niche on the market	0.1917	*0OWbW*	0.1877	*0OWb0l*	0.1730	*0OW 1*
Follower	The firm is a follower on the market	0.3163	*0O1b/*	0.2464	*0OWP/*	0.2494	*0OWW*
Challenger	The firm is a challenger on the market	0.2898	*0ObW*	0.3138	*0O1*	0.3056	*0O1P/*
Leader	The firm is the leader on the market	0.2023	*0O0Po*	0.2521	*0OW1*	0.2719	*0O b*
	Local incentives						
Rivals have an e-commerce website	Competitors have an e-commerce website	0.3101	*0O1/1*	0.3668	*0O /W*	0.4292	*0Oobb*
	Market power						
Local renown	The firm's brand is well-known at a local level	0.1661	*0OW/W*	0.1218	*0OWlW*	0.1169	*0OWP1*
National renown	The firm's brand is well-known at the national level	0.4965	*0O00/*	0.5215	*0Oooo*	0.5169	*0O00W*
International renown	The firm's brand is well-known at an international level	0.3375	*0OlWP*	0.3567	*0Olo*	0.3663	*0O /W*
	External support						
Supplier support	30% or more of the firm purchases transacted with a single supplier	0.3012	*0Obo0*	0.2751	*0O 1o*	0.3258	*0O1o/*
Guild chamber platform	The firm knows the guild chamber's platform "Handwriek-online" that give ICTs information	0.1979	*0OWb 1*	0.1963	*0OWblb*	0.2202	*0OP o*
Entreprises.lu	The firm knows the portal "entreprises.lu"	0.3242	*0O1 W*	0.3711	*0O W*	0.3775	*0O bW*
CASES	The firm knows the portal for information security CASES	0.0769	*0O11b*	0.0931	*0Oo0*	0.0899	*0O 1W*
Electronic signature law	The firm knows the electronic signature law	0.2783	*0O W*	0.3152	*0O1 o*	0.3326	*0OlPl*
Electronic commerce committee	The firm knows the "electronic commerce" committee	0.0972	*0OolW*	0.1132	*0OWPl0*	0.1303	*0OWWP*
Electronic commerce code	The firm knows the electronic commerce code	0.1484	*0OWbl*	0.1705	*0OW1W*	0.1865	*0OWb00*
OLAS	The firm knows the Luxembourg Office of Accreditation	0.1104	*0OWPW*	0.1261	*0OWW/*	0.1416	*0OWo0*
Quality label	The firm knows the quality label "Luxembourg e-commerce certified"	0.3110	*0O1WP*	0.3481	*0Ol1l*	0.3775	*0O bW*
Governmental initiatives score	Sum of the knowledge of governmental initiatives (0-8)	1.5442	*PObl/*	1.7335	*POb//*	1.8562	*/OP01*

9 22 7 0

(-4 KFH(3 *) 4 3 / 3 (* * (D -)

	8 * 4B		8 * 4	
	* H F	8 (/F F	* H F	8 (/F F
Constant	-10.0877**		-0.7490*	
	(0.4408)		(0.4119)	
(4 (3 *)				
Firm resources				
Small	-0.6769**	-0.2294	-0.6028**	-0.2083
	(0.2904)		(0.2853)	
Mid	-0.4973*	-0.1907	ns	
	(0.2931)			
Large	Ref.		Ref.	
Multi-unit organisation	0.2486*	0.0882	0.2357*	0.0844
	(0.1430)		(0.1408)	
Group	ns		ns	
ICTs training	0.2499**	0.0910	0.2502**	0.0917
	(0.0987)		(0.0971)	
Business characteristics				
Industry	Ref.		-	
Construction	ns		-	
Trade	0.2593*	0.0934	-	
	(0.1531)			
Horesca	10.1769***	0.3006	-	
	(0.2881)			
Transports	0.3540**	0.1226	-	
	(0.1786)			
Computing	10.0972***	0.2910	-	
	(0.3180)			
Services	0.3941**	0.1368	-	
	(0.1694)			
Household	-		ns	
Exports	ns		ns	
Dependence with a consumer	ns		ns	
Attitudes and past experiences with ICTs				
Intranet	0.1904*	0.0703	0.1886*	0.0700
	(0.1013)		(0.1003)	
Extranet	ns		ns	
Mail service	ns		ns	
Videoconference	ns		ns	
Electronic forum	0.4114*	0.1398	0.4640**	0.1569
	(0.2198)		(0.2169)	
Electronic working group calendar	0.3328***	0.1186	0.3414***	0.1223
	(0.1272)		(0.1257)	
Project scheduler	ns		ns	
LAN	0.6257**	0.2446	0.5035*	0.1970
	(0.3067)		(0.2970)	
Low connection	Ref.		Ref.	
Low DSL	0.3907***	0.1401	0.3926***	0.1416
	(0.1051)		(0.1037)	

High speed DSL	0.4135*** (0.1087)	0.1476	0.4444*** (0.1079)	0.1591
Reordering system	ns		ns	
Invoicing system	ns		ns	
Logistic system	ns		ns	
Suppliers system	ns		ns	
Customers system	ns		ns	
EDI	ns		ns	
1 7 (4 (3 *)				
Competition				
Limited competition	ns		ns	
Intense competition	Ref.		Ref.	
Very intense competition	ns		ns	
Leader	0.3606*** (0.1385)	0.1269	0.3266** (0.1356)	0.1164
Challenger	ns		ns	
Follower	Ref.		Ref.	
Niche	ns		ns	
Local incentives				
Rivals have an e-commerce website	0.3095*** (0.0961)	0.1115	0.3501*** (0.0940)	0.1264
Market power				
Local renown	Ref.		Ref.	
National renown	0.2793** (0.1223)	0.1031	0.2580** (0.1216)	0.0959
International renown	ns		ns	
External support				
Supplier support	ns		ns	
Guild chamber	ns		-0.2289** (0.1156)	-0.0871
Entreprises.lu	0.1942** (0.1058)	0.0709	0.1997* (0.1033)	0.0733
CASES	ns		ns	
Electronic signature	ns		ns	
Electronic commerce committee	ns		ns	
Electronic commerce code	ns		ns	
OLAS	ns		ns	
Quality label	ns		ns	
6 6	-611.25955		-625.55055	
W . * :	0.1889		0.1699	
* 3 4A34())	72.08%		70.67%	

9 22 7 J 0

(- 4 LFH(3 *) 4 3 / %)) A* 5*45 % .) * (D -)

	* (4)3*		* %(*)3*		1 3*%% 3)3*	
	8 * 4B	8 * 4	8 * 4B	8 * 4	8 * 4B	8 * 4
4 3 %* 4	/	/	W*))*	W*))*	/	/
	* H F	* H F	* H F	* H F	* H F	* H F
Constant	0.6456** (0.2984)	0.5794** (0.2905)	ns	ns	ns	ns
			(4 (3 *)			
			Firm resources			
Small	ns	ns	ns	ns	ns	ns
Mid	ns	ns	ns	ns	ns	ns
Large	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Multi-unit organisation	ns	ns	ns	ns	ns	-0.1684* (0.1010)
Group	0.0931* (0.0520)	0.0927* (0.0520)	ns	0.0950* (0.0568)	0.1413* (0.0814)	0.1437* (0.0814)
ICTs training	ns	ns	ns	ns	ns	ns
			Business characteristics			
Industry	Ref.		Ref.		Ref.	
Construction	-0.1797** (0.0879)	-	-0.2231** (0.0983)	-	-0.3362** (0.1418)	-
Trade	ns	-	ns	-	ns	-
Horesca	0.3374*** (0.1181)	-	0.2949** (0.1312)	-	0.4775*** (0.1806)	-
Transports	ns	-	ns	-	ns	-
Computing	ns	-	ns	-	ns	-
Services	ns	-	ns	-	ns	-
Household	-	0.1156** (0.0580)	-	0.1411** (0.0637)	-	ns
Exports	-0.0014* (0.0008)	ns	ns	ns	-0.0043*** (0.0013)	-0.0034*** (0.0013)
Dependence with a consumer	-0.1456** (0.0598)	-0.1481** (0.0596)	-0.1148* (0.0662)	-0.1116* (0.0651)	-0.2383** (0.0967)	-0.2427** (0.0960)
			Attitudes and past experiences with ICTs			
Intranet	ns	ns	ns	ns	ns	ns
Extranet	ns	ns	ns	ns	ns	ns
Mail service	ns	ns	ns	ns	-0.2388* (0.1381)	ns
Videoconference	ns	ns	ns	ns	ns	ns
Electronic forum	ns	ns	ns	ns	ns	ns
Electronic working group calendar	ns	ns	ns	ns	ns	ns
Project scheduler	0.1724*** (0.0665)	0.1546** (0.0671)	0.1429* (0.0739)	0.1257* (0.0735)	0.2148** (0.1034)	0.1786* (0.1043)
LAN	ns	ns	ns	ns	ns	ns
Low connection	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Low DSL connection	ns	ns	ns	ns	ns	ns
High speed DSL	ns	ns	ns	ns	ns	ns
Reordering system	ns	ns	ns	ns	ns	ns
Invoicing system	ns	ns	ns	ns	ns	ns
Logistic system	ns	ns	ns	ns	ns	ns

Suppliers system	ns	ns	ns	ns	ns	ns
Customers system	0.1600**	0.1793**	ns	ns	0.3386***	0.3587***
	(0.0739)	(0.0748)			(0.1130)	(0.1149)
EDI	ns	ns	ns	ns	ns	ns
Characteristic of the website						
Age of the site	0.0215*	0.0321**	ns	0.0263*	0.0341*	0.0543***
	(0.0130)	(0.0129)		(0.0142)	(0.0205)	(0.0204)
1 7 (4 (3 *)						
Competition						
Limited competition	ns	ns	ns	ns	ns	ns
Intense competition	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Very intense competition	ns	ns	ns	ns	-0.1466**	-0.1547**
					(0.0728)	(0.0731)
Leader	0.1922***	0.1706**	0.1816**	0.1661**	ns	ns
	(0.0686)	(0.0686)	(0.0767)	(0.0757)		
Challenger	0.1049*	ns	ns	ns	ns	ns
	(0.0632)					
Follower	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Niche	ns	ns	ns	ns	ns	ns
Local incentives						
Rivals have an e-commerce website	0.1970***	0.2267***	0.1401***	0.1675***	0.3198***	0.3587***
	(0.0447)	(0.0443)	(0.0500)	(0.0488)	(0.0701)	(0.0694)
Market power						
Local renown	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
National renown	ns	ns	ns	ns	ns	ns
International renown	0.1459*	0.1946**	ns	0.1992**	0.2513*	0.3313**
	(0.0834)	(0.0846)		(0.0931)	(0.1327)	(0.1347)
External support						
Supplier support	0.1083**	0.1244**	ns	0.0976*	0.1697**	0.1871**
	(0.0491)	(0.0488)		(0.0537)	(0.0763)	(0.0760)
Guild chamber	0.1240**	ns	0.1438**	ns	ns	ns
	(0.0591)		(0.0660)			
Entreprises.lu	ns	ns	ns	ns	ns	ns
CASES	ns	ns	ns	ns	ns	ns
Electronic signature	0.1388**	0.1365**	0.1620**	0.1604**	0.1568*	0.1516*
	(0.0567)	(0.0569)	(0.0627)	(0.0621)	(0.0894)	(0.0897)
Electronic commerce committee	ns	ns	ns	ns	ns	ns
Electronic commerce code	ns	ns	ns	ns	ns	ns
OLAS	ns	ns	ns	ns	ns	ns
Quality label	ns	ns	ns	ns	ns	ns
	0.0285*	0.0365**			0.0764*	0.0962**
	(0.0152)	(0.0157)			(0.0387)	(0.0405)
6 6	-1435.1982	-1445.3247	-1218.1286	-1226.3128	-1080.8537	-1092.1419
6 U 4 ** (*) * 0	chi2 (1) = 4.16 Prob≥chi2 = 0.02	chi2 (1) = 6.69 Prob≥chi2 = 0.005	chi2 (1) = 0.00 Prob≥chi2 = 0.500	chi2 (1) = 8.6e-07 Prob≥chi2 = 0.500	chi2 (1) = 4.84 Prob≥chi2 = 0.014	chi2 (1) = 7.37 Prob≥chi2 = 0.003

Reexamining the Effects of the Most-Favored-Nation (MFN) Provision in Input Prices on R&D Incentives:
The Effect of the MFN on Access Charge Agreements

J.Y Kim [1] and J. Nahm [2]

Abstract

We examine the effect of the most-favored-nation provision in access charge agreements on downstream networks' R&D incentives. Contrary to the previous literature, we show that the effect depends on the extent of substitutability between downstream networks if they compete in two-part tariffs. When a network lowers its marginal cost, it entails two conflicting effects on its access fee, the market share effect rewarding the low-cost network in addition to the standard elasticity effect penalizing it. If network substitutability is high enough, the former dominates the latter, and thus downstream firms will choose a higher cost technology under the MFN provision.

JEL classification number: L11

Key Words: MFN provision, price discrimination, two-part tariff, access charge

[1] Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaemunku, Seoul 130-701, Korea, (Tel) +822-961-0986.
[2] Hong Kong University of Science and Technology, Department of Economics, HKUST, Clearwater Bay, Kowloon, Hong Kong, (Tel) +852-2358-7626.

1 Introduction

Competing firms in the downstream market often buy their inputs from a common input supplier (upstream firm). Examples abound. Long-distance call service providers must purchase access service from a local carrier. Publishers supply books to many competing retail bookstores.

One important policy issue is whether or not upstream firms should be allowed to charge discriminatory input prices. Banning discriminatory wholesale prices dates back to the Robinson-Patman Act of 1936 in the United States.[1] As a prominent example, the U.S. Telecommunications Act of 1996 requires all incumbent local exchange carriers to provide non-discriminatory access to all requesting carriers.[2] Under the most-favored-nation (MFN) clause, a competitive local exchange carrier (CLEC) is allowed to adopt for itself an agreement that another CLEC has negotiated or received through arbitration. The main goal of the MFN clause has been considered as promoting fair competition by preventing upstream firms from blockading entry of certain downstream firms.

In this paper, we examine the effect of the MFN provision on the R&D incentive of downstream firms. We consider two downstream firms that compete with each other by offering a two-part tariff. They can undertake R&D activities that determine their marginal costs, whereupon they purchase an essential input from an upstream firm. After observing the costs of downstream firms, the upstream firm sets its prices for the essential inputs. By comparing the market outcomes in the discriminatory-price regime and in the uniform-price regime, we can derive implications of the MFN policy.

Assuming that downstream firms compete as Cournot duopolists, previous studies (DeGraba (1990), Choi (1995)) explore the effects of an upstream monopolist's price discrimination on the downstream firms' cost-reducing R&D and find that the effect of the MFN clause on downstream firms' R&D incentives is positive. As DeGraba (1990) shows, if downstream firms have different constant marginal costs, the price-discriminating monopolist will charge

[1]Several papers (e.g., Katz (1987), DeGraba (1990), Yoshida (2000), and Valletti (2003)) examine the welfare effects of third-degree price discrimination by an input monopolist.

[2]See §251.c.2. This clause is also included in the WTO agreement on trade in Telecommunications services.

a higher input price for the lower-cost firm, since a lower-cost firm has a more inelastic demand for the input. Thus, the lower-cost firm is 'penalized,' and its cost advantage is partially offset by the upstream firm's discriminatory input prices. Therefore, the downstream firms will choose a higher marginal cost technology in the discriminatory-price regime than they would in the uniform-price regime. In the context of government tariffs in international trade, Choi (1995) also obtains a similar result that a lower marginal cost technology will be chosen under the MFN clause than under discriminatory tariffs.

However, by assuming that the two downstream firms compete by offering a two-part tariff, we can obtain the opposite result. In our paper, the two downstream firms compete in a Hotelling-product-differentiation model by offering a two-part tariff. As in a Cournot model, a lower-cost firm has a more inelastic demand for an input, which induces the upstream firm to 'penalize' the lower-cost downstream firm under the discriminatory-price regime. This is the standard elasticity effect. However, we show that if one firm has a lower marginal cost than the other, a customer of the lower-cost downstream firm, by buying more units, contributes more to the upstream firm's profits than a customer of the other firm does. Thus, the upstream firm's profit increases with the lower-cost firm's market share. We call this 'market share effect.' Because of the market share effect, the upstream firm wants to increase the lower-cost firm's market share, which rewards a firm achieving the lower cost. Thus, when the two firms compete in a two-part tariff, we need to compare these two effects to determine the overall effect of MFN on the downstream firm's R&D incentive.

We show that if the substitutability between the two downstream products is high (or the degree of product differentiation is low), the market share effect dominates the elasticity effect, and, consequently, the effect of MFN on the cost-reducing R&D is negative. That is, the downstream firms will choose a higher marginal cost technology under the MFN provision than they will under the discriminatory-price regime. This contrast with the previous literature comes from a difference in competition modes: while downstream firms in DeGraba (1990) and Choi (1995) compete in Cournot fashion, downstream firms in our paper compete in two-part tariffs.

It deserves reexamining the effect of MFN under two-part tariffs, since competition by two-part tariffs is quite common especially in IT industries including telecommunicaitons, wireless Internet, and membership discount retailers such as shopping clubs etc. In particular, our analysis can be applied to regulatory issues of access charges in telecommunications. For example, some regulatory authorities recommend an access charge by a bottleneck sector to be based not on the individual costs of downstream firms, but on their representative ones, which can be regarded as another form of MFN provision. Our results suggest that the effect of the regulation may be positive or negative, depending on the substitutability of downstream services.

There are several papers on the effect of the input price on the R&D incentive especially in the telecommunications industry. Among others, Gans (2001) derives the access pricing formula inducing the optimal timing to invest in essential facilities. DeGraba (2001) asserts that a positive access charge would cause a bias in favor of the technology with high marginal costs because the access charge reflects only the marginal cost. Kim and Lim (2004) argue that the revenue-sharing rule for interconnection charges may give a firm a perverse incentive to raise its own cost in order to grab a higher share of the revenues. Valletti and Cambini (2005) show that firms tend to collude by underinvesting in quality in a two-way interconnection model of telecommunications. None of them, however, addresses the issue associated with MFN.

This paper is organized as follows. In Section 2, we set up the model. In Section 3, as a preliminary step, we consider a benchmark case in which market shares of downstream firms are given, and reconfirm the standard positive effect of MFN on the downstream firms' R&D incentive. In Section 4, we show that the effect of MFN can be negative if we consider their competition for market shares. In Section 5, we derive some welfare implications of MFN. In Section 6, we discuss the consequences of some variations in assumptions. Policy implications and conclusion follow in Section 7.

2 Model

We consider two downstream firms $i = 1, 2$, that sell horizontally differentiated products or service. We assume that they are located at the two extremes of an Hotelling line of length one ($x_1 = 0, x_2 = 1$). Consumers are heterogeneous along the line, and their location (type) is denoted by x. A consumer can be a customer of either firm 1 or firm 2 but not both. We assume that a customer of a firm must purchase products (or service) only from the firm. When a consumer located at x becomes a customer of firm i, it incurs her "transportation costs" $t|x - x_i|$, and accordingly her utility is given by

$$u_0 - t|x - x_i| + u(q_i),$$

where u_0 is the base utility from the service, q_i is the number of units to be purchased by a customer of firm i and $u(q_i)$ is her utility from the purchases. We assume that u_0 is large enough that the two firms cover the whole market. We discuss the case in which the market is not fully covered in Section 6.

The two firms charge two-part tariffs

$$T_i(q_i) = F_i + p_i q_i, i = 1, 2,$$

where F_i is the fixed fee and p_i is the unit price for a product or the usage fee. A customer of firm i chooses q_i in order to maximize $u(q_i) - p_i q_i$. Thus, the optimal quantity to be purchased is $q_i^* = D(p_i)$, where $D(\cdot) = u_q^{-1}(\cdot)$. We assume that the demand function satisfies standard assumptions such as downward sloping and concavity, i.e., $D'(\cdot) < 0$ and $D''(\cdot) \leq 0$.[3] We denote the indirect utility function by $v(p_i) \equiv \max_{q_i}\{u(q_i) - p_i q_i\}$. A consumer remains as a customer of firm i as long as F_i is less than or equal to $v(p_i)$.

Given (F_i, p_i), consumers choose their firm affiliation. Ignoring the transportation cost, the net surplus a consumer enjoys from subscribing to firm i is

$$w_i = u_0 + v(p_i) - F_i.$$

Only the difference between w_i and w_j determines the market share. Firm i's market share is

$$\alpha_i = 1/2 + \sigma(w_i - w_j),$$

[3]The condition $D'' \leq 0$ implies that the ordinary marginal revenue, $pD'(\cdot) + D(\cdot)$, is decreasing, which ensures that the ordinary second-order condition is satisfied. The decreasing marginal revenue is also a standard assumption for the existence of a unique and symmetric pure strategy Nash Equilibrium in a Cournot game.

where $\sigma = 1/(2t)$.

In order to produce a unit of good, each firm must purchase a unit of input from an upstream firm $i = 0$, thereby incurring the marginal input production cost c_0 to firm 0. Producing a unit of the final good incurs firm i an additional marginal cost c_i. The size of c_i is realized as a result of R&D efforts by firm i. The relation between the R&D efforts e_i and the marginal cost c_i is given by $e_i = \phi(c_i)$, where $\phi' < 0$ and $\phi'' > 0$. That is, a lower marginal cost can be achieved when firm i spends more on its R&D. We assume that $\phi(c_i)$ is convex enough that the profit function of firm i becomes concave.

The sequence of decision-making is assumed as follows. First, two firms independently decide how much they invest in the cost reducing R&D. Then, upstream firm 0 chooses input prices a_i, $i = 1, 2$. Third, each downstream firm i selects its two-part tariff schedule (F_i, p_i). Fourth, consumers decide where to buy their products. Lastly, consumers choose the quantity q_i to be purchased from their host firm i.

3 Benchmark Analysis

As a benchmark, we will consider the case that each firm has a fixed number of users, n_i, $i = 1, 2$. In other words, we assume for the time being that the affiliation decisions of consumers are already made. The analysis for this benchmark case will turn out to be useful to later analysis.

Since firm i needs to pay a_i for the input, the total marginal cost for firm i is $a_i + c_i$. The optimal two-part tariff of firm i will be to set unit usage fee at its total marginal cost and to extract the consumer's surplus by its fixed fee, that is, $(F_i^*, p_i^*) = (v(a_i + c_i), a_i + c_i)$.

In the case of discriminatory pricing (DP), each customer to firm i buys $D(a_i + c_i)$ units, and the total demand for the input by firm i is $n_i D(a_i + c_i)$. Thus, the profit of the upstream firm is

$$\pi_0 = \sum_{i=1}^{2} n_i (a_i - c_0) D(a_i + c_i).$$

Since there is no competition between the two downstream firms, firm 0 maximizes its profit piecewise under DP regime. The first-order condition implies that the optimal input prices

a_i^* must satisfy

$$D(a_i^* + c_i) + (a_i^* - c_0)D'(a_i^* + c_i) = 0, \text{or } \frac{a_i^* - c_0}{a_i^*} = \frac{1}{\epsilon_{a,i}}, \tag{1}$$

where $\epsilon_{a,i}$ is the demand elasticity with respect to a_i. The optimal input price characterizes the classical Lerner formula. Since the two markets are separated, a_i^* depends only on c_i, not on c_j. Let $a^*(c_i)$ denote the firm 0's optimal input price charged to firm i under DP regime. We have the following lemma.

Lemma 1 *a_i^* decreases in c_i, but the total marginal cost $a^*(c_i) + c_i$ increases in c_i.*

Proof. See the appendix.

The crucial point is that firm 0's optimal input price charged to firm i depends on c_i as well as on c_0. This is because the demand for the input depends on c_i. In general, if firm i lowers its marginal cost, firm 0 will increase a_i, which partially offsets the decrease in c_i and penalizes a lower-cost firm. The intuition is quite similar to DeGraba's (1990)[4]: Since a lower cost firm has a less elastic demand for an input, an upstream firm charges a higher input price for the lower-cost firm.

In the case of uniform pricing (UP) under the MFN provision, firm 0 chooses a common input price $\tilde{a}$. Thus, firm 0's profit is

$$\pi_0 = \sum_{i=1}^{2} n_i(\tilde{a} - c_0)D(\tilde{a} + c_i).$$

The first-order condition implies that the optimal charge $\tilde{a}$ must satisfy

$$\sum_{i=1}^{2} n_i \left[D(\tilde{a}^* + c_i) + (\tilde{a}^* - c_0)D'(\tilde{a}^* + c_i)\right] = 0, \text{ or } \frac{\tilde{a}^* - c_0}{\tilde{a}^*} = \frac{1}{\bar{\epsilon}_a}, \tag{2}$$

where $\bar{\epsilon}_a = -\frac{\partial \bar{D}(\tilde{a})}{\partial \tilde{a}} \frac{\tilde{a}}{\bar{D}(\tilde{a})}$ and $\bar{D}(\tilde{a}) = \sum_{i=1}^{2} n_i(\tilde{a} - c_0)D(\tilde{a} + c_i)$ is the overall demand for the input. This is again the standard Lerner formula.

The optimal common input price depends on both c_1 and c_2. Let $\tilde{a}^*(c_1, c_2)$ denote the firm 0's optimal input price under UP. We have

[4]There are two distinctions from DeGraba (1990). First, unlike DeGraba, the markets for consumers are separated in this model. So, DeGraba's intuition can be applied more clearly. Second, instead of a linear demand function assumed in DeGraba, we consider a more general demand function by assuming $D'' \leq 0$.

Lemma 2 $\widetilde{a}^*(c_1, c_2)$ *decreases in* c_i.

Proof. See the appendix.

Lemma 2 implies that when one of downstream firms lowers its marginal cost by investing in R&D, firm 0 increases the common input price.

The following lemma tells us the relative size of $\partial a^*(c_i)/\partial c_i$ and $\partial \widetilde{a}^*(c_i, c_j)/\partial c_i$.

Lemma 3 $|\partial a^*(c_i)/\partial c_i| > |\partial \widetilde{a}^*(c_1, c_2)/\partial c_i|$.

Proof. See the appendix.

Suppose that firm i reduces its cost. Then, it is in the interest of firm 0 to raise the common input price. However, since a fall in the cost increases only the firm's demand for the input (not the other firm's demand) under the UP regime, the input price is increased by a less amount than under the DP regime. In other words, the penalizing effect is less in the UP regime than in the DP regime.

Now, let us consider the R&D incentives under each regime. Since there is a one-to-one correspondence between e_i and c_i, we will consider firm i's R&D decision in terms of its choice of c_i. Each firm makes its R&D choice taking into account the effect of c_i on input prices.

By substituting the optimal two-part tariffs and the optimal input prices, we get the reduced profit function of firm i as $\pi_i(\mathbf{c}, a(\mathbf{c}))$, where $\mathbf{c} = (c_1, c_2)$. In the R&D stage, each firm maximizes

$$\Pi_i(\mathbf{c}, a(\mathbf{c})) = \pi_i(\mathbf{c}, a(\mathbf{c})) - \phi(c_i).$$

Under the DP regime, the first-order conditions are given as follows;

$$\frac{\partial \pi_i}{\partial c_i} + \frac{\partial \pi_i}{\partial a_i}\frac{\partial a_i^*}{\partial c_i} - \phi'(c_i) = 0, \ i = 1, 2. \tag{3}$$

On the other hand, under the UP regime, the first-order conditions are as follows;

$$\frac{\partial \pi_i}{\partial c_i} + \frac{\partial \pi_i}{\partial a_i}\frac{\partial \widetilde{a}^*}{\partial c_i} - \phi'(c_i) = 0, \ i = 1, 2. \tag{4}$$

Since $\widetilde{a}^*$ depends on both c_1 and c_2 under UP, firm i's total marginal cost per unit depends not only on its own cost, but also on the other firm's cost. Thus, firm i's optimal R&D choice depends on the other firm's R&D choice. We assume that $\mid \frac{\partial^2 \Pi_i}{\partial c_i^2} \mid > \mid \frac{\partial^2 \Pi_i}{\partial c_i \partial c_j} \mid$ to ensure that there is a unique Nash equilibrium. Since the two firms are symmetric ex ante, the unique equilibrium is symmetric. Lemma 3 leads to Proposition 1.

Proposition 1 *When the market shares of downstream firms are fixed, they achieve a lower cost under UP than under DP. That is,* $c_i^* > \widetilde{c}_i^*$.

Proof. See the appendix.

This proposition says that the MFN provision gives a stronger incentive for R&D investments in the absence of competition between downstream firms. The intuition behind this is as follows. When a firm lowers its cost, firm 0 increases its input price. Under MFN, however, the input price cannot be raised very much because the high input price is applied to the other firm as well. Thus, the penalty effect is alleviated, which yields high rewards for an R&D investment. Hence, each firm has a stronger incentive to make cost-reducing investments.

4 General Analysis

In this section, we drop the temporary assumption that the market share of each downstream firm is fixed, and consider the competition between downstream firms for market shares.

4.1 Retail Pricing

Since the net surplus w_i directly determines market shares, it is analytically convenient to imagine that firms compete in p_i and w_i rather than in p_i and F_i. Each firm i thus solves

$$\pi_i = \max_{p_i, w_i} \alpha_i(w_i, w_j)[(p_i - a_i - c_i)D(p_i) + u_0 + v(p_i) - w_i].$$

The first order condition with respect to p_i is

$$D(p_i^{**}) + (p_i^{**} - a_i - c_i)D'(p_i^{**}) + v'(p_i^{**}) = 0.$$

Using the fact that $D'(p_i) < 0$ and $v'(p_i^{**}) = -D(p_i^{**})$, we have

$$p_i^{**} = a_i + c_i.$$

Since each firm sets its usage fee at its total marginal cost, firm i's profit function becomes

$$\pi_i = \max_{w_i} \alpha_i(w_i, w_j)[u_0 + v(a_i + c_i) - w_i].$$

By using $\partial\alpha_i/\partial w_i = 1/(2t)$, the first order condition with respect to w_i yields

$$w_i^{**} = \frac{1}{2}w_j + \frac{1}{2}(u_0 + v(p_i^{**}) - t), \ i = 1, 2. \tag{5}$$

Solving (5) simultaneously leads to $w_i^{**} = u_0 + \frac{2}{3}v(p_i^{**}) + \frac{1}{3}v(p_j^{**}) - t$. Thus, we have $F_i^{**} = \frac{1}{3}(v(p_i^{**}) - v(p_j^{**})) + t$, and accordingly, $\alpha_i^* = \frac{1}{2} + \frac{1}{6t}(v(p_i^{**}) - v(p_j^{**}))$.

To summarize, when two firms compete in a two-part tariff, each firm sets its variable fee at the total marginal cost $(a_i + c_i)$ and its fixed fee at $\frac{1}{3}(v(p_i^{**}) - v(p_j^{**})) + t$. Consequently, we have $\alpha_i^{**} = \frac{1}{2} + \frac{1}{6t}(v(p_i^{**}) - v(p_j^{**}))$ and $\pi_i^{**} = \alpha_i^{**} F_i^{**}$. If the two firms have the same total marginal cost, we have $\alpha_i^* = 1/2$, and each firm earns $t/2$, which depends only on the extent of product differentiation between the two products.

4.2 Pricing for the Input

4.2.1 Uniform Pricing

When firm 0 sets its price at $\tilde{a}$, each firm will respond by setting its variable fee at $\tilde{a} + c_i$. Thus, a customer of firm i will buy the unit of $D(\tilde{a} + c_i)$, which contributes to firm 0's profit by $(\tilde{a} - c_0)D(\tilde{a} + c_i)$. Firm 0's profit is then as follows;

$$\pi_0 = \sum_{i=1}^{2} \alpha_i(\tilde{a} - c_0)D(\tilde{a} + c_i).$$

Suppose that $c_i < c_j$. Since $\tilde{a} + c_i < \tilde{a} + c_j$, a customer of firm i, by buying more units, generates more profits for firm 0 than does a customer of firm j. Thus, if the two firms have different marginal costs, firm 0's profit increases as the market share of the lower-cost firm gets higher.

As shown in the previous subsection, the market shares of downstream firms depend on the input price. To elaborate, an increase in $\widetilde{a}$ increases the variable fee, in turn decreasing the indirect utility of buying products from firm i by $-D(\widetilde{a}+c_i)$. Accordingly, we have $\partial\alpha_i/\partial\widetilde{a} = (1/6t)((v_i - v_j)/\partial\widetilde{a}) = (1/6t)(D(\widetilde{a}+c_j) - D(\widetilde{a}+c_i))$. If $c_i < c_j$, we have $\partial\alpha_i/\partial\widetilde{a} < 0$. That is, when firm 0 increases the input price, it shrinks the market share of the lower-cost firm.[5]

Thus, when firm 0 changes $\widetilde{a}$, it has two effects on π_0. First, $\widetilde{a}$ affects the volume of consumption of each user by changing p_i since $p_i = \widetilde{a} + c_i$. Second, it changes the firms' market shares by affecting the indirect utility from subscribing to each downstream firm. Formally, the first-order condition with respect to $\widetilde{a}$ yields

$$\begin{aligned}\frac{\partial\pi_0}{\partial\widetilde{a}} &= \sum_{i=1}^{2}\alpha_i\left[D(\widetilde{a}^{**}+c_i) + (\widetilde{a}^{**}-c_0)D'(\widetilde{a}^{**}+c_i)\right] \\ &\quad -\frac{1}{6t}(\widetilde{a}^{**}-c_0)[D(\widetilde{a}^{**}+c_1) - D(\widetilde{a}^{**}+c_2)]^2 = 0. \qquad (6)\end{aligned}$$

The first term corresponds to the standard effect, given exogenous market shares. The second term measures the market share effect.[6] When the two downstream firms have different marginal costs, if firm 0 increases the input price, it shifts consumers from a firm with a lower cost to a firm with a higher cost, which reduces firm 0's profits.

4.2.2 Discriminatory Pricing

Under the DP regime, firm 0 maximizes its profits by charging two (possibly) different input prices a_i, $i = 1, 2$. Since each firm sets its variable fee at $a_i + c_i$, firm 0's profit is as follows;

$$\pi_0 = \sum_{i=1}^{2}\alpha_i(a_i - c_0)D(a_i + c_i).$$

As in the UP case, input prices affect not only the consumption volume of each product, but also the downstream firms' market shares. An increase in a_i decreases the indirect utility

[5] If $c_i = c_j$, a change in $\widetilde{a}$ does not affect the market shares.

[6] If $\widetilde{a}$ is increased by one unit, $(1/6t)(D(\widetilde{a}+c_j) - D(\widetilde{a}+c_i))$ consumers switch from firm i to firm j where $c_i < c_j$. When a consumer switches from firm i to j, she will change her volume of consumption from $D(a+c_i)$ to $D(a+c_j)$. Thus, the total changes in the volume of consumption are $-(1/6t)(\widetilde{a}-c_0)[D(\widetilde{a}+c_i) - D(\widetilde{a}+c_j)]^2$.

from subscribing to firm i by $-D(a_i+c_i)$, which leads to $\partial\alpha_i/\partial a_i = (1/6t)(\partial(v_i-v_j)/\partial a_i) = -(1/6t)D(a_i+c_i)$. This implies that $(1/6t)D(a_i+c_i)$ consumers switch from firm i to firm j when a_i is increased by one unit.

Formally, the first-order conditions with respect to a_1 and a_2 yield

$$\frac{\partial \pi_0}{\partial a_i} = \alpha_i[D(a_i^{**}+c_i)+(a_i^{**}-c_0)D'(a_i^{**}+c_i)] - \frac{1}{6t}D(a_i^{**}+c_i)[(a_i^{**}-c_0)D(a_i^{**}+c_i) - (a_j^{**}-c_0)D(a_j^{**}+c_j)] = 0. \tag{7}$$

Again, the first term is the standard elasticity effect, given the market shares, and the second term measures the market share effect.

As in the UP regime, if the two downstream firms have different marginal costs, a subscriber to the lower-cost firm contributes more to firm 0's profit than does a subscriber to the higher-cost firm.

Lemma 4 *If $c_i < c_j$, the upstream firm's profit per customer of firm i is equal to or larger than that of the upstream firm per customer of firm j. That is, $(a_i^{**}-c_0)D(a_i^{**}+c_i) \geq (a_j^{**}-c_0)D(a_j^{**}+c_j)$.*

Proof. See the appendix.

We know that the lower-cost firm has a lower elasticity of demand for inputs than does the other firm, which induces firm 0 to charge a higher input price for the lower-cost firm. In particular, when there is no competition, if $c_i < c_j$, we have $a_i^* > a_j^*$. However, as Lemma 4 indicates, the second term of (7) is positive (negative *resp.*) for the higher-cost (lower-cost *resp.*) firm. That is, when firm 0 increases an input price for the lower-cost firm, some consumers switch from the lower-cost firm to the higher-cost. Clearly, this has a negative effect on firm 0's profit. Hence, in order to increase the lower-cost firm's market share, firm 0 has an incentive to lower the input price for the lower-cost firm or to increase the input price for the higher-cost firm. Thus, we may have the case in which firm 0 charges a lower input price for the lower-cost firm.[7] For instance, when $D(p_i) = 4 - p_i$, $c_0 = 0$, $c_1 = 0.5$, $c_2 = 0.51$ and $t = 0.5$, we have $a_1^{**} < a_2^{**}$.

[7] The relative size of a_i^{**} and a_j^{**} is, in general, ambiguous.

4.2.3 Comparison

Exploiting the fact that in the symmetric equilibrium the two firms will achieve the same marginal cost in the R&D stage, we will explore how input prices react to c_i when $c_1 = c_2$.

Lemma 5 *When $c_i = c_j$, we have (i) $\partial\widetilde{a}^{**}/\partial c_i < 0$ and (ii) $\partial a_i^{**}/\partial c_i + \partial a_j^{**}/\partial c_i = 2\partial\widetilde{a}^{**}/\partial c_i < 0$.*

Proof. See the appendix.

Under DP, when there is a change in c_i, the upstream firm has an incentive to change both a_1 and a_2 in order to affect the market shares through a_1 and a_2. Under the assumption that two downstream firms compete in a Cournot fashion, Choi (1995) shows that the sum of changes in input prices in response to changes in c_i is the same under both UP and DP. Lemma 5(ii) shows that this is still the case when the two downstream firms compete by offering a two-part tariff in a Hotelling model.

Suppose that $\partial a_j^{**}/\partial c_i < \partial a_i^{**}/\partial c_i$. Then, Lemma 5(ii) implies that there are only two cases; either $\partial a_j^{**}/\partial c_i < 0 < \partial a_i^{**}/\partial c_i$ or $\partial a_j^{**}/\partial c_i < \partial a_i^{**}/\partial c_i < 0$. That is, when firm i lowers its cost, firm 0 may either reduce a_i and increase a_j, or increase a_i and increase a_j by a larger amount. In other words, when firm i reduces its cost, the difference in the total marginal cost between the two firms under the DP regime $(a_j + c_j - a_i - c_i)$ becomes larger than the cost difference $(c_j - c_i)$, which implies that, in comparison with the UP case,[8] the lower-cost firm is rewarded by discriminatory input prices.

On the other hand, suppose that $\partial a_i^{**}/\partial c_i < \partial a_j^{**}/\partial c_i$. Then, either $\partial a_i^{**}/\partial c_i < 0 < \partial a_j^{**}/\partial c_i$ or $\partial a_i^{**}/\partial c_i < \partial a_j^{**}/\partial c_i < 0$. That is, when firm i lowers its cost, firm 0 may reduce a_j and increase a_i, or increase a_j and increase a_i by a larger amount. In this case, when firm i reduces its cost, the difference in the total marginal cost between the two firms $(a_j + c_j - a_i - c_i)$ becomes smaller than the cost difference $(c_j - c_i)$, which implies that a lower-cost firm is penalized by discriminatory input prices in comparison with the UP case. Therefore, the relative size of $\partial a_i^{**}/\partial c_i$ and $\partial a_j^{**}/\partial c_i$ determines whether the UP rewards or penalizes the lower-cost firm.

[8] Under UP, the total marginal cost differential $(\widetilde{a} + c_j - \widetilde{a} - c_i)$ remains the same as the marginal cost differential $(c_j - c_i)$.

Let us compare the relative size of $\partial a_i^{**}/\partial c_i$ and $\partial a_j^{**}/\partial c_i$. Lemma 6 shows that their relative size depends on t.

Lemma 6 *When $c_i = c_j$, there is a critical $\widehat{t} > 0$ such that $\partial a_i^{**}/\partial c_i > \partial a_j^{**}/\partial c_i$ if and only if $t < \widehat{t}$.*

Proof. See the appendix.

The intuition behind Lemma 6 goes as follows. We identified two effects of a change in c_i, the elasticity effect and the market share effect. It is imperative to combine these two effects to determine the consequence of the MFN clause on downstream firms' R&D incentives. If the degree of product differentiation is low, it is easier for firm 0 to shift consumers from the lower-cost firm to the higher-cost firm. Thus, if t is small, the upstream firm optimally increases the lower-cost firm's market share by enlarging the total marginal cost differential more than the cost differential.

4.2.4 R&D Decisions

In the R&D stage, each firm maximizes

$$\Pi_i(\mathbf{c}, a_1(\mathbf{c}), a_2(\mathbf{c})) = \pi_i(\mathbf{c}, a_1(\mathbf{c}), a_2(\mathbf{c})) - \phi(c_i).$$

Note that a downstream firm's profit depends on the input price charged to the other firm as well, because it affects the market share.

If we assume that $\mid \frac{\partial^2 \Pi_i}{\partial c_i^2} \mid > \mid \frac{\partial^2 \Pi_i}{\partial c_i \partial c_j} \mid$ as in Section 3, the unique Nash equilibrium must be symmetric. Let c^{**} and $\widetilde{c}^{**}$ denote the marginal costs in the symmetric equilibrium under the DP regime and the UP regime, respectively. Then, c^{**} and $\widetilde{c}^{**}$ must satisfy the corresponding first-order condition;

$$\Phi_i(c^{**}) \equiv \frac{\partial \pi_i}{\partial c_i} + \frac{\partial \pi_i}{\partial a_i}\frac{\partial a_i^{**}}{\partial c_i} + \frac{\partial \pi_i}{\partial a_j}\frac{\partial a_j^{**}}{\partial c_i} - \phi'(c^{**}) = 0, \tag{8}$$

$$\widetilde{\Phi}_i(\widetilde{c}^{**}) \equiv \frac{\partial \pi_i}{\partial c_i} + \frac{\partial \pi_i}{\partial a_i}\frac{\partial \widetilde{a}^{**}}{\partial c_i} + \frac{\partial \pi_i}{\partial a_j}\frac{\partial \widetilde{a}^{**}}{\partial c_i} - \phi'(\widetilde{c}^{**}) = 0. \tag{9}$$

Lemma 7 shows that the relative size of c^{**} and $\widetilde{c}^{**}$ depends on the relative size of $\partial a_i^{**}/\partial c_i$ and $\partial a_j^{**}/\partial c_i$ and this leads to our main result (Proposition 2).

Lemma 7 (Choi (1995)) *$\tilde{c}^{**} > c^{**}$ if and only if $\partial a_i^{**}/\partial c_i > \partial a_j^{**}/\partial c_i$.*

Proof. See the appendix.

Proposition 2 *Downstream firms achieve a higher marginal cost under UP than under DP if and only if $t < \hat{t}$. That is, we have $\tilde{c}^{**} > c^{**}$ if and only if $t < \hat{t}$.*

Proof. This follows immediately from Lemma 6 and Lemma 7.

Proposition 2 implies that the effect of MFN on the R&D investments depends on the degree of substitutability between the two firms. In particular, if the substitutability is high (or products are not so much differentiated), the MFN provision weakens firms' cost-reducing R&D incentives. If the degree of the product differentiation is low, it is easier for firm 0 to shift consumers from the lower-cost to the higher-cost firm. That is, the higher the substitutability is, the larger the market share effect is. If the market share effect dominates the elasticity effect, the lower-cost firm is rewarded, which increases the R&D incentive of downstream firms. However, if the substitutability is low enough, the market share effect is dominated by the elasticity effect, and the low-cost firm is penalized, which reduces firms' R&D incentives.

5 Welfare Analysis

The cost-saving R&D activity of downstream firms affects social welfare in the following three ways. First, in the symmetric equilibrium, the gross profits of downstream firms, π_i^{**}, depend only on the extent of t, as Section 4.1 shows. Thus, as they spend more on R&D activities, their profits get lower. Second, as Section 4.1 shows, consumers' net surplus w_i increases when the equilibrium variable fee gets lower. Thus, if downstream firms invest more on R&D activity, thereby lowering variable fees, consumers' surpluses are increased. Third, in the symmetric equilibrium, the upstream firm's profit, which is $D(a_i + c_i)(a_i - c_0)$, decreases in c_i. It implies that the upstream firm's profit gets higher, as downstream firms invest more on R&D. To summarize, we have

Proposition 3 *Higher R&D activities of downstream firms lead to higher consumer welfare and higher profits of the upstream firm, but lower profits of the downstream firms.*

Proof. omit.

One implication of Proposition 3 is that the upstream firm prefers to adopt the MFN clause if and only if t is higher than $\widehat{t}$, since it can get higher profits then as downstream firms invest more. This reconfirms the well-known result of time inconsistency in literature on input price discrimination. In this model, two choice variables (a_1, a_2) are available to the upstream firm under DP, while only one choice variable $\widetilde{a}$ is available under UP. Therefore, the upstream firm is weakly better off under DP than under UP, once the downstream firm makes cost-saving investments. However, as DeGraba (1990) and Choi (1995) pointed out, the upstream firm might have an *ex ante* incentive to adopt MFN to induce downstream firms to make higher investments if t is higher than $\widehat{t}$. On the other hand, if t is less than $\widehat{t}$, the downstream firms invest more under DP, implying that it is optimal for the upstream firm not to adopt MFN *ex ante*. Therefore, in this case, the upstream firm faces no time-inconsistency problem.

Now, to investigate the effect of MFN on the overall social welfare, let $SW(c)$ be the social welfare, that is, the sum of the total consumer surpluses and the profits of the three firms, when the marginal costs of downstream firms are c. Let c^{SW} be the maximizer of $SW(c)$. When downstream firms have the same marginal cost c, their gross profits do not depend on c, implying that $\frac{d\pi_i}{dc} = 0$. Thus, c^{SW} satisfies the following first-order condition

$$\frac{dSW(c^{SW})}{dc} = \left[\frac{dCS}{dc} + \frac{d\pi_0(c^{SW})}{dc}\right] - 2\phi'(c^{SW}) = 0, \tag{10}$$

where CS is the total consumer surplus. Under the assumption that the demand function is linear, we find that $c^{SW} < \widetilde{c}^{**}$ by comparing (9) and (10). Since $\widetilde{c}^{**} < c^{**}$ when $t > \widehat{t}$ by Proposition 2, the cost under UP is closer to the socially optimal level than the cost under DP. This implies that the social welfare is higher under UP than under DP when t is higher than $\widehat{t}$.

Proposition 4 *When the demand function is linear, it is socially optimal for the upstream firm to adopt the MFN clause, as long as it is privately optimal for the upstream firm to do* *so.*

Proof. See the appendix.

This proposition has an implication that the upstream firm should not be discouraged from adopting the MFN clause voluntarily from the social point of view. It is, however, to be stressed that Proposition 4 does not imply that it is always socially optimal to adopt the MFN clause.

6 Discussion

A. Cournot Competition

We compare our result with DeGraba (1990) and Choi (1995). They analyze the effect of the MFN clause on downstream firms' R&D incentive when the downstream firms engage in Cournot competition. In particular, Choi (1995) shows that in the Cournot competition $\partial a_i^{**}/\partial c_i$ must be smaller than $\partial a_j^{**}/\partial c_i$. That is, in Cournot competition, an upstream monopolist's discriminatory input price always penalizes a lower-cost downstream firm.

Suppose that two firms engage in Cournot competition. The two firms' output levels are denoted by q_1 and q_2, respectively. If $c_1 < c_2$, firm 1 has less elastic demand for the input, and accordingly the upstream firm sets $a_1 > a_2$. Then, a customer of downstream firm 1 brings the upstream monopolist the profit of $(a_1 - c_0)$, which is larger than $(a_2 - c_0)$. Thus, the upstream firm has an incentive to increase firm 1's market share by reducing a_1 or increasing a_2. However, when the upstream firm reduces a_1, firm 1 will not expand q_1 that much since its profit margin, $p(q_1 + q_2) - a_1 - c_1$, decreases in q_1. Thus, the increase in firm 1's market share is very limited. Consequently, the elasticity effect dominates the market share effect.

B. Bertrand Competition

In this subsection, we examine the R&D incentive of downstream firms in a Bertrand competiton.

Under the MFN clause, when an upstream monopoly charges $\tilde{a}$ for the input price, the two downstream firms' total marginal costs are $\tilde{a}+c_1$ and $\tilde{a}+c_2$, respectively. In the Bertrand competition, only a lower-cost firm can make positive sales, and the equilibrium retail price is equal to the second lowest marginal cost. If $c_1 < c_2$, the equilibrium retail price is $\tilde{a} + c_2$, and the input monopolist's profit is $(\tilde{a} - c_0)D(\tilde{a} + c_2)$, where the demand function is given by $q = D(p)$. Thus, the monopoly chooses $\tilde{a}^*$ maximizing $(\tilde{a} - c_0)D(\tilde{a} + c_2)$. The profit of the lower-cost downstream firm is then $(c_2 - c_1)D(\tilde{a}^* + c_2)$, which is positive.

However, under DP, the input monopolist can charge two input prices, a_1 and a_2. If $c_1 < c_2$, the equilibrium retail price is a_2+c_2,[9] and the input monopolist's profit is $(a_1-c_0)D(a_2+c_2)$, and the lower-cost downstream firm's profit is $(a_2 + c_2 - a_1 - c_1)D(a_2 + c_2)$. The joint profit is $(a_2 + c_2 - c_0 - c_1)D(a_2 + c_2)$. Let p^m denote the joint-profit maximizing retail price, i.e., $p^m = \arg\max(p - c_0 - c_1)D(p)$. If the input monopolist sets $a_2 = p^m - c_2$ and $a_1 = a_2 + c_2 - c_1$, the input monopolist takes the maximized whole industry profits, and the lower-cost downstream firm earns zero. That is, even though a downstream firm achieves a lower cost than the other firm, its profit would be zero. This implies that downstream firms have no incentive to reduce costs under DP. Since the low-cost downstream firm can have a positive profit under UP, downstream firms have a stronger incentive to invest in R&D under the UP regime, as in Cournot competition.

This suggests that if downstream firms compete by linear prices, the MFN clause will have a positive effect on the R&D incentive of downstream firms even if the products are differentiated but the substitutability between the products is very high. That is, unlike two-part tariffs, linear pricing does not, in general, result in a negative effect of MFN on the R&D incentive. The intuition can be provided as follows. In the case of two-part tariffs, the upstream firm's choice of a_i, $i = 1, 2$, directly implies variable fees set to $a_i + c_i$. On the other hand, in the linear pricing case, the upstream firm's choice induces another round of retail price competition between downstream firms. We may call this the third effect of retail price competition. More specifically, since the upstream firm must be concerned about

[9]There are three cases; (i) $a_1 + c_1 < a_2 + c_2$ (ii) $a_1 + c_1 = a_2 + c_2$ (iii) $a_1 + c_1 > a_2 + c_2$. However, it can be easily shown that if $c_1 < c_2$, (a_1, a_2) satisfying (ii) or (iii) is dominated by the pair of input prices satisfying (i).

selling more overall, apart from selling more to the low-cost downstream firm, the firm will find it in its interest to induce fierce competition between downstream firms by choosing a_1 and a_2 such that $a_1 + c_1$ and $a_2 + c_2$ are sufficiently close. Hence, downstream firms will not have enough incentive to reduce their production cost under the DP regime. However, since the third effect does not exist in the case of two-part tariffs, the upstream firm, without caring about the extra constraint,[10] can choose a_1 and a_2 in such a way that it rewards the low cost firm by taking only the elasticity effect and the market share effect into account.

C. Partial Coverage

We have assumed that u_0 is large enough that the two firms cover the whole market in the Hotelling model. However, if the "transportation cost" is too high or if prices are too high, the market is not fully covered, and some consumers subscribe to neither firm. Then, the two firms cover only consumers close to their locations, and both firms behave as a local monopolist. In this case, there is no market share effect, and we can easily extend the argument of Proposition 1 to assert that the effect of MFN on firms' R&D incentive is positive.

7 Policy Implication and Conclusion

We have analyzed the effects of the MFN provision on downstream firms' R&D incentives when the downstream firms compete by offering a two-part tariff. We have shown that the MFN provision can reduce a downstream firm's R&D incentive if downstream products are not differentiated very much.

This analysis can be applied to many situations especially in telecommunications industry, such as competitive long-distance carriers' access to the local exchange carrier and competitive mobile carriers' access to the fixed network. We admit of course that our result is not directly applicable to those access issues, because the government usually regulates the access charge to the bottleneck facility, implying that discriminatory access charges are

[10]When $t > 0$, this constraint that $a_1 + c_1 \approx a_2 + c_2$ does not have to be satisfied exactly, but roughly satisfied in the sense that it cannot be ignored.

disallowed. However, our analysis suggests that the current regulation rule can have a negative effect on social welfare if downstream products are not so much differentiated. As another limitation of our model, we did not consider two-way interconnection by assuming away the possibility of the bottleneck sector's access to the competitive sector. However, even if we allow the possibility of two-way access, the main result will remain unaffected as long as the upstream firm sets a uniform price for outgoing calls to firm 1 and firm 2, because the upstream firm's pricing cannot affect the consumers' decision of network affiliation nor downstream firms' pricing. If the upstream firm can charge different prices, it can penalize/reward a low-cost downstream firm either by access charges or by retail prices. Since it can still employ retail prices for penalizing/rewarding even under the MFN provision, the effect of the MFN provision will be presumably alleviated.

We believe that the applicability of our analysis is not restricted to the telecommunications industry, but can be extended to more broadly defined network industries including electrical power generators' access to a power transmission grid, natural gas providers' access to pipelines, etc. We look forward to richer implications of this model in those contexts.

Appendix

Proof of Lemma 1:

By using $p_i^* = a_i^* + c_i$, equation (1) can be written as

$$D(p_i^*) + (p_i^* - c_i - c_0)D'(p_i^*) = 0.$$

By total differentiating it, we get

$$\frac{dp_i^*}{dc_i} = \frac{D'(p_i^*)}{2D'(p_i^*) + (p_i^* - c_i - c_0)D''(p_i^*)}.$$

Since $D' < 0$, $D'' \leq 0$ and $p_i^* - c_i - c_0 > 0$, we have $0 < dp_i^*/dc_i \leq 1/2$. Since $p_i^* = c_i + a_i$, the fact that $dp_i^*/dc_i \leq 1/2$ implies that $da_i^*/dc_i < 0$. $\|$

Proof of Lemma 2:

By total differentiation of (2), we get

$$\frac{d\tilde{a}^*}{dc_i} = -\frac{n_i D'(\tilde{a}^* + c_1) + n_i(\tilde{a}^* - c_0)D''(\tilde{a}^* + c_1)}{2n_1 D'(\tilde{a}^* + c_1) + 2n_2 D'(\tilde{a}^* + c_2) + (\tilde{a}^* - c_0)[n_1 D''(\tilde{a}^* + c_1) + n_2 D''(\tilde{a}^* + c_2)]}.$$

Since $D'' \leq 0$ and $\widetilde{a}^* - c_0 > 0$, we have $d\widetilde{a}^*/dc_i < 0$. $\|$

Proof of Lemma 3: This is immediate from the following claim.

Claim 1 *$a^*(\overline{c}) < \widetilde{a}^*(c_i, c_j) < a^*(\underline{c})$, for all c_i and c_j such that $c_i \neq c_j$, where $\underline{c} = \min\{c_i, c_j\}$ and $\overline{c} = \max\{c_i, c_j\}$.*

Proof. When both firms have the same costs, firm 0 does not benefit from charging different input prices. That is, if $c_i = c_j$, then $\widetilde{a}^*(c_i, c_j) = a^*(c_i)$. When the two firms have different costs, we know from Lemma 1 that $a^*(\underline{c}) > a^*(\overline{c})$. Since $\partial\widetilde{a}^*(c_i, c_j)/\partial c_i < 0$ from Lemma 2 (and $\partial\widetilde{a}^*(c_i, c_j)/\partial c_j < 0$ by symmetry), we have $a^*(c_j) = \widetilde{a}^*(c_j, c_j) < \widetilde{a}^*(c_i, c_j) < \widetilde{a}^*(c_i, c_i) = a^*(c_i)$. $\|$

Proof of Proposition 1:

Equations (3) and (4) hold at c_i^* and $\widetilde{c}_i^*$, respectively. Since $\partial\pi_i/\partial a_i < 0$ and $|\partial a^*(c_i)/\partial c_i| > |\partial\widetilde{a}(c_i, c_j)/\partial c_i|$, the left hand side of equation (4) is negative at c_i^*. Since $\pi_i - \phi(c_i)$ is concave, we have $c_i^* > \widetilde{c}_i^*$. $\|$

Proof of Lemma 4:

Since $p_i = a_i + c_i$, firm 0' profit function can be written in terms of the final variable fees, p_1 and p_2

$$\pi_0 = \alpha_1(p_1 - c_0 - c_1)D(p_1) + \alpha_2(p_2 - c_0 - c_2)D(p_2).$$

Suppose that $c_1 = c_2$. Then, firm 0 sets its price maximizing $(p - c_0 - c_2)D(p)$. Let $\widetilde{p}$ denote the optimal price. By setting p_1 and p_2 at $\widetilde{p}$, firm 0's profit becomes $(\widetilde{p} - c_0 - c_2)D(\widetilde{p})$. Suppose that c_1 gets lower than c_2. By the envelope theorem, we know that $\partial\pi_0/\partial c_i = -\alpha_i D(\widetilde{p}) < 0$, which implies that the firm's overall profit must increase when c_1 gets lower. Let p_1^{**} and p_2^{**} denote the firm's optimal prices at the new cost structure c_1 and c_2, where $c_1 < c_2$. The firm's profit at (c_1, c_2) is $\alpha_1^{**}(p_1^{**} - c_0 - c_1)D(p_1^{**}) + \alpha_2^{**}(p_2^{**} - c_0 - c_2)D(p_2^{**})$. Suppose that $(p_1^{**} - c_0 - c_1)D(p_1^{**}) < (p_2^{**} - c_0 - c_2)D(p_2^{**})$. Because $(p_2^{**} - c_0 - c_2)D(p_2^{**}) < (\widetilde{p} - c_0 - c_2)D(\widetilde{p})$, we have that $\alpha_1^{**}(p_1^{**} - c_0 - c_1)D(p_1^{**}) + \alpha_2^{**}(p_2^{**} - c_0 - c_2)D(p_2^{**}) < (\widetilde{p} - c_0 - c_2)D(\widetilde{p})$. That is, the firm's overall profit would get lower. This is a contradiction. Therefore, if $c_i < c_j$, then $(a_j^{**} - c_0)D(a_j^{**} + c_j) \leq (a_i^{**} - c_0)D(a_i^{**} + c_i)$. $\|$

Proof of Lemma 5:

(i) For simpler notation, let q_i denote $D(p_i)$. Since $\tilde{a}^{**} + c_i = \tilde{a}^{**} + c_j$, we have $q_1 = q_2$. By total differentiation of (6) and using $q_1 = q_2$, we have

$$\frac{d\tilde{a}^{**}(c_1, c_2)}{dc_1} = -\frac{\alpha_1 q_1' + \alpha_1(\tilde{a}^{**} - c_0)q_1''}{2\alpha_1 q_1' + 2\alpha_2 q_2' + (\tilde{a}^{**} - c_0)[\alpha_1 q_1'' + \alpha_2 q_2'']}.$$

Since $D' < 0$, $D'' \leq 0$ and $\tilde{a}^{**} - c_0 > 0$, we have $d\tilde{a}^{**}(c_i, c_j)/dc_i < 0$.

(ii) We let q_i denote $D(a_i^{**} + c_i)$ for simpler notation. Total differentiation of (7) gives

$$\begin{pmatrix} \frac{\partial^2 \pi_0}{\partial a_1^2} & \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} \\ \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} & \frac{\partial^2 \pi_0}{\partial a_2^2} \end{pmatrix} \begin{pmatrix} \frac{\partial a_1^{**}}{\partial c_1} \\ \frac{\partial a_2^{**}}{\partial c_1} \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 \pi_0}{\partial a_1 \partial c_1} \\ -\frac{\partial^2 \pi_0}{\partial a_2 \partial c_1} \end{pmatrix}.$$

First, note that $\frac{\partial \pi_0}{\partial a_1} = \frac{\partial \pi_0}{\partial c_1} + \alpha_1 q_1$. Thus, we have that $\frac{\partial^2 \pi_0}{\partial c_1 \partial a_1} = \frac{\partial^2 \pi_0}{\partial a_1^2} - A$ and $\frac{\partial^2 \pi_0}{\partial c_1 \partial a_2} = \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} - B$, where $A = \alpha_1 \frac{\partial q_1}{\partial a_1} - \frac{q_1}{6t} q_1$ and where $B = \frac{q_2}{6t} q_1$. By Cramer's rule, we have

$$\begin{aligned}
\frac{\partial a_1^{**}}{\partial c_1} &= \frac{1}{H}\left[\frac{\partial^2 \pi_0}{\partial a_2^2}\left(-\frac{\partial^2 \pi_0}{\partial a_1^2} + A\right) + \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2}\left(\frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} - B\right)\right], \\
\frac{\partial a_2^{**}}{\partial c_1} &= \frac{1}{H}\left[\frac{\partial^2 \pi_0}{\partial a_1 \partial a_2}\left(\frac{\partial^2 \pi_0}{\partial a_1^2} - A\right) + \frac{\partial^2 \pi_0}{\partial a_1^2}\left(-\frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} + B\right)\right] \\
&= \frac{1}{H}\left[\frac{\partial^2 \pi_0}{\partial a_1^2} B - \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} A\right],
\end{aligned}$$

where $H = \frac{\partial^2 \pi_0}{\partial a_1^2}\frac{\partial^2 \pi_0}{\partial a_2^2} - (\frac{\partial^2 \pi_0}{\partial a_1 \partial a_2})^2$.

Under the UP regime, firm 0 must set $a_1 = a_2$. Thus, firm 0's optimization problem can be written in terms of Lagrangian function $L = \pi_0(a_1, a_2) + \lambda(a_1 - a_2)$. The corresponding first order conditions are

$$\begin{aligned}
L_1 &= \frac{\partial \pi_0}{\partial a_1} + \lambda = 0, && (11) \\
L_2 &= \frac{\partial \pi_0}{\partial a_2} - \lambda = 0, && (12) \\
L_3 &= a_1 - a_2 = 0. && (13)
\end{aligned}$$

Total differentiation of equations (11), (12) and (13) gives

$$\begin{pmatrix} \frac{\partial^2 \pi_0}{\partial a_1^2} & \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} & 1 \\ \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} & \frac{\partial^2 \pi_0}{\partial a_2^2} & -1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial a_1^{**}}{\partial c_1} \\ \frac{\partial a_2^{**}}{\partial c_1} \\ \frac{\partial \lambda}{\partial c_1} \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 \pi_0}{\partial a_1 \partial c_1} \\ -\frac{\partial^2 \pi_0}{\partial a_2 \partial c_1} \\ 0 \end{pmatrix}.$$

By Cramer's rule, we get

$$\frac{\partial \widetilde{a}^{**}}{\partial c_i} = \frac{1}{\overline{H}}\left(\frac{\partial^2 \pi_0}{\partial a_1^2} - A + \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} - B\right),$$

where $\overline{H} = -\left(\frac{\partial^2 \pi_0}{\partial a_1^2} + \frac{\partial^2 \pi_0}{\partial a_2^2} + 2\frac{\partial^2 \pi_0}{\partial a_1 \partial a_2}\right)$.

Since $\frac{\partial^2 \pi_0}{\partial a_1^2} = \frac{\partial^2 \pi_0}{\partial a_2^2}$ when $c_1 = c_2$, we have $\frac{\partial a_1^{**}}{\partial c_i} + \frac{\partial a_2^{**}}{\partial c_i} = 2\frac{\partial \widetilde{a}^{**}}{\partial c_i}$. $\|$

Proof of Lemma 6:

From the proof of Lemma 5(ii), we know that

$$\frac{\partial a_1^{**}}{\partial c_1} = \frac{1}{H}\left[\frac{\partial^2 \pi_0}{\partial a_2^2}\left(-\frac{\partial^2 \pi_0}{\partial a_1^2} + A\right) + \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2}\left(\frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} - B\right)\right],$$

$$\frac{\partial a_2^{**}}{\partial c_1} = \frac{1}{H}\left[\frac{\partial^2 \pi_0}{\partial a_1^2}B - \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2}A\right],$$

where A and B are defined in the proof of Lemma 5(ii). Hence,

$$\frac{\partial a_1^{**}}{\partial c_1} - \frac{\partial a_2^{**}}{\partial c_1} = \frac{(A - B)}{\frac{\partial^2 \pi_0}{\partial a_1^2} - \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2}} - 1.$$

Now, we have

$$\frac{\partial^2 \pi_0}{\partial a_1^2} - \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} = \alpha_1\left[2q_1' + (a_1^{**} - c_0)q_1''\right] - \frac{2q_1}{6t}\left[q_1 + q_2 + 2(a_1^{**} - c_0)q_1'\right], \tag{14}$$

where $q_i = D(a_i^{**} + c_i)$. In the symmetric case, we have $q_1 = q_2$. Also, since the market share effect is zero in the symmetric case, a_1^{**} is set such that $q_1 + (a_1^{**} - c_0)q_1' = 0$. Thus, equation (14) is reduced to

$$\frac{\partial^2 \pi_0}{\partial a_1^2} - \frac{\partial^2 \pi_0}{\partial a_1 \partial a_2} = \alpha_1[2q_1' + (a_1 - c_0)q_1''].$$

Therefore, we have

$$\frac{\partial a_1^{**}}{\partial c_1} - \frac{\partial a_2^{**}}{\partial c_1} = -\frac{\alpha_1[q_1' + (a_1 - c_0)q_1''] + (q_1/6t)(q_1 + q_2)}{\alpha_1[2q_1' + (a_1 - c_0)q_1'']}.$$

The denominator is negative since $q_1' < 0$, $q_1'' < 0$. The sign of nominator is positive if and only if $t < \widehat{t} = -\frac{q_1(q_1+q_2)}{6\alpha(q_1' + (a_1 - c_0)q_1'')}(> 0)$. That is, $\partial a_i^{**}/\partial c_i > \partial a_j^{**}/\partial c_i$ if and only if $t < \widehat{t}$. $\|$

Proof of Lemma 7:

We let

$$\Psi \equiv \left(\frac{\partial \pi_i}{\partial a_i}\frac{\partial a_i^{**}}{\partial c_i} + \frac{\partial \pi_i}{\partial a_j}\frac{\partial a_j^{**}}{\partial c_i}\right) - \left(\frac{\partial \pi_i}{\partial a_i}\frac{\partial \widetilde{a}}{\partial c_i} + \frac{\partial \pi_i}{\partial a_j}\frac{\partial \widetilde{a}}{\partial c_i}\right).$$

Using Lemma 5(ii), we have

$$\Psi = \frac{1}{2}\left(\frac{\partial a_1^{**}}{\partial c_1} - \frac{\partial a_2^{**}}{\partial c_1}\right)\left(\frac{\partial \pi_1}{\partial a_1} - \frac{\partial \pi_1}{\partial a_2}\right).$$

By the envelope theorem, we know that $\frac{\partial \pi_1}{\partial a_1} < 0$ and that $\frac{\partial \pi_1}{\partial a_2} > 0$. Thus, $\frac{\partial \pi_1}{\partial a_1} - \frac{\partial \pi_1}{\partial a_2} < 0$. Therefore, $\Psi < 0$ if and only if $\frac{\partial a_1^{**}}{\partial c_1} - \frac{\partial a_2^{**}}{\partial c_1} > 0$. Equation (8) and (9) hold at $c = c^{**}$ and $c = \widetilde{c}^{**}$ respectively. Suppose $\Psi < 0$, i.e., $\frac{\partial a_1^{**}}{\partial c_1} - \frac{\partial a_2^{**}}{\partial c_1} > 0$. This implies that $\widetilde{\Phi}_i(c^{**}) > \Phi_i(c^{**}) = 0$. Since $\pi_i - \phi(c_i)$ is concave, we have $\widetilde{c}^{**} > c^{**}$. Thus, $\widetilde{c}^{**} > c^{**}$ if and only if $\frac{\partial a_i^{**}}{\partial c_i} > \frac{\partial a_j^{**}}{\partial c_i}$ ||

Proof of Proposition 4:

Since $\widetilde{c}^{**} < c^{**}$ when $t > \widehat{t}$, we have only to show that $c^{SW} < \widetilde{c}^{**}$. Let us compare equations (9) and (10). In equation (10), by the envelope theorem, we have $\frac{d\pi_0(c)}{dc} = -q_i$. Also, we have $\frac{dCS}{dc} = \frac{\partial CS}{\partial c} + \frac{\partial CS}{\partial a_i}\frac{\partial a_i}{\partial c}$. When the two firms lower the cost by the same amount, we have $\frac{\partial a_i}{\partial c} = -\frac{1}{2}$ in the linear demand case, and $\frac{\partial CS(c)}{\partial c} = \frac{\partial CS(a_i)}{\partial a_i} = -q_i$ since $p_i = a_i + c_i$. Hence, equation (10) boils down to $-\frac{3}{2}q_i - 2\phi'(c^{SW}) = 0$.

Now, consider equation (9). We have $\frac{\partial \pi_i}{\partial a_i} = -\frac{2}{3}q_i$, and $\frac{\partial \pi_i}{\partial a_j} = \frac{1}{6}q_i$. Also, since $\pi_i = \alpha((p_i - a_i - c_i)q_i + F_i)$, $\frac{\partial \pi_i}{\partial c_i} = -\frac{2}{3}q_i$ in the symmetric equilibrium by using $\frac{\partial \alpha_i}{\partial c_i} = -\frac{1}{6t}q_i$, $\alpha = 1/2$ and $F_i = t$. Since $\frac{\partial \widetilde{a}}{\partial c_i} = -\frac{1}{4}$ in the linear demand case, (9) is reduced to $-\frac{13}{24}q_i - \phi'(\widetilde{c}^{**}) = 0$. Therefore, the comparison between (9) and (10) tells us that $c^{SW} < \widetilde{c}^{**}$. Thus, from Proposition 3, the result is immediate. ||

References

[1] Armstrong, M., 1998, Network Interconnection in Telecommunications, *Economic Journal* 108, 545-564

[2] Carter, M. and J. Wright, 1999, Interconnection in Network Industries, *Review of Industrial Organization* 14, 1-25

[3] Choi, J. 1995, Optimal Tariff and the Choice of Tariff vs. the 'Most Favored Nation' clause, *Journal of International Economics* 38, 1995, 143-160

[4] DeGraba, P., 1990, Input Market Price Discrimination and the Choice of Technology, *American Economic Review* 80, 1246-1253

[5] DeGraba, P., 2001, Efficient Interconnection for Competing Networks When Customers Share the Value of a Call, Mimeo

[6] Gans, J., 2001, Regulating Private Infrastructure Investment: Optimal Pricing for Access to Essential Facilities, *Journal of Regulatory Economics* 20, 167-189

[7] Katz, M.L., 1987, The Welfare Effects of Third-degree Price Discrimination in Intermediate Goods Market, *American Economic Review* 77, 154-167

[8] Kim, J.-Y. and Y. Lim, 2004, Revenue-Sharing Rule for Interconnection Charges. *Japanese Economic Review* 55, 298-310

[9] Laffont, J-J., P. Rey and J. Tirole, 1998, Network Competition: I. Overview and Nondiscriminatory Pricing, *Rand Journal of Economics* 29, 1-37

[10] Valletti, T., 2003, Input Price Discrimination with Downstream Cournot Competitors, *International Journal of Industrial Organization* 21, 969-988

[11] Valletti, T. and C. Cambini, 2005, Investments and Network Competition, *Rand Journal of Economics* 36, 446-467

[12] Yoshida, Y., 2000, Third-degree Price Discrimination in Input Markets: Output and Welfare. *American Economic Review* 90, 240-246

International University in Germany

School of Business Administration

Access Regulation and the Adoption of VoIP

Paul de Bijl,
Martin Peitz

Working Paper 43/2006

Bruchsal, May 2006

ISSN 1613-6691 (print version)
ISSN 1613-6705 (online version)

Prof. Dr. Martin Peitz is Professor of Economics and Quantitative Methods at *International University in Germany*, Bruchsal.

Paul de Bijl, Ph.D. is Research Coordinator of TILEC and CentER Fellow at *Tilburg University*.

Contact: martin.peitz@i-u.de, pdebijl@uvt.nl

Access Regulation and the Adoption of VoIP[1]

Paul W. J. de Bijl[2]
Tilburg University

Martin Peitz[3]
International University in Germany and University of Mannheim

31 May 2006

[1]Financial support from ENCORE (Economics Network for Competition and Regulation) is gratefully acknowledged. We would like to thank Georg Götz; the Dutch telecommunications and postal authority OPTA, in particular, Jorn van Steenis and Elbert Jan van Veldhuizen; and seminar participants at OPTA, University of Antwerp, Tilburg University, and the 2006 meeting of the Industrial Economics Section of the Verein für Socialpolitik for helpful comments and discussions. Only the authors are responsible for the contents of this paper and any errors it may contain.

[2]TILEC (Tilburg Law and Economics Center) and CentER for Economic Research, Tilburg University, PO Box 90153, 5000 LE Tilburg, Netherlands, also affiliated with the ENCORE Fellows Network; pdebijl@uvt.nl.

[3]School of Business Administration, International University in Germany, Campus 3, 76646 Bruchsal, Germany, also affiliated with ENCORE and CESifo; Martin.Peitz@i-u.de.

Abstract

The introduction of packet-switched telephony in the form of VoIP raises concerns about current regulatory practice. Access regulation has been designed for traditional telephony on PSTN networks. In this paper we analyze the effect of access regulation of PSTN networks on the adoption of a new technology in the form of VoIP. In particular, we show that with endogenous consumer choice between PSTN and VoIP telephony, higher prices for terminating access to the PSTN network make VoIP less likely to succeed and lead to lower profits of operators that offer exclusively VoIP telephony.

JEL-Classification: L96, L51, L13.

Keywords: telecommunications, voice over broadband (VoB), voice over Internet protocol (VoIP), entry, access, regulation, imperfect competition.

1 Introduction

With the emergence of voice telephony based on the Internet protocol, generally known as "voice over IP" or VoIP, the telecommunications landscape is rapidly changing. This new technology, which is fundamentally different from telephony over the PSTN (public switched telephone network), is providing a new impetus to local loop unbundling (LLU) and also stimulates entry into telephony markets by cable operators. Without even mentioning software applications for voice telephony that run completely over the Internet, it is clear that incumbent operators are facing a serious threat. In particular, they face the question of whether they should milk the PSTN as long as possible, or introduce VoIP quickly and at low prices, at the cost of cannibalizing PSTN revenues, in the hope of at least partially deterring entry of new operators.

In this paper we explore a situation of imperfect competition between an incumbent and an entrant. While the incumbent, with a history in PSTN telephony, is assumed to have a complete local access network, the entrant is either a cable operator with a full-coverage broadband network or a newcomer who uses LLU to reach end-users. The incumbent offers PSTN (public switched telephone network) voice telephony to one segment of customers, as well as VoIP services to another segment, while the entrant only offers VoIP services in the latter segment. We distinguish two set-ups. In the first one, the relative size of these segments is exogenously given, so that there is no migration from PSTN to VoIP services. In the second set-up, we allow for endogenous migration between the segments, so that consumers actually choose between staying with the PSTN network versus adopting VoIP services.[1]

In this set-up, we explore the nature of competition in a market for voice telephony between an incumbent trying to balance its tactics with regard to PSTN and VoIP telephony, and an entrant without ties to the past. In addition, we explicitly focus on the effects of regulation on the market. Accordingly, there is a close link to regulatory practice. National regulatory authorities (NRAs) are currently struggling with the question whether they should restrict the incumbent's activities with regard to VoIP, or refrain from intervening so that the market will determine how this new technology will develop. Regulation may be necessary in order to prevent anticompetitive behavior, but on the other hand, intervening may easily distort the development and adoption processes of innovation.

[1] In both cases, we assume that there is full coverage, that is, all consumers make a purchase. See De Bijl and Peitz (forthcoming) for an analysis of partial market coverage in a setting that focuses on unbundling rather than terminating access.

In this paper, we mainly focus on regulation of terminating access. We will also discuss the effects that the retail price in the market for PSTN telephony may have on market outcomes in the VoIP segment. Accordingly, we will not so much be considering regulation of the VoIP market itself. Instead, we look at the broader regulatory picture, which may partly be motivated by considerations of a universal service obligation in the PSTN market, or by market power with regard to call termination on the PSTN. Within the European regulatory framework for communications markets, such considerations—which may be legitimate in a relatively isolated context—can easily trigger regulatory interventions, such as regulation of the incumbent's access price. However, because of network interconnection, regulators should be aware of the effects that they may have on emerging markets. For a recent policy document on this, see OECD (2006).[2] Our aim is to articulate some of the most salient side effects of current regulatory interventions.

The assumption of imperfect competition means that both operators have some market power; such a situation is more realistic than the more stylized set-up with a competitive fringe that needs to purchase the essential input from the incumbent. In practice, entry immediately tends to generate some discipline on incumbents. In such a situation, supposing that VoIP operators use "bill-and-keep" or some other predetermined scheme for call termination, we analyze the competitive effects of terminating access at the PSTN network. In particular, we consider access that is not priced at its underlying marginal cost level. We will summarize our results in the concluding section of the paper.

Literature review. Terminating access in telecommunications networks has been recently analyzed in situations in which operators need mutual access. This literature on two-way access includes the seminal papers by Armstrong (1998) and Laffont, Rey and Tirole (1998); overviews are provided by Laffont and Tirole (2000), Armstrong (2002), De Bijl and Peitz (2002), and Vogelsang (2003). Our paper builds on that literature by analyzing the emergence of VoIP networks in a PSTN environment, and in which the PSTN and VoIP networks are interconnected. In such an environment, "metering"of call traffic may still make sense on the PSTN, but this may be different for VoIP calls. Hence the natural starting point is to consider the case in which VoIP operators do not charge for call termination.

Note that technically speaking, we look at a two-way access problem. However, when VoIP operators do not charge for access, our set-up boils down to a one-way access problem. Hence

[2] We will see later that some of the views in OECD (2006), on the link between the PSTN access price and the intensity of competition in the VoIP market, may lead to the wrong conclusions.

our paper directly connects to the literature on one-way access, which has typically focused on access problems in which either all firms or at least the downstream entrants do not have market power in the retail segment. In the former case there is pure Bertrand competition, and in the latter case, downstream entrants form a (perfectly) competitive fringe. In such settings, when the incumbent's retail price is assumed to be fixed, the efficient component pricing rule (ECPR) has been proposed as the socially optimal pricing method for setting the incumbent's access price (see Baumol, 1983, and Willig, 1979). The ECPR says that to obtain productive efficiency, the access price should be equal to the marginal cost of access plus the incumbent's opportunity cost of providing access (the incumbent's lost profits due to entry). If the incumbent and the entrant offer perfect substitutes, this rule reduces to the margin rule according to which the incumbent's lost profit equals its lost retail revenues (Armstrong, 2002). Accordingly, the vertically integrated incumbent can increase its profits by granting access to a more efficient downstream entrant. Since foreclosure is unprofitable if an entrant is more efficient, there is no need to regulate the incumbent's access price. However, the logic behind the ECPR has been challenged in several adaptations and extensions; see Armstrong (2002) for a thorough overview.

Access pricing in a situation of imperfect competition (with regulated access prices) has been analyzed by Laffont and Tirole (1994, 1996), Armstrong and Vickers (1998), Lewis and Sappington (1999) and De Bijl and Peitz (forthcoming), among others. Laffont and Tirole (1994) analyze the implementation of Ramsey prices via a global price cap. Apparently, imperfect competition complicates the appropriate use of a global price cap, because supply-side and demand-side effects have to be taken into account simultaneously. In a full information world, in which only the access price is regulated, the idea is to replace retail regulation by competition. Here, access prices can be used as a regulatory instrument to affect retail price levels. If, in particular, the regulator can set different rates for a bottleneck owner and a non-integrated competitor, the regulator may want to subsidize the competitor at the margin to increase competitive pressure (Ebrill and Slutsky, 1990; and Lewis and Sappington, 1999). In addition, in an asymmetric market the regulator may want to use the access price to favor the more efficient firm (Armstrong and Vickers, 1998; Lewis and Sappington, 1999; and De Bijl and Peitz, forthcoming). The reason behind such a policy is that the more efficient firm would otherwise obtain a market share that is less than socially optimal. Note that such a bias becomes less pronounced in a more competitive market (Lewis and Sappington, 1999).

Foros (2004) analyzes a competitive situation which is somewhat related to ours. To consider the retail market for Internet access, he models a situation of a vertically integrated firm controlling both local access and providing broadband access, and a downstream Internet retailer. The integrated firm can invest in the capacity of local connections, and given the outcome of that decision, the regulator chooses an access price. The firms compete in a Cournot fashion in the retail market. The focus of the paper is mainly on regulation as a way to induce the integrated firm to invest efficiently and to deter it from foreclosing the market. Hence, Foros' results can be seen as complementary to ours.

The economics literature has also looked at bypass possibilities (see e.g. Armstrong, Doyle and Vickers, 1996; Laffont and Tirole, 1994, 1996). Note, however, that although VoIP can substitute PSTN telephony, it does not allow for full bypass as long as some consumers stay with the PSTN. In our model, with full penetration of VoIP the access problem has disappeared whereas at any intermediate situation, terminating access to PSTN remains essential and a bypass possibility is not available.

With respect to our modeling framework, this paper also contributes to the literature on multiproduct firms in a multidimensional product space. This literature has been reviewed in Manez and Waterson (2001).[3] In this paper we provide a tractable model of multiproduct competition which allows for endogenous formation of market segments and can be solved analytically. However, our model has the property that integration does not affect the incentives with respect to prices (in the VoIP segment). Our framework may prove to be useful for other applications in industrial organization.

The structure of this paper is as follows. Section 2 provides some illustrative background information on terminating access in relation to VoIP. In section 3, we explore a model in which the group of consumers purchasing VoIP services is exogenously given. Section 4 analyzes the case in which migration from PSTN to VoIP services is endogenously determined. Section 5 concludes the paper and summarizes the results.

2 Terminating access and VoIP in practice

In this section, we provide some background on IP-based services, wholesale access to local access networks, and call termination. Also, based on this background information we clarify the focus of our analysis, as it is beyond the scope of this paper to provide a comprehensive analysis of VoIP. We abstract from LLU regulation in order to focus on regulation of

[3] Contributions include Katz (1984), Lal and Matutes (1989) and Canoy and Peitz (1997).

terminating access.[4]

2.1 IP-based telephony

The Internet Protocol (IP) is a data protocol, based on packet switching rather than circuit switching, that is used for routing and carriage of messages over the Internet. As any type of electronic information can be transported in packets, IP can also be applied to transport voice calls.

The number and variety of IP-based telephony service propositions are large and increasing, and terms like VoIP, IP telephony and Internet telephony are often used interchangeably. In order to clarify the focus of our paper, it is useful to recapitulate the main types of IP-based telephony:

1. *IP-based transport in the core of traditional networks:* At the level of long-distance backbones, PSTN operators have been supplementing and replacing traditional circuit-switched technology with IP-based technology. To reach end-users, they may still use traditional (i.e., not upgraded to DSL) copper wires connections.

2. *IP-based transport at the edges of traditional networks, allowing for IP-based offerings from traditional operators:* PSTN operators may upgrade their local connections to digital subscriber lines (DSL), enabling broadband Internet access.[5] Such connections may also be used to offer VoIP telephony. This type of VoIP is also know as voice over broadband (VoB) or voice over DSL (VoD). Operators may offer Internet access and VoIP services as a bundle.

3. *IP-based offerings from cable operators:* Cable operators may adapt their local lines so that they can carry high-speed two-way traffic, enabling broadband Internet access as well as VoIP telephony. Again, these offerings may be sold as a bundle.

4. *IP-based offerings from entrants without local networks base on LLU:* If the incumbent's local network is unbundled, entrants without their own local loops can lease unbundled local lines from the incumbent and offer broadband Internet access or VoIP services to

[4] For an overview of the development of LLU throughout Europe and the European regulatory framework, see De Bijl and Peitz (2005).

[5] DSL is a technique that increases the available frequency spectrum on copper wires, so that more data can be sent through a line.

end-users. This type of VoIP is also know as voice over broadband (VoB) or voice over DSL (VoD).

5. *IP-based offerings from entrants without local networks based on bitstream access:* Entrants without their own local loops can purchase "bitstream "access from the incumbent or an entrant using the incumbent's unbundled DSL connections, and offer broadband Internet access or VoIP services to end-users.

6. *IP-based offerings through an ISP:* End-users, subscribing to an ISP, may purchase VoIP from or via that ISP.

7. *Fully IP-based "next generation networks":* Operators may roll out new networks or upgrade existing ones to create completely IP-based networks. An example is BT's 21st Century Network.

8. *VoIP over the Internet:* This is IP-based telephony that is purely Internet-based. Consumers with Internet access can download free, peer–to-peer based, voice telephony software, enabling them to make free calls to consumers with the same software installed on their computers (computer-to-computer calls). A well-known example of this software is Skype. Calls to subscribers of other telecoms networks (computer-to-phone calls) are also possible, although they may be charged, as termination on other networks may be costly.

9. *IP-based private branch exchanges (PBXs):* Corporate customers may, for in-house telecommunications services on their local and wireless access networks (LAN and WAN), use IP-enabled switches. Traditionally, in-house switches were circuit-switched. Note that with an IP-PBX, calls to the outside world may be transformed into circuit-switched calls, depending on the nature of the network that a customer subscribes to.

Experts expect that the current variety will remain for the foreseeable future, if only because of the wide diversity in the ways that end-users have access to services.[6]

In this paper, we restrict our focus to competition between an incumbent offering IP-based services (type 2 or 7) and an entrant with or without a local network (type 3, 4 or 5). Note that for type 7 (the incumbent upgrading to a next generation network), we assume in our analysis that it is still in the process of upgrading its network from a PSTN to an all-IP

[6] Ofcom (2006, p. 8-9).

network, so that during this transition, it offers both PSTN and IP-based telephony. In case of a LLU-based entrant (type 4), we abstract from problems associated with the setting of the wholesale lease price of local loops. In our analysis, we assume that if a customer switches to a LLU-based entrant, he or she completely substitutes the PSTN service with the entrant's VoIP service. Hence our analysis captures "naked" DSL (also known as "standalone" DSL), a service proposition in which an entrant provides only a broadband Internet connection based on DSL (typically priced at a flat rate) by leasing only the broadband part of the frequency spectrum of the copper wire. Accordingly, the narrowband part of the line is no longer used. Finally note that in our models, we implicitly allow for IP-based backbones (type 1), as we do not specify the nature of long-distance backbones.

2.2 Terminating access

Public telecoms networks, whether PSTN or IP-based, must interconnect with one another, so that users can be reached irrespective of the network that they subscribe to. The process that makes this possible, network interconnection, consists of the mutual provision of terminating access. Traditionally, operators charge each other for call termination. Charging for call termination is typically done on a per-minute basis. An alternative to charging for access is "bill-and-keep" (or "reciprocal settlement-free termination"), a system in which calls are terminated without access payments between operators. The emergence of VoIP may radically change operators' wholesale deals on call termination.[7]

With IP-based telephony, the rationale behind termination charges is undermined, as the marginal cost of call termination is drastically reduced, and VoIP calls are often not metered anymore. Nevertheless, calls from an entrant's VoIP network to the incumbent's PSTN network are delivered at a traditional circuit-switched interconnection point or through a "gateway", which allows for straightforward identification of incoming calls, and, hence, for termination charges. Accordingly, for calls from an IP-network to a PSTN, a VoIP provider may have to pay for call termination. Such charges create a perceived marginal cost for the VoIP provider, which possibly translates into a strictly positive per-minute price for this type of call. In the case of calls from one IP-based network to another, operators may find it more efficient to implement bill-and-keep, in line with the packet-based nature of VoIP that, to a certain extent, eliminates the logic of metering incoming calls.

In the basis case of our analysis, we suppose that the incumbent charges for call termina-

[7] See Analysys (2004) for an overview of possible business models for call termination.

tion on its PSTN, and that no termination charges are used for other types of calls. This is in line with the observation that the marginal cost for termination at the PSTN is typically seen as being strictly positive, whereas call termination on IP-based networks comes virtually without a cost. It is straightforward to consider different wholesale pricing schemes in our models, though.[8]

3 Exogenously given consumer segments

3.1 The model

There are two firms, an incumbent (operator 1) and an entrant (operator 2). The incumbent is assumed to have a complete local access network. The incumbent's network can be used for PSTN-based telephony as well as IP-based telephony (VoIP). For instance, its local connections have been upgraded to allow for Digital Subscriber Line (DSL) technology, and its (long-distance) backbone to an IP-based network. The entrant uses only IP-based technology to offer voice services. The entrant may be a cable operator with a full-coverage broadband network. Alternatively, it may be using LLU to reach end-users, that is, it leases unbundled local connections from the incumbent. In the latter case, we assume that the line rental of the local loop is regulated at a cost-based level, so that the entrant is on an equal footing as the incumbent. This assumption allows us to abstract from regulatory issues that stem from LLU, and focus solely on terminating access.

Consumers are heterogeneous with respect to their reluctance to use a new rather than an established technology (with some abuse of language, we will sometimes refer to consumer groups as "old" and "new" segments). The incumbent offers PSTN-based voice telephony to customers with little technological savvy (the old segment), as well as VoIP to the new segment which is open to a new technology, while the entrant only aims at the latter segment by offering VoIP. The group of old consumers is of size λ_0 and the other consumer group of size λ. The total size of the market is normalized to 1, so that $\lambda_0 + \lambda = 1$. More precisely, there is a continuum of consumers with mass 1. A possible interpretation is that consumers in the old segment are narrowband users, whereas consumers in the new segment are broadband users.

[8] Such a structure of access prices may be a good approximation of the outcome of negotiations between VoIP and PSTN operators. As stated in an OECD report (OECD, 2006, p. 26), "...it seems likely in reality that VoIP operators might not charge PSTN operators for IP termination while PSTN operators would still charge VoIP operators for the same call in the opposite direction, due to the VoIP providers' weaker negotiation power."

Throughout this section, we assume that consumers cannot "migrate" from one segment to the other, while the segment sizes are exogenously given (this assumption will be dropped in the subsequent section).

All networks are interconnected, so that any consumer can make calls to any other consumer. To allow for a call from one operator's network to the other's network, the first operator must purchase a wholesale service called "terminating access" from the second one. We assume that the marginal cost of call termination on a VoIP network is 0,[9] and that operators do not charge for call termination to a customer subscribing to a VoIP service. This is in line with the tendency of VoIP providers to use "bill-and-keep" arrangements for call termination, and with the fact that interconnection typically has already been settled at the underlying level of Internet service providers. The marginal cost of call termination on the PSTN network is $c > 0$, and the incumbent charges a termination charge a for call termination to its PSTN customers. To keep the number of parameters small without loss of generality, we set all other costs equal to zero.

Access price a is set by a regulator.[10] Since we do not explicitly model the regulator as a player, access price a is an exogenous parameter in our model.

By supposing that all consumers have an identical, inelastic demand to make calls once they have a subscription, each consumer will make a given number of calls. Without loss of generality, we normalize this number to 1.[11]

The retail price in the old segment is assumed to be given by p_0. For instance, it is set by the regulator or it may be determined by the presence of a competitive fringe in PSTN telephony (e.g. carrier-select based competitors competing on price).[12] Thus we can treat p_0 as a parameter. In the new segment, the operators compete by setting flat fees. Operator i's retail price for VoIP telephony is denoted by p_i, $i = 1, 2$. Note that implicitly, all per-minute

[9] The "true" marginal costs of electronic communications are virtually zero. Nevertheless, in practice, operators allocate fixed costs to traffic, and hence may partly treat these costs as marginal costs when setting prices. Thus, what we call the marginal cost of call termination is in fact the traffic-dependent cost of call termination. These costs are substantially lower for IP networks than for PSTN networks, and therefore we set them at 0 for VoIP calls.

[10] For instance, the regulator has determined that the incumbent has "significant market power "(SMP) in the wholesale access market, and because of that, and in line with the regulatory framework that is in place, applies price controls (this illustration corresponds to the situation in EU member states).

[11] See De Bijl and Peitz (2002) for a more elaborate specification.

[12] Even if no actual retail regulation is in place the threat of regulatory intervention may effectively be reflected by an upper bound on the PSTN retail price. When this upper bound is binding our actual analysis applies.

prices are 0.

In a more elaborate model, one could incorporate that consumers have elastic demand to make calls, or to have access to the Internet, in addition to the demand for a subscription. Such extensions lead to additional interactions between the operators, for instance because there is call traffic between the networks—see De Bijl and Peitz (2002) for an inclusion of call traffic. Nevertheless, the present model is rich enough to capture some crucial elements of the strategic interaction between PSTN and VoIP providers.[13]

When a consumer makes a call, the receiving consumer may be any other consumer with equal probability, independent of the network they are subscribed to. This implies that calling patterns are balanced, that is, the volumes of on-net and off-net calls are proportionate to market shares. This assumption, which is common in the literature on competition in telecommunications markets, simplifies the analysis and should be seen as the natural benchmark.

Market shares in the segment of the new technology depend on the retail prices, and are denoted by $s_i(p_1, p_2)$, $i = 1, 2$. We assume that an operator's market share is decreasing in its own price and increasing in the price of its rival. Furthermore, we assume that market shares only depend on the price difference $p_2 - p_1$. This assumption is satisfied for quasi-linear preferences when consumers have identical demand functions. With full participation, total market demand is fixed. For an example see below. Figure 1 illustrates the set-up of the model.

The property that market share changes continuously with price implies that firms have market power. Consumers do not consider the services provided by the two firms as perfect substitutes and therefore do not necessarily go for the lowest price. In reality, imperfect substitutes seem to be common in telecommunications (as well as other services markets), for instance due to heterogeneity in brand recognition, corporate images, and consumer switching costs. Also, services offered by operators are offered in different bundles with other services: if the bundles are not the same, they will be considered as imperfect substitutes.

[13] the present model may be seen as an approximation of a model with positive usage charges but rather inelastic demand. The demand assumptions simplify the analysis considerably and allows us to focus on participation decisions, abstracting from usage intensity. In the light of the enormous variety in non-linear contracts, the case of flat fees in a world of "simple"demand structures provides a natural benchmark. Traditionally, operators have set two-part tariffs for PSTN telephony, while at present, operators seem to be inclined to set flat fees for VoIP services (possibly combined with linear prices for calls terminating on the PSTN). Related to the emergence of IP-based telephony, there seems to be a trend towards flat fees for all voice telephony services.

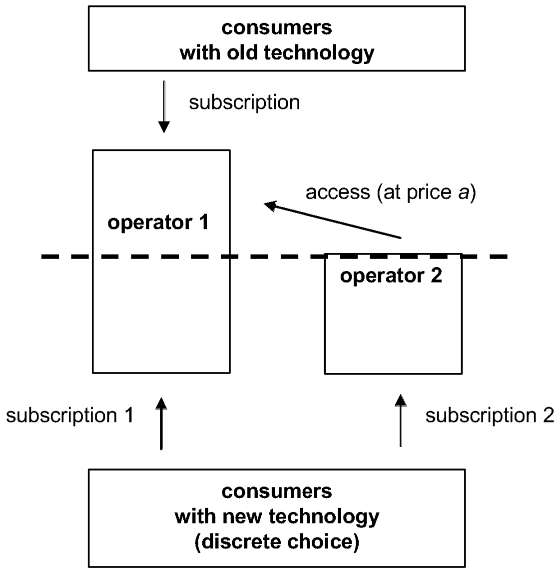


Figure 1: Illustration of the model.

Profit functions are as follows. Firm 1's profits can be written as

$$\pi_1(p_1, p_2; a, p_0) = \lambda_0[p_0 - \lambda_0 c] + \lambda[s_1(p_1, p_2)(p_1 - \lambda_0 c) + s_2(p_1, p_2)\lambda_0(a - c)], \tag{1}$$

and firm 2's profits as

$$\pi_2(p_1, p_2; a) = \lambda s_2(p_1, p_2)(p_2 - \lambda_0 a). \tag{2}$$

These profit functions reflect the volumes of on-net and off-net traffic between operator 1's PSTN network and both operators' VoIP networks—volumes that are proportionate to market shares—as well as the wholesale payments for calls terminating on the PSTN network.

A special case of our general model is obtained by assuming that the networks are horizontally differentiated. Suppose, for instance, that consumers are uniformly distributed on the interval $[0, 1]$. Firm 1 is located at location $y_1 = 0$ on the interval, and firm 2 at $y_2 = 1$. A consumer located at z buying from firm i incurs a disutility $-\theta|y_i - z|$. Note that a higher value of parameter θ corresponds to more differentiation between the networks. A consumer at z buys from firm 1 if $v_1(p_1, p_2) - \theta z > v_2(p_1, p_2) - \theta(1-z)$, where $v_i(p_1, p_2)$ denotes the conditional indirect utility of a network at the ideal location z. Market shares then satisfy $s_i(p_1, p_2) = \frac{1}{2} + (v_i(p_1, p_2) - v_j(p_1, p_2))/(2\theta)$, where $j \neq i$. This is a simple Hotelling specification which has also been widely used in models on two-way access (see e.g. Laffont, Rey and Tirole, 1998; Armstrong, 1998; and the survey by Armstrong, 2002).

Structure of the game and equilibrium: The structure of the model is as follows:
$t = 0$: The regulator sets access price a and retail price p_0, or alternatively, the latter price is determined by a competitive fringe in the retail price for PSTN voice telephony.
$t = 1$: Operators choose their prices for VoIP voice services in order to maximize profits.
$t = 2$: Consumers observe retail prices and make purchasing decisions, based on utility maximization. Consequently, market shares and profit levels are realized.

We are interested in a Nash equilibrium in prices (p_1^*, p_2^*), which is defined in such a way that given its rival price, neither firm has an incentive to change its own price. That is, each operator's price p_i^* maximizes profits $\pi_i(p_1, p_2; a, p_0)$ when p_j^* is given, $i \neq j$. Accordingly, given the equilibrium price of the competitor, the profit maximization problem of operator 1 can be written as

$$\max_{p_1} \pi_1(p_1, p_2^*; a, p_0), \tag{3}$$

while operator 2 maximizes

$$\max_{p_2} \pi_2(p_1^*, p_2; a, p_0). \tag{4}$$

3.2 Analysis

Suppose that there exists a unique pair (p_1^*, p_2^*) which solves problems (3) and (4) simultaneously (hence it constitutes an equilibrium). We are then interested in which way a change in regulatory policy with regard to the access price a affects market outcomes. Hence, consider the following increase of the termination charge: $a' = a + \Delta a$, where $\Delta a > 0$. We can then show that this increase is passed through to consumers. Market shares in equilibrium, as well as the entrant's profits, are unaffected. However, we will see that the incumbent benefits in two ways: (i) it can charge a higher mark-up in the retail market, and (ii) it receives higher revenues from calls that terminate on its PSTN network. Consumers in the new segment are worse off, as the they face higher retail prices by both networks. We will now explore the underlying mechanism in detail.

Given the new access price a', we claim that equilibrium retail prices are $p_1^{**} = p_1^* + \lambda_0 \Delta a$ and $p_2^{**} = p_2^* + \lambda_0 \Delta a$. Our proof consists of establishing that for each operator i, p_i^{**} is the solution of the maximization problem of operator i.

Operator 1: Given the new access price a', the incumbent's profit can be written as

$$\pi_1(p_1, p_2; a', p_0) = \lambda_0(p_0 - \lambda_0 c) + \lambda[s_1(p_1, p_2^{**})(p_1 - \lambda_0 c) + (1 - s_1(p_1, p_2^{**}))\lambda_0(a + \Delta a - c).$$

Provided that the competing operator sets $p_2^{**} = p_2^* + \lambda_0 \Delta a$, the incumbent's market share satisfies $s_1(p_1, p_2^{**}) = s_1(p_1 - \lambda_0 \Delta a, p_2^*)$ because they only depend on price differences. Hence operator 1's profit can be rewritten as

$$\lambda_0(p_0 - \lambda_0 c) + \lambda[s_1(p_1 - \lambda_0 \Delta a, p_2^*)(p_1 - \lambda_0 \Delta a - \lambda_0 c) + \lambda_0(1 - s_1(p_1 - \lambda_0 \Delta a, p_2^*))(a - c) + \lambda_0 \Delta a].$$

With a change of variable $\widetilde{p}_1 \equiv p_1 - \lambda_0 \Delta a$, the incumbent's maximization problem becomes

$$\max_{\widetilde{p}_1} \lambda_0(p_0 - \lambda_0 c) + \lambda[s_1(\widetilde{p}_1, p_2^*)(\widetilde{p}_1 - \lambda_0 c) + \lambda_0(1 - s_1(\widetilde{p}_1, p_2^*))(a - c)] + \lambda\lambda_0 \Delta a \tag{5}$$

Clearly, p_1^* is the solution to this problem because, apart from the constant $\lambda\lambda_0\Delta a$, it is exactly the same as problem (3). Since $\widetilde{p}_1 \equiv p_1 - \lambda_0 \Delta a$ we have shown that $p_1^{**} = p_1^* + \lambda_0 \Delta a$, provided that $p_2^{**} = p_2^* + \lambda_0 \Delta a$. Moreover, notice that the increase in the access price leads to an increase in the incumbent's profits by $\lambda\lambda_0\Delta a$.

Operator 2: Given the new access charge a', operator 2's profit can be written as

$$\pi_2(p_1, p_2; a') = \lambda s_2(p_1^{**}, p_2)(p_2 - \lambda_0(a + \Delta a))$$

Provided that the competing operator sets $p_1^{**} = p_1^* + \lambda_0 \Delta a$, the entrant's market share satisfies $s_2(p_1^{**}, p_2) = s_2(p_1^*, p_2 - \lambda_0 \Delta a)$. Hence, using the change of variable $\widetilde{p}_2 = p_2 - \lambda_0 \Delta a$ the maximization problem of the non-integrated network's profit can be written as

$$\max_{\widetilde{p}_2} s_2(p_1^*, \widetilde{p}_2)(\widetilde{p}_2 - \lambda_0 a) \tag{6}$$

Clearly, p_2^* is the solution to this problem because it is equivalent to problem (4). Since $\widetilde{p}_2 \equiv p_2 - \lambda_0 \Delta a$ we have shown that $p_2^{**} = p_2^* + \lambda_0 \Delta a$, provided that $p_1^{**} = p_1^* + \lambda_0 \Delta a$.

Hence, we have established the following result:

Result 3.1. *Consider an increase in access price for call termination on the PSTN network. As a consequence, the incumbent's profits increase, while the entrant's profits are unaffected.*

Both operators pass on the access price increase to consumers by charging a higher retail price for VoIP telephony. Market shares remain the same.

The result says that rents are redistributed from consumers using the new technology to the bottleneck owner of the old technology. It is instructive to take another look at the above result. An access price increase by Δa works affects prices in the same way as a per-user cost increase of the new technology (of magnitude $\lambda_0 \Delta a$). This can be seen as follows. The profits of operator 2 are equal to $\lambda s_2(p_1, p_2)(p_2 - \lambda_0 a + \lambda_0 \Delta a)$, that is, the profit function has the same form as with access price a and costs $\lambda_0 \Delta a$. The profit function of operator 1 becomes

$$\pi_1(p_1, p_2; a + \Delta a, p_0) = \lambda_0(p_0 - c) + \lambda[s_1(p_1, p_2^{**})p_1 + \lambda_0(1 - s_1(p_1, p_2^{**}))(a + \Delta a) - \lambda_0 c].$$

The profit-maximizing price p_1 when p_2 is given, is determined by the first-order condition of profit maximization:

$$\frac{\partial s_1(p_1, p_2)}{\partial p_1} p_1 - \lambda_0 \frac{\partial s_1(p_1, p_2)}{\partial p_1}(a + \Delta a) + s_1(p_1, p_2) = 0,$$

which is equivalent to

$$\frac{\partial s_1(p_1, p_2)}{\partial p_1}(p_1 - \lambda_0 \Delta a) - \lambda_0 \frac{\partial s_1(p_1, p_2)}{\partial p_1} a + s_1(p_1, p_2) = 0.$$

This equation is also the first-order condition of profit maximization given access price a and per-used costs $\lambda_0 \Delta a$ for the new technology. Hence, a access price increase for accessing the old technology is passed on to consumers (which increases the expected cost of providing service for firm 2 by $\lambda_0 \Delta a$) in exactly the same way as a cost increase for a the new technology by $\lambda_0 \Delta a$. The only difference between an access price increase and a cost increase is that the owner of the essential facility, that is, firm 1, benefits from an access price increase because the associated “downstream”cost increase generates revenues “upstream”at the essential facility. All consumers using the new technology suffer. In terms of consumer behavior this suggests that providing an access rule that is beneficial to the network owner of the old technology is likely to discourage consumers to move to the new technology. We will return to this issue when we analyze a model with endogenous consumer migration.

While a high access price is not desirable for consumers, firm 2's after entry profits are neutral with respect to the access price. Hence, in this simple model, entry incentives are not affected by the level of the access price. However, for high retail prices (which are due to a high access price), at some point the participation or incentive constraint becomes binding (for some consumers). When that happens, there is no full pass-through of access payments to the consumers of firm 2. Rather firm 2 will have to reduce its profit margin, making entry less attractive. See also the next section, which explicitly considers the incentive constraint of consumers to adopt the new technology.

In the present context an analysis of total surplus is straightforward. Provided that the market is symmetric, the socially desirable market share for each operator is 50%. This is an equilibrium outcome for any access price such that the participation constraint of consumers is not violated (and the technology choice by consumers is exogenous). However, if the market is not fully symmetric, strategic behavior between firms typically does not lead to an implementation of a socially optimal outcome. In particular, if one network is more attractive than the other on average, then the equilibrium market share of the less attractive network is socially excessive.[14]

4 Consumers choosing between PSTN and VoIP

4.1 The model

In the previous section, we assumed that a fraction of λ_0 consumers stay with the PSTN-based technology, while the remaining fraction of consumers λ adopt VoIP; these fractions were exogenously given ($\lambda_0 + \lambda = 1$). In this section we look at the case in which consumers can decide to switch from PSTN to VoIP. If they do so, they can choose between the VoIP offerings of the incumbent and the entrant.

As before, the marginal cost of call termination on the PSTN network is $c > 0$, while the incumbent charges a termination charge a for call termination to its PSTN customers. Consumers have identical, inelastic demand for one unit of telephony services, while calling patterns are balanced. Market shares in VoIP services are denoted by $s_i(p_1, p_2)$, $i = 1, 2$. We extend on our earlier model specification as follows.

Consumers' utility functions: Consumer tastes are described by types (y, t), uniformly distributed on $[0, 1] \times [0, 1]$. The y dimension describes preferences for operator 1 versus operator 2 (or their brands), and the t dimension reflects consumers' inclinations towards VoIP versus PSTN. A straightforward interpretation is that y captures consumers' loyalty towards operator 1, independent of the service that they purchase. With regard to the other

[14]However, in difference to Armstrong and Vickers (1998) and Lewis and Sappington (1999), the access price does not affect market shares and therefore is ineffective. See also the model in the following section.

dimension of a consumer's type, if a consumers has type t close to 0 this means that he is more inclined to adopt VoIP, whereas a consumer with t close to 1 is rather reluctant to adopt VoIP. The distance between the addresses of the products and consumer types give the disutility of consumers for the particular offerings, as will be specified below. VoIP services are "located" at points $(0,0)$ and $(1,0)$, and the PSTN service at $(0,1)$ (in fact, with a properly adjusted U_0 the latter could be any point for which the second coordinate is 1). Note that in our setting y not only plays a role when consumers choose between VoIP services, but also when consumers decide whether purchase PSTN or VoIP services.[15]

Consumers either subscribe to the PSTN service offered by the incumbent firm or to one of the two VoIP offerings. A consumer who purchases PSTN services derives utility $r + U_0 - \tau(1-t) - \theta y - p_0$, where r is the basic utility from telephony and $U_0 \in \mathbb{R}$ is interpreted as a technology-specific utility of PSTN-services relative to VoIP services (which may also include the firm-specific utility, see below). Parameters τ and θ measure the degree of heterogeneity among consumers: a large τ corresponds to a low substitutability between PSTN and VoIP, and a large θ corresponds to a large degree of differentiation between the operators.

A consumer who purchases VoIP services from firm 1 derives utility $r + U_1 - \tau t - \theta y - p_1$ where $U_1 \in \mathbb{R}$ can be interpreted as a brand or firm-specific utility that captures the asymmetry between operators. Similarly, a consumer who purchases VoIP services from firm 2 derives utility $r - \tau t - \theta(1-y) - p_2$. We will implicitly assume that all consumers make a purchase; hence, parameters are such that there is always a technology available that delivers sufficient gross utility.

Fulfilled expectations: We will be assuming that before consumers learn the prices of VoIP services, they have certain beliefs about the prices that they can expect, and based on these beliefs, they figure out whether to go for VoIP or stick with PSTN. This may correspond to a situation in which consumers—before they actively start searching for information about a recently introduced product that they are interested in—have already had some exposure to some information about that good, for instance through friends and relatives, articles in newspapers and magazines, and advertisements that they may have passively observed. Hence they are aware of the existence of the product, and, based on the various pieces of information that they have received, they form expectations about its price. Now suppose that based on her beliefs and preferences, a consumer decides that she wants to buy VoIP services. She

[15] This does not affect our results in any important way.

will then start searching more actively, in order to learn actual prices. Also, she will make comparison between the incumbent's virtues compared to those of the competitor. We may allow for consumers to return to PSTN if they find out that VoIP is too expensive compared to its benefits. However, if beliefs concerning prices are correct, this will not happen. In our model, we do not explicitly incorporate underlying processes of advertising, belief formation and search behavior, but we capture the essence by requiring that in equilibrium, beliefs must be fulfilled. Accordingly, consumers make their migration decision, that is, whether or not to adopt VoIP, *before* VoIP prices are searched.[16]

An alternative way of understanding this specification is to argue that the decision to migrate to VoIP involves a certain level of commitment, as the effort to make a first comparison between PSTN and VoIP services has been sunk (and possibly some equipment has been replaced), whereas prices can be adjusted in a more flexible way. In other words, due to search and learning costs, consumer migration to the new technology involves more commitment than setting prices.[17]

Consumers have identical beliefs about VoIP prices. Moreover, since we restrict the analysis to pure strategies, a belief function can be described by a function that, for each firm i, attaches probability 1 to one particular price level $\widehat{p}_i$, and probability 0 to all other prices $p_i \neq \widehat{p}_i$. To simplify the notation, we will not explicitly define these belief functions, but more simply summarize beliefs by $\widehat{p}_1$ and $\widehat{p}_2$.

Profit functions: The profit functions have the same structure as specified in the previous section. Because of endogenous consumer migration, they have to be adapted as follows:

$$\begin{aligned}\pi_1(p_1,p_2;a,p_0) &= \lambda_0(p_0,\widehat{p}_1,\widehat{p}_2)[p_0-\lambda_0(p_0,\widehat{p}_1,\widehat{p}_2)c]+\lambda(p_0,\widehat{p}_1,\widehat{p}_2)\\ &\quad\times[s_1(p_1,p_2)(p_1-\lambda_0(p_0,\widehat{p}_1,\widehat{p}_2)c)+s_2(p_1,p_2)\lambda_0(p_0,\widehat{p}_1,\widehat{p}_2)(a-c)],\\ \pi_2(p_1,p_2;a,p_0) &= \lambda(p_0,\widehat{p}_1,\widehat{p}_2)s_2(p_1,p_2)(p_2-\lambda_0(p_0,\widehat{p}_1,\widehat{p}_2)a),\end{aligned}$$

Structure of the game and equilibrium: The model that we analyze then has the following

[16] Another possibility would be that consumers decide after observing prices *and* their taste parameters. While we consider such a specification a valid alternative, such a model becomes very cumbersome to work with.

[17] An alternative specification would be to assume that consumers do have price information before making their migration decision (but do not yet know their taste parameter with respect to operators). This would give the entrant more possibilities to penetrate the market with its VoIP services and would introduce an additional strategic dimension into the problem. Namely, the incumbent could make the VoIP segment on average less attractive by increasing its price.

structure:

$t = 0$: The regulator sets access price a and PSTN price p_0, observed by all.[18]

$t = 1$: Each consumer learns his or her preference parameter $t \in [0, 1]$, reflecting an individual's inclination towards PSTN versus VoIP. All consumers form expectations about VoIP prices $\widehat{p}_1$ and $\widehat{p}_2$.

$t = 2$: Given their preferences and beliefs, consumers decide whether to go for PSTN or VoIP. At the same time, the operators (simultaneously) set VoIP prices p_1 and p_2.

$t = 3$: Each consumer learns his or her preference parameter $y \in [0, 1]$, reflecting an individual's inclination towards operator 1 versus operator 2. Consumers observe prices p_1 and p_2 and make purchase decisions, that is, they choose VoIP telephony from the incumbent or from the entrant if they opted for VoIP at $t = 2$. Otherwise, they choose PSTN telephony from the incumbent.

We solve for fulfilled expectation equilibrium, that is, (i) each firm maximizes its profits while taking consumers' beliefs and its rival's strategy as given; (ii) based on their beliefs $\widehat{p}_1$, $\widehat{p}_2$ consumers choose the utility maximizing technology. Subsequently, at stage 2, given the prices set by the firms, they choose the utility maximizing operator provided they adopted the new technology; and (iii) in equilibrium, consumers' beliefs are fulfilled, so that equilibrium prices p_1^* and p_2^* satisfy $p_1^* = \widehat{p}_1$ and $p_2^* = \widehat{p}_2$.[19]

Alternative representation. It is important to note that the set-up with fulfilled beliefs is not necessary to keep the model tractable. We would obtain the same results if we solved for subgame perfect equilibria in the game in which stage $t = 2$ is split into two separate stages:

$t = 2a$: Given their preferences consumers decide simultaneously whether to go for PSTN or VoIP.

$t = 2b$: The operators (simultaneously) set VoIP prices p_1 and p_2.

In this alternative formulation, one does not need to introduce consumer beliefs about VoIP prices. At stage 2, consumers maximize their utility given the decision of all other consumers. At this stage consumers' utility levels depend indirectly on the decision of the other consumers, because these subsequently determine the equilibrium prices that are charged.

[18] In an extension of the game, we will later consider the case in which the incumbent chooses the price for the PSTN service.

[19] In their seminal paper Katz and Shapiro (1984) solve for fulfilled expectations equilibria in a market with network effects.

Surplus levels. To be able to discuss the effects of regulation on consumer surplus and welfare, we provide the formulas for calculating various surplus levels in the model. The aggregate surplus of PSTN users is equal to:

$$\begin{aligned} CS^{\text{PSTN}} &= \int_{\lambda^*}^{1}\int_{0}^{1}(r+U_0-\tau(1-t)-\theta y-p_0)\mathrm{d}y\mathrm{d}t \\ &= [(r+U_0-\tau-p_0-\frac{1}{2}\theta)(1-\lambda^*)+\frac{1}{2}\tau(1-(\lambda^*)^2)]. \end{aligned}$$

The aggregate surplus of subscribers to operator 1's VoIP service is equal to:

$$\begin{aligned} CS_1^{\text{VoIP}} &= \int_{0}^{\lambda^*}\int_{0}^{s_1(p_1^*,p_2^*)}(r+U_1-\tau t-\theta y-p_1^*)\mathrm{d}y\mathrm{d}t \\ &= [(r+U_1-p_1^*)s_1(p_1^*,p_2^*)-\frac{1}{2}\theta s_1(p_1^*,p_2^*)^2]\lambda^*-\frac{1}{2}\tau s_1(p_1^*,p_2^*)(\lambda^*)^2. \end{aligned}$$

The aggregate surplus of subscribers to operator 2's VoIP service is equal to:

$$\begin{aligned} CS_2^{\text{VoIP}} &= \int_{0}^{\lambda^*}\int_{s_1(p_1^*,p_2^*)}^{1}(r-\tau t-\theta(1-y)-p_2^*)\mathrm{d}y\mathrm{d}t \\ &= [(r-\theta-p_2^*)(1-s_1(p_1^*,p_2^*))+\frac{1}{2}\theta(1-s_1(p_1^*,p_2^*)^2)]\lambda^*-\frac{1}{2}\tau(1-s_1(p_1^*,p_2^*))(\lambda^*)^2 \\ &= [(r-p_2^*)s_2(p_1^*,p_2^*)-\frac{1}{2}\theta s_2(p_1^*,p_2^*)^2]\lambda^*-\frac{1}{2}\tau s_2(p_1^*,p_2^*)(\lambda^*)^2. \end{aligned}$$

Let $CS^{\text{VoIP}} = CS_1^{\text{VoIP}} + CS_2^{\text{VoIP}}$ Aggregate consumer surplus is equal to $CS = CS^{\text{PSTN}} + CS_1^{\text{VoIP}} + CS_2^{\text{VoIP}}$. Producer surplus is equal to aggregate profits: $PS = \pi_1^* + \pi_2^*$. Welfare is then defined as the sum of consumer and producer surplus: $W = CS + PS$.

4.2 Equilibrium analysis

We start by looking at consumers' choices at the last stage, $t = 3$, for those consumers who have chosen to adopt VoIP. The consumer who is indifferent between the two VoIP services is located at location $\overline{y}$, given by $U_1 - \theta\overline{y} - p_1 = -\theta(1-\overline{y}) - p_2$. All consumers characterized by parameter $y < \overline{y}$ subscribe to operator 1's service, and all others to operator 2. Accordingly, if a fraction λ demands VoIP services, then the total demand for VoIP offered by firm 1 is

$$\lambda s_1(p_1,p_2) = \lambda\left(\frac{1}{2}+\frac{U_1}{2\theta}+\frac{p_2-p_1}{2\theta}\right).$$

Note that if $U_1 \geq \theta$, operator 2 must price below operator 1 to capture any market share. This corresponds to a situation in which there is vertical quality differentiation between the two operators and where operator 1 offers higher quality. Correspondingly, if $U_1 \leq -\theta$ operator 2 offers higher quality.

At $t = 2$, consumers expect prices $\widehat{p}_1$ and $\widehat{p}_2$. At this stage, they have learned their locations t but they do not yet know their addresses y. Hence, the expected utility of a consumer of type t who intends to migrate to VoIP is as follows:

$$\begin{aligned}
& \int_0^{s_1(\widehat{p}_1,\widehat{p}_2)} [r + U_1 - \tau t - \theta y - p_1]\mathrm{d}y + \int_{s_1(\widehat{p}_1,\widehat{p}_2)}^1 [r - \tau t - \theta(1-y) - p_2]\mathrm{d}y \\
= \; & s_1(\widehat{p}_1,\widehat{p}_2)r + s_1(\widehat{p}_1,\widehat{p}_2)U_1 - s_1(\widehat{p}_1,\widehat{p}_2)\tau t - \theta\frac{s_1(\widehat{p}_1,\widehat{p}_2)^2}{2} - s_1(\widehat{p}_1,\widehat{p}_2)\widehat{p}_1 \\
& + s_2(\widehat{p}_1,\widehat{p}_2)r - s_2(\widehat{p}_1,\widehat{p}_2)\tau t - \theta\frac{s_2(\widehat{p}_1,\widehat{p}_2)^2}{2} - s_2(\widehat{p}_1,\widehat{p}_2)\widehat{p}_2 \\
= \; & r - \tau t - s_1(\widehat{p}_1,\widehat{p}_2)(\widehat{p}_1 - U_1) - s_2(\widehat{p}_1,\widehat{p}_2)\widehat{p}_2 - \frac{\theta}{2}[s_1(\widehat{p}_1,\widehat{p}_2)^2 + s_2(\widehat{p}_1,\widehat{p}_2)^2] \\
= \; & r - \tau t - \frac{\theta}{2} \\
& - s_1(\widehat{p}_1,\widehat{p}_2)(\widehat{p}_1 - U_1) - s_2(\widehat{p}_1,\widehat{p}_2)\widehat{p}_2 - \frac{\theta}{2}[s_1(\widehat{p}_1,\widehat{p}_2)^2 + s_2(\widehat{p}_1,\widehat{p}_2)^2 - 1] \\
= \; & r - \tau t - \frac{\theta}{2} - \widetilde{p}(\widehat{p}_1,\widehat{p}_2),
\end{aligned}$$

where

$$\widetilde{p}(\widehat{p}_1,\widehat{p}_2) \equiv s_1(\widehat{p}_1,\widehat{p}_2)(\widehat{p}_1 - U_1) + s_2(\widehat{p}_1,\widehat{p}_2)\widehat{p}_2 + \frac{\theta}{2}[s_1(\widehat{p}_1,\widehat{p}_2)^2 + s_2(\widehat{p}_1,\widehat{p}_2)^2 - 1].$$

This function $\widetilde{p}(\widehat{p}_1,\widehat{p}_2)$ will be called the "adjusted average price" for VoIP services. Compared to the average price for VoIP services, it is adjusted in order to take into account the potentially asymmetric utility level U_1 as well as the expected reduction in utility from not consuming the ideal product specification. It is straightforward to show that $\widetilde{p}(\widehat{p}_1,\widehat{p}_2)$ can be simplified into

$$\widetilde{p}(\widehat{p}_1,\widehat{p}_2) = s_1(\widehat{p}_1,\widehat{p}_2)(\widehat{p}_1 - U_1) + s_2(\widehat{p}_1,\widehat{p}_2)\widehat{p}_2 - \theta s_1(\widehat{p}_1,\widehat{p}_2)s_2(\widehat{p}_1,\widehat{p}_2).$$

The expected utility derived from staying with the PSTN service is $r + U_0 - \tau(1-t) - \frac{\theta}{2} - p_0$. Accordingly, at $t = 2$, the location $\overline{t}$ of the consumer who, given his beliefs about VoIP prices, is indifferent between PSTN and VoIP services, is implicitly defined by

$$r + U_0 - \tau(1-\overline{t}) - \frac{\theta}{2} - p_0 = r - \tau\overline{t} - \frac{\theta}{2} - \widetilde{p}(\widehat{p}_1,\widehat{p}_2).$$

Therefore, the fraction of consumers opting for VoIP services, that is, all consumers located at $t < \overline{t}$, is given by

$$\lambda(p_0,\widehat{p}_1,\widehat{p}_2) = \frac{1}{2} + \frac{p_0 - U_0 - \widetilde{p}(\widehat{p}_1,\widehat{p}_2)}{2\tau}.$$

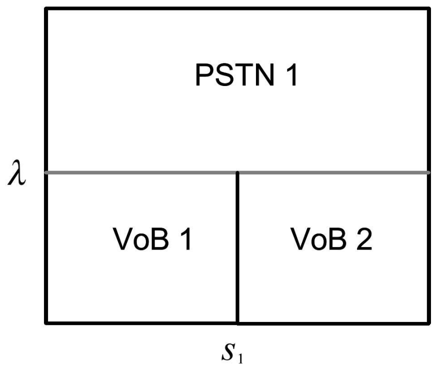


Figure 2: Segments and market shares for PSTN and VoIP telephony.

The fraction of consumers staying with the PSTN network is then, by definition, equal to $\lambda_0(p_0, \widehat{p}_1, \widehat{p}_2) = 1 - \lambda(p_0, \widehat{p}_1, \widehat{p}_2)$.

Figure 2 illustrates the endogenous segmentation of the market, and the division of the VoIP segment among the incumbent and the entrant.

At $t = 2$, for given consumer beliefs, $\lambda(p_0, \widehat{p}_1, \widehat{p}_2)$ and $\lambda_0(p_0, \widehat{p}_1, \widehat{p}_2)$ are fixed. This simplifies the profit functions. Note that at this stage, again because expectations are given, also function $\widetilde{p}(\widehat{p}_1, \widehat{p}_2)$ can be treated as a constant. Thus we write $\widetilde{p} = \widetilde{p}(\widehat{p}_1, \widehat{p}_2)$, and for given consumer choices with regard to PSTN versus VoIP, it can be shown that the Nash equilibrium at $t = 2$ is characterized by the following prices:

$$p_1(p_0, \widetilde{p}) = \frac{3\theta + U_1}{3} + \frac{a}{2} + \frac{\widetilde{p} - p_0 + U_0}{2\tau} a, \quad (7)$$

$$p_2(p_0, \widetilde{p}) = \frac{3\theta - U_1}{3} + \frac{a}{2} + \frac{\widetilde{p} - p_0 + U_0}{2\tau} a. \quad (8)$$

Some interim observations can be made from equations (7)-(8) under the assumption that $\widetilde{p}$ is fixed. Clearly, if the VoIP services are closer substitutes (θ smaller), then lower prices result. Brand loyalty or superior performance of firm 1's VoIP services ($U_1 > 0$) translate into a higher price p_1. Finally, provided that the last term in the pricing equations is sufficiently small, a higher access price translates into higher prices. Furthermore, firm 2's price-cost margin is not affected by the access price since $p_2 = \frac{3\theta - U_1}{3} + \lambda_0(p_0, \widehat{p}_1, \widehat{p}_2)a$. Hence, for given expectations the neutrality result, which was derived in the previous section, still holds. For given expectations, the present model is a special case of the model analyzed in the previous section.

Still given the assumption that $\widetilde{p}$ is fixed, we also observe that a higher price in the PSTN segment translates into *lower* prices for VoIP services. This is due to the cost effect that a higher p_0 will lead to less demand for PSTN services, which reduces the likelihood that subscribers to operator 2's VoIP service make use of terminating access to the PSTN network. This corresponds to lower perceived costs for operator 2, and hence, a more competitive outcome. The reverse holds for the adjusted average VoIP price $\widetilde{p}$, and for the fixed-utility advantage of PSTN compared to VoIP services, U_0.

The prices in (7)-(8) do not yet characterize an equilibrium outcome. To be an equilibrium, beliefs must be confirmed in equilibrium, that is, the above solution must satisfy $p_1(p_0, \widetilde{p}) = \widehat{p}_1$ and $p_2(p_0, \widetilde{p}) = \widehat{p}_2$. If we define

$$\begin{aligned} g(\widetilde{p}) \equiv\ & s_1(p_1(p_0,\widetilde{p}), p_2(p_0,\widetilde{p}))(p_1(p_0,\widetilde{p}) - U_1) + s_2(p_1(p_0,\widetilde{p}), p_2(p_0,\widetilde{p}))p_2(p_0,\widetilde{p}) \\ & -\theta s_1(p_1(p_0,\widetilde{p}), p_2(p_0,\widetilde{p}))s_2(p_1(p_0,\widetilde{p}), p_2(p_0,\widetilde{p})), \end{aligned}$$

then the equilibrium value $\widetilde{p}^*$ is defined as a fixed point of $g(\cdot)$. It is straightforward to verify that $g(\cdot)$ is linear in $\widetilde{p}$, so that there exists a unique fixed point:

$$\widetilde{p}^* = \frac{18a\theta(\tau - p_0 + U_0) + (27\theta^2 - 18U_1\theta - U_1^2)\tau}{18\theta(2\tau - a)}.$$

The interpretation of this solution is that the quality-adjusted average price for VoIP services is decreasing in the utility of firm 1's VoIP services U_1 (for θ not too small). For $a > 0$, it is increasing in the utility of PSTN services U_0, and decreasing in the price of the competitive segment p_0. The latter two properties can be explained by the fact that an decrease in $U_0 - p_0$ makes migration to VoIP more attractive, everything else equal, as explained above. Therefore, in an equilibrium outcome, pricing in the VoIP segment becomes more competitive (and therefore, $\widetilde{p}^*$ falls).

We restrict our analysis to moderate levels of the terminating access price. This is a reasonable restriction in the light of the fact that the underlying marginal costs of call termination are very small or even negligible in reality.

Assumption: $a < 2\tau$.

Substituting the constant $\widetilde{p}^*$ into (7)-(8), we obtain the equilibrium size of the PSTN segment:

$$\lambda_0^* = \lambda_0(p_0, p_1^*, p_2^*) = \frac{9\theta[4(\tau - p_0 + U_0) + 3\theta] - U_1^2 - 18U_1\theta)}{36\theta(2\tau - a)}. \tag{9}$$

Next, we obtain equilibrium prices for VoIP services:

$$\begin{aligned} p_1^* &= \theta + \frac{U_1}{3} + \lambda_0^* a \\ &= \frac{a\{9\theta[4(\tau - p_0 + U_0) - \theta] - 30\theta U_1 - U_1^2\} + 24\theta(3\theta + U_1)\tau}{36\theta(2\tau - a)}, \end{aligned} \tag{10}$$

$$\begin{aligned} p_2^* &= \theta - \frac{U_1}{3} + \lambda_0^* a \\ &= \frac{a\{9\theta[4(\tau - p_0 + U_0) - \theta] - 6\theta U_1 - U_1^2\} + 24\theta(3\theta - U_1)\tau}{36\theta(2\tau - a)}. \end{aligned} \tag{11}$$

Note that if an equilibrium exists, it is unique (given by (10)-(11)). One can check that in order for p_1^* and p_2^* to be profit-maximizing prices, the following second-order condition for profit maximization has to be satisfied:

$$-\frac{U_1^2 + 18U_1\theta + 9\theta(4\tau - 4a + 4p_0 - 3\theta - 4U_0)}{36\theta^2(2\tau - a)} < 0.$$

This condition is equivalent to $\lambda^* > 0$, and will therefore always be satisfied in an interior equilibrium outcome.[20] Using that $a < 2\tau$, it can be rewritten as $a < \frac{1}{36\theta}U_1^2 + \frac{1}{2}U_1 + p_0 - U_0 - \frac{3}{4}\theta + \tau$.

The insight from the previous section with respect to prices in the VoIP segment is still valid. In the Hotelling specification we obtain the markup due to product differentiation θ corrected by a term that reflects the asymmetry that is introduced due to U_1 plus the marginal cost for operator 2 due to termination on the PSTN segment.

Operator 2's profits in an equilibrium are equal to:

$$\pi_2^* = \frac{(3\theta - U_1)^2[9\theta(4\tau + 4p_0 - 4U_0 - 4a - 3\theta) + U_1^2 + 18U_1\theta]}{648\theta^2(2\tau - a)}$$

For $U_1 = 0$ the expression reduces to

$$\pi_2^* = \frac{\theta(4\tau + 4p_0 - 4U_0 - 4a - 3\theta)}{8\theta(2\tau - a)}$$

The equilibrium expression for firm 1's profit is somewhat more involved.

[20] In what follows, we implicitly assume that the solution to the system of first-order conditions for profit maximization characterizes the equilibrium outcome that we discuss.

Remark: Since we are interested in the migration from PSTN to VoIP we analyze only interior equilibria. Here we comment on the possibility that full migration to VoIP is an equilibrium. We will see that for a range of parameter constellations there exist multiple equilibria. For a given price p_0, a consumer of type $t = 1$ has expected utility from PSTN-telephony equal to $r + U_0 - \theta/2 - p_0$. In equilibrium his expected utility for VoIP telephony would be $r - \tau - ((s_1^*)^2 + (s_2^*)^2)\theta/2 + s_1^*(U_1 - p_1^*) - s_2^* p_2^*$. Note that in an interior equilibrium a consumer of type $t = 1$ must strictly prefer PSTN. However, in such an equilibrium VoIP prices are higher than in a situation in which all consumers have migrated. Therefore, denoting equilibrium values for $a = 0$ with superscript 0 the condition for the existence of an equilibrium in which all consumers have migrated to VoIP is $r + U_0 - \theta/2 - p_0 < r - \tau - ((s_1^{*0})^2 + (s_2^{*0})^2)\theta/2 + s_1^{*0}(U_1 - p_1^{*0}) - s_2^{*0} p_2^{*0}$.

4.3 Comparative statics

4.3.1 The PSTN terminating access price

The focus of this paper is to understand the impact of regulatory decisions in the PSTN segment on market outcomes in the VoIP segment. For this we derive comparative statics results in the regulated access price a. Note that for a given number of VoIP customers a higher access charge implies that the entrant faces higher perceived marginal costs and the incumbent a higher opportunity cost to attract customers in the VoIP segment. This shifts the reaction curve of both operators outward. Since products are strategic complements retail prices are inflated.[21] In the previous section, we showed that the neutrality of firm 2's profits resulted from the property that both firms' equilibrium prices increase by the increase in opportunity costs due to a higher a. This cost increase (compared to the case $a = 0$) is equal to $\lambda_0 a$, which is the expected access payment incurred by the entrant. Since $p_1 - p_2$ was not affected if a increased, market shares s_i remained the same. Note that, in the present model with endogenous segment size, higher perceived costs are also passed on to consumers. In particular, $p_2^* - \lambda_0 a = \theta - U_1/3$, which, again, is independent of a and market shares s_i^* are independent of a. However, firm 2's profits are not neutral to the access price. The reason is that consumers, anticipating higher VoIP prices, become more reluctant to migrate to VoIP,

[21] Note that in standard models of price competition with differentiated products firms offer strategic complements. This gives rise to monotone comparative statics properties (see e.g. Vives, 1990, and Milgrom and Roberts, 1990). For a recent overview of the literature on strategic complementarites see e.g. Vives (2005). For a first application of the theory to telecommunications markets see Peitz (2005). In our simple price competition model, VoB prices are indeed strategic complements.

that is $\partial\lambda_0^*/\partial a > 0$.[22] Formally, taking derivatives of the expressions reported in equations (10) and (11), we obtain for $i = 1, 2$,

$$\begin{aligned}\frac{\partial p_i^*}{\partial a} &= \lambda_0^* + \frac{\partial \lambda_0^*}{\partial a} a \\ &= \lambda_0^* + \frac{a\lambda_0^*}{2\tau - a} \\ &> 0.\end{aligned}$$

Note also that prices respond more strongly to changes in the access price if the access price is already high. Formally, p_i^* is convex in a:

$$\frac{\partial^2 p_i^*}{\partial a^2} = \frac{\lambda_0^*}{2\tau - a} + \frac{(\lambda_0^* + \frac{a\lambda_0^*}{2\tau - a})(2\tau - a) + a\lambda_0^*}{(2\tau - a)^2} > 0.$$

Consider now the change of firm 2's equilibrium profit in response the change in the access price. The equilibrium profit of firm 2 is decreasing in the access price because

$$\frac{\partial \pi_2^*}{\partial a} = \frac{\partial \lambda^*}{\partial a} \left[s_2(p_1^*, p_2^*)(p_2^* - \lambda_0^* a) \right] < 0$$

Our main comparative statics result can be summarized as follows.

Result 4.1. *For a given PSTN price, a higher access price for call termination on the PSTN network leads to*

(i) a smaller customer base for VoIP telephony.

(ii) higher prices for VoIP telephony, and

(iii) lower profits for operator 2.

This result not only holds for our particular specification but more generally. Consider profit functions as in equations (1) and (2) that depend on p_i, λ and α. The required properties are that (P1) for given a and λ, prices are strategic complements; (P2) for given λ, a higher access prices increases marginal profits; and (P3) higher retail prices lead to a lower penetration of VoIP. Properties (P1) and (P2) have been shown in the model with exogenous shares λ_0 and λ. Thus, for given λ_0 and λ an increase in a leads to higher prices. Since this is anticipated, given these higher prices more consumers stay with PSTN. This feeds back into higher expected costs for operator 2 and marks a new round in which firms increase their price. Hence, in equilibrium (provided that it exists), prices increase and the market penetration of VoIP decreases.

[22] From equation (9) it can be directly seen that indeed $\partial\lambda_0/\partial a > 0$ whenever $\lambda_0 > 0$.

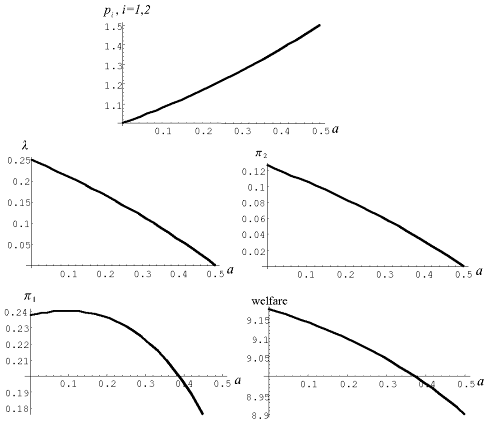


Figure 3: Illustration of the equilibrium outcome when the PSTN retail price is exogenously given.

We will illustrate the equilibrium properties with some diagrams based on a numerical example.[23] Suppose that $p_0 = 0.25$, $\theta = \tau = 1$, $r = 10$, $U_0 = U_1 = 0$, and $c = 0.1$. The condition for an interior solution then requires that $a < 0.5$. We will therefore look at the implications for $a \in [0, 0.5]$.[24] Figure 3 contains various illustrations of the equilibrium properties.

As illustrated in figure 3, firm 1's profit may be partly increasing and partly decreasing in the access price a. This is the case for p_0 sufficiently large. For small values of access price a, a PSTN consumer is then in expectations more valuable for firm 1 than a VoIP consumer. Thus, an increase in a which shifts consumers from the VoIP to the PSTN segment is profit increasing. This explains why firm 1's profits are initially increasing in a. This no longer holds for larger a. The reason is that for larger a, competition in the VoIP segment is more relaxed so that, for retail prices in the VoIP segment above a certain level, a consumer in the

[23] Obtained by using *Mathematica* software.

[24] We have checked the robustness of the effects on surplus levels by varying parameter levels. This confirmed that qualitatively, the observations discussed above do not seem to depend on the parameter levels (within reasonable bounds).

VoIP segment is in expectations more valuable than a consumer in the PSTN segment. Firm 1 may therefore obtain a larger profit with a lower access price since this implies a larger VoIP segment. This result suggests that it is not necessarily in the interest of firm 1 to lobby for a high access charge. In particular, if p_0 is sufficiently small, then firm 1's profit is globally decreasing in a (note: this is not illustrated in the figure). The reason is the following: With a higher access price a consumers expect the VoIP segment to be less competitive. Therefore only few consumers decide to migrate to the VoIP segment. Since the PSTN-segment is not very profitable, firm 1 would be better off if many consumers would migrate. To the extent that firm 1 can influence U_0 it has no incentive to improve the quality of PSTN services. Rather the opposite is true, since it would like to convince consumers to move to the VoIP segment.

More generally, one can observe that the larger p_0, the larger the profit-maximizing access price. Again the argument is that relaxed competition in the VoIP segment (and thus a smaller market share of VoIP) is in the interest of firm 1 if retail price regulation in the PSTN is less strict (to the effect that PSTN customers are more valuable).

Figure 3 shows that total welfare is decreasing in a, which is somewhat surprising as the policy implication is that a should be lower than marginal cost. However, note that we assumed that the demand for calls is perfectly inelastic. This implies that retail prices above or below perceived marginal costs do not affect participation. This, in turn, implies that all welfare results are completely driven by the division of the market among PSTN and VoIP, and in the VoIP segment, the division between the two operators. For the specific parameter values that we chose, it turns out that the welfare-maximizing size of the VoIP segment is as large as possible (under the restriction that $a \geq 0$). Note that in general VoIP prices are increasing in a. Therefore all consumers are necessarily (weakly) worse off after an increase of the access price.

4.3.2 The PSTN retail price

Recall our assumption that the retail price for PSTN services is regulated. While this is no longer an appropriate description in those countries in which retail regulation has been phased out, it still can be used as a useful benchmark since various forms of wholesale regulation affect retail prices in the PSTN segment (e.g. resale competition limit the incumbent's market power in the retail market). Thus the fixed price for PSTN that we assume in our model, can be seen as a simplification of situations in which the PSTN price is less flexible than VoIP prices, for instance due to regulatory measures that lead to unbundling and resale-based competition in the PSTN segment.

It is interesting to see how our results depend on the level of the PSTN retail price. We can make a number of observations, mostly based on (10)-(11):

Result 4.2.
(i) A higher price for PSTN telephony leads to a larger customer base for VoIP telephony.
(ii) Provided that the PSTN access price is positive, a higher price for PSTN telephony leads to lower prices for VoIP telephony.
(iii) Provided that the PSTN access price is zero, a higher price for PSTN telephony does not affect prices for VoIP telephony.
(iv) A higher price for PSTN telephony increases the entrant's profits.

Let us discuss these observations in some more detail. Observation (i) is not surprising. If PSTN telephony becomes more expensive, more consumers will switch to VoIP.

Observations (ii)-(iii) can be explained as follows. The mechanism behind the effect

$$\frac{\partial p_i^*}{\partial p_0} = \frac{\partial \lambda_0^*}{\partial p_0} a < 0, \ i = 1, 2,$$

is that an increase in p_0 reduces the size of the segment of PSTN customers, which in turn reduces the probability that a customer of operator 2 makes a call to the PSTN network. Hence, because of the reduction in expected access payments to the incumbent, operator 2's perceived marginal cost is reduced. The result is a more competitive outcome in the VoIP segment, as has been explained before. Thus a higher price in the PSTN segment leads to a lower prices in the VoIP segment so that products across segments are strategic substitutes. Note, however, that if the incumbent's access price for termination on the PSTN network is zero, then the entrant's perceived marginal cost remains unaffected if the number of PSTN customers decreases.

We remark that an increase in p_0 has the same effect on prices in the VoIP segment as an increase in the fixed utility of PSTN telephony. More precisely, a larger value for U_0 increases the customer base for PSTN services, and hence inflates the entrant's perceived marginal cost. Therefore,

$$\frac{\partial p_i^*}{\partial U_0} > 0, \ i = 1, 2.$$

To understand observation (iv), note that the equilibrium profit of firm 2 is increasing in the price of the PSTN-segment,

$$\frac{d\pi_2^*}{dp_0} = -\frac{d\pi_2^*}{dU_0} = \frac{(3\theta - U_1)^2}{18\theta(2\tau - a)} > 0.$$

The reason is that such a higher price leads to more migration to the VoIP segment. This effect is reinforced because such migration leads to lower perceived costs of firm 2 and thus, with fulfilled expectations, makes the VoIP segment more attractive. Due to the expansion of the VoIP segment, the entrant benefits from such a change.

Finally, we turn to the comparative statics properties of firm 1's profits. Firm 1's profit is increasing in p_0 for p_0 small. This is hardly surprising since a high retail price in the PSTN segment directly feeds into profits. A possibly countervailing effect is that firm 1 loses market share in the retail market. However, as long as a consumers in the VoIP segment is in expectations more valuable than a consumer in the PSTN segment, migration from PSTN to VoIP is good news for firm 1. For large p_0 the effect is reversed. Thus, for a given a there is a finite profit-maximizing retail price for PSTN telephony. In principle, the regulator can refrain from regulating the retail price p_0 in the PSTN segment even if, as in our model, firm 1 maintains a monopoly position in that segment. The reason is that although firm 1 wants to milk its PSTN customers it cannot price too high in order not to loose consumers to VoIP. This confirms that VoIP offers by firm 2 give rise to some disciplining effect on the firm 1's PSTN offers. We will return to this situation below.

4.4 Access regulation when the PSTN retail price is endogenous

As an extension of the model, consider two ways in which the PSTN price p_0 may be endogenous.

Resale competition in the PSTN segment: Suppose that due to regulation of the incumbent's originating access price (equal to a_0), there is intense competition in the PSTN segment.[25] The calls that the competitive fringe of PSTN entrants without local networks generate, terminate either on the incumbent's PSTN network, or on one of the VoIP networks. Hence each of these entrants faces a perceived marginal cost of $\lambda_0 a + a_0$. Perfect competition in the PSTN retail segment then drives down the incumbent's retail price to $p_0 = \lambda_0 a + a_0$.

In the model, the determination of p_0 takes place as follows. For given access prices, equilibrium prices p_1^* and p_2^* have been determined. Then p_0 is obtained as the fixed point

[25] See also De Bijl and Peitz (2002), ch. 5.

of $p_0 = \lambda_0^*(p_0, p_1^*, p_2^*)a + a_0$. Notice that the operators, when choosing prices in the VoIP segment, do not take into account that p_0 affects the size of the VoIP segment. The reason is that at the moment they choose VoIP prices, consumers already have made the decision whether to migrate to VoIP or not.

The effect of access price a is now as follows. The access price a does not affect the size of the VoIP segment λ (nor does it affect market shares within that segment). An increase in a leads to higher retail prices in both segments and thus reduces both CS^{PSTN} and CS^{VoIP}. It increases the incumbent's profits while leaving the entrant's profits unaffected. Overall, welfare W is not affected. Accordingly, the model's outcomes are similar to the outcomes of the model in section 3, where we assumed that there was no migration between technologies (but distributional effects are typically different).

Sequential price setting: As an alternative to a situation of resale competition in the PSTN segment, consider the case in which operator 1 is be a monopolist in the PSTN segment and is able to set a profit-maximizing price p_0. In particular, we assume that operator 1 chooses p_0 after all parameters, including the access price, are set, but before the rest of the game evolves. Hence the incumbent takes into account the equilibrium prices from competition in the VoIP segment by using backward induction. This sequential timing of moves can be motivated by the fact that the incumbent is less flexible in setting PSTN prices than in setting VoIP prices. For instance, because of universal service obligations, the incumbent may have to notify the regulator, or will need regulatory approval, for changes in PSTN prices.

Given the PSTN access price and the equilibrium outcome in the VoIP segment, operator 1 chooses p_0 to maximize profits $\pi(p_1^*, p_2^*; a, p_0)$. One can show that this profit function is concave in p_0. More precisely, it is quadratic, so that finding the profit-maximizing PSTN price is relatively straightforward. It can be shown that

$$\begin{aligned} p_0^* &= \frac{1}{144\tau\theta}[36a^2\theta - 3a(U_1^2 + 10U_1\theta - 3\theta(-4c + 4U_0 + \theta - 4\tau)) \\ &\quad +2\tau(U_1^2 - 6U_1\theta + 9\theta(4c + 4U_0 + 5\theta + 4\tau))]. \end{aligned}$$

The following result, which we state without proof, confirms our Result 4.1 for the case that the PSTN retail price is not fixed:

Result 4.3. *For endogenous p_0, a higher access price for call termination on the PSTN network leads to*

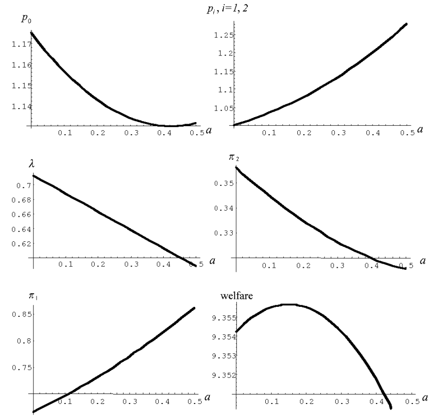


Figure 4: Illustration of the equilibrium outcome when the PSTN retail price is endogenously determined.

(i) a smaller customer base for VoIP telephony.
(ii) higher prices for VoIP telephony, and
(iii) lower profits for operator 2.

We illustrate the equilibrium properties with some diagrams based on numerical calculation, when $\theta = \tau = 1$, $r = 10$, $U_0 = U_1 = 0$, and $c = 0.1$.[26] See figure 4. The PSTN price p_0^* may be decreasing or increasing in access price a. For the chosen parameter values, it is convex in a: first decreasing, but increasing for relatively high values of a.

The observation that $p_0^*(a)$ is U-shaped can be explained as follows. Note that the incumbent makes profits from selling wholesale access in the PSTN segment, and from selling retail services both in the PSTN and the VoIP segment. First, recall that earlier we saw that an increase in the access price (for given p_0) led to higher prices in the VoIP segment. This implies that for a given p_0, a higher access price leads to a smaller VoIP segment. However, the incumbent is free to adjust its price. Note that for low levels of the access price, an increase feeds only weakly into higher VoIP prices. Since PSTN customers are very valuable at low levels of a, it is profit maximizing for the incumbent to reduce p_0. At higher levels of

[26] Again obtained by using *Mathematica* software.

a an increase feeds strongly into higher VoIP prices, as VoIP consumers are, in expectation, rather valuable for the incumbent. He therefore reduces the reduction of the size of the VoIP segment (the reduction that would occur with a constant p_0) by increasing its PSTN price.

Note that in general, a higher terminating access price a leads to larger PSTN segment. For relatively low levels of the access price, the incumbent decreases its PSTN price to benefit from an even larger PSTN segment. This implies that PSTN consumers benefit from a moderately high access price. VoIP customers suffer and there is less migration to VoIP. To the extent that our model approximates current telecommunications markets in which the incumbent enjoys market power in the PSTN segment, our findings run counter to the view that a higher access charge would lead to a decline of the PSTN segment.[27]

Figure 4 shows that total welfare is concave in a. To understand this, recall that all welfare results are driven by the division of the market among PSTN and VoIP, and in the VoIP segment, the division between the two operators.[28] The regulator could in theory use the access price to implement the optimal split between the PSTN and VoB segment, which explains why the optimal access price is, in general, not equal to marginal cost. The optimal split depends on the parameters of the model. With the parameters of our numerical example the welfare-maximizing allocation does not have the property that $\lambda = 1/2$ for two reasons. First, the PSTN segment is more costly to operate, and second, since there is more variety in the VoIP segment, VoIP offerings tend to better fit consumers' tastes. In our example, the welfare-maximizing allocation can be shown to be $\lambda = 0.675$.

5 Conclusion

In this paper, we explored competition between an incumbent offering both PSTN and VoIP telephony, and an entrant active only in the VoIP segment. Our analysis has shed light on the effects of regulation in one segment on competition in another, unregulated segment, and has focused on cost effects of access price regulation. Given the publication of reports such as OECD (2006), this type of analysis is urgently called for.

We looked at two different settings. In the first setting, which can be seen as a benchmark case, we assumed that the size of the customer segment interested in VoIP was exogenously given (hence consumer migration between the segments was not possible). In this simple

[27] For instance, a recent OECD report (OECD, 2006, p. 28) contains a statement that can be interpreted as such.

[28] Welfare in a model with full participation only depends on price differences and not on price levels.

model with inelastic demand an increase of the PSTN terminating access price increases the incumbent's profits, but as it does not affect the entrant's profits, market entry is independent of access regulation.

In the second setting, we endogenized consumers' choices for PSTN versus VoIP services. We used a set-up with the interpretation of fulfilled expectations about retail prices in the VoIP segment (this interpretation was not necessary, though). For instance, before consumers learn the prices of VoIP services, they already have certain beliefs about expected prices, which help them to figure out whether to go migrate to VoIP. Accordingly, consumers are already aware of the existence of a new technology, and make their migration decision before prices are actively searched. Another interpretation is that the decision to migrate to a new technology requires some commitment, as the effort to do so is sunk before the process of actively searching prices starts. Focusing on fulfilled-beliefs equilibria allowed us to keep the analysis tractable, while at the same time incorporating consumers' migration decisions in a realistic way.

Below we shortly discuss our main results and make some comments on the application of CPP versus RPP. We then conclude by addressing a couple of limitations of our analysis which suggest avenues for further research.

Regulation of the PSTN retail price: The PSTN retail price only affects competition in the VoIP segment if the consumers' technology adoption decision is endogenous. An important result of the analysis is that, as long as the PSTN access price is positive, a higher price for PSTN telephony leads to lower prices for VoIP telephony. Only for an access price equal to zero, the retail price level of PSTN telephony does not affect retail prices for VoIP telephony. These results illustrate the links between different telephony networks—links that should not be ignored by regulators. These links have been explored in more detail. Suppose that an entrant in the VoIP market faces a positive access price. This access price may or may not reflect marginal cost levels; it is only important that this access price is positive. Then a lower regulated PSTN price leads to a smaller customer base for VoIP telephony and softens price competition among VoIP operators.

Note that if a regulator allows an integrated incumbent to include a mark-up for common costs in its access charge, which is typically the case, then the access price will be above the marginal cost level. The result is less migration to VoIP. In addition, if a universal service obligation forces the incumbent to price PSTN telephony at a low level, VoIP retail prices become inflated and the adoption of VoIP will be slowed down even more.

Access regulation: Access regulation on the PSTN network affects the VoIP market. For instance, if the incumbent charges for call termination on the PSTN and VoIP networks use bill-and-keep, then a higher access price for call termination on the PSTN network leads to a smaller customer base for VoIP telephony. In the context of access price regulation it is important that regulators take into account these linkages between different market segments, and that regulation within one segment may have spillover effects to other segments.

In markets in which the PSTN retail price is *not* regulated (and in which an incumbent enjoys market power), a higher access price leads to higher VoIP retail prices (as in the regulated case) but tends to lead to lower retail prices in the PSTN segment. This suggests that regulation has winners and losers: consumers of the "old" technology are the winners from a high access price (which can be seen as a protective measure for the old technology) and consumer of the "new" technology are the losers.

RPP versus CPP: Our analysis has been carried out under the "calling party pays" (CPP) principle. If, however, the "receiving party pays" (RPP) principle is applied, then the user of the PSTN network has to pay for terminating access of the calls that he receives. This implies that the perceived cost of the VoIP operator no longer contains any payment for terminating access. In that case, if the PSTN operator charges a higher price to his PSTN customers for receiving calls, he makes PSTN telephony less attractive and will therefore reduce his PSTN customer base. Thus the comparative static results with respect to market share are reversed. Furthermore, under fulfilled expectations the price level of the VoIP segment is independent of the relative success of VoIP.

This suggests that with the appearance of VoIP the rationale for applying the case for applying RPP is strengthened. More generally, with the coexistence of different technologies, technology-specific costs have to be attributed to users, ignoring the issue of markups. Under RPP, costs of PSTN telephony are attributed to PSTN users whereas under CPP, they are partly borne by VoIP users. However, to the extent that some of the costs arise due to social obligations, in particular, universal service obligations, the application of RPP may put an excessive burden on PSTN users. This burden may become unbearable if the PSTN segment shrinks drastically in size.

Our modelling strategy has been to isolate the cost effects of access price regulation abstracting from two important aspects. First, integration in our model with a given PSTN price is neutral to competition. Thus cannibalization is not an issue. This result is due to the particular timing, because an integrated firm with a regulated PSTN price cannot commit to a high VoIP price (which would avoid cannibalization). However, the firm can possibly commit not to offer VoIP services at all. If we introduce this possibility in our model, there is a range of parameters where this is indeed the profit maximizing solution for the incumbent. By not offering VoIP services it relaxes competition in the VoIP segment thus making consumers reluctant to migrate to VoIP. Also note that if we endogenize the PSTN price, our model no longer has the property that integration is neutral to competition. An integrated firm takes profits from retail in the VoIP segment into account and adjusts the PSTN price accordingly.

Second, in our model predation is not an issue. Predation tends to make the incumbent more aggressive in the VoIP market as an attempt to maintain its customer base. Such predatory behavior arises in dynamic models, in particular with consumer switching costs. We leave it for future research to analyze predatory behavior in the context of VoIP.

References

[1] Analysys (2004), "IP Voice and Associated Convergent Services", Final Report for the European Commission, 28 January.

[2] Armstrong, M. (1998), "Network Interconnection in Telecommunications", *Economic Journal* 108, 545-564.

[3] Armstrong, M. (2002), "The Theory of Access Pricing and Interconnection", in: M. Cave, S. Majumdar , and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, Amsterdam: North Holland.

[4] Armstrong, M., Doyle, C. and J. Vickers (1996), "The Access Pricing Problem: A Synthesis", *Journal of Industrial Economics* 44, 131–150.

[5] Armstrong, M. and J. Vickers (1998), "The Access Pricing Problem with Deregulation: A Note", *Journal of Industrial Economics* 46, 115–121.

[6] Baumol, W. (1983), "Some Subtle Issues in Railroad Regulation", *International Journal of Transport Economics* 10, 341-355.

[7] Canoy, M. and M. Peitz (1997), "The Differentiation Triangle", *Journal of Industrial Economics* 45, 305-324.

[8] De Bijl, P.W.J. and M. Peitz (2002), *Regulation and Entry into Telecommunications Markets*, Cambridge University Press, Cambridge, UK.

[9] De Bijl, P.W.J. and M. Peitz (2005), "Local Loop Unbundling in Europe: Experience, Prospects and Policy Challenges", *Communications and Strategies* 57, 33-57.

[10] De Bijl, P.W.J. and M. Peitz (forthcoming), "Local Loop Unbundling: One-Way Access and Imperfect Competition", in: R. Dewenter and J. Haucap (eds.), *Access Pricing: Theory and Practice*, Elsevier Science.

[11] Ebrill, L. and S. Slutsky (1990), "Production Efficiency and Optimal Pricing in Intermediate-Good Regulated Industries", *International Journal of Industrial Organization* 8, 417-442.

[12] Foros, O. (2004), "Strategic Investments with Spillovers, Vertical Integration and Foreclosure in the Broadband Access Market", *International Journal of Industrial Organization* 22, 1-24.

[13] Katz, M. (1984), "Firm Specific Differentiation and Competition among Multiproduct Firms", *Journal of Business* 56, S149-166.

[14] Katz, M. and C. Shapiro (1985), "Network Externalities, Competition and Compatibility", *American Economic Review* 75, 424–440.

[15] Lal, R. and C. Matutes (1989), "Price Competition in Multiproduct Duopolies", *RAND Journal of Economics* 20, 516-537.

[16] Laffont, J.-J., P. Rey, and J. Tirole (1998), "Network Competition: I. Overview and Nondiscriminatory Pricing", *RAND Journal of Economics* 29, 1-37.

[17] Laffont, J.-J., and J. Tirole (1994), "Access Pricing and Competition", *European Economic Review* 38, 1673-1710.

[18] Laffont, J.-J., and J. Tirole (1996), "Creating Competition Through Interconnection: Theory and Practice", *Journal of Regulatory Economics* 10, 227-256.

[19] Laffont, J.-J. and J. Tirole (2000), *Competition in Telecommunications*, MIT Press, Cambridge, MA.

[20] Lewis, T.R., and D.E.M. Sappington (1999), "Access pricing with unregulated downstream competition", *Information Economics and Policy* 11, 73-100.

[21] Manez, J. A. and M. Waterson (2001), "Multiproduct Firms and Product Differentiation: A Survey", University of Warwick Working paper no 594.

[22] Milgrom, P. and J. Roberts (1990), "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities", *Econometrica* 58, 1255-1277.

[23] OECD (2006), "The Policy Implications of Voice over Internet Protocol", report to the Working Party on Telecommunication and Information Services Policies, DSTI/ICCP/TISP(2005)3/FINAL.

[24] Ofcom (2006), "Regulation of VoIP Services", Statement and further consultation, 22 February.

[25] Peitz, M. (2005), "Asymmetric Access Price Regulation in Telecommunications Markets", *European Economic Review* 49, 341-358 .

[26] Vives, X. (1990), "Nash Equilibrium with Strategic Complementarities", *Journal of Mathematical Economics* 19, 305-321.

[27] Vives, X. (2005), "Complementarities and Games: New Developments", *Journal of Economic Literature* 43, 437-479.

[28] Vogelsang, I. (2003), "Price Regulation of Access to Telecommunications Networks", *Journal of Economic Literature* 41, 830–862.

[29] Willig, R. (1979), "The Theory of Network Access Pricing", in: Trebing, H. (ed.), *Issues in Public Utility Regulation*, Michigan State University, East Lansing, MI.

The Competitiveness of Upstream Markets with Vertically Integrated Firms*

Marc Bourreau† Johan Hombert‡ Jerome Pouyet§ Nicolas Schutz¶

March 5, 2006

Abstract

We analyze price competition between vertically integrated firms which interact on a downstream and an upstream markets. The upstream market allows downstream-only firms to buy inputs needed to compete on the final market. The upstream product offered by the vertically integrated firms is perfectly homogenous whereas the downstream market is differentiated. The setting is relevant to analyze the role of upstream markets in network industries, the licensing of (substitute) innovations or patents to potential rivals in oligopolistic markets, or the impact of resale markets.

As common sense suggests, there always exists an equilibrium in which upstream prices are equal to the unit cost of providing the intermediate product. However, even in a static framework, two other equilibria appear: one in which the upstream market becomes endogenously monopolistic, and another one in which vertically integrated firms set identical prices strictly larger than their average cost but strictly smaller than the joint profit-maximizing upstream price.

In light of these results, we then examine the determinants of the competitiveness of the upstream markets. An application to the analysis of wholesale markets in telecommunications is proposed.

Journal of Economic Literature Classification Number: L13, L51.
Keywords: Upstream market, downstream market, integrated firms.

*Particular thanks to Marie-Laure Allain, Bernard Caillaud, David Ettinger, Bruno Jullien, Bernard Salanié and Jean Tirole for helpful discussions during this work.

†ENST and CREST-LEI. Address: ENST, Département EGSH, 46 rue Barrault, 75634 Paris Cedex 13, FRANCE. Tel: +33(0)145817246. Fax: +33(0)145659515. E-mail: marc.bourreau@enst.fr.

‡PSE. Address: Paris-Jourdan Sciences Economiques, 48 Boulevard Jourdan, 75014, Paris, FRANCE. E-mail: johan.hombert@gmail.com.

§Author for correspondence. Ecole Polytechnique and CEPR. Address: Department of Economics, Ecole Polytechnique, 91128 Palaiseau Cedex, FRANCE. Phone: +33(0)169332646. Fax: +33(0)169333858. E-mail: jerome.pouyet@polytechnique.edu

¶PSE. Address: Paris-Jourdan Sciences Economiques, 48 Boulevard Jourdan, 75014, Paris, FRANCE. E-mail: nicolas.schutz@gmail.com.

1 Introduction

In the telecommunications industry, two types of competitors are observed. "Facility-based" firms roll out proprietary networks, which requires substantial investments, and rely mainly on their own infrastructures to provide services to end customers. Contrary to facility-based firms, "service-based" firms do not (or only partially) invest in facilities but lease access to the networks of facility-based firms to offer services on retail markets.[1]

For instance, in the broadband market, broadband cable networks, the broadband units of incumbent local exchange carriers (ILECs) and the competitive local exchange carriers (CLECs) which have installed modems (DSLAM) in the local exchanges of ILECs can be viewed as facility-based firms. They compete with service-based entrants, which use bitstream access or resale offers as inputs for their broadband services. In the mobile telephony market, mobile network operators (facility-based firms) compete with "mobile virtual network operators" (service-based firms), which do not own a mobile network but lease access to existing mobile networks.

Since service-based firms need access to the networks of facility-based firms to offer services to end customers and cannot enter in the downstream market otherwise, an important policy question is whether competing facility-based firms are likely to offer access to their networks on a voluntary basis. In other words, in the absence of regulation, when facility-based competition is in place, is a wholesale market likely to emerge?

There is no clear consensus on this question. Some regulators question that a wholesale market can emerge. For instance, in 2005, the French regulatory authority, ARCEP, proposed to regulate the conditions of access to mobile networks for MVNO, as it believed that mobile networks would not give access to their facilities otherwise.[2] In some European countries (e.g., in Denmark, France, Italy, the UK),[3] bitstream access offers are regulated on such grounds too.[4] But others challenge this view. For instance, Hazlett (2005) argues that unregulated infrastructure-based competition should lead to the emergence of a wholesale market. According to him, this was observed in the mobile telephony market in the US.[5] The UK regulatory authority, OFCOM, in its review of the wholesale broadband market (OFCOM, 2004), considers that, in an unregulated environment, facility-based firms would lease access to their infrastructures: "*Under competitive market conditions, both cable and BT would have an incentive to offer a wholesale product.*" (OFCOM, 2004, §2.154, p. 49).

Once a wholesale market has emerged, will it be competitive? The usual presump-

[1] Note that facility-based firms might lease some network elements to ILECs, and that service-based firms might have to install some telecommunications equipments. But, roughly, a facility-based firm *builds* more than it *leases*, and a service-based firm *leases* more than it *builds*.

[2] ARCEP finally withdrew its project (see: ARCEP, 2005).

[3] See ERG (2005).

[4] In a recent theory-oriented paper, Woroch (2004) assumes too that infrastructure-based firms do not voluntarily give access to their facilities to service-based firms.

[5] "*The arrival of physical networks drove free entry into the wholesale market, not vice versa*" (Hazlettt, 2005, p. 32).

tion is that it will be indeed the case: under competitive conditions, two facility-based firms would have incentives to cut prices to gain wholesale revenues until the marginal cost is attained. For instance, in the US, according to Hausman and Sidak (2005), the Federal Communications Commission believed that facility-based competition would lead to the development of a competitive wholesale market.[6] According to OFCOM (2004), too: "*In a competitive market, cable's and BT's upstream (network) and downstream (retail) divisions would each earn a normal return*". However, evidence in the mobile industry suggests that wholesale prices granted to MVNOs remain relatively high. In the broadband market too, there is no evidence of a highly competitive wholesale market.

The objective of the paper is to provide a formal framework to address these questions, that is, to study under which conditions a wholesale market is likely to emerge, and, if it emerges, whether it is likely to be competitive.

Though we shall apply our setting to the telecommunications industry, our framework remains general. "Facility-based" firms are vertically integrated firms which interact both on a downstream (retail) market and an upstream (wholesale) market, whereas the "service-based" firms operate on the downstream market only.

We analyze the competition between two vertically integrated firms and a potential pure downstream rival. The rival firm can enter in the downstream market only it gets access to the upstream good. The vertically integrated firms compete on the upstream market with homogeneous goods, whereas the firms which are active on the downstream market compete with differentiated goods.

We show that the wholesale market is active at the equilibrium, that is, the downstream rival firm is not foreclosed at the equilibrium. Integrated firms have two motivations to give access to their upstream unit to the rival downstream firm. First, providing a wholesale product generates additional revenues. Second, and more fundamentally, serving the rival downstream firm with a relatively high upstream price mitigates competition on the downstream market.

As common sense suggests, there is an equilibrium in which the wholesale market is competitive, that is, the vertically integrated firms set prices equal to marginal cost for the upstream input. However, our game has other non-competitive equilibria. First, there is an infinity of equilibria in which the integrated firms share the upstream market at an upstream price that is higher than the marginal cost and lower than the joint profit maximizing upstream price. Second, there are equilibria in which the upstream market becomes endogenously monopolistic. This latter type of equilibrium is weakly preferred by the two vertically integrated firms.

Hence, even though a wholesale market is likely to emerge, the same economic forces

[6] "*Competition among facilities-based providers to supply network elements to future generations of CLECs would decrease the price of those network elements. The next generation of CLECs would, in turn, pass those savings along to end users in the form of lower retail prices. At some point in the process, the regulator could, in theory, withdraw and allow a competitive market for inputs to discipline the price of retail service.*" (Hausman and Sidak, 2005, p. 19).

that favor its development will mitigate the intensity of competition.

Our paper is related to the literature on network competition, which has developed following the pioneer work of Laffont, Rey, and Tirole (1998a, 1998b). In contrast with this literature, we do not study the interconnection strategies of infrastructure-based firms. We focus on the competition between infrastructure- and service-based firms in wholesale and retail markets, and study the emergence and competitiveness of this wholesale market. The closest works to ours in the literature are the ones that study the conditions of access to the network of a facility-based firm for a service-based firm.[7] However, these papers do not consider competition between facility-based firms, and hence do not study the wholesale market.

Our paper is related to the literature on ex post licensing strategies too. In this literature, it has been shown that a patent holder (an "essential facility" owner), which operates on a downstream market, has incentives to refuse to license to a pure downstream competitor when the innovation is drastic (for instance, see: Tirole, 1988). Our model extends this analysis to a situation in which two competing patent holders can sell a license to a pure downstream firm. In our setting too, the pure downstream firm can enter in the market only if it gets a license from one of the patent holders. We show that, when patent holders set royalties rates, at the equilibrium, the pure downstream firm gets a license. However, with fixed fees only, there is an equilibrium in which the patent holders refuse non-cooperatively to offer a license to the downstream firm.

The rest of the paper is organized as follows. We begin by describing the model in Section 2. In Section 3, we solve for the equilibrium and derive the main results of the paper. In Section 4 we highlight the main mechanisms of the model and provide some extensions. We provide an application to the broadband and mobile markets in the telecommunications industry in Section 5. Finally, we conclude.

2 The Model

We consider two vertically-related markets: firms active in the downstream market sell services to end-users. In order to provide the final products, these firms need an upstream input (on a one-to-one basis). We now describe the building blocks of our model.

The firms. There are two vertically integrated firms, denoted by 1 and 2, and sometimes referred as i and j in the following, where $i \neq j$ is always implicitly assumed. There is also one pure downstream competitor, denoted by e. Each vertically integrated firm is composed of an upstream and a downstream unit which produce the upstream input and the final good respectively. The pure downstream competitor may be active on the downstream market if it can get access to the upstream input from either of the vertically integrated firms.

[7] See, among others, De Bjil and Peitz (2005).

The downstream market. We consider a model of competition along the lines of Salop (1979).[8] A unit mass of consumers is uniformly distributed around a circle of unit length; consumers have unit demands for a service or good. A consumer located in x who buys from operator $k \in \{1, 2, e\}$ located in x_k incurs a cost $t|x - x_k|$, where t is the usual transportation cost and is a proxy for the intensity of competition at the downstream level. Its net utility is $u_k - p_k - t|x - x_k|$, where u_k and p_k denote respectively the gross utility associated to the final good offered by firm k and the price charged by that firm on the downstream market.

Throughout most of the paper, we shall consider that operators provide consumers with the same gross utility: $u_k \equiv u$, $\forall k$. u is also assumed to be large enough to ensure that the market remains covered in all equilibrium configurations studied later on. Firm k incurs a unit downstream cost per consumer c_{dk} due to advertising, marketing, billing, etc; we again assume that $c_{dk} \equiv c_d$, $\forall k$.[9] Finally, firms are located symmetrically on the circle.

At this stage, we can derive the demands of final services for each operator.[10] For instance, the downstream demand faced by firm 1 is:

$$D_1(p_1, p_2, p_e) = \frac{1}{3} - \frac{1}{2t}(2p_1 - p_2 - p_e).$$

The downstream demands addressed to firms 2 and e are obtained by permuting indices.

The upstream market. In order to provide one unit of final service, firm e must get access to the upstream service produced by either of the vertically integrated firms.

We denote by a_i the upstream price offered by vertically integrated firm i to the pure downstream firm, $i \in \{1, 2\}$. We consider that the upstream services offered by firm 1 and firm 2 are perfectly homogeneous. Let c_{ui} be the unit cost of providing the upstream service for firm i; we assume that $c_{ui} \equiv c_u$, $\forall i$.[11]

Timing. The sequence of decision making is as follows:

- In a first stage, the vertically integrated firms offer simultaneously and non-cooperatively their prices on the upstream market.
- In a second stage, the pure downstream competitor decides which upstream offer(s) to accept; concomitantly, firms choose non-cooperatively their prices in the downstream market.

[8] This specification of the demand functions allows to obtain simple closed-form solutions. Our results carry over to the case of a linear variable demand (detailed computations are available at the following url: `http://ceco.polytechnique.fr/home/pouyet`).

[9] Subscript 'd' stands for 'downstream'.

[10] See Tirole (1988) for instance.

[11] Subscript 'u' stands for 'upstream'.

We look for the subgame-perfect Nash equilibria (in pure strategies) of this two-stage game and proceed as usual by backward induction.

Profits. Given that the choices of downstream prices and the choice of the upstream provider(s) by the pure downstream firm are simultaneous, the choice of upstream provider(s) has no strategic impact on the product market competition: hence, firm e elects as its upstream provider the cheapest vertically integrated firm; if prices in the upstream market are equal, then we consider that firm e shares its demand across the two upstream suppliers.[12]

The profit of the pure downstream firm is given by:

$$\pi_e = [p_e - \min\{a_1, a_2\} - c_d]\, D_e.$$

The vertically integrated firm i earns profit from the downstream market and, possibly, from the upstream market; its profit can be expressed as follows:

$$\pi_i = [p_i - c_u - c_d]\, D_i + [a_i - c_u]\left[\mathbb{I}_{\{a_i<a_j\}} D_e + \mathbb{I}_{\{a_i=a_j\}} \frac{1}{2} D_e\right].$$

Notice that if the lower upstream price coincides with the upstream marginal cost, i.e., $\min\{a_1, a_2\} = c_u$, then all firms would compete on a level-playing field in the downstream market.

3 Upstream-Downstream Competition

We proceed in two steps: first, we consider the competition on the downstream market for given upstream prices set by the vertically integrated firms. Second, we focus on the competition between integrated firms on the upstream market.

3.1 Competition on the downstream market

Depending on the level of upstream prices, three cases must be considered: the demand for upstream services is supplied by one firm only (i.e., either $a_1 < a_2$ or $a_2 < a_1$); both vertically integrated firms are active on the upstream market (i.e., $a_1 = a_2$); none of the vertically integrated firms supply the upstream market, thereby implying that the pure downstream firm e is foreclosed from the downstream market.

A unique supplier on the upstream market. Suppose that $a_i < a_j$ so that firm i is the unique supplier of upstream services for firm e. For the sake of clarity, the analysis

[12] We comment and refine this argument in Section 4. In particular, we investigate an alternative timing in which firm e chooses its upstream supplier(s) before the downstream price competition stage.

proceeds by considering that the price of the upstream product is larger than its marginal cost: $a_i \geq c_u$; we show later on that $a_i < c_u$ actually never arises at equilibrium.

Firm i derives profit from both upstream and downstream levels and its problem is:

$$\max_{p_i} (p_i - c_d - c_u)D_i + (a_i - c_u)D_e.$$

By contrast, firm j and firm e earn profit from the downstream market only. However, they have different unit costs for the upstream product since the integrated firm j produces the upstream good in-house at unit cost c_u, whereas the pure downstream firm e buys this good at unit price $a_i \geq c_u$. Their problem can thus be written respectively as follows:

$$\max_{p_j} (p_j - c_d - c_u)D_j,$$
$$\max_{p_e} (p_e - c_d - a_i)D_e.$$

Solving the set of necessary and sufficient first-order conditions yields the equilibrium downstream prices. In that case with firm i as the unique upstream supplier, denote by $p_k^{(i)}$ the equilibrium price set by firm k and $\pi_k^{(i)}(a_i)$ its total profit.[13]

The following inequalities hold:

$$p_e^{(i)}(a_i) \geq p_i^{(i)}(a_i) \geq p_j^{(i)}(a_i).$$

Intuitively, the pure downstream competitor e is less competitive than the other firms since it must buy the upstream product at a price a_i larger than the upstream marginal cost. Since firm i internalizes that a consumer lost on the downstream market can be recovered on the upstream market, it behaves less aggressively than the rival integrated firm j on the downstream market.[14]

It is interesting to focus on the impact of the upstream price on the firms' profits.

Lemma 1. (When being the unique upstream supplier is individually optimal) *Let $\{i, j\} = \{1, 2\}$. $\pi_i^{(i)}(.)$ is non-monotonic strictly concave and $\pi_j^{(i)}(.)$ is strictly increasing. There exists $a_* > c_u$ such that $\pi_i^{(i)}(a_i) \geq \pi_j^{(i)}(a_i)$ iff $c_u \leq a_i \leq a_*$ (with equalities only for $a_i = c_u$ and $a_i = a_*$).*

Proof. See Appendix A.1. □

Intuitively, if the upstream price is below the average cost, each vertically integrated firm prefers not to supply the upstream market. More surprisingly, a vertically integrated firm prefers to be the sole supplier of the upstream market if the upstream price is not

[13]Superscripts in parentheses refer to the firms which actively supply the upstream market, if any.

[14]When the upstream price is too high, $\pi_e^{(i)}(a_i)$ becomes negative; define thus $a_{\max} = \sup\{a_i | \forall k, \pi_k^{(i)}(a_i) > 0\}$. Similarly, when the upstream price is too low, some profits may become negative; define thus $a_{\min} = \inf\{a_i | \forall k, \pi_k^{(i)}(a_i) > 0\}$. In the following, to streamline the analysis, we always implicitly assume that $a_{\min} < a < a_{\max}$.

too high. Indeed, still assuming that firm i is the unique upstream supplier, consider an increase in the upstream price a_i:

(i) The price on the upstream market affects the unit cost of firm e on the downstream market. This raise-the-rival's-cost effect implies that competition on the downstream market is softened since firm e is led to behave less aggressively.[15] In addition, since prices are strategic complement, both integrated firms react by increasing their downstream prices.

(ii) The upstream revenue $(a_i - c_u)D_e$ earned by the unique upstream provider is affected; typically, a high upstream price tends to contract the demand for the upstream good.

(iii) The unique upstream provider's behavior on the downstream market changes: indeed, since firm i fully internalizes the revenue from the upstream market, it tends to behave less aggressively on the downstream market to 'protect' firm e.

Since firm j earns profit from the downstream market only, it always benefits from an increase of the price that prevails on the upstream market. Indeed, this deteriorates the competitiveness of firm e (see (i)) and leads firm i to behave less aggressively (see (iii)). For these reasons, firm j's profit is strictly increasing in the upstream price. By contrast, firm i's profit is non-monotonic strictly concave. Indeed, it benefits from the raise-the-rival's-cost effect (see (i)), but a large upstream price tends to reduce its upstream revenue (see (ii)). As a result, from the viewpoint of a vertically integrated firm, it is preferable to be the unique upstream provider only if the price that prevails on the upstream market is not too large.

Let us now define the 'unique provider monopoly upstream price':[16]

$$a_m = \arg\max_{a_i} \pi_i^{(i)}(a_i).$$

Loosely speaking, a_m relates to the best upstream price set by a non-cooperative vertically integrated firm that would be the sole provider on the upstream market. The following lemma is shown in Appendix.

Lemma 2. (Upstream market's curse) *The profit of a unique supplier of upstream services at the unique provider monopoly upstream price is smaller than the profit of the vertically integrated firm which does not supply the upstream market:* $\pi_i^{(i)}(a_m) < \pi_j^{(i)}(a_m)$, $i \neq j \in \{1, 2\}$.

[15] Formally, firm e's best-response is characterized by:

$$R_e(p_i, p_j, a_i) = \arg\max_{p_e} (p_e - c_d - a_i)D_e = \frac{1}{2}\left[a_i + c_d + \frac{t}{3} + p_i + p_j\right],$$

which increases in a_i.

[16] Since $\pi_i^{(i)}(.)$ is non-monotonic strictly concave, a_m is well-defined and characterized by a first-order condition.

Proof. See Appendix A.2. □

In terms of upstream prices, this result amounts to $a_m > a_*$. Therefore, if the vertically integrated firm i supplies the upstream market and sets the most profitable upstream price, i.e., $a_i = a_m$, then firm j has no incentive to undercut firm i on the upstream market. This result already highlights one of the specificities of price competition on the upstream market: some firms on this market might prefer not to compete at all since the loss in terms of foregone upstream revenues might be more than offset by higher downstream profits due to a softer downstream competition. In the following, we shall show that this intuition although correct is also incomplete.

Two vertically integrated firms supply the upstream market. This case arises when $a_1 = a_2 \equiv a$. Again, we give the intuitions in the case $a \geq c_u$. As expected, the equilibrium final prices are such that $p_e^{(1-2)}(a) \geq p_1^{(1-2)}(a) = p_2^{(1-2)}(a)$. Let us denote by $\pi_k^{(1-2)}(a)$ the profit of firm k in that case.[17] Then, we obtain the following lemma.

Lemma 3. (When sharing the upstream market is individually optimal) *Let* $\{i, j\} = \{1, 2\}$. *There exist* $\underline{a}$ *and* $\overline{a}$ *such that* $c_u < \underline{a} < \overline{a} = a_m$ *and:*

- *for* $a \leq c_u$, $\pi_i^{(i)}(a) \leq \pi_i^{(1-2)}(a) \leq \pi_i^{(j)}(a)$;
- *for* $c_u \leq a \leq \underline{a}$, $\pi_i^{(j)}(a) \leq \pi_i^{(1-2)}(a) \leq \pi_i^{(i)}(a)$;
- *for* $\underline{a} \leq a \leq \overline{a}$, $\max\{\pi_i^{(j)}(a), \pi_i^{(i)}(a)\} \leq \pi_i^{(1-2)}(a)$;
- *for* $a \geq \overline{a}$, $\pi_i^{(i)}(a) \leq \pi_i^{(1-2)}(a) \leq \pi_i^{(j)}(a)$

(where all weak inequalities can be replaced by strict ones).

Proof. See Appendix A.3. □

In the case of a unique upstream provider, Lemma 1 states that the vertically integrated firms value differently an increase in the upstream price: in a nutshell, the upstream provider benefits from the upstream profit while the other vertically integrated firm benefits from a stronger competition moderation effect.

Consider now that vertically integrated firms share the upstream market and set the same upstream price a. When $a \leq \underline{a}$, a vertically integrated operator is mostly interested in deriving additional revenues from the upstream market. When $a \geq \overline{a} = a_m$, a reverse conclusion holds: this firm prefers letting its rival be the unique upstream provider in order to benefit from a stronger moderation of the downstream competition. For intermediate values of the upstream price, $\underline{a} \leq a \leq \overline{a}$, sharing the upstream market is the most profitable strategy for a vertically integrated firm: it allows to earn some upstream profit and contributes to softening the downstream competition.

[17] As in the unique upstream provider case, we restrict attention to values of a such that all profits are positive.

In our model, the following inequality holds:

$$\pi_i^{(1-2)}(a) \geq \frac{1}{2}\pi_i^{(i)}(a) + \frac{1}{2}\pi_i^{(j)}(a).$$

This implies first that $\pi_i^{(1-2)}(a_*) \geq \pi_i^{(i)}(a_*) = \pi_i^{(j)}(a_*)$. Hence, from the viewpoint of the non-cooperative behavior of a vertically integrated firm on the upstream market, in the neighborhood of a_*, each integrated firm prefers to share the upstream market with its rival than either to exit or to be the sole upstream supplier. Therefore supra-competitive equilibria may emerge, as will be confirmed later on.[18]

Second, consider that integrated firms collude on the upstream market (but still behave non-cooperatively on the downstream market). The previous inequality shows that the best upstream collusive agreement involves sharing the upstream market at an upstream price defined by $a_c = \arg\max_a \pi_1^{(1-2)}(a) + \pi_2^{(1-2)}(a)$.[19] In the sequel, we shall show that such a perfect collusive outcome never arises at equilibrium.

Foreclosure of the pure downstream competitor. Suppose that the vertically integrated firms set prices on the upstream market such that the demand for upstream services is nil. This arises when a_1 and a_2 are both larger than the threshold value $a_{\max}$ defined in footnote 14. In this situation, the pure downstream firm e is foreclosed from the downstream market and the profits of firms 1 and 2 are given by duopoly profits $\pi_i^{(\emptyset)}$, $i = 1, 2$.

In Appendix we show the following lemma.

Lemma 4. (No equilibrium foreclosure) *The profit of vertically integrated firm i under foreclosure is lower than when it is the sole supplier of the upstream market at the upstream monopoly price: $\pi_i^{(\emptyset)} < \pi_i^{(i)}(a_m)$.*

Proof. See Appendix A.4. □

A direct implication of this lemma is that foreclosure of the pure downstream competitor never arises at equilibrium. To build the intuition, consider that firm j is not active

[18] One could legitimately question the generality of this result. From the typical IO literature, it is reasonable to assume that $\pi_i^{(i)}(.)$ is non-monotonic strictly concave and $\pi_i^{(j)}(.)$ strictly increases in the upstream price. However, the relative position of $\pi_i^{(1-2)}(.)$ w.r.t. the average of $\pi_i^{(i)}(.)$ and $\pi_i^{(j)}(.)$, and therefore the existence of a zone of upstream prices such that vertically integrated firms prefer to share the upstream market at this upstream price, depends in a complex way on the specification of the demand functions. Nevertheless, we show that our results carry over to other settings.

[19] $\pi_i^{(1-2)}(.)$ is non-monotonic strictly concave and thus admits an interior maximum. Notice that $a_c > a_m$. Indeed, a unique upstream provider faces the following trade-off: on the one hand, a high upstream price tends to lower the upstream profit, on the other hand, it helps softening its competitive environment at the downstream level. When both vertically integrated firms share the upstream market, the first effect tends to become less important, since each provider earns only one half of the upstream profit. In addition, for a given upstream price, the competitive pressure it faces on the downstream market becomes less intense, since the other integrated firm now internalizes the impact of its downstream pricing on its upstream profit, and therefore becomes less aggressive. The way both effects are modified implies that a_c is greater than a_m.

on the upstream market and that firm i decides whether to foreclose the pure downstream competitor firm or not. This decision trades off several effects: as usual, foreclosure reduces competition at the downstream level (a duopoly instead of a triopoly) at the cost of giving up the upstream profit; furthermore, removing the inefficient pure downstream firm tightens the competitive pressure exerted by firm j (since downstream prices are strategic complements).

Figure 1 offers a graphical representation of Lemmas 1-4.

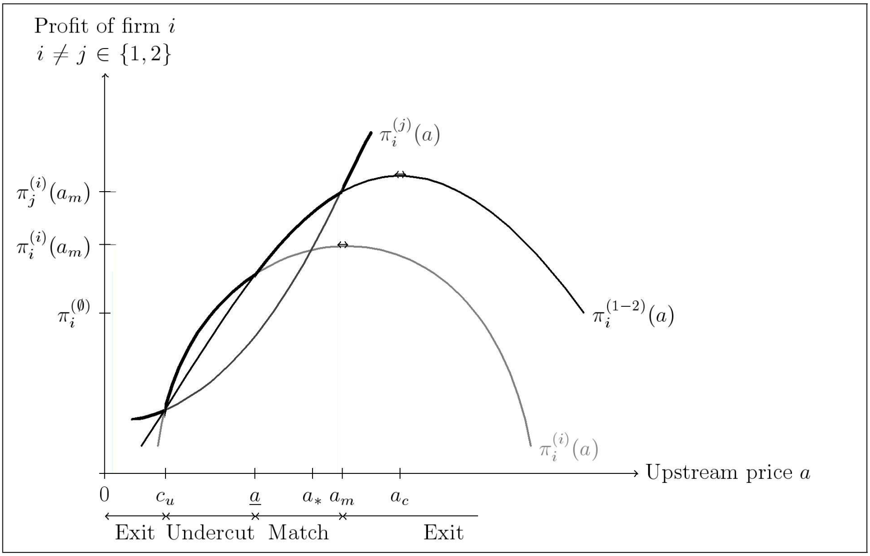


Figure 1: When only one integrated firm supplies the upstream market, it earns more profit than the other integrated firm if and only if $a \in [c_u, a_*]$ (Lemma 1). At upstream price a_m, firm i prefers that its rival be the sole supplier of the upstream market (Lemma 2) If firm i sets $a \in [\underline{a}, \overline{a}]$, firm j prefers to share the upstream market with its rival rather than undercut or exit (Lemma 3). A vertically integrated firm prefers to be the unique upstream provider at the upstream monopoly price than exclude the pure downstream competitor (Lemma 4).

3.2 Competition on the upstream market

It remains to determine the outcome of the competition at the upstream level. Given the upstream price a_j set by its rival on the upstream market, vertically integrated firm i can adopt three different strategies:

- It can *undercut* its rival on the upstream market, i.e., $a_i < a_j$; in this case, it becomes the unique provider on the upstream market and earns $\pi_i^{(i)}(a_i)$.

- It can *match* its rival on the upstream market, i.e., $a_i = a_j \equiv a$; in this case, it shares the upstream demand and earns $\pi_i^{(1-2)}(a)$.

- It can *exit* the upstream market by setting $a_i > a_j$, in which case it earns $\pi_i^{(j)}(a_j)$.

A quick inspection of Figure 1 provides the intuitions for the different equilibria that arise in our game: indeed, the upper envelope of the profit functions drawn in Figure 1 (the thick dotted curve) allows to determine the optimal strategy of firm i for a given upstream price set by rival firm j.

If firm j sets $a_j \in (c_u, \underline{a})$ then firm i has an incentive to undercut; but firm j prefers in turn to undercut firm i. Intuitively, this process continues until both vertically integrated firms offer upstream services at the upstream unit cost c_u.

By contrast, if firm j sets an upstream price $a_j \in [\underline{a}, \overline{a} = a_m]$ then firm i prefers to match and share the upstream market. Given the symmetry of firms i and j, this implies that there exists a continuum of equilibria in which vertically integrated firms share the upstream market at a common price $a \in [\underline{a}, a_m]$.

Finally, suppose that firm i sets $a_i = a_m$; then, firm j is exactly indifferent between sharing the upstream market or exiting. As a result, there exists an equilibrium in which the upstream market is supplied by one firm only at the unique upstream provider monopoly price.

The next proposition rigorously states these intuitions.

Proposition 1. *There exist three types of subgame-perfect Nash equilibria in pure strategies:*

- *A Bertrand-like outcome in which vertically integrated firms offer upstream services at the upstream unit cost:* $a_1 = a_2 = c_u$.

- *Supra-competitive outcomes in which vertically integrated firms offer upstream services at the same upstream price* $a \in [\underline{a}, a_m]$.

- *Monopoly-like outcomes in which only one vertically integrated firm offers upstream services at the unique upstream provider monopoly price* a_m.

There exist no other equilibria.

Proof. First, we show that there exists an equilibrium in which integrated firms share the upstream market at common price c_u. Consider an upward deviation by firm i: $a_i > c_u$. Then, firm j becomes the unique supplier of the upstream market and firm i's profit is $\pi_i^{(j)}(c_u) = \pi_i^{(1-2)}(c_u)$ given Lemma 3. Firm i is thus indifferent between sharing the upstream market at a price c_u and letting firm j serve the upstream market. Consider a downward deviation by firm i: $a_i < c_u$. Its profit is now $\pi_i^{(i)}(a_i) < \pi_i^{(i)}(c_u) = \pi_i^{(1-2)}(c_u)$, since $\pi_i^{(i)}$ is increasing for $a_i < a_m$; thus a downward deviation is not profitable.

Second, we show that, for all $a \in [\underline{a}, a_m]$, there exists an equilibrium in which integrated firms share the upstream market at common price a. Consider an upward deviation by firm i: $a_i > a$. Then, firm i's profit is $\pi_i^{(j)}(a)$ which is smaller than $\pi_i^{(1-2)}(a)$ given Lemma 3. Consider a downward deviation by firm i: $a_i < a$. Then, firm i's profit is $\pi_i^{(i)}(a_i)$ which is smaller than $\pi_i^{(i)}(a)$, since $\pi_i^{(i)}$ is increasing for $a_i < a_m$. $\pi_i^{(i)}(a)$ being itself smaller than $\pi_i^{(1-2)}(a)$ given Lemma 3, we deduce that a downward deviation is not profitable.

Third, we show that there exists an equilibrium in which integrated firm i supplies the upstream market at price a_m and firm j sets a price larger or equal to $a_{\max}$. Given Lemma 4, firm i does not want to foreclose the pure downstream firm e. Given that it is the unique upstream supplier, setting $a_i = a_m$ is indeed optimal for firm i. Consider now a deviation by firm j. If $a_j > a_m$, its profit is unchanged. If $a_j = a_m$, it earns $\pi_j^{(1-2)}(a_m) = \pi_j^{(i)}(a_m)$. If $a_j < a_m$, since $\pi_j^{(j)}$ is increasing for $a_j < a_m$, it earns $\pi_j^{(j)}(a_j) < \pi_j^{(j)}(a_m)$, itself smaller than $\pi_j^{(i)}(a_m)$ given Lemma 2. Hence no deviation by firm j is strictly profitable.

We now prove that there exist no other subgame perfect equilibria in pure strategies.

Claim 1. *It is not possible that the upstream market is supplied at an upstream price $a < c_u$.*

Proof. If firm i is the unique supplier of the upstream market with $a_i < c_u$, then it is a strictly profitable deviation for it to set $a_i + \epsilon$, with a small $\varepsilon > 0$ such that $a_i + \varepsilon < \min\{a_j, c_u\}$, since $\pi_i^{(i)}$ is strictly increasing for $a_i < a_m$. If both integrated firms share the upstream market at common price $a_i = a_j < c_u$, then is is a strictly profitable deviation for firm i to exit the upstream market since $\pi_i^{(j)}(a_i) > \pi_i^{(1-2)}(a_i)$ given Lemma 3. □

Claim 2. *It is not possible to have a unique upstream provider at price c_u.*

Proof. Suppose that firm i is the unique upstream supplier at price c_u and that firm j sets a price $a_j > c_u$. Then it is a strictly profitable deviation for firm i to set $c_u + \varepsilon$ with a small $\varepsilon > 0$ such that $c_u + \varepsilon < \max(a_j, a_m)$, since π_i^i is strictly increasing for $a_i < a_m$. □

Claim 3. *It is not possible that the upstream market is supplied at an upstream price $a \in (c_u, \underline{a})$.*

Proof. If firm j is the sole supplier of the upstream market at a price $a_j \in (c_u, \underline{a})$, then it is a strictly profitable deviation for firm i to undercut, since $\pi_i^{(i)}(a_j - \varepsilon) > \pi_i^{(j)}(a_j)$ for a sufficiently small ε given Lemma 3. Again, if both integrated firms share the upstream market with $a_i = a_j \in (c_u, \underline{a})$, it is a strictly profitable deviation for firm i to undercut. □

Claim 4. *It is not possible to have a unique upstream provider at a price $a \in [\underline{a}, a_m)$.*

Proof. If firm j is the sole provider of the upstream market at a price $a_j \in [\underline{a}, a_m)$, then it is a strictly profitable deviation for firm i to match and share the upstream market, since $\pi_i^{(1-2)}(a) > \pi_i^{(j)}(a)$ given Lemma 3. □

Claim 5. *It is not possible that the upstream market is supplied at an upstream price $a > a_m$.*

Proof. If firm i is the sole supplier of the upstream market at a price $a_i > a_m$, then it is a strictly profitable deviation for it to set $a_i = a_m$. If both integrated firms share the upstream market with $a_i = a_j > a_m$, it is a strictly profitable deviation for firm i to exit the upstream market, since $\pi_i^{(j)}(a) > \pi_i^{(1-2)}(a)$ given Lemma 3. □

Claim 6. *It is not possible that the upstream market is not supplied by any integrated firm.*

Proof. Immediate from Lemma 4. □

This concludes the proof of the proposition. □

In our model, the upstream providers offer perfectly homogeneous products. Hence, it seems rather intuitive that the usual Bertrand result emerges: competition on the upstream market is fierce and drives upstream prices to the upstream unit cost. Thus, competition on the upstream market allows the pure downstream firm to compete on a level playing field with both vertically integrated firms on the downstream market. However, as shown by the previous proposition, other equilibria exist, which are much less competitive.

Indeed, non-cooperative competitors may be able to implement the unique upstream provider monopoly price, i.e., a partially collusive outcome, on the upstream market. This radical departure from the Bertrand outcome stems from the fact that the upstream competitors are also present on the downstream market. Hence, when a vertically integrated firm sets its price for its upstream services, it trades off the profit derived from that market with the impact on the competition at the downstream level. If, say, firm i sets an upstream price in $[\underline{a}, a_m]$, then firm j finds it optimal to exactly match firm i and to share the upstream demand. Undercutting would imply that firm i derives no upstream profit; in turn, this would make firm i a more aggressive competitor on the downstream market, a non profitable strategy for upstream prices in $[\underline{a}, a_m]$.[20]

Notice that the equilibria in which vertically integrated firms share the upstream market at a price $a \in [\underline{a}, a_m]$ can be ranked: both vertically integrated firms unambiguously prefer to implement a larger common upstream price.

Finally, there also exists an equilibrium in which only one vertically integrated operator serves the upstream market at the upstream monopoly price. When, say, firm i is the unique upstream supplier at price a_m, firm j turns out to be exactly indifferent between

[20] Suppose that the vertically integrated firms have different cost structures (either for the upstream or the downstream markets). Then, in line with Lemma 3, we would be led to define thresholds for each vertically integrated firm: $\underline{a}_i$ and $\overline{a}_i$, $i \in \{1, 2\}$. If the costs difference is not too large, then $\max\{\underline{a}_1, \underline{a}_2\} < \min\{\overline{a}_1, \overline{a}_2\}$ and there still exist equilibria in which firms share the upstream market at a supra-competitive common price.

exiting and sharing the upstream market. However, firm i prefers that firm j shares the market and becomes a less aggressive competitor on the downstream market. Again, in the standard Bertrand game, the decision to exit the market is always weakly dominated (since the firm then earns no profit); in our model, such a strategy can be optimal for a vertically integrated firm since it allows to benefit from a less aggressive rival on the downstream market.

We conclude this discussion with the following proposition.

Proposition 2. *The partially-collusive-like equilibrium in which $a_1 = a_2 = a_m$ is weakly preferred to any other equilibrium by both vertically integrated firms, and strictly preferred by at least one of them.*

Proof. Immediate. □

Hence, the following presumption emerges from our analysis: vertically integrated firms which compete on an upstream market are likely to reach a partially collusive outcome; the competitiveness of the upstream market should not be taken for granted. Though there exists a potential competitive pressure on the upstream market, it is unlikely that this pressure alone drives the upstream price down, even absent any form of tacit or overt collusion.

It is also interesting to highlight the reason that makes full collusion on the upstream market impossible to implement by the vertically integrated firms. Remember that a_c is the upstream price that maximizes the sum of the vertically integrated firms' profits. If the vertically integrated firm i sets a_c then its integrated rival has an incentive to overbid, that is, to exit the upstream market. Again the logic of the standard Bertrand competition game is reversed: with respect to the collusive joint optimum, a firm has an incentive not to undercut its rival on the upstream market but rather to offer a strictly larger upstream price.

4 Discussions and Extensions

In this section we highlight the main mechanisms driving our model thanks to several thought experiments. We also provide some extensions and comparative statics results.

4.1 Fundamental Mechanisms

Pure upstream competitor. Consider the scenario in which a pure upstream competitor, that is, a firm which is active only in the upstream market, might provide the intermediate service to the pure downstream firm. That upstream firm has no direct interactions with the downstream market; thus, its best-response is to undercut its rivals as long as the upstream price remains above the average cost. Moreover, if the upstream competitor becomes the unique upstream provider, then an integrated operator has an incentive to undercut since this steals upstream profits without affecting the other firms' best-responses on the downstream market. In other words, undercutting the pure upstream firm has not the adverse interaction effect of making integrated competitors more aggressive on the downstream market.

As a result, only the Bertrand-like equilibrium can emerge when there is at least one pure upstream competitor. This result highlights the fundamental role of upstream-downstream interactions in our model.

The role of upstream pricing. Throughout our analysis, we have considered that vertically integrated firms could meter the usage of their intermediate services made by the pure downstream competitor. Hence, unit prices (or, analogously, royalties or access charges) can be implemented. In some contexts, like patent licensing, integrated firms might only be able to impose a fixed fee for the usage of their intermediate input. Denote by P_i^u the fixed fee offered by vertically integrated firm i for the access to the upstream technology. A direct consequence of this type of upstream pricing is that the pure downstream competitor buys the intermediate service from one integrated firm only.

In this context, firm e's problem is now:

$$\max_{p_e}(p_e - c_d)D_e - \min\{P_1^u, P_2^u\},$$

provided that it earns a positive profit at equilibrium. Clearly, the upstream fees do not affect the downstream outcome. Therefore, the only equilibrium in which the downstream firm is not foreclosed from the downstream market is such that vertically integrated firms set upstream fees so as to exactly cover their costs of providing the upstream services.

Hence, this upstream pricing may appear as pro-competitive. However, it modifies the incentives to foreclose the pure downstream firm. Start from a situation where this firm is foreclosed, and consider the new trade-off faced by an integrated firm deciding whether or not to serve the upstream market. On the one hand, a fixed fee does not allow to relax competition on the downstream market. On the other hand, this pricing allows to capture the whole profit of a pure downstream firm. It turns out that in our setting, when integrated firms use fixed fees, there exists an equilibrium in which the pure downstream competitor is foreclosed.[21]

What are the competitive impacts of the nature of upstream pricing? Thanks to unit pricing firms can manipulate their rivals' costs and relax competition on the downstream market. This creates strong interactions between both markets, giving rise to non-competitive equilibria with an active downstream firm. By contrast, fixed upstream fees do not give rise to these upstream-downstream interactions, but create conditions favorable to foreclosure.

[21] Computations are straightforward and left to the reader.

Strategic nonparticipation. Suppose that the vertically integrated firms can credibly commit, say in stage zero, not to serve the upstream market before the price competition stage on this market. In this case, the Bertrand-like outcome cannot be part of a subgame-perfect equilibrium. Indeed, if it were the case, an integrated firm would rather prefer not to participate to the upstream market in stage zero, thereby triggering a higher upstream price at stage one, and softening downstream competition in stage two.

This is reminiscent of the analysis in Jehiel and Moldovanu (1996) who study bidders' incentives to participate in an auction with negative externalities and derive a similar result. In fact, our price competition game on the upstream market can be viewed as an auction run by the pure downstream firm in which bidders are the vertically integrated firms. In Jehiel and Moldovanu's paper, the payoff earned by the bidders depends on who wins the auction; hence it may be interesting not to participate in order to influence the outcome of the auction. In our model, the externality depends on the level of upstream prices; a bidder may decide not to participate to avoid low-price equilibria, such as the Bertrand outcome.

This variation of our setting reinforces our result on the emergence of the monopoly-like equilibrium. Indeed, it would exist even if the upstream market curse (see Lemma 2) did not hold, provided that the decision not to participate to the upstream market is a binding commitment.

Strategic choice of the upstream provider(s). In our basic setting, we make two assumptions about the choice of the upstream provider(s). First, when $a_1 = a_2$, the pure downstream firm splits its upstream demand in equal shares across the two suppliers. Second, the upstream provider(s) is (are) chosen concomitantly with the downstream prices. Let us examine the role of these assumptions.

First, when $a_1 = a_2$, the pure downstream firm is ex post indifferent between any sharing out of its upstream demand. However, its choice can affect its the upstream providers' behavior. Consider a variation of our model in which firm e can decide how to split its upstream demand when upstream prices are equal. To simplify, we restrict its last stage strategy to constant proportional sharing rules. If it purchases a fraction $\alpha \in [0,1]$ from firm i when $a_i = a_j = a$, then denote by $\pi_i^{(1-2)}(a,\alpha)$ (resp. $\pi_j^{(1-2)}(a_1-\alpha)$) the profit of firm i (resp. j). Lemma 3 can be extended by defining new thresholds as functions of α. There exist $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ such that $c_u < \underline{a}(\alpha) \leq a_* \leq \overline{a}(\alpha)$, and the profits $\pi_i^{(i)}(.)$, $\pi_i^{(1-2)}(.,\alpha)$ and $\pi_i^{(j)}(.)$ are ordered as in Lemma 3 between these thresholds. $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ are decreasing, $\underline{a}(1/2) = \underline{a}$, $\overline{a}(1/2) = \overline{a}$ and $\underline{a}(0) = \overline{a}(1) = a_*$. Applying this result, we obtain that the three types of perfect-subgame equilibria shown in Proposition 1 are conserved, and that the range of upstream prices sustaining a collusive-like equilibrium shrinks to $[\max\{\underline{a}(alpha_i), \underline{a}(1-\alpha)\}, \min\{\overline{a}(\alpha), \overline{a}(1-\alpha)\}] = [\underline{a}(\min\{\alpha, 1-\alpha\}), \overline{a}(\max\{\alpha, 1-\alpha\})]$. It reaches its maximal size $[\underline{a}, \overline{a}]$ for $\alpha = 1-\alpha = 1/2$; it degenerates to the single point a_* for $\alpha = 0$ or $\alpha = 1$. Hence, our main result, the existence of collusive-like equilibria, does not crucially rely on our assumption of symmetric split of the upstream demand when upstream prices are equal.

Second, in our basic setting, the pure downstream firm determines simultaneously its price on the final market and which integrated firm to buy the intermediate input from. This choice is concomitant to the determination of downstream prices by the integrated firms. As a consequence, the choice of upstream provider has no strategic impact on competition that takes place on the downstream market. Let us now consider a variation of this baseline scenario, in which the decision to choose a provider is made before the downstream price competition game. Let denote by $\alpha \in [0,1]$ the share of the upstream demand that is bought from integrated firm i. Given upstream prices, the pure downstream firm's program is:

$$\max_{\alpha\in[0,1]} \pi_e^*(\alpha, a_i, a_j) \equiv [p_e^*(\alpha) - c_d - \alpha a_i - (1-\alpha)a_j]\, D_e(p_e^*(\alpha), p_i^*(\alpha), p_j^*(\alpha)),$$

where $p_i^*(\alpha)$, $p_j^*(\alpha)$ and $p_e^*(\alpha)$ are the equilibrium downstream prices for a given α. The choice of α has a direct effect on firm e's markup, which makes it willing to buy from the cheapest upstream provider. Interestingly, this choice has now a strategic impact on the downstream outcome. At this level of generality, this strategic effect cannot be signed. However, in our context, it turns out that the direct effect always dominates the strategic one, so that firm e always buy the intermediate input from the cheapest provider. Whether this feature continues to hold in other settings would be worth investigating: indeed, if it were not the case, this would mean that an upstream buyer would accept to buy the same intermediate input at different prices in order to alter strategically the final product market competition.

Differentiated upstream market. A natural question to ask is whether our results continue to hold when the intermediate inputs offered by vertically integrated firms on the upstream market are not perfectly substitutable. Suppose for instance that the share of the upstream demand that is served by firm i is given by $\alpha_i(a_i, a_j) = 1/2 - (a_i - a_j)/(2\tau)$, where $\tau > 0$ is the degree of differentiation between vertically integrated firms on the upstream market.[22] Straightforward computations first confirm that our results are indeed robust.[23] As soon as the ratio t/τ is large enough, then two symmetric equilibria exist, one in which upstream prices are rather low, an another in which upstream prices are rather high. Second, when the ratio t/τ becomes small enough there exists a unique symmetric equilibrium with upstream prices converging to the unit upstream cost as this ratio tends towards 0. A tentative conclusion would therefore be that as competition intensifies on the downstream market (i.e., as t goes to 0), vertically integrated firms compete more and

[22]Even though inputs are now heterogenous, we maintain the assumption that one unit of input from firm i, $i = 1, 2$, allows firm e to produce one unit of final product and that the choice of input provider does not affect the base quality of the final product.

[23]Computations are available from the authors upon request.

more fiercely on the upstream market. The conventional wisdom expressed by regulators is thus somewhat reversed: the development of competition of downstream markets is key to the emergence of competitive upstream markets.

Existence of equilibria. Notice that monopoly-like equilibria are less likely to exist than collusive-like equilibria. Indeed, assume that a monopoly-like equilibrium exists, then it must be that $a_m > a_*$ (otherwise, undercutting would be optimal). Using:

$$\pi_i^{(1-2)}(a_*) \geq \frac{1}{2}\pi_i^{(i)}(a_*) + \frac{1}{2}\pi_i^{(j)}(a_*) = \pi_i^{(i)}(a_*) = \pi_i^{(j)}(a_*),$$

we obtain that $a_1 = a_2 = a_*$ is part of a collusive-like equilibrium, since deviating cannot yield more than $\max\{\pi_i^{(i)}(a_*), \pi_i^{(j)}(a_*)\}$. Therefore, in the sequel, we will say that the upstream market is not competitive when a collusive-like equilibrium exists.

4.2 The Determinants of Effective Upstream Competition

It is obvious that fierce competition on the upstream market enhances competition on the downstream market. In turn, we will see that characteristics of the downstream market affect the outcomes that can arise on the upstream market. These results provide some insights on the strategies of firms.

Downstream differentiation strategy. It is straightforward to see that as final products become less differentiated (t goes to 0), both the values of $\underline{a}$ and $\overline{a}$ tend to the upstream unit cost c_u (the analytical expressions of $\underline{a}$ and $\overline{a}$ are given in Appendix). As competition intensifies on the downstream market, the upstream market becomes more competitive (in the following sense: whatever the equilibrium configuration, the upstream price tends to the marginal cost of the intermediate service).

Consider now the other polar case in which final products are strongly differentiated so that each firm is a local monopoly on the downstream market (which arises when the transportation cost t is large enough with respect to the base utility level u). Interactions between firms on the downstream market disappear, implying that the upstream market outcome does not help softening competition on the downstream market. As a consequence, vertically integrated firms compete head-to-head on the upstream market and, at the unique equilibrium, they share the upstream demand at the upstream unit cost.

To summarize, there is a non-monotonic link between downstream differentiation and the competitiveness of the upstream market. On the one hand a high degree of differentiation softens competition on the downstream market. On the other hand, differentiation reduces interactions between both markets and makes non-competitive outcomes on the upstream market less likely.

Incentives to quality degradation/improvement. Straightforward calculations show that an increase in the downstream cost of firm k (c_{dk}) is equivalent to a decrease of the same amount in the gross utility provided to its final consumers (u_k). It seems reasonable to assume that integrated firms can voluntarily damage the upstream good, which in turn affects the quality of the pure downstream firm, and hence its cost efficiency. To keep the analysis as simple as possible we assume that both integrated firms adopt the same level of quality degradation.

Comparative statics on c_{de} allows to study the incentives to degrade or improve the quality of the upstream good provided to firm e. The intuition goes as follows: the more efficient the pure downstream competitor is, the larger is its downstream market share, thus the more important is the potential upstream profit of an integrated firm with respect to its downstream profit. This leads to balanced conclusions. On the one hand, sharing the upstream market becomes more profitable relative to exiting it. This allows a higher non-cooperative common upstream price $\overline{a}$, and thus higher profits for the vertically integrated firms, as long as the collusive-like equilibrium with $a_1 = a_2 = \overline{a}$ exists. On the other hand, sharing the upstream market becomes less profitable relative to supplying it alone. This exacerbates the incentives of integrated firms to undercut their rival in order to be the sole supplier of the upstream market. When the pure downstream competitor is efficient enough, the former effect becomes so important that the partially collusive equilibrium disappears. In this case the Bertrand outcome becomes the only Nash equilibrium.

To conclude, as long as quality differences are not too large (c_{de} not too different from $c_{d1} = c_{d2}$), vertically integrated firms have incentives to improve the quality of the upstream good in order to raise the upstream price. However when the pure downstream firm is too efficient ($c_{de} - c_{d1}$ inferior to a certain threshold), only the Bertrand-like equilibrium can arise, so that the integrated firms want to degrade the upstream good in order to sustain non-cooperative non-competitive upstream prices.

Market structure. We now address the impact of market structure, i.e., the number of firms of each type, on competition intensity. Basically, the conventional wisdom states that more competitors leads to an increased competitive pressure. However, we argue that the presence of upstream-downstream interactions requires a more careful analysis.

The condition for the upstream market to be competitive is that there are no equilibria in which all integrated firms share the upstream market at a common upstream price, because each of them would rather undercut to the 'unique provider monopoly upstream price' and capture the whole upstream demand. As already emphasized, starting from a situation in which all integrated firms share the upstream market, the decision to undercut trades off two effects. It allows to get more upstream profits, at the cost of making the other integrated firms more aggressive on the downstream market. We now discuss informally the impact of the number of integrated and pure downstream firms on the (dis)incentives to undercut.

Consider first an increase in the number on pure downstream firms. On the one hand, the additional upstream profit of an undercutting integrated firm increases relative to its downstream profit. On the other hand, it may worsen the negative impact on its downstream profit, since the competition moderation effect of sharing the upstream market is larger with more pure downstream firms. Overall whether the entry of pure downstream firms enhances competition on the upstream market or not is ambiguous.

The entry of new vertically integrated firms has an ambiguous impact on the relative importance of upstream and downstream profits, since both profits decrease. Nevertheless one might expect the marginal profit-stealing effect of one additional integrated competitor to be greater when there are few integrated firms, since undercutting allows to get an additional fraction of the upstream market equal to $1 - 1/(\#$ of integrated firms$)$, which is concave. The competition moderation effect is ambiguously affected by the number of integrated firms. On the one hand, a higher number of integrated competitors implies that each firm supplies a smaller fraction of the upstream market, which make them less accommodating on the downstream market. On the other hand, an undercutting integrated firm faces suddenly a larger number of aggressive rivals. Again, the overall effect of the entry of vertically integrated firms is a priori undetermined, though we suspect a pro-competitive impact when there are initially few integrated firms.[24]

Predation or intense competition? We now go back to our baseline three-firm case. If the pure downstream competitor is inefficient enough (c_{de} high relative to $c_{d1} = c_{d2}$), the positions of profit curves in the neighborhood of $a = c_u$ are reversed (inequalities in the first two items of Lemma 3 are reversed). In particular, if a vertically integrated firm, say 1, offers an upstream price just below the unit cost c_u, firm 2's best response is to undercut and supply the whole upstream demand. The intuition is the following: if $a_1 < c_u$, firm 1 makes upstream losses. Therefore it is induced to behave aggressively on the downstream market in order to minimize the pure downstream operators's market share. Firm 2 prefers undercutting on the upstream market. Indeed, since the pure downstream is assumed to be inefficient, firm 2's upstream losses are small enough to be outweighed by the benefit of avoiding the aggressive behavior of firm 1.

In this case the thresholds $\underline{a}$, a_* and $\overline{a}$ are below c_u. Following the guidelines of the proof of Proposition 1 we can show that there exist a Bertrand-like equilibrium and a set of 'super-competitive' equilibria, in which vertically integrated firms share the upstream demand at a common price $a \in [\overline{a}, \underline{a}]$.

Throughout most of the paper we have emphasized that interactions between upstream and downstream markets can lead to non-competitive outcomes, in which upstream prices

[24]Numerical computations confirm this intuition. For example, when allowing for downstream cost differences between pure downstream and vertically integrated firms, the constellation of cost parameters for which there exists a collusive-like equilibrium has a maximal size for an intermediate number of integrated firms, and shrinks for extreme values.

are above unit costs without collusion. This paragraph illustrates that in some cases (an inefficient pure downstream firm), they can also lead to super-competitive outcomes, in which upstream prices are below unit prices without predation.

5 Applications: Wholesale Markets in Telecommunications

The development of wholesale markets and the competitiveness of these markets are key policy issues in the telecommunications industry. We provide below an analysis of wholesale markets in the broadband and the mobile telephony markets. We then discuss some implications of our results in terms of regulatory policy.

5.1 The Broadband Market

Broadband services enable users to access the Internet at high-speed rates. In most industrialized countries, broadband is developing fast. For instance, there were 37.9 million broadband lines in the US and 33.4 million broadband lines in the European Union in 2004, compared to 28.2 million lines and 19.4 million lines, respectively, one year earlier.[25]

Different technologies can deliver broadband, but two platforms dominate local broadband markets worldwide: the cable modem platform (which is the most widely used in the US)[26] and the copper-based digital subscriber line (DSL) platform (which is the dominant technology in the EU).[27]

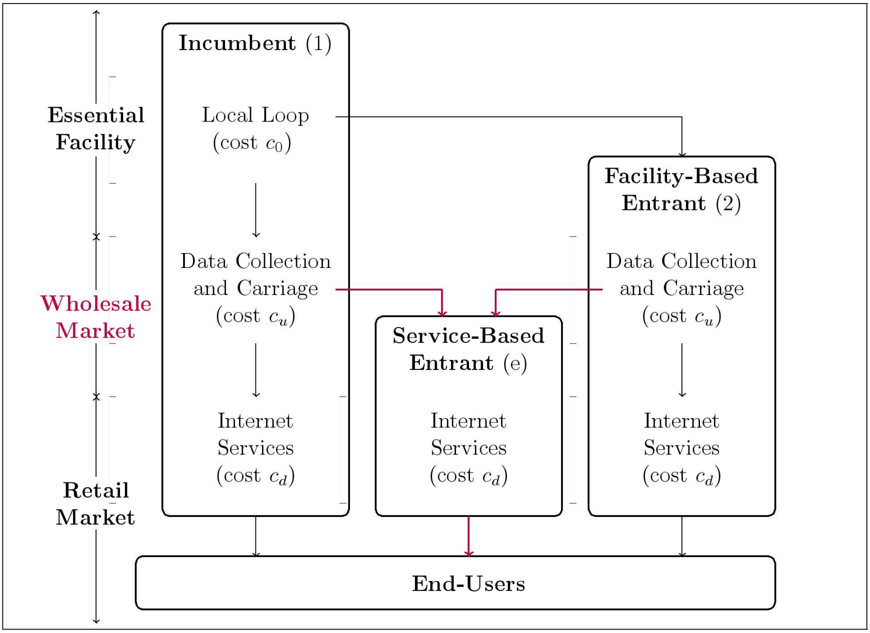
The broadband market has a three-tiered structure: the local access network market provides the local access input for wholesale services; the wholesale broadband market provides the input for broadband services; and finally, the retail market delivers broadband services to consumers. Facility-based firms either own a local access network (such as, a copper network or a cable network) or get unbundled access to the local loop of the ILEC.[28] In general, they have to upgrade the local network to offer broadband services. Service-based firms buy a wholesale broadband service to facility-based firms (if available). This is represented in Figure 1.

[25] See FCC (2005a) and EC (2006a). Data for the US indicate the number of business and residential broadband lines as of December 2003 and December 2004. Data for the EU are valid as of October 2003 and 2004.

[26] According to FCC (2005a), in 2004, cable providers served 56.4 per cent and ADSL 36.5 per cent of the broadband lines in the US. The remaining Internet access lines were delivered through wireless, fiber, etc.

[27] According to EC (2006a), in 2004, DSL platforms served 80.4 per cent of the broadband lines in the European Union, compared to 16.8 per cent of lines provided by cable and 2.8 per cent by other access technologies.

[28] Unbundled access to the local loop is a regulatory measure which obliges the ILEC to give access to its copper lines that connect the customer premises to the local exchanges. It enables entrants to install DSL equipments to offer broadband services to end customers.



The wholesale broadband market. There are two main types of wholesale offers. With resale, the service-based firm buys the service of the infrastructure-based firm on a wholesale basis, and sells this service under its brand name; however, possibilities of differentiation are limited, since the two services are identical. With bitstream access, the service-based firm buys data interconnection to the infrastructure-based firm's broadband network; some differentiation from the infrastructure-based firm's services is possible. Bitstream access can be delivered on different platforms, and in particular, on the cable modem and DSL platforms.[29]

Though there is no legal obligation on cable networks to open their facility to rivals, wholesale broadband offers over cable can be found in Europe,[30] the US,[31] Canada and Israel.[32] Wholesale offers on DSL platforms exist in many countries. Whereas in some countries (e.g., Denmark, France or Italy), bitstream access over the ILEC's network

[29]For instance, the European Regulators Group (ERG) considers that "it appears quite possible that a data over cable network will be able to provide each and every one of the key elements of bitstream access functionality" (ERG, 2005, p. 16).

[30]For instance, in the UK, one of the two cable operators has signed a wholesale broadband agreement with a third-party ISP (see Ofcom, 2004, §3.15, p. 69).

[31]In the US, there are a few examples of wholesale broadband offers over cable. For instance, following its merger with AOL Time Warner signed wholesale agreements with Earthlink and other few ISPs (see Hausman and Sidak (2005) for other examples). "Open access" of cable networks gave rise to a heated debate around 2000.

[32]See also ERG (2005).

is regulated, in other countries (e.g., Austria or Finland) it is subject to commercial negotiation. Wholesale services might also be provided by (DSL) infrastructure-based CLECs. For instance, in France, in 2005, two DSL network operators, Cegetel and Neuf Telecom, owned between 30 and 50 percent of the national wholesale broadband market.[33]

The broadband market corresponds well to the assumptions of our setting; there is limited or no horizontal differentiation between wholesale broadband offers, whereas some degree of differentiation is observed in the retail market. Besides, consistently with our modeling, in some countries, pure downstream firms appear to purchase the wholesale product from different suppliers. For instance, this used to be the case for AOL in France.

Though we gave examples of wholesale broadband offers, neither in the US nor in the EU, there is evidence of a competitive wholesale market. This is consistent with our results; though a broadband wholesale market is likely to emerge, there are high chances that it will not be competitive.

Asymmetric regulation of the wholesale market. The broadband market provides an example of asymmetric regulation on the wholesale market. Indeed, in some countries, ILECs have been obliged to offer a bitstream access offer and/or this offer is regulated.

Is regulation likely to dominate the unregulated outcome? From another point of view, could deregulation lead to a “competitive” wholesale market? For instance, in France, the regulation of the ILEC’s wholesale broadband offer has been removed to favor the development of competition in the wholesale broadband market.[34]

It might be argued that a relatively low regulated price for the ILEC’s wholesale service hinders entry in the wholesale market, and hence, that deregulation could stimulate entry. Our setting suggests that this intuition is shortsighted. Indeed, if the price of the ILEC’s wholesale product is relatively high, infrastructure-based competitors have no incentives to enter in the wholesale market; they prefer to remain outside of this market. This is by setting a relatively low price for the ILEC’s wholesale product that the regulator can stimulate competition in the wholesale market.

Local loop unbundling and the wholesale broadband market. In our setting, we assumed that the vertically integrated firms have an infrastructure of their own. In the broadband market, however, some facility-based firms lease unbundled access to the local loop of the ILEC. What is the effect of the regulated unbundling rate on the market equilibrium? In our setting, it has a neutral role, since the total demand is assumed to be fixed. In other settings, the unbundling rate could introduce asymmetries between firms, which would modify the equilibrium in a non trivial way.

[33]See: DGCCRF, Decision C2005-44 related to the merger between Neuf Telecom and Cegetel.

[34]The French ILEC’s wholesale broadband offer was subject to the approval of the French regulatory authority (ARCEP) until Decision n°05-281 of 28 July 2005 removed the approval procedure.

5.2 The Mobile Market

Together with broadband the mobile telephony market is one of the most dynamic telecommunications market. In 2004, there were 184.7 million mobile users in the US and 383 million users in the EU. In the EU, the average penetration rate in the 25 Member States in 2005 was above 90 per cent.[35]

However, due to very high fixed (spectrum license and network) costs, entry in the mobile markets has been limited, and most national markets currently host only a few mobile network operators (MNOs). For instance, in most European countries, there are typically 3 to 4 MNOs offering 2G services.[36] But since 1999 a new class of service-based competitors has emerged, known as "mobile virtual network operators" (MVNOs).[37]

MVNOs do not have a spectrum license, and lease access to the networks of MNOs.[38] In general, MVNOs position themselves as "low cost" service providers (e.g., Easy Mobile in the UK) or target specific market segments (e.g., teenagers for Virgin Mobile in the UK or NRJ Mobile in France).

It is often argued that, though MVNOs represent a competitive threat for MNOs, MNOs have incentives to contract with MVNOs. The first motivation is to use spare wholesale capacity (in particular, following 3G investments) to generate additional wholesale revenues. The second motivation is to address market niches that the existing mobile firms do not reach very effectively (like teenagers or specific ethnic groups).

The wholesale mobile market. There are MVNOs in most national markets. According to Hazlett (2005), in 2005, there were around 20 MVNOs in the US. The two main MVNOs were TracFone which served over 3.5 million customers and Virgin Mobile USA which served over 3 million customers. In the European Union, there were a total of 214 MVNOs in 14 Member States out of 25 (EC, 2006b).

In most countries contracts between MNOs and MVNOs are commercial agreements, and access charges are typically set on a retail minus basis. In some countries (e.g., in Norway), mobile operators have been forced to give access to their networks to MVNOs. In some other countries (e.g., Hong Kong and Ireland), 3G licenses included the obligation to give access to MVNOs. Finally, in some countries (e.g., France and Ireland), regulators have been willing to regulate conditions of access for MVNOs, but with no success so far.[39]

The MVNO case fits well with our model. MNOs represent vertically-integrated firms, where MVNOs represent pure-downstream firms, which lease access to MNOs on a whole-

[35]See: FCC (2005b) and EC (2006a). As for the EU, there were 332 million mobile users in the 15 Members States and 383 in the 25 Member States.

[36]In 2005, there were 79 2G mobile operators in the EU (EC, 2006a).

[37]In 1999, the first MVNO, Virgin Mobile, entered in the UK market.

[38]Some MVNOs are merely resellers, hence do not have any infrastructure. Others (e.g., "full" MVNOs) own some mobile network elements, which gives them higher possibilities of differentiation.

[39]The Irish decision was eventually annulled while the French regulation project was rejected by the European Commission.

sale upstream market. Besides, the wholesale mobile services offered by MNOs are relatively identical, whereas there is differentiation on the downstream market.[40]

Should MNOs be obliged to open their networks? Currently, in most countries, MNOs are not obliged to give access to their networks to MVNOs. The rationale for not imposing an obligation is that, so far, MVNOs have been able to enter in most markets without regulatory intervention. Our model suggests that this position is correct; a wholesale mobile market can be expected to emerge in an unregulated environment. We point out that MNOs are willing to contract with MVNOs not only because it generates additional wholesale revenues, but also because it contributes to moderating competition on the retail market.

Competition in the wholesale mobile market. The structure of the wholesale mobile market varies across countries. In some countries, the wholesale market is almost monopolized. For instance, this is the case in Finland, Norway, the Netherlands and the UK, where the leading suppliers control more than 85 per cent of the wholesale mobile market. In other markets, there is some degree of competition in the wholesale market. For instance, in Denmark, the leading supplier, TDC Mobil, has 58 per cent of the wholesale market.[41] These different patterns of competition in the wholesale market are consistent with our analysis. We found that there are different potential market equilibria: some in which the wholesale market is monopolized; others in which a few firms may be active (though, in the case, the wholesale market is not necessarily competitive).

5.3 Implications for Regulation

Price cap. Though a wholesale market is likely to emerge without regulatory intervention, competition might fail to develop. One possibility to stimulate competition in the wholesale market would be to introduce a sufficiently low price cap on the price of the upstream good of at least one of the firms, and let the competitive process drive the wholesale prices to the average costs.

Indeed, a price cap, a_r, set at a level such that $a_r < \underline{a}$ implies that the Bertrand-like outcome becomes the unique equilibrium. If the price is such that $\underline{a} \leq a_r < a_m$, then the monopoly-like equilibria disappear; collusive-like equilibria in which firms share the wholesale market are preserved and are strictly preferred by the wholesale providers to the Bertrand-like equilibrium. Finally, if the cap is too high, that is when $a_r \geq a_m$, then regulation has almost no impact: only the equilibrium in which the incumbent exits the wholesale market does not survive.

[40] Though there is no horizontal differentiation on the wholesale market, there might be vertical differentiation;

[41] Source: David Spector, "What exactly is the Airtours retaliation criterion? A case study: the analysis of the French wholesale mobile market", Panel on collective dominance in mobile telephony ACE meeting, Copenhagen, December 1, 2005.

Structural separation. It is sometimes advocated that another solution to promote competition in telecommunications industry consists in separating vertically the incumbent operator. In that scenario, an upstream division independently sets the price of the wholesale offer without taking into account the impact on the downstream division, and reciprocally. In the broadband market, in some countries, this scenario has been observed.[42]

In our model, under vertical separation, the independent upstream unit has no direct interaction with the retail market and thus compete head-to-head on the wholesale market. Given the access price set by the other facility-based firm, the upstream unit always prefers to supply alone the wholesale market at that price since it is not affected by the intensity of competition on the retail market. As a result, only the Bertrand-like equilibrium can emerge under vertical separation of the incumbent.

Mergers. A merger between facility-based firms can affect the market equilibrium in a profound way. Consider the simplest case of two facility-based firms which would decide to merge. According to our setting, before the merger, the wholesale market is active; there are multiple equilibria, one of which is competitive. After the merger, a unique facility-based firm remains; a standard analysis suggests that it decides to foreclose the pure-downstream firm to protect its monopoly profits.

Now, consider that there are three facility-based firms and a unique pure downstream firm. It can shown that in this situation, only the competitive equilibrium is maintained. If two facility-based firms merge (and firms relocate), a new set of supra-competitive equilibria appears. Hence, following the merger, it is possible that the market switches from a competitive to a supra-competitive equilibrium.

This analysis highlights that a merger can not only affect the competition *per se*, but also through the change of market equilibrium.

The merger in 2005 of Neuf Telecom and Cegetel, two French infrastructure-based companies, provides a good illustration of this discussion. Before the merger, these two firms were competing in the broadband wholesale market with the ILEC, France Télécom (they had between 30 and 40 per cent of this market). The merger led to a reduction of the number of competitors in the wholesale broadband market from three to two. As a condition for the merger, the French Ministry of Economy required Neuf Telecom and Cegetel to maintain their activity in the wholesale broadband market for a period of three. Indeed, it was suspected that these two firms would exit the wholesale market after the merger otherwise.

[42] For instance, in France, some years ago, the ILEC's Internet service provider, Wanadoo, was independent from its parent company, France Télécom.

6 Conclusion

A Appendix

A.1 Proof of Lemma 1

Define $\tilde{p}_l = p_l - c_u - c_d$, $l \in \{1, 2, e\}$ and $\tilde{a}_i = a_i - c_u$, $i \in \{1, 2\}$; define accordingly the vector of margins $\tilde{p} \equiv (\tilde{p}_1, \tilde{p}_2, \tilde{p}_e)$. Consider that firm i is the unique provider on the upstream market.

Then simple manipulations lead to:

$$\begin{aligned}
\pi_i^{(i)}(\tilde{a}_i, \tilde{p}) &= \tilde{p}_i D_i(\tilde{a}_i, \tilde{p}) + \tilde{a}_i D_e(\tilde{a}_i, \tilde{p}),\\
\pi_j^{(i)}(\tilde{a}_i, \tilde{p}) &= \tilde{p}_j D_j(\tilde{a}_i, \tilde{p}),\\
\pi_e^{(i)}(\tilde{a}_i, \tilde{p}) &= (\tilde{p}_e - \tilde{a}_i) D_e(\tilde{a}_i, \tilde{p}).
\end{aligned}$$

Solving for the system of first-order conditions associated to the downstream price competition stage, we obtain:

$$\tilde{p}_e^{(i)} = \frac{t}{3} + \frac{7\tilde{a}_i}{10} \geq \tilde{p}_i^{(i)} = \frac{t}{3} + \frac{\tilde{a}_i}{2} \geq \tilde{p}_j^{(i)} = \frac{t}{3} + \frac{3\tilde{a}_i}{10}. \tag{1}$$

Profits are thus equal to:

$$\begin{aligned}
\pi_i^{(i)}(\tilde{a}_i) &= \tfrac{t}{9} + \tfrac{\tilde{a}_i(-30\tilde{a}_i + 50t)}{100t},\\
\pi_j^{(i)}(\tilde{a}_i) &= \tfrac{t}{9} + \tfrac{\tilde{a}_i(9\tilde{a}_i + 20t)}{100t},\\
\pi_e^{(i)}(\tilde{a}_i) &= \tfrac{t}{9} + \tfrac{\tilde{a}_i(9\tilde{a}_i - 20t)}{100t}.
\end{aligned} \tag{2}$$

The difference $\pi_i^{(i)}(\tilde{a}_i) - \pi_j^{(i)}(\tilde{a}_i)$ is a second order polynomial expression in $\tilde{a}_i$ which admits two roots: 0 and $\frac{10t}{13}$. This expression is positive (resp. negative) between (resp. outside) the roots. This gives $a_* \equiv c_u + c_d + \frac{10t}{13}$.

A.2 Proof of Lemma 2

Define $\tilde{a}_m = \arg\max_{\tilde{a}_i} \pi_i^{(i)}(\tilde{a}_i)$ (computed previously). Straightforward computations show that: $\tilde{a}_m = \frac{5t}{6}$; hence, $a_m = c_u + c_d + \frac{5t}{6} > a_*$.

A.3 Proof of Lemma 3

When the upstream market is supplied by both vertically integrated firms at common price a, profits are given by:

$$\begin{aligned}
\pi_i^{(1-2)}(\tilde{a}, \tilde{p}) &= \tilde{p}_i D_i(\tilde{a}, \tilde{p}) + \tfrac{\tilde{a}}{2} D_e(\tilde{a}, \tilde{p}),\\
\pi_j^{(1-2)}(\tilde{a}, \tilde{p}) &= \tilde{p}_j D_j(\tilde{a}, \tilde{p}) + \tfrac{\tilde{a}}{2} D_e(\tilde{a}, \tilde{p}),\\
\pi_e^{(1-2)}(\tilde{a}, \tilde{p}) &= (\tilde{p}_e - \tilde{a}) D_e(\tilde{a}, \tilde{p}),
\end{aligned}$$

where $\tilde{p}_l = p_l - c_u - c_d$, $l \in \{1, 2, e\}$, and $\tilde{a} = a - c_u$.

Solving for the system of first-order conditions associated to the downstream price competition stage, we obtain:

$$\tilde{p}_e^{(1-2)} = \frac{t}{3} + \frac{7\tilde{a}}{10} \geq \tilde{p}_1^{(1-2)} = \tilde{p}_2^{(1-2)} = \frac{t}{3} + \frac{2\tilde{a}}{5}. \tag{3}$$

Profits are thus equal to:

$$\begin{array}{l} \pi_1^{(1-2)}(\tilde{a}) = \pi_2^{(1-2)}(\tilde{a}) = \frac{t}{9} + \frac{\tilde{a}(-9\tilde{a}+35t)}{100t}, \\ \pi_e^{(1-2)}(\tilde{a}) = \frac{t}{9} + \frac{\tilde{a}(9\tilde{a}-20t)}{100t}. \end{array}$$

Comparing these profits with those obtained when the upstream market is served by one firm only (given by (2)) is straightforward and allows to prove the inequalities claimed in Lemma 3; this also gives $\underline{a} \equiv c_u + c_d + \frac{5t}{7}$ and $\overline{a} \equiv c_u + c_d + \frac{5t}{6} = a_m$.

A.4 Proof of Lemma 4

Consider the final prices given by (1) and (3). Given upstream prices a_1 and a_2, the final demand that faces firm e is equal to $\frac{1}{3} - \frac{3}{10t} \min\{a_1 - c_u, a_2 - c_u\}$ (whether a_1 and a_2 are equal or not). This shows that firm e is foreclosed if and only if $\min\{a_1, a_2\} \geq c_u + c_d + \frac{10t}{9} \equiv a_{\max}$. In this case, the final demand addressed to firm $i \in \{1, 2\}$ is $D_i^{(\emptyset)} = \frac{1}{2} + \frac{p_i - p_j}{t}$, $j \neq i \in \{1, 2\}$, and the profits are given by:

$$\pi_i^{(\emptyset)}(\tilde{p}_1, \tilde{p}_2) = \tilde{p}_i D_i^{(\emptyset)}(\tilde{p}_1, \tilde{p}_2),$$

where $\tilde{p}_i = p_i - c_u - c_d$.

Solving for the system formed by the first order conditions immediately leads to the corresponding equilibrium final margins: $\tilde{p}_1^{(\emptyset)} = \tilde{p}_2^{(\emptyset)} = \frac{t}{2}$. Profits are therefore equal to: $\pi_1^{(\emptyset)} = \pi_2^{(\emptyset)} = \frac{t}{4}$. This profit must be compared with the profit of a unique upstream supplier i at the upstream price a_m, which is equal to $\pi_i^{(i)}(a_m) = \frac{23t}{72}$ (see the proof of Lemma 1), and thus strictly lower than $\pi_i^{(\emptyset)}$.

B References

Fudenberg, D. and J. Tirole, 1984, "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look", *American Economic Review*, 74(2): pp. 361-366.

Jehiel, P. and B. Moldovanu, 1996, "Strategic Nonparticipation", *RAND Journal of Economics*, 27(1): pp. 84-98.

Diffusion of Electronic Business Technologies

Kai Sülzle*

Ifo Institute for Economic Research at the University of Munich & Dresden University of Technology

July 2006

Preliminary and incomplete draft version.
Please do not cite or circulate without author's permission

Abstract

This paper presents a duopoly model of technology diffusion that accounts for the specifics of electronic business technology. A leader's and adopter's profits from new technology usage depend on its development and implementation cost and both firms' adoption timing. The leader benefits from quick adoption by the follower while the latter prefers either late or no adoption. This is due to a delayed first-mover benefit, reflecting a standard setting capability of a pioneer in electronic business technology adoption. We further discuss interfirm adoption subsidies as viable tool to overcome delayed follower adoption.

JEL-classification numbers: O31, L1
Keywords: Diffusion, Network Effects, Innovation

*The author thanks participants of the Conference on the Economics of Information and Communication Technologies, Paris, June 2006 for helpful comments.
Address: Ifo Institute for Economic Research, Poschingerstr. 5, 81679 Munich, Germany
Phone: +49 (89) 9224 1282; Fax: +49 (89) 9224 1460; e-mail: suelzle@ifo.de

1 Introduction

The timing and nature of new technology adoption are fundamental issues of firms' business performance. In particular in the last decade, adoption of upcoming electronic business technologies played a crucial role as innovative enabler for professional activities and relationships. Although not all the evidence on technology adoption is conclusive, there is broad agreement on two points: First, new technologies are never adopted by all potential users simultaneously.[1] Second, the adoption decision crucially depends on the amount of improvement which the new technology offers over any previous technology, its development and implementation costs and the adoption decision of related firms. The latter can either be suppliers, customers, collaborators or competitors.

The present paper contributes to the theory of new technology adoption by focussing on the peculiarities of technology diffusion in a duopolistic framework. This happens, by introducing three specific effects which are especially observable in diffusion processes of electronic business technologies: First, a *standard setting effect* implies a first-mover advantage for the first firm in an industry that adopts to a new technology; second, the *follower effect* benefits second-movers since their cost of adoption are (usually) lower than for the first-mover; third, the *network effect* implies higher benefits of an applied technology, conditional on its current and future users. Dependent on the interplay of these effects, firms adoption timing implies either simultaneous early adoption, sequential adoption or no adoption. An additional focus of the analysis is on the commonly observable subsidization activities of technology leading firms, that benefit indirectly from the adoption of related business partners as for example suppliers. Firms maximize their profits from an innovation does not only depend on the innovation itself, but on the amount they produce, the costs they bear, and the price they can charge for their product. The decision to adopt an innovation is an investment decision that involves costs in the expectation of future - rather than immediate - rewards which is based on efficiency gains in their production activities. Firms are usually in competitive situations - their gains from an innovation depend on the behavior of their competitors and their costumers. However, consistent

[1] See e.g. Götz (1993).

with the literature on inter-firm diffusion of innovations, the adoption of a new technology is treated as an investment decision and an all-or-nothing event, abstracting from the level of utilization of the new technology within each firm. When the adoption of a new technology, e.g. e-business technology, requires large set-up costs, a firm's decision to adopt or not depends on the comparison of gains and losses, associated with the use and installation of the new technology.[2]

Related Literature: While there are many industry-specific and innovation-specific case studies of the adoption of new technologies, the theoretical literature on the adoption of electronic business activities is sparse. Hoppe (2002) and Geroski (2000) provide excellent surveys on both the theoretical literature on new technology adoption and patent races.

Most theoretical contributions have a common base in the seminal work by Reinganum (1981) who provides a duopoly model of technology adoption. In her model a change in market concentration may speed or slow technology adoption if firms make once-and-for-all commitments to their eventual adoption dates. Fudenberg and Tirole (1985) extend this work by studying situations in which firms decide at any point in time whether to adopt a cost-reducing new technology, knowing that adoption costs decline over time. By assumption, the increase in profits due to innovation is greater for the first follower than for the second. This potential first-mover advantage stimulates preemption up to a point where the extra flow profit for the first mover just equals the extra costs of speeding up adoption.

Götz (1993) analyzes the adoption and diffusion of a new technology in a market for a differentiated product with monopolistic competition, showing a positive relationship between firm size and speed of adoption. Additionally, he identifies a rank effect, stating that potential users differ with respect to the (expected) returns from adoption. Further, a stock effect implies a dependency of firms' adoption payoffs on the stock of firms already using a new technology.

[2]See Chen (1996).

Such stock effects imply an asymmetry in payoffs of adoptions, which give rise to differing rather than uniform adoption dates in both market and planner solutions.

One of the first contributions dealing with technology adoption in the presence of network externalities offer Katz and Shapiro (1986). They study the dynamics of industry evolution in a market with technological change where two inherently incompatible technologies are subject to network externalities. They show that a potential second-mover advantage may result in subgame-perfect equilibria without preemption and payoff equalization. In their model, payoffs to different firms are asymmetric. They further state that in such a setup, network effects are a crucial feature. They dispose over two fundamental effects: first, the relative attractiveness today of rival technologies is influenced by their sales histories: a given product is more attractive the larger is the in-place base of consumer using that product. Second, in the presence of network externalities, a consumer in the market today also cares about the future success of the competing products.

The closest related contribution is Benoit (1985). In a duopoly model he derives that a technology leader's expected profits from innovation are not monotonic in the cost of innovation, given that successful innovation is probabilistic. He further shows that an increase in the innovation cost may cause followers to adopt to the innovation.

The remainder of the paper is structured as follows: Section 2 introduces the basic model. Section 3 analyzes its outcome. In section 4 we modify the basic model to the e-business case. Section 5 discusses the possibility of interfirm innovation cost subsidies as an extension. Section 6 concludes the paper.

2 The basic Model

The basic framework builds on the duopoly model by Benoit (1985). Consider two firms using the current best-practice technology. The two firms are not necessarily competitors but could also be vertically related players in a value chain or industry.[3] Both firms decide upon innovation and adoption of a new electronic business technology (in the following: e-business technology), dependent on the expected benefits from using the new technology, its implementation cost and the adoption behavior of the other firm.

The first firm, which will be labeled L as *leader*, has the know-how and resources for the innovation of a new e-business technology. It faces the decision, whether to innovate and implement the new technology or not.[4] When L chooses to develop the new technology, it incurs a fixed cost $F(\geq 0)$ and the new technology will be implemented or take effect after $N(\geq 0)$ periods.

The second firm is labeled A as *adopter*. If L does not develop a new e-business technology, A does not get informed about that. Contrarily, if L decides to innovate and implement this new technology, A becomes aware of that only when the e-business technology takes effect. This happens N periods after L makes it investment in the development of the new technology. The follower then has the following possibilities of technology adoption: it can match the new e-business technology at a cost F with an implementation lag N or decide not to adopt the new technology.[5]

In the following, we exclusively deal with benefit resulting from technology usage, hence we do not look at any product market in a later stage. The initial benefit of technology use, i.e. when both firms apply the old technology is normalized to 0. During any single period in which L has innovated but not

[3]Note, this papers focusses exclusively on the costs and benefits associated with the innovation and adoption of a new technology. An analysis of a downstream product market is not an issue.

[4]L could be interpreted as technology leader in an industry or value chain. This does not necessarily mean that this is a dominant firm.

[5]Initially we assume, that the cost F for technology implementation and development are the same for both firms. This could be justified that L has higher costs for product development while A incurs higher implementation costs for e.g. staff schooling, since the technology is not produced in house. We will relax this assumption later on.

(yet) the adopter A, the leader L earns $\Pi_L(1, X)$ and A earns $\Pi_A(1, X)$, with X being a random variable. If both firms have implemented the e-business technology, L earns $\Pi_L(2, X)$ while A earns $\Pi_A(1, X)$.[6] To keep things simple, we assume that X can take one of two values: with probability P it takes the value x_s and with probability $(1 - P)$ it takes the value x_f. x_f stands for "e-business technology is a failure" and x_s for "e-business technology is a success". Both firms are risk-neutral and maximize the expected present values of their profits with the discount rate $\delta \in (0; 1)$.

The leader prefers a successful innovation to a failed innovation, and a successful innovation to no innovation (before considering fixed costs). Since there is a rivalry between the two firms, if the innovation is "successful", L is better off in any single period when it is the only firm that uses the new e-business technology. Hence, the according profits denote as:

$$\begin{aligned} \Pi_L(1, x_s) &> \Pi_L(1, x_f) \\ \Pi_L(1, x_s) &> 0 \\ \Pi_L(1, x_s) &> \Pi_L(2, x_s) \end{aligned} \tag{1}$$

If the new e-business technology is a success, A is better off also adopting than not to match the new technology (before considering fixed costs). If A also adopts, it is better off, when the new technology is a success than a failure. Given that A has not adopted, it is better off if the innovation is a failure than if it is a success. These assumption formally read as:

$$\begin{aligned} \Pi_A(2, x_s) &> \Pi_A(1, x_s) \\ \Pi_A(2, x_s) &> \Pi_A(2, x_f) \\ \Pi_A(1, x_s) &\leq \Pi_A(1, x_f) \end{aligned} \tag{2}$$

Once a firm has implemented the new technology, there is no way to reverse this decision, i.e. the firm stays with the new technology in any case.[7] It is

[6] As Benoit (1985) we do not specify the process from which these payoffs arise since there is no need to do so. The only requirement is that this process is relatively stable, so that when innovation and adoption occurs, the situation changes in a predictable way.

[7] Technically this is $\Pi_L(1, x_f) \geq 0, \Pi_A(2, x_f) \geq \Pi_A(1, x_f)$. Practically this is that the added cost of reversing the investment is prohibitive.

further assumed that $K \leq N$ periods after the first firm (L) adopts the new technology, the realization of X is revealed to both firms. Before these K periods, the firms receive no new information on X.[8] Accordingly, $\tilde{\Pi}_L(1, X)$ and $\tilde{\Pi}_A(1, X)$ denote the single period profits before the end of the K periods, when firms learn the realization of X. The following analysis of the model builds on the mutual best responses of the two firms, based on the respective present values, associated with their decision to innovate/adopt or not. Figure 1 below gives a graphical illustration of the possible time structures, where t denotes the time periods.

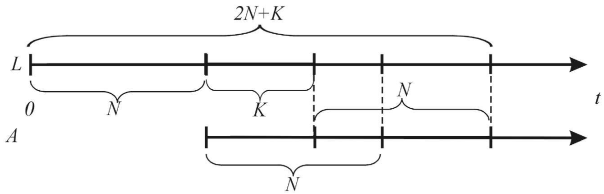


Figure 1: Possible time structures of the game

2.1 The Adopter's Decision Problem

Given that L innovates and implements the new e-business technology, the adopter A has three choices: never adopt the new technology, wait for K periods and then adopt, or adopt immediately. In the latter case A immediately incurs a fixed cost F and the new technology will be implemented after N periods. If A waits for K periods, it can determine afterwards whether the innovation is successful or not and then innovate if it is. Accordingly, the leader L has to build an expectation about the adopter's response when deciding whether to innovate or not. It is assumed that L correctly assess A's beliefs about the innovation of the e-business technology. By backwards induction, we determine the subgame perfect equilibrium of the game where L decides upon innovation or not and A decides upon its response. The notation for present

[8] See Benoit (1985) for a justification of these assumptions.

values of the corresponding payoffs are depicted in table 1 below. We only depict those present values for the case that (at least) L innovates since in the case where no firm adopts the new technology we assume the corresponding payoffs from adoption to be zero.

		"adopter"		
		never adopt	adopt immediately	wait K periods
		$V_{i;n}^{A}$	$V_{i;a}^{A}$	$V_{i;w}^{A}$
"leader"	innovate	$V_{i;n}^{L}$	$V_{i;a}^{L}$	$V_{i;w}^{L}$

Table 1: Notation and allocation of payoff present values

We first analyze A's decision problem in order to determine the corresponding present values.

2.1.1 A never responds

If A decides to never respond to an innovation by L, the present value of this strategy is given by:

$$V_{i;n}^{A} \equiv \sum_{i=0}^{K-1} \delta^i E\left[\tilde{\Pi}_A(1,X)\right] + \sum_{i=K}^{\infty} \delta^i E\left[\mathit{\Pi}_A(1,X)\right], \tag{3}$$

where $E[\ \ ]$ is the expected value operator. During the first K periods after the implementation of the new e-business technology by L, A receives the expected single period payoff $E\left[\tilde{\Pi}_A(1,X)\right]$ with L using the new technology alone. After K periods, both firms learn the realization of X, but since A did (and does) not adopt, the adopter receives the expected payoff single period payoff $\Pi_A(1,X)$ in every subsequent period (see figure 2 below).

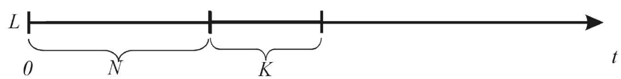


Figure 2: Timing, when A never adopts

2.1.2 A adopts immediately

If A adopts immediately after it notices that L has innovated (i.e. N periods after L decided to innovate), the corresponding present value is defined as:

$$V_{i;a}^{A} \equiv -F + \sum_{i=0}^{K-1} \delta^i E\left[\tilde{\Pi}_A(1,X)\right] + \sum_{i=K}^{N-1} \delta^i E\left[\mathit{\Pi}_A(1,X)\right] + \sum_{i=N}^{\infty} \delta^i E\left[\mathit{\Pi}_A(2,X)\right]. \quad (4)$$

The fixed cost F has to be incurred immediately with the decision to adopt. Again, during the first K periods after the implementation of the new e-business technology by L, A receives the expected single period payoff $E\left[\tilde{\Pi}_A(1,X)\right]$ with L using the new technology alone. After K periods, both firms learn the realization of X, and A receives the expected payoff single period payoff $\Pi_A(1,X)$ until A's new e-business technology is also implemented (which happens N periods after the adoption decision). Afterwards, when both firms apply the new technology, A receives the payoff $\Pi_A(2,X)$ in every period that follows (see figure 3 for an illustration of the timing from A's perspective.).

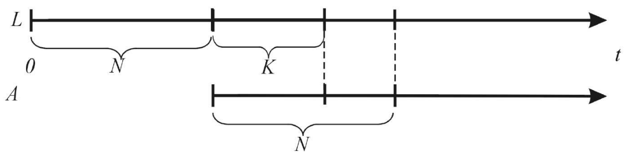


Figure 3: Timing, when A adopts immediately

2.1.3 A waits K periods till adoption

Finally, A can choose to wait K periods (until the type of X is revealed) and then adopt (or not in case that the innovation is a failure). The present value of this strategy is given by:

$$\begin{aligned} V_{i;w}^{A} \equiv & \sum_{i=0}^{K-1} \delta^i E\left[\tilde{\Pi}_A(1,X)\right] + (1-P)\sum_{i=K}^{\infty} \delta^i \mathit{\Pi}_A(1,x_f) \\ & +P\left(-\delta^K F + \sum_{i=K}^{N+K-1} \delta^i \mathit{\Pi}_A(1,x_s) + \sum_{i=N+K}^{\infty} \delta^i \mathit{\Pi}_A(2,x_s)\right) \quad (5) \end{aligned}$$

Once again, A receives the expected single period payoff $E\left[\tilde{\Pi}_A(1,X)\right]$ in the first K periods after the implementation by L. Then, if the new technology is a failure - which happens with probability $(1-P)$ - A does not adopt the new technology and hence receives $\Pi_A(1,x_f)$ in every following period afterwards. Otherwise, if the new technology is a success, A chooses to adopt and incurs the present value of the fixed costs F. Further, A receives $\Pi_A(1,x_s)$ as long as the implementation of the new technology did not occur yet. After the implementation (i.e. N periods after A decided to adopt), A receives $\Pi_A(2,x_s)$ in all future periods (see figure 4 below).

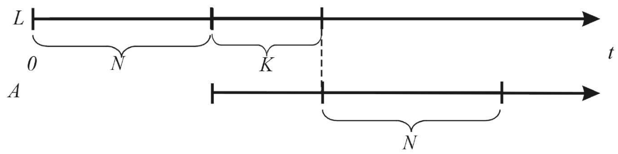


Figure 4: Timing, when A waits for K periods with adoption decision

The adopter's choice depends upon which of the above expressions is the greatest. It will never respond, iff:

$$F > \frac{\delta^N \left(E\left[\mathit{\Pi}_A\left(1,X\right)\right] - E\left[\mathit{\Pi}_A\left(2,X\right)\right]\right)}{\delta - 1} \equiv \tilde{F} \tag{6}$$

Accordingly, if (6) does not hold, the follower will wait K periods, if

$$\begin{aligned} F\left(P\delta^K - 1\right) \quad > \quad & \frac{\delta^N}{\delta - 1} E\left[\mathit{\Pi}_A\left(1,X\right) - \mathit{\Pi}_A\left(2,X\right)\right] + \frac{\delta^{N+K}}{\delta - 1} P\langle \mathit{\Pi}_A(1,x_s) - \mathit{\Pi}_A(2,x_s)\rangle \\ & + \frac{\delta^K}{\delta - 1}(E\left[\mathit{\Pi}_A(1,X)\right] - \langle P\mathit{\Pi}_A(1,x_s) + (1-P)(\mathit{\Pi}_A(1,x_f)\rangle), \end{aligned}$$

which reduces to

$$\begin{aligned} F(1 - P\delta^K) \quad > \quad & \frac{\delta^N}{1-\delta} E\left[\mathit{\Pi}_A\left(2,X\right) - \mathit{\Pi}_A\left(1,X\right)\right] \\ & - \frac{\delta^{N+K}}{1-\delta} P\left(\mathit{\Pi}_A(2,x_s) - \mathit{\Pi}_A(1,x_s)\right) \equiv F'(1 - P\delta^K), \end{aligned} \tag{7}$$

since $E\left[\mathit{\Pi}_A(1,X)\right] = P\mathit{\Pi}_A(1,x_s) + (1-P)(\mathit{\Pi}_A(1,x_f)$. Otherwise, if (7) does not hold, the follower prefers to adopt immediately instead of waiting K periods.

2.2 The Leader's Decision Problem

Given the choice by the follower, the leader has to choose between developing and implementing the e-business technology or not. In calculating the respective profits, L takes into account conditions (6) and (7).

2.2.1 L's response if A never adopts

If A never adopts, L will innovate, if:

$$V_{i;n}^{L} \equiv -F + \sum_{i=N}^{N+K-1} \delta^i E\left[\tilde{\Pi}_L(1,X)\right] + \sum_{i=N+K}^{\infty} \delta^i E\left[\Pi_L(1,X)\right] \quad > 0 \tag{8}$$

When L innovates and A never adopts, it incurs fixed costs F and receives the expected single period payoff $E\left[\tilde{\Pi}_L(1,X)\right]$ until the type of X is revealed (which happens after $N+K$ periods). Afterwards, L (as only user of the new technology) receives $E\left[\Pi_L(1,X)\right]$ in every subsequent period (a graphical representation gives figure 1 above). Hence, let us define the corresponding crucial fixed cost value

$$F_n \equiv \frac{E\left[\tilde{\Pi}_L(1,X)\right]\left(\delta^{N+K}-\delta^N\right)}{\delta-1} - \frac{E\left[\Pi_L(1,X)\right]\delta^{N+K}}{\delta-1}. \tag{9}$$

2.2.2 L's response if A adopts immediately

In the case that A chooses to adopt immediately, L will innovate, iff:

$$\begin{aligned} V_{i;a}^{L} \equiv\ & -F + \sum_{i=N}^{N+K-1} \delta^i E\left[\tilde{\Pi}_L(1,X)\right] + \sum_{i=N+K}^{2N-1} \delta^i E\left[\Pi_L(1,X)\right] \\ & + \sum_{i=2N}^{\infty} \delta^i E\left[\Pi_L(2,X)\right]) \quad > 0 \end{aligned} \tag{10}$$

Again, in case of innovation, L it incurs fixed costs F. In the time interval between the point in time when the new e-business is implemented at L and the realization of X is revealed (in $N+K$), the leader receives $E\left[\tilde{\Pi}_L(1,X)\right]$ per period. Now, between the revelation of X and the implementation of the new technology at A, L still is the only user of the new technology and hence receives a per period payoff $E\left[\Pi_L(1,X)\right]$. Figure 5 illustrates the timing from

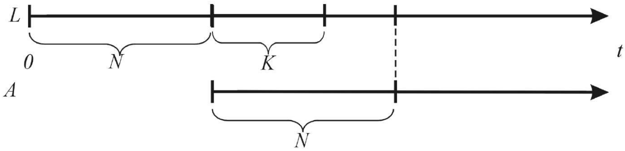


Figure 5: Timing, when adopts immediately

L's perspective. Accordingly, in this case we define the crucial fixed cost value for the innovation decision as

$$F_a \equiv \frac{E\left[\tilde{\Pi}\left(1,X\right)\right]\delta^{N+K}}{\delta-1} - \frac{E\left[\tilde{\Pi}\left(1,X\right)\right])\delta^{N}}{\delta-1} + \frac{E\left[\Pi_1\left(1,X\right)\right]\delta^{2\,N}}{\delta-1} - \frac{E\left[\Pi_1\left(1,X\right)\right]\delta^{N+K}}{\delta-1} - \frac{E\left[\Pi_1\left(2,X\right)\right]\delta^{2\,N}}{\delta-1}. \tag{11}$$

2.2.3 L's response if A waits K periods to adopt

If A chooses to wait K periods till she adopts, L will innovate, iff:

$$\begin{aligned} V^L_{i;w} \equiv & -F + \sum_{i=N}^{N+K-1} \delta^i E\left[\tilde{\Pi}_L\left(1,X\right)\right]) + (1-P)\sum_{i=N+K}^{\infty} \delta^i \Pi_L\left(1,x_f\right) \\ & + P\left(\sum_{i=N+K}^{2\,N+K-1} \delta^i \Pi_L\left(1,x_s\right) + \sum_{i=2\,N+K}^{\infty} \delta^i \Pi_L\left(2,x_s\right)\right) > 0 \end{aligned} \tag{12}$$

The intuition for the payoff till period $N+K$ is identical to the case of A adopting immediately. Since A now waits K periods to decide wether to adopt or not, A will learn which X-type the technology provides. Accordingly, if X is of type x_f, A will not adopt and therefore L receives the per period payoff $\Pi_L\left(1,x_f\right)$ in all future periods. Otherwise, if the new technology is successful (with probability P), A will adopt and L subsequently receives $\sum_{i=N+K}^{2\,N+K-1} \delta^i \Pi_L\left(1,x_s\right) + \sum_{i=2\,N+K}^{\infty} \delta^i \Pi_L\left(2,x_s\right)$ as per period payoff.

Again, we define the corresponding crucial fixed costs as

$$\begin{aligned} F_k \equiv & \frac{\delta^N - \delta^{N+K}}{1-\delta} E\left[\tilde{\Pi}_L\left(1,X\right)\right] + P\frac{\delta^{2N+K}}{1-\delta}\Pi_L\left(2,x_s\right) \\ & + P\frac{\delta^{N+K} - \delta^{2N+K}}{1-\delta}\Pi_L\left(1,x_s\right) + (1-P)\frac{\delta^{N+K}}{1-\delta}\Pi_L\left(1,x_f\right). \end{aligned} \tag{13}$$

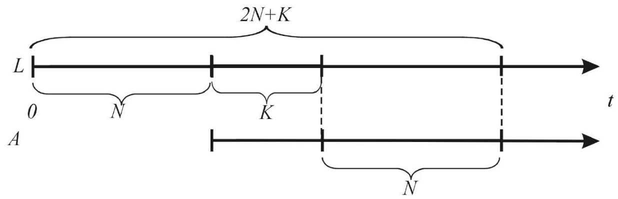


Figure 6: L's timing when A waits for K periods with adoption decision

3 Outcome of the basic Model

The approach above completely characterizes the conditions for innovation/adoption, together with the respective resulting payoffs. Depending on the specific parameter values, many innovation patterns are possible. With the basic assumptions upon the parameter values made above, the following results hold.

Proposition 1 *(see Benoit, 1985) L's expected present value of innovation is lower if A waits for K periods compared to never responding to an innovation by L. Further, L is better off when A waits for K periods than if A adopts immediately, given that:* $P(1-\delta^K)(\Pi_L(1,x_s)-\Pi_L(2,x_s))+(1-P)(\Pi_L(1,x_f)-\Pi_L(2,x_f)) > 0$.

Proof. The first statement of proposition 1 follows from (8) and (12). The second part is derived from (12) together with (10). It holds since we know from (1) that $\Pi_L(1,x_s) > \Pi_L(2,x_s)$. Then, given that $\Pi_L(1,x_f) > \Pi_L(2,x_f)$ (or if P is high enough), the condition in proposition 1 holds. ■

Benoit (1985) points out that in such a scenario L's expected profits increase, when the associated fixed cost F increase from a little below of $\tilde{F}$ to a little above that threshold value (which induces A to switch its response). Hence, in this fixed cost range, L prefers a more costly development of the new technology to a cheaper one, given that the adopter's cost increases as well. When L's payoff might increase due to increasing fixed costs, the overall outcome might change from no innovation to innovation. For simplicity, assume that in the following

$$\tilde{\Pi}_L(1,X) = \mathit{\Pi}_L(1,X). \tag{14}$$

Let us suppose, that

$$\begin{aligned} E[\Pi_L(1,X)] > \Pi_A(2,X) + \Pi_A(1,X) \quad > \quad & E[\Pi_L(1,X)] - P\delta^{N+K}[\mathit{\Pi}_L(1,X) \\ & -\mathit{\Pi}_L(2,X)], \end{aligned} \tag{15}$$

so that the following holds:

$$F_n > \tilde{F} > F_k. \tag{16}$$

Now, if F lies a little below $\tilde{F}$, A would wait before innovating and L would not innovate. When F increases such that $F \in (F_n; \tilde{F})$, A would never respond to an innovation and L would innovate. Therefore the increase in the fixed cost leads to more innovation. If further the condition in proposition 1 holds, then if:

$$\begin{aligned} H \quad \equiv \quad & E[\Pi_A(2,X) - \Pi_A(1,X)] + \delta^N E[\Pi_L(1,X)] - \delta^N E[\Pi_L(2,X)] \\ & -P\delta^K[\mathit{\Pi}_A(2,x_S) - \Pi_A(1,x_s) - E[\Pi_L(1,X)] + \delta^N E[\Pi_L(1,X)] \\ & -\delta^N E[\Pi_L(2,X)]] > 0, \end{aligned} \tag{17}$$

and

$$\begin{aligned} & H - (1-\delta^K P)[\delta^N E[\Pi_L(1,X) \\ & -\Pi_L(2,X)] - \delta^{N+k} P[\Pi_L(1,x_s) - \Pi_L(2,x_s)] < 0, \end{aligned} \tag{18}$$

then

$$F_k > F' > F_a. \tag{19}$$

If $F \in (F', F_a)$, no innovation occurs, but when F rises, such that $F \in (F_k, F')$, innovation occurs, such that the following result holds:

Proposition 2 *(see Benoit, 1985) An increase in F might lead to innovation that would not have been conducted with lower fixed costs.*

4 Diffusion of e-Business Technologies

The results from above crucially depend on the assumptions upon the parameter values made in (1) and (2). While these assumptions might hold in traditional innovation scenarios, they have to be adjusted in the context of the diffusion of electronic business technologies. Some important peculiarities of e-business technologies here are that followers benefit from late adoption since they face lower R&D costs for the development of e.g. an e-business software. Due to this peculiarity, pioneering firms usually face lower payoffs from technology use during the time of implementation. Further, such leaders in e-business adoption scenarios usually have lower benefits from technology usage, when they are the only user of a new e-business technology. This is due to network effects, the more or the earlier other firms adopt, the higher is the benefit of an applied technology to its innovator, due to the fact, that the first innovator sets the technology standard. In this regard, we have to modify the underlying assumptions as follows:

$$\begin{aligned} \Pi_L(1, x_s) &> \Pi_L(1, x_f) \\ \Pi_L(1, x_s) &> 0 \\ \Pi_L(1, x_s) &< \Pi_L(2, x_s) \end{aligned} \tag{20}$$

The leader prefers a successful innovation to a failed innovation, and a successful innovation to no innovation (before considering fixed costs). Due to the argumentation above, L is worse off in any single period when it is the only firm that uses the new e-business technology. Hence, we assume that $\Pi_L(1, x_s) < \Pi_L(2, x_s)$. Accordingly, the adopters profits follow the relationship below:

$$\begin{aligned} \Pi_A(2, x_s) &> \Pi_A(1, x_s) \\ \Pi_A(2, x_s) &< \Pi_A(2, x_f) \\ \Pi_A(1, x_s) &> \Pi_A(1, x_f) \end{aligned} \tag{21}$$

If the new e-business technology enables the leader to impose its standard, A is better off also adopting than not to match the new technology (before considering fixed costs). If A also adopts, it is worse off, when the new technology enables the leader to impose its standard than not. Given that A has not adopted, it is better off if the innovation is a failure than if it enables the leader to impose its standard. Due to this changes in the assumption with regard to the peculiarities of e-business, we get the following result that reverses proposition 1:

Proposition 3 *In the context of e-business, L's expected present value of innovation is higher if A waits for K periods compared to never responding to an innovation by L. Further, L is better off when A adopts immediately than if A waits for K periods.*

Proof. Again, the statement of proposition 3 follows from (8) and (12) together with (10). Since the underlying assumptions concerning the payoff structure are now (20) and (21), it now holds that $\Pi_L(1, x_s) < \Pi_L(2, x_s)$. Hence the opposite of proposition 1 holds. ■

Since the changed assumption also impact the respective threshold values of the fixed cost, the argument of proposition 2 is reversed as well.

Proposition 4 *An increase in F might hinder innovation activities that would have been conducted with lower fixed costs.*

Proof. The changed assumptions (20) and (21) impact on (18) such that

$$\begin{aligned} & H' - (1 - \delta^K P)[\delta^N E[\Pi_L (1, X) \\ & - \Pi_L (2, X)] - \delta^{N+k} P[\Pi_L (1, x_s) - \Pi_L (2, x_s)] > 0, \end{aligned} \tag{22}$$

since now $\Pi_L (1, x_s) < \Pi_L (2, x_s)$. Then i follows that

$$F_a > F' > F_k, \tag{23}$$

leading to the outcome in proposition 4, due to the same proceeding as in the approach to proposition 2. ■

5 Extension: Interfirm Subsidization

From the results in the last section, we can deduct, that in the context of electronic business, the leader would like the adopter to follow earlier than the adopter would like to. This is due to the standard setting network effect on the one hand side and due to high adoption cost in case of the follower. Therefore, in e-business relationships, pioneering firms often sponsor or subsidize the adoption by suppliers, customers and even competitors. It is in their interest to convince their business partners to apply their standard and to adopt the new technologies. This happens quite often e.g. by providing training, consulting or even financial support for new hard- and software investments. We introduce the possibility of subsidization by the parameter s into the duopolistic setup. This subsidy is financed by L and benefits A. Accordingly, the present values now read as

$$V_{i;a}^{A*} \equiv V_{i;a}^{A} + sF \qquad V_{i;a}^{L*} \equiv V_{i;a}^{L} - sF, \tag{24}$$

if A adopts immediately. And if A chooses to wait for K more periods, the present values now read as

$$V_{i;k}^{A*} \equiv V_{i;k}^{A} + \delta^k sF \qquad V_{i;a}^{L*} \equiv V_{i;a}^{L} - \delta^k sF. \tag{25}$$

It can be easily seen that there is a transfer possibility among the firms via the payment of s. Therefore, the leader has a tool to set an incentive for the adopter to adopt earlier than otherwise intended, which is in line with propositions 3 and 5. The following proposition summarizes this finding:

Proposition 5 *When there is the possibility of interfirm transfers, there exist parameter constellations, such that*
(a) L sponsors A so to adopt immediately instead of waiting for K periods.
(b) L sponsors A so to waiting for K periods instead of never adopting.
(c) no innovation occurs despite that L would like to sponsor A to quicken adoption.

Proof. To be completed. ■

6 Conclusion

The timing and nature of the adoption of a e-business technologies and associated processes are fundamental issues of firms' business performance in the information society. The present paper accounts for the specifics of the diffusion process of electronic business technologies which are determined the amount of fixed costs for development and implementation of a new technologies. Further, in the context of electronic business the *standard setting* capabilities of a pioneering firm (might) imply a delayed first-mover advantage. On the other hand, if the development costs are high a the *follower effect* benefits second-movers since their cost of adoption are lower than the fixed cost for e.g. R&D by the leader. Finally, *network effects* imply higher benefits of an applied technology, conditional on its current and future users. Dependent on the interplay of these forces, firms adoption timing implies either immediate adoption by the following firm, sequential adoption or no adoption, either just by the follower or by both firms.

The model at hand could be extended in various ways: one obvious extension could be the analysis in an oligopolistic setup instead of the 2-firm case. Nonetheless, the results deducted above dispose over some predictive forces for such a scenario as well. An innovator of a new technology would still dispose over standard setting capacities associated with fixed development costs. But in an oligopolistic scenario, it could be the case that, the network effect that results from an increase in the number of adopting firms somehow has an upper ceiling in terms of the total number of adopters, if we think of potential network congestion in terms of administrative and service costs.

Another interesting extension would be to endogeneously determine the number of periods the adopter would optimally choose to adopt as well as the optimal number of periods the leader would want the adopter to follow. Obviously, this extension would require some notational clarification of the interpretation of the parameter K. In the present setup, K determines the time till the true value of the innovation is revealed and simultaneously the time the adopter would wait in case that she doesn't immediately adopt. This extension will be taken up in future research.

References

Benoit, J.-P. (1985): "Innovation and Imitation in a Duopoly," *The Review of Economic Studies*, 52(1), 99–106.

Chen, Z. (1996): "New Technology, Subsidies, and Competitive Advantage," *Southern Economic Journal*, 63(1), 124–139.

Fudenberg, D., and J. Tirole (1985): "Preemption and Rent Equalization in the Adoption of New Technology," *Review of Economic Studies*, 52, 383–401.

Geroski, P. (2000): "Models of technology diffusion," *Research Policy*, 29, 603–625.

Götz, G. (1993): "Monopolistic Competition and the Diffusion of New Technology," *RAND Journal of Economics*, 30(4), 679–693.

Hoppe, H. (2002): "The Timing of New Technology Adoption: Theoretical Models and Empirical Evidence," *The Manchester School*, 70(1), 56–76.

Katz, M., and C. Shapiro (1986): "Technology Adoption in the Presence of Network Externalities," *The Journal of Political Economy*, 94(4), 822–841.

Reinganum, J. (1981): "Market Structure and the Diffusion of New Technology," *The Bell Journal of Economics*, 12(2), 618–624.

Are Digital Rights Valuable? Theory and Evidence from eBook Pricing[1]

Gal Oestreicher-Singer, Arun Sundararajan

Working Paper CeDER-06-01, Center for Digital Economy Research
Leonard N. Stern School of Business, New York University

This version: January 2006

Abstract: The effective management of digital rights is the central challenge in many industries making the transition from physical to digital products. We present a new model that characterizes the value of these digital rights when products are sold both embedded in tangible physical artifacts, and as pure digital goods, and when granting rights permitted by one's digital rights management (DRM) platform may affect the extent of digital piracy. Our model indicates that in the absence of piracy, digital rights should be unrestricted, since a seller can use its pricing strategy to optimally balance sales between physical and digital goods. However, the threat of piracy limits the extent to which digital rights should be granted: the value of digital rights is determined not only by their direct effect on the quality of legal digital goods, but by a differential piracy effect that can lower a seller's pricing power. When the latter effect is sufficiently high, granting digital rights can have a detrimental effect on value – our model indicates that this kind of effect is more likely to be observed for digital rights that aim to replicate the consumption experience of physical goods, rather than enhancing a customer's digital experience. We test the predictions of our analytical model using data from the ebook industry. Our empirical evidence supports our theoretical results, showing that four separate digital rights each have an economically significant impact on ebook prices, and establishing that the digital rights which aim to replicate physical consumption while increasing the threat of piracy are the ones that have negative impact on seller value. We also show that if the pricing of a digital good is keyed off that of an existing tangible good, optimal pricing changes for the former should be more nuanced, rather than simply mirroring changes in the price of the latter, and we discuss the effect of the technological sophistication of potential customers on optimal pricing and rights management. Our results represent new evidence of the importance of an informed and judicious choice of the different digital rights granted by a DRM platform, and provide a new framework for guiding managers in industries that are progressively being digitized.

[1]We thank Bin Gu, Ravi Mantena, Foster Provost, Jeffrey Simonoff, Yossi Spiegel, and Manuel Trajtenberg for feedback on earlier drafts and/or helpful discussions, and seminar participants at Ben-Gurion University, Haifa University, the Technion, Tel-Aviv University, the University of Connecticut, the University of Michigan and the International Conference on Information Systems for their comments. A precursor to this paper, titled "Are Digital Rights Valuable? Theory and Evidence from the eBook Industry" won the award for best overall paper at the 25th International Conference on Information Systems.

1. Introduction

This paper studies how granting digital rights influences the quality of legal and pirated digital goods, and how this affects their value to buyers accustomed to consuming these goods as physical artifacts. We present a new model of how the demand for physical, digital and pirated goods are related to pricing and rights management choices in industries being transformed by digitization, and we subsequently validate predictions from this model using data from the ebook industry.

The importance of the specific questions addressed by this paper is illustrated well by the ongoing transformation of a number of industries whose products (music, video, graphic art, magazines, newspapers, books), historically embedded in tangible artifacts, are now increasingly available as pure digital goods. These digital goods have a number of attractive properties for sellers. They can be produced at a small fraction of the cost of producing their tangible counterparts. They can be delivered easily over the Internet, often allowing sellers to bypass costly intermediaries that limit the geographic reach of their sales. They may include electronic features that enhance their users' rendering experience. Many issues related to the economics, pricing and delivery of these digital goods have been recognized and analyzed by prior research in information systems.

However, the process of digitization in these industries has been gradual. A majority of these digital goods continue to be purchased as tangible artifacts, and their pure digital counterparts are sometimes viewed as niche products. A number of factors influence the pace of this transition. The business models of the dominant firms in some of these industries are predicated on their superior production/distribution of the tangible artifact, and these firms may thus believe that a rapid shift to digital goods could threaten their dominance. If consumers are not technologically sophisticated, they are likely to view digital products as being of lower value, and possibly as inferior substitutes for the existing physical products they are accustomed to using. Furthermore, some goods are intrinsically more easily "digitizable": they lend themselves to digitization more naturally than others, on account of the nature of the physical artifact the good is embedded in. For instance, music on a CD, while viewed as a good embedded in a physical artifact, is already digital, and thus more easily converted into a pure digital good than, say, the contents of a physical book. Finally, and perhaps most importantly, digital goods are subject to piracy, which is widespread in many industries, and which affects demand for both nascent digital goods, as well as the corresponding established physical goods (music being the most visible recent example, but not the only one).

The uncertain and potentially incomplete process of the transition towards digitizing products presents sellers with difficult technological and business decisions. These relate crucially to the strategic control of the quality of digital goods and the extent to which the presence of these digital goods leads to piracy. Both of these are currently implemented through a seller's management of the *digital rights* associated with their digital goods. The technological challenges of managing digital rights are gradually being addressed by the emergence of viable industry-specific platforms for digital rights management (DRM). A different range of rights is technologically feasible under each DRM platform, and the extent of such flexibility, along with the secureness or strength of the DRM system, are important considerations when choosing a platform. However, technological feasibility and robustness aside, it is also imperative that managers *grant* the different digital rights permitted by their DRM platforms in a judicious and informed way. This is because an insufficient level of digital rights can result in premature failure of a digital initiative, as illustrated by early emusic services (Mossberg, 2003). On the other hand, granting digital rights that are too extensive raises concerns about prematurely cannibalizing established sales of physical goods[2]. More importantly, granting certain digital rights may increase the prevalence and quality of pirated substitutes for the good. For instance, the right to download (rather than stream) digital audio files increases the desirability of purchasing legal digital music, while simultaneously increasing the threat of digital piracy over file-sharing networks. The right to print a digital book encourages ebook adoption by customers used to reading printed pages, but enables the creation of pirated PDF copies of the book. The right to a backup of a digital movie on a DVD adds value to legal downloaders, but facilitates illegal secondary sales. Since different digital rights contribute differentially towards increasing the quality of legal digital goods, and towards facilitating piracy, a careful assessment of

[2]Similar issues have been recognized in the context of channel conflict caused by Web-based channels for, among other businesses, online retailing (Viswanathan, 2005) and online brokerage (Konana et al., 2000).

their relative contribution towards each is important.

To summarize, the appropriate choice of digital rights is influenced by many technological and business factors that include:

1. The relative extent to which each digital right a seller may grant buyers contributes towards an increase in the value of the digital good and the corresponding increase in the prevalence of piracy. The latter is also influenced by the strength of the industry's DRM platform.

2. The extent to which these increases affect the current revenues generated by the sale of physical goods, and the anticipated revenues from sales of digital goods.

3. The technological sophistication of potential customers, which determines what fraction of them place a relatively high value on digital goods (both legal and pirated).

4. The extent to which the physical good is "digitizable", which influences how closely digital goods are assessed as substitutes for their tangible counterparts.

We capture each of these factors in our model of sellers who choose digital rights and prices for their digital goods, while simultaneously selling corresponding physical goods at pre-specified price levels. The extent to which each digital right is granted affects the quality of the digital good, and may affect the quality of a (free) pirated substitute. Customers vary in the increases in value they enjoy on account of the granting of digital rights. Additionally, they vary in their tastes for physical and digital goods, captured by the difference in value they ascribe to a digital good and physical good of equal quality. We characterize the appropriate pricing and digital rights choices of a seller, both in the absence and presence of piracy. The results of our analytical model indicate, among other things, that:

1. The price of a digital good is increasing in the price of its corresponding physical counterpart. However, the prevalence of piracy dampens the extent to which changes in tangible good prices are mirrored in the pricing of their digital counterparts.

2. In the absence of a threat of piracy, the price of a digital good is increasing in the level of each associated digital right, and a seller should always choose to grant the highest level of rights

permitted by its DRM platform, since the potential threat of cannibalization can be effectively managed through the strategic control of price.

3. The extent to which granting digital rights affects a seller's pricing power is determined by the interplay between a direct quality effect, proportional to the increase in the quality of the legal digital good, and a differential piracy effect, proportionate to the (weighted) increase in the quality of the pirated good. If these two effects balance, then granting the right is profitable, and should increase the seller's optimal price. However, if the latter effect dominates the former substantially, granting digital rights can actually be associated with a decrease in the price of the legal digital good.

4. During early stages of the transition from tangible to digital goods, when the level of technological sophistication of a seller's consumer base is low, an increase in the level of technological sophistication may be associated with an increase in the price of digital goods relative to their tangible counterparts, although this effect is limited by the threat of piracy as a market becomes progressively more technologically sophisticated.

Our analytical results relate the prices of digital goods to the prices of their physical counterparts, the extent to which digital rights are granted to customers, the relative demand for each kind of good, and the technological sophistication of the product's target customer base. We test these results using a data set containing pricing and digital rights data for over 3,000 ebook titles, across six categories, and sold by a specialized ebook retailer. There are many reasons why the ebook industry is an excellent setting to test our theory. The digital rights associated with each ebook are determined individually by publishers rather than by a major retailer, and thus there is substantial variation in the level to which rights are granted across ebooks (the summary of variables in Table 3.1 in section 3 illustrates this well). There is also a substantial variation in prices across ebooks, and significant demand for both tangible and digital versions of a wide variety of books.

The equations we estimate are hedonic pricing indices that relate an ebook's price to the price of its tangible counterpart, the digital rights it grants, and the book's category. Our empirical findings provide strong support for the predictions of our theoretical model, showing that tangible book prices and each of four digital rights have a significant impact on ebook prices, and that digital rights explain a significant fraction of variation in ebook prices. More importantly, our results indicate that specific digital rights are associated with a significant *increase* in the threat of piracy and a corresponding reduction in a seller's pricing power, while other digital rights result in a net increase in the value a seller can derive from its legal digital goods. We also distinguish between these rights based on an intuitive categorization of them into those that aim to replicate the physical consumption experience, and those that enhance the digital consumption experience, showing that granting rights in the former set is likely to have a detrimental effect on seller value, while granting rights in the latter set is valuable.

Our paper adds to a growing literature that aims to guide managers facing piracy and rights management challenges in digital industries. Prior work has studied alternative approaches to the strategic control of digital piracy, through the strategic choice of either quality, or of pricing. Papers which model the deterrence of piracy by varying quality generally base their analysis on Mussa and Rosen's (1978) model of vertical product differentiation, and the idea is that a seller might produce a lower-priced degraded substitute for its flagship good so that piracy is either reduced or even eliminated. For instance, Poddar (2005) shows that the quality and reliability of the pirated product are among three attributes that determine whether permitting piracy is profitable or not. Alvisi and Carbonara (2002) find that in the presence of piracy and heterogenous copying costs, such differentiation arises as the optimal strategy for a monopolist to deter piracy, by diverting consumers from the pirated good to the original one. By offering legal digital products of varying quality, the monopolist can effectively discriminate between those consumers with higher and lower copying costs. Two other models that explore ideas closely related to the ones described above are those by Snir (2003), and by Wu et al. (2003). Belleflamme (2003) studies the interdependence between different producers' incentives to accommodate/deter the presence of a pirated good. The relative welfare benefits of legal and technological deterrence are studied by Png and Chen (2003) and by Takeyama (1994).

In our model, we build on the approach of many of these prior papers, preserving their notion of the pirated good as an inferior (vertically differentiated) substitute for the legal good, but expanding on it by simultaneously modeling the demand for a tangible good horizontally differentiated from its (legal and pirated) digital counterparts. The seller chooses quality by explicitly varying the digital rights it grants buyers of its legal digital good. Further, we allow the seller to use pricing in addition to digital rights management as an instrument to deter piracy. A more involved study of the strategic choice of pricing in the presence of piracy has been done by Sundararajan (2004), who shows that while in the absence of price discrimination, a seller's optimal level of protection against piracy is at the technologically maximal level, a seller who can price discriminate always chooses a strictly lower level of protection. Since his model considers only pure digital goods, we extend its results by examining the effect of having a physical substitute in addition to a pirated version, and we provide evidence for the need to restrict certain digital rights even in the absence of price discrimination.

A related stream of literature has explored the benefits of piracy to a seller. A common argument is that it increases profits for a product that displays network effects, an idea first articulated by Connor and Rummelt (1991), and discussed by many others, including Bakos et el (1999) in the context of file sharing systems ("clubs") and Shy and Thisse (1999) in a duopoly model. Prasad and Mahajan (2003) and Haruvy et el. (2004) showed that piracy is not necessarily harmful to a firm seeking to launch a new product, since it establishes an initial user base, and speeds up product diffusion. The question they then address is how much piracy should be tolerated. Gu and Mahajan (2004) show that since pirated goods are found most attractive by those customers who are most price sensitive, its presence can actually reduce the intensity of price competition, by "removing" these consumers from the market, and thereby benefiting competing providers of legal digital goods. A different benefit of piracy can be that it implicitly provides free "samples" of a product to consumers who are not perfectly informed about its characteristics, thus informing them about the value of the product, and influencing legal demand (Peitz and Waelbroeck, 2003). Chellappa and Shivendu (2005) extend this idea to model sampling and strategic pricing in the presence of a pirated good. Gayer and Shy (2003) explore a related idea in the context of digital goods enhancing the sales of tangible goods. In contrast with this line of research, we focus on the central detrimental aspect of piracy: that it generates a free substitute for legally available physical and digital goods.

Our survey of the literature indicates that the relationship between piracy, digital rights management and the strategic control of quality has not been studied empirically, and our paper thus represents the first such study. The empirical studies about digital piracy we are aware of aim primarily to estimate the effect of piracy on the demand for the legal products and on profit (for instance, Zenter, 2003; Hui and Png, 2003). An interesting prior empirical study of software piracy was by Gopal and Sanders (1998), who establish that the strength with which a government enforces its digital intellectual property laws is positively related to the robustness of the domestic software industry, an observation that is likely to generalize to other digital industries like music and video as well.

We have organized the rest of this paper as follows. Section 2 models the legal demand for a seller's tangible and digital goods, and describes the relationship between the prices of legal digital goods, their digital rights and the prices of their tangible counterparts, both in the absence of piracy and in its presence. The predictions of this model are tested in Section 3, which describes our data set, presents our empirical models, and describes the results of their estimation. Section 4 concludes with a summary of our results, some limitations of our study, and directions for future research.

2. Theory: DRM, piracy and pricing strategy

2.1. Overview of model

We model a monopoly seller who may produce two versions of a product: a physical (*tangible*) good and a *digital* good. These goods are imperfect substitutes. The quality of the tangible good s_T is exogenous, and normalized to 1. The quality s_D of the digital good is determined by the level to which the seller grants each of a set of n digital rights $\mathbf{r} = (r_1, r_2, ..., r_n)$ to its buyers:

$$s_D = S_D(r_1, r_2, ..., r_n). \tag{2.1}$$

Examples of digital rights for different digital products include the right to print an ebook, the right to backup a downloaded video file, the right to play an MP3 file on a portable device, and the right to lend an ebook. The range of feasible values for each digital right r_i is determined by an exogenously specified DRM platform used by the seller, and each right therefore has pre-specified minimum and maximum levels. For instance, if using Adobe's DRM platform for ebooks, a seller can vary the right a buyer has to print pages of an ebook by allowing unlimited printing (the highest level), allowing no printing whatsoever (the lowest level), or specifying a fixed number of pages that can be printed during each time period (an intermediate level; for example, upto 35 pages every two weeks).

Granting a buyer a higher level of digital rights increases her value from the digital good, and therefore, S_D is (weakly) increasing in each of its arguments. While the set of possible values for some digital rights may be discrete, we instead assume that each r_i can vary continuously between its minimum and maximum levels. This assumption is for analytical convenience, though relaxing it is unlikely to directionally alter our results. Altering the level of digital rights within the constraints of its DRM platform imposes no direct cost on the seller, and consequently, changes to the quality s_D of the digital good have no associated variable cost. The physical good has a constant variable cost c per unit sold; the digital good has a constant variable cost of zero.

In addition to the legal physical and digital versions of the good, there may be a *pirated* digital version available. The quality s_P of the pirated good is also determined by the level to which the seller grants the same set of n digital rights $\mathbf{r} = (r_1, r_2, ..., r_n)$ to buyers of its legal digital good:

$$s_P = S_P(r_1, r_2, ..., r_n). \tag{2.2}$$

Since an increase in the level of rights granted to legal users often facilitates the creation of higher-quality pirated versions[3], S_P is (weakly) increasing in each of its arguments. The price of the tangible good is denoted p_T, and the price of the digital good is denoted p_D; the pirated good is free.

Consumers are heterogeneous along two dimensions. The first dimension, indexed by a *taste* variable $y \in [0,1]$ represents the consumer's relative preferences for tangible versus digital goods. Holding everything else constant, a consumer indexed by a higher value of y places a higher value on a digital good and a lower value on a physical good than a consumer indexed by a lower value of y places on the same two goods. The second dimension, indexed by a *digital type* variable $\theta \in [0,1]$ represents a consumer's preferences for digital quality, or how much value a customer ascribes to an increase in digital rights. Each customer of type θ always prefers higher digital quality (or the granting of more digital rights) to lower; however, all else being equal, a customer with a higher θ always values a digital good of fixed quality more than a customer with a lower θ.

The seller does not know the taste y or type θ of any specific consumer, but knows the distribution of y and θ in the population of potential consumers. The distribution function of y is denoted $F_h(y)$, and the distribution function of θ is denoted $F_v(\theta)$. We assume that these distributions are absolutely continuous and have corresponding density functions $f_h(\theta)$ and $f_v(\theta)$. For simplicity, we also assume that these distributions are independent.

We refer to the physical version as the *tangible* good (whose associated variables have the subscript T), the legal digital version as the *digital* good (whose associated variables have the subscript D), and the pirated digital version as the *pirated* good (whose associated variables have the subscript P). Customer preferences are linearly separable in value and price, and across digital type and taste, specified using a common utility function

$$U(s, \theta, y) = u(s, \theta) - w(y),$$

where the linearly separable form is chosen for analytical convenience. The surplus a consumer gets

[3] For instance, the right to download rather than stream digital audio files increasing the threat of digital piracy over file-sharing networks, the right to print a digital book enables the creation of pirated PDF copies of the book.

from his or her purchase of the digital good is:

$$\text{Digital: } u(s_D, \theta) - w([1-y]) - p_D, \tag{2.3}$$

the surplus a consumer gets from his or her use of the free pirated good is

$$\text{Pirated: } u(s_P, \theta) - w([1-y]), \tag{2.4}$$

and the surplus a consumer gets from his or her purchase of the tangible good is

$$\text{Tangible: } u(1,1) - w(y) - p_T. \tag{2.5}$$

We use a common underlying utility function in order to maintain consistency in the variation of a customer's preferences across the three goods[4], since these goods are imperfect substitutes. Equations (2.3) and (2.4) reflect a choice of modeling a pirated good as an inferior (vertically differentiated) substitute for the legal digital good, which is consistent with a subset of the prior literature we discuss in Section 1. Equation (2.5) indicates that we model consumers with the same taste y as being *homogeneous* in their preferences for the quality of the tangible good. We do this because we wish to focus our analysis on the interplay between cannibalization of the sales of the tangible good, the threat of digital piracy and the granting of digital rights. Introducing an additional customer characteristic that models heterogeneity in preferences for tangible goods will complicate things unnecessarily. Anchoring the "common type" of customers with respect to their preferences for the quality of the tangible good at 1 simply reflects an assumption that all customers are used to using tangible goods, and this forms the benchmark for their assessment of the value of the digital good. One might think of this as benchmarking the "intrinsic quality" of the good and treating the quality of the digital and pirated goods as being measured relative to this benchmark. This is especially true for the industry (books) in which we test our theory's predictions.

Customers use exactly one version of the good, purchasing either the tangible good or the digital

[4] For those familiar with models of imperfect competition, our model includes aspects of both horizontal and vertical differentiation. The tangible good is horizontally differentiated from both the digital/pirated goods, and the digital good is vertically differentiated from the pirated good. The subscripts of the distribution functions h (horizontal) and v (vertical) are labeled accordingly for this reason.

good, or using the free pirated good. We do not choose a specific functional form for u or w, but assume they have the following properties:

1. $\frac{du}{ds} > 0, \frac{du}{d\theta} > 0, \frac{dw}{dy} > 0.$

2. $u(s, 0) \leq w(0)$: consumers of digital type 0 derive no value from digital goods at any quality level, or some customers are just not interested in purchasing digital goods.

3. $u(1, 1) - w(1) > p_T$: the market is fully covered.

4. $\frac{d^2u}{dsd\theta} > 0$: consumers of higher digital type value the same incremental increase in quality more than consumers of lower digital type.

Assumption 1 is consistent with how we have described the variation in preferences with type θ and location[5] y. The (average) slope of $w(y)$ could indicate the extent to which the good is "digitizable", by determining how closely consumers on average assess tangible and digital goods as substitutes (for example, if $w'(y) = 0$, a tangible and digital good of equal quality are perfect substitutes). Assumption 3 posits that in the absence of digital goods, all consumers purchase the tangible good. This is towards focusing our analysis on a scenario in which consumers switch from consuming a tangible good to consuming its digital counterpart (as the industry in question transforms from being "physical" to being digital), and the associated issues of cannibalization and piracy, rather than a scenario in which the digital good expands the seller's market. We discuss the implications of relaxing this assumption in Section 4. Assumption 4 is the (standard) Spence-Mirrlees condition.

The sequence of events is as follows: the seller announces their choice of r, p_T and p_D, consumers make their choices, and each party realizes its payoffs.

[5] The results of our model are unchanged if we assume that u increases in y and alter equations (2.3) through (2.5) accordingly. We have made this choice to be consistent with how utility varies with location in standard models of horizontal differentiation with "misfit" or "transportation" costs (for instance, the Hotelling model).

2.2. Choosing digital rights in the absence of piracy

Our first set of results set a benchmark by specifying the optimal choice of digital rights and pricing in the absence of piracy. Therefore, in this subsection,

$$S_P(\mathbf{r}) = 0, \tag{2.6}$$

and customers choose between the tangible good and the digital good. The price and quality of the tangible good are exogenous, and the seller chooses the level of each of its digital rights and the price of the digital good. Given these choices, demand can be characterized as follows. For those customers with location y, define $\theta_{TD}(y, p_D)$ as the type indifferent between the tangible good and the digital good when the price of the digital good is p_D. For each y, $\theta_{TD}(y, p_D)$ is thus defined as follows:

$$u(1,1) - w(y) - p_T = u(s_D, \theta_{TD}(y, p_D)) - w([1-y]) - p_D. \tag{2.7}$$

A customer with location and type (y,θ) prefers the tangible good to the digital good if $\theta < \theta_{TD}(y, p_D)$, and prefers the digital good to the tangible good if $\theta > \theta_{TD}(y, p_D)$. Under assumptions 2 and 3, it follows that

$$u(1,1) - w(0) - p_T > u(s_D, 0) - w(1) - p_D, \tag{2.8}$$

and therefore, realized demand is as summarized in Figure 2.1.

The demand for the tangible good is:

$$q_T(p_D) = F_H(y_D) + Q(y_D, 1, \theta_{TD}(y)), \tag{2.9}$$

where $y_D(p_D)$ is the taste value below which all consumers purchase the tangible good:

$$y_{TD}(p_D) = y : \theta_{TD}(y, p_D) = 1, \tag{2.10}$$

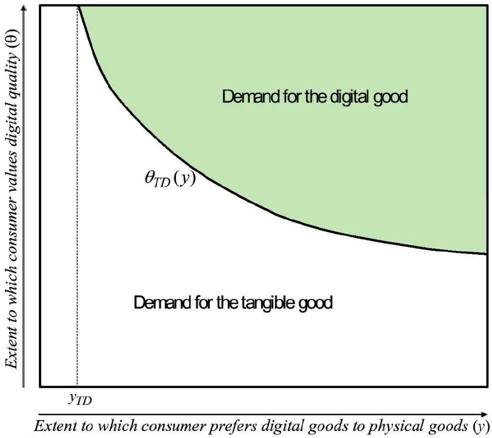


Figure 2.1: Summarizes demand for the tangible and digital good in the absence of piracy. The curve $\theta_{TD}(y)$ represents those consumers indifferent between the two goods.

and

$$Q(y_a, y_b, g(y)) \equiv \int_{y=y_a}^{y_b} \int_{\theta=0}^{g(y)} f_v(\theta) f_h(y) dy \tag{2.11}$$

is the mass of customers over the area under an arbitrary curve $\theta = g(y)$, between y_a and y_b. Correspondingly, the demand for the digital good is

$$q_D(p_D) = 1 - F_H(y_{TD}(p_D)) - Q(y_{TD}(p_D), 1, \theta_{TD}(y, p_D)). \tag{2.12}$$

Evidently, $\theta_{TD}(y)$ and y_{TD} also depends on p_T and s_D, though, to keep our equations more readable, we do not include these as arguments[6].

Based on this characterization of demand, we can derive an expression for the optimal price of the digital good, and describe corresponding optimal extent to which the seller should grant its digital rights. This analysis, which is presented in Appendix A, leads to our first two proposition:

Proposition 1. *A unit increase in the price of the tangible good leads to a corresponding unit*

[6] A parallel case involves parameter values such that $\theta_{TD}(0) < 1$. Each of its results are directionally similar to those presented in this section.

increase in the price of the digital good.

Proposition 2. *In the absence of piracy, the seller grants its customers the maximum possible level of rights permitted by its DRM platform.*

The intuition underlying the result of Proposition 2 is quite simple. In the absence of piracy, the seller can address any threat of cannibalization as effectively as is necessary through its strategic control of the price of the digital good. Its profits increase with an increase in the quality of the digital good, and therefore, it chooses digital rights to maximize the quality of its digital good. This result is especially stark because it holds when the market is fully covered; clearly, it will generalize to a model in which market expansion is possible through the introduction of the digital good, since a seller's incentives to increase s_D are stronger if it were able to expand its market in addition.

Empirically, however, we observe that sellers frequently restrict the digital rights they grant buyers. Thus, even when demand is modeled quite generally, with no restrictions on the ease with which a good can be digitized, or on the technological sophistication of the target consumer base, a model which ignores piracy yields results that seem to contradict what is observed in practice. In our next subsection, we return to the model described in section 2.1, and establish how the presence of piracy alters these predictions about pricing and rights management for digital goods.

2.3. Pricing and rights management in the presence of piracy

We now return to the complete model described in section 2.1. Granting a level of digital rights $(r_1, r_2, ..., r_n)$ to each buyer of the digital good also induces the creation of a free pirated good of quality $s_P = S_P(r_1, r_2, ..., r_n) > 0$. Buyers choose between three imperfect substitutes: the tangible good, the digital good and the pirated good. Characterizing demand completely will therefore involve analyzing three pairwise comparisons made by customers.

The first pairwise comparison, between the tangible good and the digital good, yields the functions $\theta_{TD}(y, p_D)$ and $y_{TD}(p_D)$ that are defined in (2.7) and (2.10). Next, for those customers with taste y, define $\theta_{TP}(y)$ as the customer type that is indifferent between the tangible good and the pirated good:

$$u(1,1) - w(y) - p_T = u(s_P, \theta_{TP}(y)) - w([1-y]). \tag{2.13}$$

Notice that $\theta_{TP}(y)$ does not depend on p_D, the price of the digital good. A customer with taste and type (y,θ) prefers the tangible good to the pirated good if $\theta < \theta_{TP}(y)$, and prefers the pirated good to the tangible good if $\theta > \theta_{TP}(y)$. The taste value y_{TP} below which consumers of all types $\theta \in [0,1]$ prefer the tangible good to the pirated good is defined by:

$$y_{TP} = y : \theta_{TP}(y) = 1. \tag{2.14}$$

If $\theta_{TP}(1) > \theta_{TD}(1)$, then all relevant consumers (that is, those who will not purchase the tangible good when a digital or pirated substitute is available) prefer the digital good to the pirated good, and the analysis proceeds as in Section 2.2. Similarly, if $y_{TP} < y_{TD}(p_D)$, then all relevant consumers prefer the pirated good to the digital good, and this scenario will therefore not occur, since the seller can raise its profits by simply not offering a digital good. Therefore, the only scenario of economic relevance is when the following conditions hold:

$$\theta_{TP}(1) \quad < \quad \theta_{TD}(1, p_D), \tag{2.15}$$

$$y_{TP} \quad > \quad y_{TD}(p_D), \tag{2.16}$$

and under these conditions, there is a unique value of y at which $\theta_{TP}(y) = \theta_{TD}(y, p_D)$, which we label $y_{PD}(p_D)$.

$$y_{PD}(p_D) = y : \theta_{TP}(y) = \theta_{TD}(y, p_D) \tag{2.17}$$

Clearly,

$$y_{TD}(p_D) < y_{TP} < y_{PD}(p_D), \tag{2.18}$$

and therefore, demand of each good is as depicted in Figure 2.2.

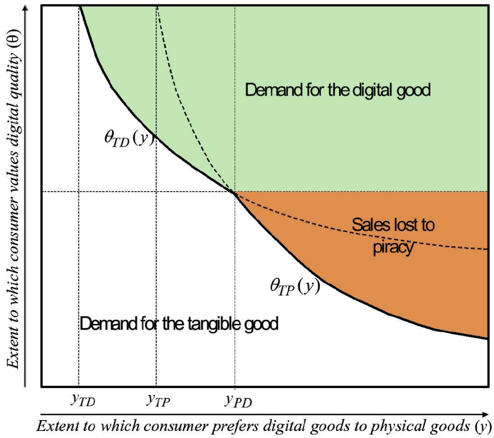


Figure 2.2: Illustrates demand for the tangible and digital goods in the presence of piracy. The curve $\theta_{TD}(y)$ defines those consumers indifferent between the tangible and digital goods, and curve $\theta_{TP}(y)$ defines those consumers indifferent between the tangible and pirated goods.

The demand for the tangible good is:

$$q_T(p_D) = F_h(y_{TD}(p_D)) + Q(y_{TD}(p_D), y_{PD}(p_D), \theta_{TD}(y, p_D)) + Q(y_{PD}(p_D), 1, \theta_{TP}(y)), \quad (2.19)$$

where the function Q was defined in (2.11), the demand for the digital good is:

$$\begin{aligned} q_D(p_D) &= F_h(y_{PD}(p_D)) - F_h(y_{TD}(p_D)) - Q(y_{TD}(p_D), y_{PD}(p_D), \theta_{TD}(y, p_D)) \quad (2.20) \\ &+ [1 - F_h(y_{PD}(p_D)][1 - F_v(\theta_{TD}(y_{PD}(p_D), p_D))] \end{aligned}$$

and the demand for the pirated good is

$$q_P(p_D) = [1 - F_h(y_{PD}(p_D))][F_v(\theta_{TD}(y_{PD}(p_D), p_D))] - Q(y_{PD}(p_D), 1, \theta_{TP}(y)) \quad (2.21)$$

This is summarized in Figure 2.2. Based on this characterization of demand, the optimal price for the digital good can be characterized. This characterization leads to our next proposition:

Proposition 3. *The optimal price of the digital good is of the form:*

$$p_D^* = [p_T - c]\delta(p_D^*) + \frac{q_D(p_D^*)}{-\left[\frac{dq_D(p_D^*)}{dp_D}\right]}, \tag{2.22}$$

where $\delta(p_D^*) < 1$. Therefore, a change in the price of the tangible good causes a less than proportionate change in the price of the digital good

The intuition for the result of Proposition 3 is illustrated in Figure 2.3. An increase in the price of the digital good shifts the curve $\theta_{TD}(y, p_D)$ upwards, reflecting an increase in the fraction of consumers who now prefer the tangible good to the digital good. The relative fractions of consumers who prefer the tangible good over the pirated good are not affected, since the curve $\theta_{TP}(y)$ does not shift. There are corresponding shifts in y_{TD} (to the right) and y_{PD} (to the left). The upward shift of $\theta_{TD}(y, p_D)$ therefore reduces demand for the digital good in two ways. First, it causes a shift in demand from the digital good to the tangible good, illustrated by the green (left) shaded area. In addition, it causes a shift in demand from the digital good to the pirated good, as illustrated by the red (right) shaded area. The discount factor $\delta(p_D^*)$, the ratio of the change in demand for the tangible good caused by this shift to the change in demand for the digital good, is therefore always less than 1 in the presence of piracy.

Proposition 3 thus indicates that a seller who offers both digital and tangible goods must approach pricing changes for the former with care. Changes in the price of the digital good must be more nuanced, rather than simply mirroring changes in the price of the tangible good. We return to this observation when discussing our empirical findings in Section 4.

Next, consider the effect of a change in any right r_i on the price of the digital good. Notice that since

$$\frac{dp_D^*}{dr_i} = \frac{dp_D^*}{ds_D}\frac{dS_D}{dr_i} + \frac{dp_D^*}{ds_P}\frac{dS_P}{dr_i}, \tag{2.23}$$

it follows that

$$\frac{dp_D^*}{dr_i} \propto \frac{dS_D}{dr_i} - \left[\frac{-dp_D^*/ds_P}{dp_D^*/ds_D}\right]\frac{dS_P}{dr_i}. \tag{2.24}$$

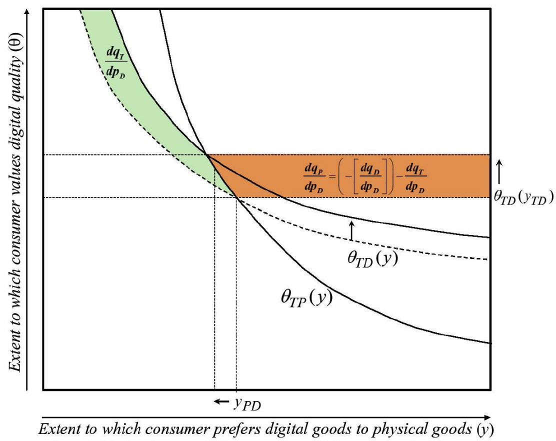


Figure 2.3: Illustrates the shift in demand from a small increase in the price p_D of the digital good. The upward shift of the curve $\theta_{TD}(y)$ reduces demand for the digital good in two ways. Part of this demand shift is captured by the tangible good (the green shaded area on the left) while the rest is lost to piracy (the red shaded area on the right). Notice that the curve $\theta_{TP}(y)$ remains unchanged by an increase in p_D, since the relative surplus from the tangible and pirated goods do not change.

(2.24) reveals that there are two separate effects that increasing the extent to which a digital right is granted has on the price of the digital good. The first is a *direct quality effect* (represented by the term $\frac{dS_D}{dr_i}$) which is positive, and proportionate to the extent to which granting the additional rights will increase the quality of the legal digital good. The second is a *differential piracy effect* (represented by the term $\left[\frac{-dp_D^*/ds_P}{dp_D^*/ds_D}\right]\frac{dS_P}{dr_i}$) which is negative, and proportionate to the extent to which granting the additional rights will increase the quality of the pirated good. Since an increase in the quality of the digital good has a positive impact on price, while an increase in the quality of the pirated good has a negative impact on price, (2.24) indicates that an increase in the extent to which a digital right is granted does not always increase the price of the digital good, or the seller's total profits, for that matter. This is in contrast with the result of Proposition 2, and an (unstated) related result, which indicates that in the absence of piracy, increasing the level to which a right is granted always increases p_D.

Furthermore, since the pirated good is an (imperfect) substitute for the legal digital good, a unit increase in the quality of the legal digital good is likely to cause a higher increase in its price than a corresponding unit decrease in the quality of the pirated good, or, it is likely that

$$\frac{dp_D^*}{ds_D} > -\left(\frac{dp_D^*}{ds_P}\right) \tag{2.25}$$

Together, (2.24) and (2.25) suggest that if granting a digital right causes an equal increase in the quality of the digital good and the quality of the pirated good, or a higher increase in the quality of the digital good than the quality of the pirated good, granting this right will be associated with an increase in the price of the digital good, since the direct quality effect dominates the differential piracy effect. In contrast, if the quality of the pirated good increases substantially more than the quality of the legal digital good on account of granting a specific right, or more precisely, if

$$\frac{ds_P}{dr_i} > \left[\frac{-dp_D^*/ds_D}{dp_D^*/ds_P}\right]\frac{ds_D}{dr_i},$$

then the differential piracy effect dominates the direct quality effect, and granting of this digital right will be associated with a *decrease* in the price of the digital good. Again, we return to this observation when interpreting our empirical results in Section 4, towards discussing the kinds of rights that are likely to fall into each category.

Finally, consider the effect of a change in the shape of the distribution of y. Suppose we index the distribution of y by a parameter α such that $F_h(y;\alpha_1)$ first-order stochastically dominates $F_h(y;\alpha_2)$ if $\alpha_1 > \alpha_2$, or

$$\alpha_1 > \alpha_2 \Rightarrow F_h(y;\alpha_2) > F_h(y;\alpha_1) \text{ for all } 0 < y < 1.$$

One might consider a market corresponding to a larger α more *technologically sophisticated* than a market with a lower value of α, since there is a higher fraction of consumers with values of y closer to 1 (recall that a consumer with a higher value of y places relatively more value on the digital good). This is illustrated in Figure 4.

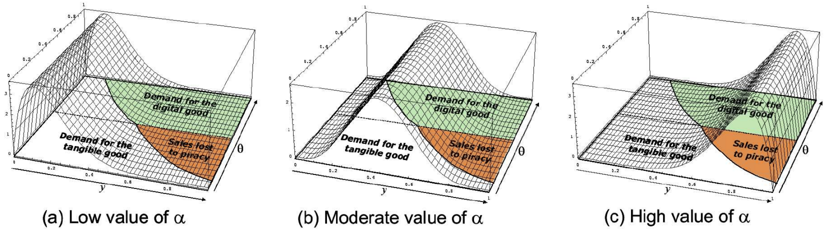


Figure 2.4: Illustrates how the mass of consumers shifts over the parameter space as a market becomes more technologically sophisticated. The two horizontal axes represent consumer taste (y) and digital type (θ), and the vertical axis represents the distribution of taste $f_h(y;\alpha)$, with a higher value of α parametrizing a distributional shift to the right.

It therefore seems natural to conclude that an increase in the technological sophistication of a market would result in a higher demand for digital goods. However, the effect of such an increase (in technological sophistication) on price is not straightforward to infer. This is because the same shift that causes an increase in the potential demand for the digital good would also cause an increase in the potential demand for the pirated good. Qualitatively, if an increase in α causes the distribution of y to move from being very right-skewed to being somewhat balanced (for instance, like a move from (a) to (b) in Figure 4), one might expect this shift to cause an increase in the price of the digital good: the primary shift in consumer mass is towards a region in which there is some potential demand for the digital good, but little potential demand for the pirated good[7]. On the other hand, if an increase in α causes the distribution of y to become very left-skewed (for instance, like a move from (b) to (c) in Figure 4), the effect on p_D is unclear. While the seller might increase the price of the digital good to benefit from the higher potential demand for that digital good, the seller's ability to increase price is limited by the larger potential demand for the pirated good, and in fact, the seller may need to lower the price of its digital good in order to ameliorate some of the loss in demand for the tangible good caused by piracy. We discuss this further in Section 4 when interpreting the variation in the price of digital goods across categories.

[7]While the positions of the curves $\theta_{TD}(y)$ and $\theta_{TP}(y)$ are unaffected by changes in $f_h(y)$, the resulting price changes may shift the curves. The figure thus illustrates what would motivate such a price change, prior to the change in price.

3. Are digital rights valuable? Evidence from the ebook industry

This section describes our data, our empirical model, and the results of its estimation, and discusses the relationship of these results to the predictions of our analytical model.

3.1. Data

We have collected the prices, digital rights and category affiliation of over 30,000 ebooks sold by a specialty Web-based ebook retailer. We have also collected data about the prices of the tangible paperback versions of a subset of these ebooks from a leading online seller of physical books. Combining the two data sets created a set of over 4,000 book titles, for which both an electronic version and a paperback version exists. The results presented are based on our data set from January 2005.

We gathered data about the prices of the tangible counterparts of our ebooks, because a substantial part of the intrinsic quality of a book is associated with its content, the popularity of its author, and so on, independent of whether its format is tangible or digital. We conjecture that information about this intrinsic quality (we use the term "quality" as it relates to a buyer's willingness to pay, rather than an assessment of the book's literary merits) is likely to be contained in the price of the tangible book, which, according to our model, influences the price of the digital good. To insure against variation in tangible book prices due to seasonal discounting or sales at the time of collection, we collected the list price of the tangible book.[8]

The ebooks in our data set are offered in up to four formats (Adobe eBook, Microsoft Reader, Palm Reader, and MobiPocket Reader). Every ebook is available in the Adobe eBook format, which was also the format whose DRM platform offers the widest range of digital rights; therefore, we choose to focus on this format. In addition to its price and category, each ebook has five digital rights associated with it: print, copy, expiry, lend and read aloud. None of the books in our data set have restricted the "expiry" right. The "lend" and "read aloud" rights have binary settings (on

[8] We also restrict ourselves to books with paperback versions for this reason, to avoid a higher price for a tangible book being on account of it being a hardcover (or some other superior) version, rather than paperback version.

The eBook you are interested in is available for purchase in the following formats:

	Adobe eBook Reader Buy this Format
Price:	$29.99
ISBN:	0672320673
Published Date:	6/19/2001
File Size:	1666kb
Security:	Settings listed below
Printing:	On, any number of pages can be printed over any number of days.
Copying:	On, any number of pages can be copied over any number of days.
Expiry:	Off
Lending:	Off
Reading Aloud:	On
Minimum Software Version:	N/A
Suitable Devices:	Palm handhelds, Tablet PCs, PCs, Macs, Laptops

The eBook you are interested in is available for purchase in the following formats:

	Adobe eBook Reader Buy this Format
Price:	$39.99
ISBN:	047101379X
Published Date:	3/22/2001
File Size:	3168kb
Security:	Settings listed below
Printing:	On, 25 pages can be printed every 10 days.
Copying:	On, 25 pages can be copied every 10 days.
Expiry:	Off
Lending:	Off
Reading Aloud:	Off
Minimum Software Version:	N/A
Suitable Devices:	Palm handhelds, Tablet PCs, PCs, Macs, Laptops

Figure 3.1: Illustrates the digital rights granted for two ebooks in our sample. Under our definition of variables associated with digital rights, the first ebook would have $PrintAll = 1, CopyAll = 1$, $Read = 1$, and $PrintPartial = 0, CopyPartial = 0, Lend = 0$.The second ebook would have $PrintPartial = 1, CopyPartial = 1$, and $PrintAll = 0, CopyAll = 0, Read = 0, Lend = 0$.

or off). The rights "print" and "copy" have three kinds of settings: unrestricted (print as many pages as one wants as frequently as needed, copy as much text of the book as one wants), none (no printing allowed, no copying allowed), and partial. There is a range of different levels of rights granted under partial, along two dimensions: the number of pages, and the frequency (for example, copy up to 25 pages every 10 days, print up to 35 pages every 7 days). The digital rights for two ebooks as depicted on the retailer's site are displayed in Figure 3.1. We have not comprehensively analyzed all the details associated with partial rights across our data set. However, all the books we examined that had partial print or copying rights specified them based on both a fixed set of pages and a fixed frequency . For simplicity, therefore, we treat all books with partial printing rights as having the same printing rights ($PrintPartial = 1$), and all books with partial copying rights as having the same copying rights ($CopyPartial = 1$). The variables associated with each of our data points are summarized on Table 1, along with some descriptive statistics of the data set, and the correlation matrix[9] is summarized in Table 3.1A.

The ebook retailer categorizes its books by subject. We focus our analysis on six categories that we believed ex ante would have target customer sets with different levels of technological sophistication: Computers, Fiction, Children's and Young Adults Fiction, Biography and Autobiography,

[9] We also computed the variance inflation factors for our data, which range from 1.2 to 4.7, thus confirming that our data does not display higher-order multicollinearity.

Variable	Brief Description	Data points	Range	Mean	SD
EPrice	ebook price	3105	1.98-165	14.68	18.01
Log[*EPrice*]	Log of ebook price	3105	0.68-5.1	2.27	0.82
TPrice	Tangible book price	3105	1.99-387	25.77	23.32
Log[*TPrice*]	Log of tangible book price	3105	0.68-5.13	2.95	0.78
Variable	**Brief Description**	**Data points**	**Range**	**% non-zero**	
CopyAll	Allows unlimited copying of text	3105	{0,1}	49%	
CopyPartial	Allows limited copying of text	3105	{0,1}	16%	
PrintAll	Allows unlimited printing of pages	3105	{0,1}	16%	
PrintPartial	Allows limited printing of pages	3105	{0,1}	14%	
Lend	Allows lending	3105	{0,1}	5.4%	
Read	Has the 'Read Aloud' feature	3105	{0,1}	39.19%	
K_1	Business category	3105	{0,1}	15%	
K_2	History category	3105	{0,1}	5%	
K_3	Fiction category	3105	{0,1}	57%	
K_4	Young category	3105	{0,1}	6%	
K_5	Biography category	3105	{0,1}	5%	

Table 3.1: Summary of variables and some descriptive statistics

Business and Economics, and History. The smallest of these categories has 164 pairs of ebooks and paperback books. Our final data set has a total of 3105 observations, where each observation corresponds to a unique title.

The first stage of our analysis investigates how the price of the ebook is related to the price of its tangible counterpart. Since we believe that a substantial part of the intrinsic quality of a book is captured in the price of the tangible version, a significant relationship between the price of the ebook and the price of the tangible book seems natural. We therefore estimate the following equation:

$$log[EPrice] = \alpha + \beta_1 \log[TPrice], \tag{3.1}$$

where we use the log transformation to account for non-constant variance. The results of this estimation are presented in Table 3.2, and confirm the expected relationship. The value of the regression coefficient β_1 (0.60) suggests partial support for Proposition 3, since it indicates that when the prices of the digital good and the tangible book are comparable, a unit increase in the price of the tangible good causes a less than unit increase in the price of the digital good. Also,

Variable	Estimated Value (SE)
α	0.51*** (0.05)
β_1	0.6*** (0.02)
$R^2 = 32.2\%, F = 1463.15$	*** significant with $p \leq 0.001$

Table 3.2: The relationship between tangible and digital book prices

the relative low R^2 indicates that there are factors beyond the price of the tangible book and the intrinsic quality it represents that influence the price of the ebook.

3.2. A hedonic price index for digital rights

Our next model estimates the value of each digital right to an ebook seller using the hedonic pricing equation:

$$\begin{aligned} log[EPrice] &= \alpha + \beta_1 \log[TPrice] + \beta_2 CopyAll + \beta_3 CopyPartial \\ &+\beta_4 PrintAll + \beta_5 PrintPartial + \beta_6 Lend + \beta_7 Read + \sum_{i=1}^{5} \gamma_i K_i \end{aligned} \tag{3.2}$$

Our model's results from Section 3 predict that a digital right whose positive direct quality effect balances or dominates the differential piracy effect will have a positive coefficient β_i associated with it. On the other hand, a right for which the differential piracy effect is substantially higher than the positive direct quality effect (that is, a digital right which is not assessed as being especially valuable by customers of the digital good, but whose granting substantially increases the quality and availability of pirated versions of the good) will have a negative coefficient β_i associated with them.

The results of this estimation are presented in Table 3.3. The signs of the coefficients are quite striking. They indicate that three of the four digital rights – copying, reading aloud and lending – are associated with a significant increase in ebook prices. However, a fourth right – printing – is associated with a significant (and substantial) *decrease* in the price of the digital version of the

Variable	Coefficient	**Estimated Value (SE)**	**Corresponding value of** e^{β_i}
Constant	α	2.235 ** (0.006)	9.267
Log[*TPrice*]	β_1	0.274 ** (0.012)	–
CopyAll	β_2	0.052 * (0.023)	1.053
CopyPartial	β_3	0.405 ** (0.042)	1.499
PrintAll	β_4	−0.338 ** (0.035)	0.713
PrintPartial	β_5	−0.394 ** (0.045)	0.674
Lend	β_6	0.132 ** (0.046)	1.141
Read	β_7	0.408 ** (0.028)	1.504
K_1 *(Business)*	γ_1	−0.376 ** (0.035)	0.687
K_2 *(History)*	γ_2	−0.808 ** (0.049)	0.446
K_3 *(Fiction)*	γ_3	−1.189 ** (0.037)	0.305
K_4 *(Young)*	γ_4	−1.485 ** (0.049)	0.227
K_5 *(Biography)*	γ_5	−0.859 ** (0.049)	0.424
$R^2 = 66.26\%, F = 507.1$		* significant with $p \leq 0.05$ ** significant with $p \leq 0.01$ *** significant with $p \leq 0.001$	

Table 3.3: The effect of digital rights on ebook prices

book.[10]

In the context of our model's predictions, these results have a straightforward interpretation. The negative effect of increasing printing rights on prices indicates that the right to print ebooks increases the value of pirated substitutes for books substantially more than it increases the value of the digital good. Recall from our model that the set of customers who purchase ebooks are those

[10] Since we estimate a semilog equation, and the rights and category are binary variables, the value of e^{β_i} is of economic significance. For instance, when *CopyPartial* = 1, the value of the price of the ebook changes by a (multiplicative) factor of =1.5, or according to our estimated price equation, ebook prices are 50% higher. When *PrintParial* = 1, the value of eBook changes by a (multiplicative) factor of 0.67, or according to our estimated price equation, ebook prices are 33 percent lower. We do not focus on the magnitude of these changes, but on their predicted direction.

whose preferences are such that they favor digital goods over physical goods, and thus, the ability to create printed copies of their ebooks is unlikely to have much of a positive direct effect on ebook quality. This right aims to embed an aspect of the physical consumption experience into the digital good, rather than enhancing digital quality. Furthermore, the PrintAll right facilitates the creation of near-perfect copies of many ebooks (although Adobe's own PDF distiller does not permit this, other free PDF-creation software like Win2PDF allow one to print an entire ebook through their distiller and create an unprotected PDF file which is almost identical to the original ebook in quality). This results in a negative differential piracy effect, which our results show dominates any positive direct quality effect[11].

In contrast, copying rights enhance the digital experience associated with purchasing an ebook, and there is alignment between the preferences of ebook buyers and the value delivered by this right. This may be especially true for reference books and textbooks. While copying may facilitate piracy, current copying rights are restricted only to text, and not to figures or images; moreover, pirated versions created by copying text lose the typesetting and layout of the original. Our results thus suggest that the direct quality effect that copying rights have on ebook quality dominate the piracy effect. This is natural, since copying is a right that enhances the digital consumption experience of the ebook, thereby increasing the digital quality valued by the customers for the good, rather than aiming to embed a characteristic of a tangible good into its digital version.

Moreover, the extent to which the former effect dominates the latter is stronger for partial copying rights. This can attributed to the fact that the positive direct effect is likely to be similar to both cases – the ability to copy relevant parts from a book to one's document. However, the ability to "copy all" will have a stronger negative differential piracy effect, since it does facilitates the creation of complete pirated ebooks, albeit of fairly low quality.

As currently implemented, the read-aloud right does not make pirating an ebook any easier or

[11] An alternative (and potentially intuitive) explanation for this result might be that granting each right does increases price, but less expensive books simply grant all rights, thereby creating the impression that the granting of the right is associated with a decrease in price. However, this is inconsistent with our empirical results, because our estimation shows that some rights are associated with an increase in price and some rights are associated with a decrease in price. We have also compared the distributions of the tangible book prices with printing rights and without, using the Kolmogorov-Smirnov test, which indicates that their distributions do not vary significantly.

more difficult. It is implemented using software that renders text into a computerized voice, and while it may permit the creation of MP3-based "books on tape," they are likely to be of fairly low quality. One can create higher-quality pirated versions by buying the book-on-tape (on which the audio rendering is of far higher quality) and digitizing its content. Similarly, the coefficients of our variable corresponding to lending rights suggest that while the right to lend may lead to a negative differential piracy effect, it also has a strong direct effect on quality that makes it overall value have a positive effect on price.

Our estimates of the variation in ebook price across categories have an interesting interpretation in light of our qualitative analysis of technological sophistication in section 2. Recall that if one considers the extent to which the distribution of taste values y is right-skewed as a measure of the technological sophistication of a market, an increase in this level may be associated with an increase in the price of digital goods relative to their tangible counterparts if technological sophistication is low. If one orders the categories in decreasing order of their estimated category coefficients (or in decreasing order of their ebook prices after controlling for tangible book price and the effects of digital rights) – Computers; Business and Economics; History; Biography and Autobiography; Fiction; and Children's and Young Adult Fiction – this resembles a partial ordering of the technological sophistication of these categories. At the very least, it seems reasonable to assert that the category with the highest adjusted ebook prices – Computers – is also the one with the most technological sophisticated market. The coefficients of the associated categorical variables are consistent with overall preferences of consumers for books being skewed towards preferring tangible books at this point in time. We do not have data that can validate that this effect is in fact due to these categories having customers who vary in their technological sophistication (for instance, from a survey of readers), although this represents an interesting direction for future research. The explanatory power of our model is fairly high (the R^2 of 66.2% is more than double that of the model that does not include digital rights and categories), thereby indicating that the extent to which digital rights are granted explains a substantial portion of the variation in ebook prices.

We have found these predicted signs of our estimated coefficients to be robust across a number of other estimations. We estimate how $Log[EPrice]$ varies with changes in $Log[TPrice]$ and digital rights (that is, without the category variables) and find that all but one of our coefficients are statistically significant, the signs of these coefficients for each digital right are the same as reported in Table 3.3, and the R^2 of 50% suggests that digital rights by themselves do increase the explanatory power of the model[12]. We have also estimated the equation (3.2) using data sets collected at other points in time, finding that the signs of each coefficient are preserved.

In the final stage of our analysis, we investigate the interaction between our categories and our digital rights, towards understanding whether granting different digital rights led to different changes in ebook prices across categories. We do so by estimating the following model:

$$Log[EPrice] = \alpha + \beta_1 Log[TPrice] + \sum_{i=2}^{7} \beta_i R_i + \sum_{i=1}^{5} \gamma_i K_i + \sum_{i=2}^{7} \sum_{j=1}^{5} \phi_{ij} R_i K_j, \qquad (3.3)$$

where we label our six digital rights variables (*CopyAll, Copy Partial,...*) R_2 through R_7 for expository convenience. Our estimates are summarized in Table 3.4.

About half of our interaction coefficients are significant at the 5% level. The main effects remained consistent with our original model, thus strengthening our findings, although the coefficients of the variables associated with CopyAll are not significant. A few additional insights also emerge from this estimation. For example, granting printing rights has the most detrimental effect on ebook price for the Computer category, perhaps reflecting a higher propensity of consumers in this category to pirate such books if it is easy to do so. Granting these rights continues to have a negative effect on ebook price across five of our six categories. The exception was the Fiction category, for which granting these rights results in a positive impact on ebook price. In contrast, this was also the category for which Lend and CopyPartial rights have a negative effect on ebook price, and the former is probably a consequence of books in this category being read once, and unlike other categories, having little "reference" value. Furthermore, the ReadAloud right has a significant and high positive effect on price for the Computers and Business categories, perhaps reflecting the

[12] In a model that estimates how $Log[EPrice]$ varies with changes in just the digital rights, two of our coefficients are not statistically significant, their signs are the same as reported in Table 3.3, and the R^2 is 37%.

	No interaction	*Business*	*History*	*Fiction*	*Young*	*Biography*
No interaction		-0.21 (0.15)	-0.41^{*} (0.18)	-0.94^{***} (0.14)	-1.15^{***} (0.21)	-0.57^{***} (0.17)
PrintAll	-0.70 *** (0.11)	0.17 (0.14)	-0.09 (0.21)	0.77*** (0.13)	0.51 * (0.21)	0.20 (0.22)
PrintPartial	-0.82 *** (0.14)	0.56** (0.18)	-0.03 (0.18)	1.21 ** (0.22)	0.80 * (0.34)	0.26 (0.20)
CopyAll	-0.02 (0.12)	-0.16 (0.14)	-0.03 (0.18)	0.06 (0.12)	0.07 (0.20)	0.10 (0.17)
CopyPartial	0.29* (0.12)	0.01 (0.15)	0.38** (0.18)	-0.44 * (0.2)	0.03 (0.33)	0.31 (0.19)
Lend	0.30 * (0.10)	-0.34* (0.13)	0.68* (0.38)	-0.54 ** (0.18)	-0.31 (0.20)	0.43 (0.31)
Read	1.05 *** (0.15)	-0.28 (0.17)	-0.54 ** (0.19)	-0.86 *** (0.16)	-0.92 *** (0.22)	-0.81 *** (0.20)
Log[TPrice]	0.24 *** (0.01)					
Constant	2.07 *** (0.15)					
$R^2 = 68.6\%$	* significant with $p \leq 0.05$ ** significant with $p \leq 0.01$ *** significant with $p \leq 0.001$					

Table 3.4: How digital rights and categories interact. The coefficients in the first column (labeled "No interaction") correspond to the estimates of α and the β_i coefficients, in the first row correspond to the estimates of the γ_i coefficients, while the other entries are estimates of the ϕ_{ij} coefficients. For example, the entry in the row labeled PrintPartial and the column labeled Business corresponds to the coefficient estimated for the variable PrintPartial*Business.

positive direct effect granting this kind of read-aloud right has on quality for these category (the fact that the voice is digital and lacks intonation may not matter as much for computer or business books as it does for books in the Fiction).

The intuitive explanations for the magnitude and signs of these coefficients might be of independent interest for the publishing industry, but from our point of view, they merely serve to further validate the link between the effects our theoretical model ascribes to changes in digital rights, and the presence of estimated coefficients that reflect these effects in our data.

4. Conclusions and ongoing work

We have presented a model characterizing the choice of digital rights and pricing for digital goods offered by a firm who also sells a physical version of the digital good, the granting of digital rights for which may also lead to an increase in digital piracy. The predictions of this model are validated by our empirical results, and suggest important new guidelines for managers in industries that are progressively being digitized. As the pace of industry transformation by information technology accelerates over the coming years, managing such transitions will become central to continued business success across a wider variety of industries, thereby increasing the value of IS research that contributes to our understanding of this transformational power of IT (Agrawal and Lucas, 2005), and of how the Internet affects markets and industry structure (Ellison and Ellison, 2005). We hope our study makes such a contribution.

Summarizing our key results:

1. In the absence of a piracy threat, digital rights are always valuable through their direct effect on increasing the quality of digital goods. Any issues of cannibalization of the sales of physical goods can be effectively addressed by a strategic choice of pricing.

2. When granting digital rights in the presence of digital piracy, the value of these rights is governed by two opposing effects: *a direct quality effect*, or how much the granting of the right increases willingness to pay for the legal digital goods, and a *differential piracy effect,* that measures how much it reduces the relative surplus of consuming legal tangible and digital goods through its inducing digital piracy.

– When the direct quality effect balances the differential piracy effect, then granting the right is beneficial to the seller. It raises prices and profits from the sales of digital goods. This is more likely to be associated with rights that enhance the digital consumption experience of the good, rather than replicating the physical consumption experience, since consumers who purchase the digital good belong to the segment who value digital quality. If they did value the physical consumption experience, these consumers would have purchased the tangible good in the first place.

– When the differential quality effect associated with the right exceeds the direct quality effect significantly, granting the right may affect the seller's pricing power adversely, since consumers who prefer digital goods over tangible goods need to be offered sufficient surplus to ensure that they do not simply resort to using a pirated copy of the good. We conjecture that this kind of right is generally likely to be one that attempts to *replicate* the physical consumption experience digitally, for two reasons. First, such rights do not contribute to those aspects of consumption that the segment who purchases digital goods values, and the direct quality effect is therefore likely to be low. Furthermore, rights that attempt to replicate the physical consumption experience digitally typically involves providing an opportunity to render the digital good into a *tangible* artifact (onto a CD, onto paper), and facilitating such rendering is likely to be associated with an increase in the quality and distribution of pirated copies, since it involves relaxing, to some extent, technological protection that prevent copying.

3. In testing our theory's predictions using pricing and digital rights data for over 3000 ebooks, we find that each of four separate digital rights have both a statistically and economically significant impact on ebook prices, after controlling for tangible book price levels and category. Our results also show that a right which aims to replicate physical consumption characteristics – printing – is the one with a negative impact on ebook prices, while rights enhancing the consumption experience by exploiting the fact that the book is now a digital good – copying and reading aloud via digital audio – have a positive impact on prices. This provides empirical support for the discussion in (2) above.

4. The variation in ebook prices we observe across categories suggests that an increase in technological sophistication of consumers is associated with an increase in the price of legal digital goods. Such a possibility is discussed qualitatively by our theoretical model as being associated with an industry in an early stage of transformation from tangible artifacts to digital goods, which seems consistent with the book publishing industry. Technological sophistication is a fairly broad term, however, and there are many aspects to it that our model and data have not captured. Studying this relationship more closely is an interesting direction for future research, since it may

prove important to choosing the timing of the transition, to the extent that sellers in an industry have control over it.

There are many other directions for future research suggested by our paper. We assume that tangible book prices are exogenous, which, apart from being simpler, reflects a model consistent with the industry our data is from. Our conversations with publishers have suggested they still price the tangible book independently, and then key the ebook price off this. However, a model of the simultaneous choice of tangible and digital pricing in the presence of piracy would be a useful extension, especially as digital goods gain prominence in more industries. While most digital goods in other industries do not currently have the rich variation in rights across products that made our empirical study of ebook pricing viable, they are likely to in the near future, and may present the opportunity for studies similar to ours across other industries like music and broadcast video. There are also indications that new forms of pricing are forthcoming in the ebook industry, most notably pay-per-use pricing from Amazon.com and Google. As this industry matures, it would be interesting to examine how our empirical findings evolve, and this is a direction of research we hope to pursue in the future.

5. References

1. Agrawal, R. and Lucas, H., 2005. The Information Systems Crisis: Focusing on High-Visibility and High-Impact Research. *MIS Quarterly* 29, 381-398.

2. Bhattacharjee, S., Gopal, R. and Sanders, G., 2003. Digital Music And Online Sharing: Software Piracy 2.0? *Communications of the ACM* 46 (7), 107-111.

3. Bakos, Y., Brynjolfsson, E. Lichtman., D., 1999. Shared Information Goods. *Journal of Law and Economics* 42, 117–156.

4. Belleflamme, P., 2002. Pricing Information Goods in the Presence of Copying. Mimeo, Queen Mary University of London.

5. Berndt, E., 1990. *The Practice of Econometrics.* Addison-Wesley Press

6. Chellappa, R., and Shivendu, S., 2003. Managing Piracy: Pricing and Sampling Strategies for Digital Experience Goods in Vertically Segmented Markets. *Information Systems Research* (forthcoming).

7. Connors, K., and Rummelt, R., 1991. Software Piracy: An Analysis of Protection Strategies. *Management Science* 37, 125-139.

8. Ellison, G. and Ellison, S., 2005. Lessons about Markets from the Internet. *Journal of Economic Perspectives* 19 (2), 139-158.

9. Haruvy, E., Mahajan, V., and Prasad, A., 2004. The Effect of Piracy on the Market Penetration of Subscription Software. *Journal of Business* 77, S81-S108.

10. Gayer, A., and Shy, O., 2003. Internet and Peer-to-Peer Distributions in Markets for Digital Products. *Economics Letters* 81, 51-57.

11. Gu, B., and Mahajan, V., 2004. The Benefits of Piracy: A Competitive Perspective. *Sixteenth Workshop on Information Systems and Economics.*

12. Gopal, R., and Sanders, L., 1998. International Software Piracy: Analysis of Key Issues and Impacts. *Information Systems Research* 9, 380-397.

13. Hui, K., Png., I., 2003, Piracy and the Legitimate Demand for Recorded Music. *Contributions to Economic Analysis and Policy* 2(1).

14. Konana, P., N. Menon, S. Balasubramanian, 2000. The Implications of Online Investing. *Communications of the ACM* 43 (1), 34-41.

15. Mossberg, W. 2003. Apple's new service beats illegal free sites. The Wall Street J. (April 30) D1.

16. Mussa, M., and Rosen, S., 1978. Monopoly and Product Quality. *Journal of Economic Theory* 18, 301-317.

17. Peitz, M., and Waelbroeck, P., 2003. Piracy of Digital Products: A Critical Review of the Economics Literature. CESifo Working Paper Series No. 1071

18. Png, I., and Chen, Y., 2003. Information Goods Pricing and Copyright Enforcement: Welfare Analysis. *Information Systems Research* 14, 107-123.

19. Poddar, S., 2004. Software Piracy: When Protection is Optimal to the Software Developer. Working Paper 0312, National University of Singapore.

20. Prasad, A., and Mahajan, V., 2003. How Many Pirates should a Software Firm Tolerate? An Analysis of Piracy Protection on the Diffusion of Software. *International Journal of Research in Marketing* 20, 337-353.

21. Shy, O., Thisse., J. F., 1999. A Strategic Approach to Software Piracy. *Journal of Economics and Management Strategy* 8, 163–190

22. Snir, E. M., 2003. The Record Industry in an Era of File Sharing: Lessons from Vertical Differentiation. In *Proceedings of the 24th International Conference on Information Systems*, Seattle, Washington, USA..

23. Sundararajan, A., 2004. Managing Digital Piracy: Pricing and Protection. *Information Systems Research* 15, 287-304.

24. Takeyama, L. N., 1994. The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Network Externalities. *Journal of Industrial Economics* 42, 155-166.

25. Viswanathan, S., 2005. Competing Across Technology-Differentiated Channels: The Impact of Network Externalities and Switching Costs. *Management Science* 51, 483-496.

26. Zentner, A., 2003. Measuring the Effect of Music Downloads on Music Sales. http://home.uchicago.edu/~alezentn/musicindustrynew.

A. Appendix: Proofs

The proofs in this appendix use the following additional notation:

$$\Delta Q(y_a, y_b, g(y)) \equiv \int_{y=y_a}^{y_b} \left[f_v(g(y)) f_h(y) \frac{dg(y)}{dp_D} \right] dy. \tag{A.1}$$

The proof of Proposition 1 follows from the following lemma:

Lemma 1. *The optimal price of the digital good, p_D^*, is uniquely defined by the following equation:*

$$p_D^* = [p_T - c] + \frac{q_D(p_D^*)}{-\left[\dfrac{dq_D(p_D^*)}{dp_D}\right]}. \tag{A.2}$$

where:

$$\frac{q_D(p_D^*)}{-\left[\dfrac{dq_D(p_D^*)}{dp_D}\right]} = \frac{1 - F_H(y_D(p_D^*)) - Q(y_D(p_D^*), 1, \theta_{TD}(y, p_D^*))}{\Delta Q(y_D(p_D^*), 1, \theta_{TD}(y, p_D^*))} \tag{A.3}$$

is the ratio of demand for the digital good to the slope of the demand for the tangible good, evaluated at the optimal price.

Proof. First, differentiating both sides of (2.9) and (2.12) with respect to p_D and using (A.1) yields:

$$\frac{dq_T(p_D^*)}{dp_D} = \Delta Q(y_{TD}(p_D^*), 1, \theta_{TD}(y, p_D)) \tag{A.4}$$

$$\frac{dq_D(p_D^*)}{dp_D} = -\Delta Q(y_{TD}(p_D^*), 1, \theta_{TD}(y, p_D)) \tag{A.5}$$

The monopolist chooses p_D^* to solve:

$$\max_{p_D} [p_T - c] q_T(p_D) + p_D q_D(p_D). \tag{A.6}$$

The first-order condition for (A.6) is:

$$[p_T - c]\frac{dq_T(p_D)}{dp_D} + p_D \frac{dq_D(p_D)}{dp_D} + q_D(p_D) = 0, \tag{A.7}$$

which in conjunction with (A.4), (A.5) and (2.12) yields equations (A.2) and (A.3). ■

Proof of Proposition 1

If the demand for the digital good is positive, then

$$[1 - F_H(y_D(p_D^*))] > Q(y_D(p_D^*), 1, \theta_{TD}(y, p_D^*)),$$

and thus the expression in (A.3) is strictly positive.

Proof of Proposition 2

For a fixed level of digital quality s_D, denote the optimal price of the digital good as $p_D^*(s_D)$, and the corresponding demand for the digital good as $q_D(p_D^*(s_D), s_D)$ The seller's problem of choosing digital rights is therefore equivalent to determining the level s_D that maximizes profits:

$$\max_{s_D} \Pi(s_D) \equiv [p_T - c][1 - q_D(p_D^*(s_D), s_D)] + [q_D(p_D^*(s_D), s_D)], \tag{A.8}$$

and then choosing a combination of rights that implements this level of quality. However, notice that using the envelope theorem, it follows that:

$$\frac{d\Pi(s_D)}{ds_D} = [p_D^*(s_D) - (p_T - c)] \frac{\partial q_D(p_D^*(s_D), s_D)}{\partial s_D}. \tag{A.9}$$

From (A.2) the first term on the RHS is strictly positive. Furthermore, differentiating both sides of (2.7) with respect to s_D and rearranging yields:

$$\frac{d\theta_{TD}(y, p_D)}{ds_D} = \frac{u_1(s_D, \theta_{TD}(y, p_D))}{-u_2(s_D, \theta_{TD}(y, p_D))} < 0 \tag{A.10}$$

for all y and p_D, since $u_1(s,\theta) > 0$ and $u_2(s,\theta) > 0$. (A.10) in turn implies that

$$\frac{\partial q_D(p_D^*(s_D), s_D)}{\partial s_D} > 0, \tag{A.11}$$

which, in conjunction with (A.9) implies that the seller's profits are strictly increasing in s_D. Since

$$s_D = S_D(r_1, r_2, ..., r_n), \tag{A.12}$$

and S_D is strictly increasing in each of its arguments, the result follows.

Our next lemma characterizes the optimal price of the digital good in the presence of piracy.

Lemma 2. *The optimal price p_D of the digital good is:*

$$p_D^* = [p_T - c]\delta(p_D^*) + \frac{q_D(p_D^*)}{-\left[\dfrac{dq_D(p_D^*)}{dp_D}\right]}, \tag{A.13}$$

where:

$$\delta(p_D^*) = \frac{\Delta Q(y_{TD}(p_D^*), y_{PD}(p_D^*), \theta_{TD}(y, p_D^*))}{\Delta Q(y_{TD}(p_D^*), y_{PD}(p_D^*), \theta_{TD}(y, p_D^*)) - \left[\frac{d\theta_{TP}(y_{PD}(p_D^*))}{dy}\frac{dy_{PD}(p_D^*)}{dp_D}\right][[1-F_h(y_{PD}(p_D^*))]f_v(y_{PD}(p_D^*))]} \tag{A.14}$$

is the discount on the digital good induced by piracy, and:

$$\frac{q_D(p_D^*)}{-\left[\dfrac{dq_D(p_D^*)}{dp_D}\right]} = \frac{1-F_h(y_D(p_D^*)) - Q(y_D(p_D^*), y_{PD}(p_D^*), \theta_{TD}(y, p_D^*))}{\Delta Q(y_D(p_D^*), y_{PD}(p_D^*), \theta_{TD}(y, p_D^*)) - \left[\frac{d\theta_{TP}(y_{PD}(p_D^*))}{dy}\frac{dy_{PD}(p_D^*)}{dp_D}\right][[1-F_h(y_{PD}(p_D^*))]f_v(y_{PD}(p_D^*))]} \tag{A.15}$$

is the ratio of demand for the digital good to the slope of the demand for the tangible good, evaluated at the optimal price.

Proof. Differentiating both sides of (2.19) and (2.20) with respect to p_D, cancelling out common terms, and simplifying using (A.1) yields:

$$\frac{dq_T(p_D^*)}{dp_D} = \Delta Q(y_{TD}(p_D^*), y_{PD}(p_D^*), \theta_{TD}(y, p_D)) \tag{A.16}$$

$$\frac{dq_D(p_D^*)}{dp_D} = -\Delta Q(y_{TD}(p_D^*), y_{PD}(p_D^*), \theta_{TD}(y, p_D^*)) + \left[\frac{d\theta_{TP}(y_{PD}(p_D^*))}{dy}\frac{dy_{PD}(p_D^*)}{dp_D}\right][[1 - F_h(y_{PD}(p_D^*))]f_v(y_{PD}(p_D^*))] \tag{A.17}$$

The monopolist chooses p_D^* to solve:

$$\max_{p_D}[p_T - c]q_T(p_D) + p_D q_D(p_D). \tag{A.18}$$

The first-order condition for (A.18) is:

$$[p_T - c]\frac{dq_T(p_D)}{dp_D} + p_D\frac{dq_D(p_D)}{dp_D} + q_D(p_D) = 0, \tag{A.19}$$

or

$$p_D = \frac{[p_T - c]\frac{dq_T(p_D)}{dp_D}}{-\left[\frac{dq_D(p_D^*)}{dp_D}\right]} + \frac{q_D(p_D^*)}{-\left[\frac{dq_D(p_D^*)}{dp_D}\right]}$$

which in conjunction with (A.16) and (A.17) yields equations (A.14) and (A.15). ■

Proof of Proposition 3

By the definition of $y_{PD}(p_D)$,

$$\theta_{TP}(y_{PD}(p_D), p_D) = \theta_{TD}(y_{PD}(p_D), p_D) \tag{A.20}$$

Differentiating both sides of (A.20) with respect to p_D and rearranging yields:

$$\frac{dy_{PD}(p_D)}{dp_D} = \frac{\frac{d\theta_{TD}(y_{PD}(p_D),p_D)}{dp_D}}{\left(\frac{d\theta_{TP}(y_{PD}(p_D),p_D)}{dy} - \frac{d\theta_{TD}(y_{PD}(p_D),p_D)}{dy}\right)} \tag{A.21}$$

Now, differentiating both sides of (2.7) with respect to p_D and rearranging yields

$$\frac{d\theta_{TD}(y, p_D)}{dp_D} = \frac{1}{u_2(s_D, \theta_{TD}(y, p_D))} > 0, \tag{A.22}$$

since $u_2(s, \theta) > 0$. Next, (A.20) in conjunction with the fact that $u_{12}(s, \theta) > 0$ implies that

$$u_2(s_P, \theta_{TP}(y, p_D)) < u_2(s_D, \theta_{TD}(y, p_D)). \tag{A.23}$$

Finally, differentiating both sides of (2.7) and (2.13) with respect to y and rearranging yields:

$$\frac{d\theta_{TD}(y, p_D)}{dy} = \frac{w_1(y) + w_1(1-y)}{-u_2(s_D, \theta_{TD}(y, p_D))}, \tag{A.24}$$

$$\frac{d\theta_{TP}(y, p_D)}{dy} = \frac{w_1(y) + w_1(1-y)}{-u_2(s_P, \theta_{TP}(y, p_D))}, \tag{A.25}$$

which, using (A.23) and the fact that $w_1(y) > 0$ implies that

$$\left[\frac{d\theta_{TP}(y_{PD}(p_D), p_D)}{dy} - \frac{d\theta_{TD}(y_{PD}(p_D), p_D)}{dy}\right] < 0. \tag{A.26}$$

(A.21), (A.22) and (A.26) imply that

$$\frac{dy_{PD}(p_D)}{dp_D} < 0., \tag{A.27}$$

(A.22) and (A.27) therefore imply that

$$-\left[\frac{d\theta_{TP}(y_{PD}(p_D^*))}{dy}\frac{dy_{PD}(p_D^*)}{dp_D}\right][[1 - F_h(y_{PD}(p_D^*))]f_v(y_{PD}(p_D^*))] > 0 \tag{A.28}$$

so long as there is non-zero demand for the pirated good. Also, from (A.1),

$$\Delta Q(y_{TD}(p_D^*), y_{PD}(p_D^*), \theta_{TD}(y, p_D^*)) > 0.$$

Inspecting the definition of $\delta^*(p_D)$ in (A.14), this implies that the denominator of the RHS of (A.14) is strictly greater than its numerator, which completes the proof.

Two-sided markets with intra-group competition*

Paul Belleflamme[†] Eric Toulemonde[‡]

June 15, 2006

Abstract

(to be written) PRELIMINARY

JEL classification codes: L11, L13, L23

Keywords: two-sided markets, vertical oligopoly

*We are grateful to Jean J. Gabszewicz, Pierre Picard and Xavier Wauthy for useful remarks and suggestions on an earlier draft.

[†]CORE and IAG, Université catholique de Louvain, 34 Voie du Roman Pays, B-1348 Louvain la Neuve, Belgium. Email: belleflamme@core.ucl.ac.be. Tel: +32 10 478291. Fax: +32 10 474301.

[‡]University of Namur, 8 Rempart de la Vierge, B-5000 Namur, Belgium. Email: Eric.Toulemonde@fundp.ac.be. Tel: +32 81 725313. Fax: +32 81 724840.

1 Introduction

(to be adapted and completed)

In many economic and social situations, the launch of a new venture requires the combination of a number of complementary assets with the following characteristics: (i) the providers of one asset are better off when the number of providers of another asset increases, as it makes the new venture more likely to succeed (*network effect*); (ii) providers of the same asset may compete with one another (*competition effect*); (iii) because the assets are in short supply, asset providers must be diverted from other existing ventures (*poaching effect*). The paper examines to which extent an independent entity can profitably launch such a venture.

This issue is relevant in a variety of contexts. Consider for instance the following two assets: wheat cultivated by farmers and bakeries that transform wheat into bread. Suppose that an intermediary (e.g., a local government) wants to divert farmers towards the production of rapeseed that would be converted into biofuel by distilleries. In each venture (wheat-bakeries and rapeseed-distilleries), (i) farmers who cultivate wheat (rapeseed) are better off if the demand by bakeries (distilleries) is large, and bakeries (distilleries) are better off if the supply of wheat (rapeseed) is large; (ii) wheat (rapeseed) producers compete for the demand from bakeries (distilleries) and bakeries (distilleries) compete for the supply of wheat (rapeseed). Two similar examples are the industrial conversion of a region and the settlement of a new colony. As far as industrial conversion is concerned, the assets are workers trained for a particular industry and the firms operating in that industry. (i) Workers have larger incentives to train in one industry if this industry is large, whereas firms are more likely to operate in an industry if they can easily hire trained workers. (ii) Workers compete for jobs and firms compete within each industry. In the colony example, settlers must be diverted from the home country. To attract inhabitants, it might be worthwhile to grant some of them exclusive concessions so that they can expect large earnings (property (ii)). However, inhabitants must also find other persons who produce various goods in the colony, and the colony will be more attractive if these other goods are sold in large supply by many producers (property (i)).

Another example involves the following two assets: firms from the logistics industry and buildings specifically designed to attract these firms in a particular area.[1] Competition (among firms and among owners of the buildings) and network effects (between firms and buildings) are clearly an issue. Also, firms are poached from other areas and buildings are designed specifically for one industry. Firms are willing to settle in a specific area if they *quickly* find specific buildings; but landowners are reluctant to launch the construction of specific buildings if they are not sure to find tenants, which hinders the development of the logistics industry.

In the previous examples, the new venture is likely to be promoted by some public institution. In other instances, the intermediary is a private institution. Consider for example the launch of a new business school. The "assets" are academics from different disciplines. They are complementary in the sense that a business school producing research and teaching only in, say, marketing would probably fail to attract any student. As for property (i), academics from one discipline are made better off if the number of colleagues from other disciplines increases, as this makes the school more attractive for students and increases the opportunities for research cross-fertilization. As for property (ii), there might be some form of competition among academics of the same discipline (as they have to share budgets, or compete for promotions). Finally, property (iii) is satisfied as a key for success will be to hire senior people, who will have to be poached from other universities.

More generally, the launch of a new business (combining various complementary inputs) or of a new market (on which providers and users of goods, services or information are matched) is often subject to network, competition and poaching effects. The business literature has mainly put the emphasis on the network effect. For instance, analysts of electronic markets stress that "liquidity" is essential for success: "To succeed, e-hubs must attract both buyers and sellers quickly, creating liquidity at both ends" (Kaplan and Sawhney, 2000); "the first pillar of e-market success is building liquidity" (Brunn *et al.*, 2002). However, the other two effects also play a crucial

[1]This illustration is derived from a real-life example: see the Belgian newspaper *Le Soir*, "Deux chantiers à l'aéroport" by P. Bodeux, March 10, 2006.

role in determining the business strategy an intermediary should follow in order to profitably launch new ventures of this type: on the one hand, the poaching effect imposes a cost on the achievement of liquidity and, on the other hand, the competition effect makes liquidity less crucial.

The interplay between the three effects can be described as follows. As assets have to be poached from existing ventures, the intermediary will have to find a way to induce asset providers to join the new venture. The competition prevailing among providers of the same asset might facilitate such endeavour. Indeed, asset providers might benefit from participating to a new venture together with a smaller group of agents of their own type. Yet, the fact that asset providers' utilities positively depend on the number of providers of other assets is likely to make it harder for the intermediary to attract any asset; this is due to the so-called "chicken-and-egg problem" (providers of one asset will not join until a sufficient number of providers of other assets also join, and vice versa...). In the presence of these various forces, which fee structure should the intermediary put in place? Should he subsidize one or the other type of asset providers? Does it matter if he attracts one type before the other? How many providers of each asset should he attract? Is the liquidity essential for the success, as is stressed in the business literature?

Questions of this sort are at the center of a recent literature in economics, which examines *multi-sided markets*. Seminal contributions are Rochet and Tirole (2003), Evans (2003), Caillaud and Jullien (2003) and Armstrong (2004). Rochet and Tirole (2004) propose a useful introduction and road map to this flourishing literature. Our contribution to this literature is to analyse a model that allows us to derive, endogenously, the payoffs of all types of agents, and to give a precise structure to the various externalities (network, competition and poaching effects) that exist between agents.[2] For

[2]Papers that deal with the competition effect usually consider competition between sellers but not between buyers. For instance, Rochet and Tirole (2002) and Schmalensee (2002) provide a formal analysis of the credit card payment industry, where merchants compete with one another but cardholders do not. Similarly, in Nocke, Peitz and Stahl (2004) sellers compete on the market for differentiated products, which are sold to independent consumers.

that purpose, we develop a model with successive oligopolies. This strategy restricts somewhat the analysis, but it allows us to give precise microeconomic foundations for the different forces at work. In particular, it is possible to trace precisely how the payoff of an agent depends on the number of agents of his own type on each venture, and on the number of agents of a different type on each venture. In other words, all payoffs are derived endogenously and the network, competition and poaching effects are clearly identified.

(description of the results)

The rest of the paper is organized as follows. In Section 2 we lay out the model. In Sections 3 and 4, we study in turn the cases where buyers produce independent varieties and where they produce homogenous varieties. We conclude and propose some directions for future research in Section 5.

2 The model

We consider two groups of homogeneous agents, denoted 1 and 2, with respectively $N_1 \geq 3$ and $N_2 \geq 3$ agents.[3] When the game starts, the two groups interact on a "platform" whose access is supposed to be free. Then, an intermediary considers launching a competing platform. Agents who switch to this new platform can interact only with the agents who have switched along. That is, we preclude multi-homing: no interaction can take place among agents affiliated with different platforms. As a result, the benefits agents derive on a platform depend only on the number of agents who are active on this platform. Formally, we denote by $\pi_i(n_i, n_j)$ the gross benefit for an agent of type i from interacting on a platform with n_i agents of its own type and n_j agents of the other type ($i \neq j \in \{1, 2\}$).[4] We assume that the benefit functions exhibit *positive inter-group externalities*:

$$\pi_1(n_1, n_2 + 1) > \pi_1(n_1, n_2) \text{ and } \pi_2(n_1 + 1, n_2) > \pi_2(n_1, n_2).$$

[3]We exclude the cases where $N_i = 2$ because they bring complications without adding any insight to our analysis. These complications are due to the fact that our model is discrete and that, with $N_i = 2$, a single agent represents half of her group, which gives her an excessive influence.

[4]Implicit in this formulation is that the two platforms are "technically equivalent".

The benefit functions may also exhibit *negative intra-group externalities*:

$$\pi_1(n_1+1,n_2) \leq \pi_1(n_1,n_2) \text{ and/or } \pi_2(n_1,n_2+1) \leq \pi_2(n_1,n_2).$$

The first effect results from the indirect network externalities usually observed in two-sided markets: more agents on one side of the market increases the utility of agents on the other side of the market.[5] The second effect translates the idea that agents may compete with one another within a particular group and may therefore prefer, all other things being equal, to be on a platform with fewer of their group mates.

The impacts of this second effect, which we broadly refer to as "rivalry", have not been systematically analyzed so far in the literature on two-sided markets. To fill this gap, we contrast three cases according to whether the negative intra-group externalities are observed in the two groups (*full rivalry*), in only one group (*partial rivalry*), or in none of the groups (*no rivalry*). Moreover, in each case, we consider two potential timing of moves by the two groups.

- In the **sequential switching game**, we assume the following order of moves: in stage 1, the intermediary sets a membership fee A_1 for agents of group 1 (which corresponds to a fixed registration charge for accessing the new platform).[6] In stage 2, agents of group 1 simultaneously choose whether to switch to the new platform. In stage 3, the intermediary sets the membership fee A_2 for agents of group 2 and in stage 4, agents of group 2 choose whether to switch to the new platform.

- In the **simultaneous switching game**, there are only two stages: in stage 1, the intermediary sets the membership fees A_1 and A_2; in stage

[5] As agents derive benefits from interacting with agents of the other group, it seems reasonable to assume that $\pi_1(n_1,0) = \pi_2(0,n_2) = 0$. We make this assumption to square with the model we use below to endogenize the benefit functions. All our results would hold if we assumed instead that $\pi_1(n_1,0) > 0$ and $\pi_2(0,n_2) > 0$.

[6] We exclude (variable) usage fees for the following reasons. Popović(2002, p. 15) invokes the firms' "reluctance to be charged every time they decide to transact", while Rochet and Tirole (2004, p. 19) argue that the platform might be unable to tax the interaction properly: "Buyers and suppliers may find each other and trade once on a B2B exchange, and then bypass the exchange altogether for future trade".

2, both groups of agents simultaneously decide whether to switch or not to the new platform.

Combining these different possibilities, we cater for a large array of real-life applications. For instance, sequential switching is a natural assumption in several categories of two-sided markets where most agents of one side of the market arrive before most agents of the other side.[7] Also, intermediaries might not always have the possibility to commit to their fee structure; they will thus set the fee for one group of agents after seeing the switching decision of the other group of agents, as assumed in the sequential switching game. (complete with the list of examples given in the intro. and give examples where rivalry does or does not exist) From an analytical point of view, our aim is to understand why these various settings may sometimes lead to radically different equilibria.

In all settings, our equilibrium concept is a refinement of subgame perfection. Because of positive network externalities, multiple Nash equilibria in pure strategies can occur in the stages where one group or the other decides to switch. As there is no obvious way to select among these equilibria on some a priori basis,[8] we require that the intermediary set fees in such a way that a unique Nash equilibrium ensues. In other words, we follow Segal (1999 and 2004) by imposing *unique implementation* (complete and check references).

We carry out the major part of the analysis with the generic benefit functions $\pi_i(n_1, n_2)$. However, to be able to completely solve the models, we may sometimes need to adopt specific functions. To this end, we add an additional layer to the games in order to generate endogenous profit functions for the two groups. In particular, we identify the two groups as respectively sellers and buyers of an intermediary input; we then model the industry as a successive vertical Cournot oligopoly (the sellers produce the input, which the buyers then transform into a potentially differentiated final

[7] For example, Hagiu (2004) points that "in the software and videogame markets, most application developers join platforms (operating systems and game consoles) before most users do."

[8] In general, coexisting equilibria cannot be Pareto-ranked as either the preferences of the groups diverge, or they converge but are at odds with the intermediary's interest.

product). We detail the derivation of the profit functions in Appendix 7.1.

3 No rivalry

This is clearly the simplest case. Because the benefits for each agent only depend on the number of agents of the other group with whom she can interact, we write the benefit functions simply as $\pi_1(n_2)$ and $\pi_2(n_1)$. We solve the sequential and simultaneous games in turn by the method of backward induction. The same result holds in both games. the intermediary adopts a divide-and-conquer strategy which consists in subsidizing the group of agents that make the smaller benefits and taxing away the larger benefits of the other agents.

3.1 Sequential switching game

Recall that group 1 moves before group 2 and that the intermediary cannot commit to his fee structure. That is, the agents in group 2 make their switching decision upon observing the fee set by the intermediary, as well as what the agents of group 1 have decided beforehand. On the other hand, the agents in group 1 are able to rationally anticipate how their decisions will shape the subsequent decisions made by the intermediary and by the agents in group 2.

Consider **stage 4** supposing that n_1 agents of group 1 have moved in stage 2 and that the intermediary set a fee A_2 in stage 3. A first immediate result is that we cannot have a Nash equilibrium with $0 < n_2 < N_2$ agents of group 2 switching. Indeed, such an equilibrium would require that on both platforms, no agent has an incentive to move to the other platform; that is, we would need that $\pi_2(n_1) - A_2 \geq \pi_2(N_1 - n_1)$ and $\pi_2(N_1 - n_1) > \pi_2(n_1) - A_2$, which are clearly incompatible.[9] The intuition is very simple: as agents are identical and have independent payoffs, they all make the same decisions. There are thus two potential Nash equilibria: either (i) no agent

[9] To avoid indeterminacies, we adopt the following tie-breaking rule: if the membership fee makes the agent just indifferent between the two platforms, then the agent chooses the new platform.

switches iff $A_2 > \pi_2(n_1) - \pi_2(N_1 - n_1)$, or (ii) all N_2 agents switch iff $A_2 \leq \pi_2(n_1) - \pi_2(N_1 - n_1)$.

In **stage 3**, the intermediary's objective is to set the highest fee that attracts all N_2 agents of group 2. However the intermediary cannot credibly make losses at this stage. Hence, the intermediary is unable to attract any agent of group 2 if the highest fee that attracts all agents of group 2 turns out to be negative. From the above discussion, this fee is[10]

$$A_2 = \pi_2(n_1) - \pi_2(N_1 - n_1),$$

which is positive as long as $n_1 \geq N_1/2$. In other words, the optimal number of agents the intermediary attracts is

$$n_2^*(n_1) = \begin{cases} N_2 & \text{if } n_1 \geq N_1/2, \\ 0 & \text{otherwise.} \end{cases}$$

Regarding **stage 2**, let us first show that we cannot have a Nash equilibrium with $0 < n_1 < N_1$ agents of group 1 switching. Such an equilibrium requires that the following two conditions be satisfied:

$$\begin{cases} \pi_1(n_2^*(n_1)) - A_1 \geq \pi_1(N_2 - n_2^*(n_1 - 1)), \\ \pi_1(N_2 - n_2^*(n_1)) > \pi_1(n_2^*(n_1 + 1)) - A_1 \end{cases} \tag{1}$$

Note that the reasoning is more complicated than in stage 4 as by moving from one platform to the other, an agent of group 1 rationally anticipates that her move may modify the number of agents of group 2 that the intermediary will subsequently attract. In other words, there are values of n_1 for which $n_2^*(n_1) \neq n_2^*(n_1 - 1)$ or $n_2^*(n_1) \neq n_2^*(n_1 + 1)$, meaning that we have to check for all possibilities. As demonstrated in Table 1, the conditions in (1) can never be met simultaneously, which proves our result.

As in stage 4, there are thus two potential equilibria. All agents of group 1 make the same decisions: they either all stay with the existing platform, or they all switch to the new one:

$$n_1^* = \begin{cases} N_1 & \text{if } A_1 \leq \pi_1(N_2), \\ 0 & \text{if } A_1 > -\pi_1(N_2). \end{cases}$$

[10] If we had adopted the alternative tie-breaking rule that indifferent agents stay on the existing platform, the maximal fee the intermediary could set would be $\pi_2(n_1) - \pi_2(N_1 - n_1) - \varepsilon$, which would unnecessarily complicate the analysis.

	$n_2^*(n_1-1)$	$n_2^*(n_1)$	$n_2^*(n_1+1)$	(1)
$0 < n_1 < \frac{N_1}{2} - 1$	0	0	0	$\begin{cases} -A_1 \geq \pi_1(N_2), \\ \pi_1(N_2) > -A_1 \end{cases}$
$n_1 = \frac{N_1}{2} - 1$	0	0	N_2	$\begin{cases} -A_1 \geq \pi_1(N_2), \\ 0 > -A_1 \end{cases}$
$n_1 = \frac{N_1}{2}$	0	N_2	N_2	$\begin{cases} 0 \geq A_1, \\ A_1 > \pi_1(N_2) \end{cases}$
$\frac{N_1}{2} < n_1 < N_1$	N_2	N_2	N_2	$\begin{cases} \pi_1(N_2) \geq A_1, \\ A_1 > \pi_1(N_2) \end{cases}$

Table 1: Stage 2 equilibrium

Note that we have simultaneous equilibria for all $-\pi_1(N_2) < A_1 \leq \pi_1(N_2)$.

We can now consider **stage 1**. As we explained above, we require unique implementation on the part of the intermediary. As the situation with $n_1 = 0$ is clearly unprofitable, the intermediary has to set a fee such that $n_1 = N_1$ is the unique equilibrium in stage 2. That is, we must have $A_1 = -\pi_1(N_2)$. This is the largest fee (i.e., the lowest subsidy) that makes sure that all agents of group 1 will switch and will then be followed by all agents of group 2. The question remains as to whether the intermediary can extract positive profits from this divide-and-conquer strategy. It is so if $\Pi_I^{(1)} \equiv -N_1\pi_1(N_2) + N_2\pi_2(N_1) > 0$. Alternatively, the intermediary could target group 2 first and obtain a profit of $\Pi_I^{(2)} \equiv -N_2\pi_2(N_1) + N_1\pi_1(N_2)$. It is clear that there is necessarily one of the two options that allows the intermediary to secure a positive profit.[11]

3.2 Simultaneous switching game

Groups 1 and 2 are now supposed to move at the same time, after observing the two fees set by the intermediary. Regarding **stage 2**, we can apply the same reasoning as in the last stage of the sequential game and conclude that all agents in one group will act the same; in other words, there cannot be a Nash equilibrium in which some agents of group i stay on the exist-

[11] If $\Pi_I^{(1)} > \Pi_I^{(2)}$, then $\Pi_I^{(1)} > 0$, and vice versa. In the particular perfect symmetric case with $N_1 = N_2$ and $\pi_1(\cdot) = \pi_2(\cdot)$, the intermediary makes no profit (but no losses either).

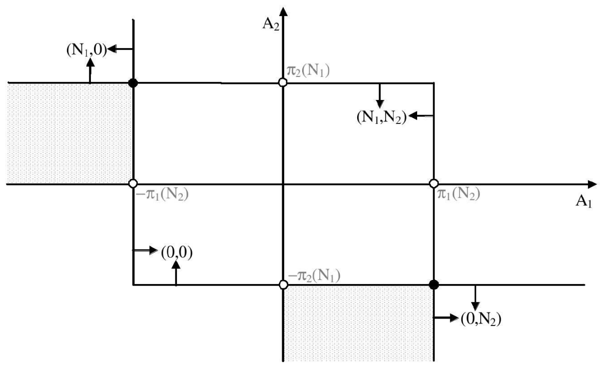


Figure 1: Simultaneous switching game with no rivalry

ing platform while the other agents move to the new platform. Therefore, the four possible equilibria are defined by the following conditions and are represented in Figure 1:

$$(0,0) \text{ if } \left\{ \begin{array}{c} A_1 > -\pi_1(N_2) \\ A_2 > -\pi_2(N_1) \end{array} \right. \quad (N_1,0) \text{ if } \left\{ \begin{array}{c} A_1 \leq -\pi_1(N_2) \\ A_2 > \pi_2(N_1) \end{array} \right.$$

$$(0,N_2) \text{ if } \left\{ \begin{array}{c} A_1 > \pi_1(N_2) \\ A_2 \leq -\pi_2(N_1) \end{array} \right. \quad (N_1,N_2) \text{ if } \left\{ \begin{array}{c} A_1 \leq \pi_1(N_2) \\ A_2 \leq \pi_2(N_1) \end{array} \right.$$

Moving now to **stage 1**, we observe first that among the four possible equilibria, only one may be profitable for the intermediary, namely (N_1, N_2).[12] To implement it as a unique and profitable equilibrium, the intermediary must choose A_1 and A_2 in the shaded areas of Figure 1. The intermediary's profits write as $N_1A_1 + N_2A_2$. As they increase in A_1 and A_2, there are two potential optima: $[A_1, A_2] \in \{[-\pi_1(N_2), \pi_2(N_1)], [\pi_1(N_2), -\pi_2(N_1)]\}$. It is clear that one of the two yields positive profits (and the other one yields losses). Therefore, the optimum is the one giving positive profits and cor-

[12]In the $(0,0)$ case, no agent moves and thus no profit can me made; in the other two cases, no fee can be extracted from the group which doe not move while subsidies have to be paid to the other group.

responds to the same divide-and-conquer strategy adopted in the sequential switching game.

We summarize our results in the following proposition.

Proposition 1 *In the absence of intra-group externalities, the intermediary always finds a profitable way to enter whether the groups of agents move sequentially or simultaneously. The optimal divide-and-conquer strategy consists in attracting all agents of one group by paying them a subsidy of* $A_i = -\pi_i(N_j)$*, and in taxing away the benefits of the other group with a fee of* $A_j = \pi_j(N_i)$.

4 Partial rivalry

In case of partial rivalry, rival agents are willing to isolate from each other. This makes it easier to attract a few set of rival agents to the new platform because the intermediary offers them a way to differentiate. However this also makes it more difficult to attract all rival agents. Also, if the platform attracts a few set of rival agents only, the platform becomes less attractive for the non rival agents. Since attractiveness of the platform is related to the revenues that the intermediary can extract from the agents, it is not clear that rivalry on one side makes it easier to launch the new platform.

Rivalry on one side also multiplies equilibria in the simultaneous game. For instance, we will show that there exists a subsidy to the rival agents and a fee to the non rival agents that support two equilibria: one with $n_1 > N_1/2$ rival agents and no non rival agents, and the other with all agents of both types. Consider the first equilibrium. Despite the absence of non rival agents, the n_1 rival agents are attracted to the new platform because of the subsidy. The other $N_1 - n_1$ rival agents prefer to stay on the old platform where competition is lower and where they can meet the N_2 non rival agents. The non rival agents prefer to trade with few rival agents on the old platform rather than to switch to the new platform and pay a positive fee. This equilibrium is not profitable for the intermediary as he has to subsidize rival agents. Under the second equilibrium, the new platform attracts all non rival agents, which makes this platform attractive for rival agents who switch despite the increase in competition. Moreover, non rival agents also switch to this platform because all rival agents are active there.

By contrast, the sequential game eliminates the non profitable equilibrium with n_1 rival agents and no non rival agent, which makes entry of the new platform more likely to be profitable. Indeed, suppose that the intermediary has attracted first $n_1 > N_1/2$ rival agents by granting them a subsidy. Thus, the new platform is more attractive than the old for the non rival agents. Hence, the intermediary always has the opportunity to set a positive fee that attracts some non rival agents in the last stages of the game. Moreover, it is not profitable for the intermediary to set a fee that does not attract any non rival agent. The fact that the intermediary does not commit to the fee of the group targeted second when the game starts eliminates an equilibrium in which he would attract none of the firms of the second group. As a result, sequentiality may make entry of the new platform profitable whereas simultaneity does not.

The presence of negative intra-group externalities in one of the two groups complicates the analysis in a number of ways: first, the group of rival agents might partition at equilibrium; second, the order of moves now crucially matters in the sequential switching game; third, the sequential and simultaneous switching games become strikingly different; finally, to reach a complete solution of the two games, we need to impose some additional structure on the general benefit functions, or even use some specific benefit functions.

4.1 Sequential switching with rivalry in the first group

We consider first the sequential game in which the intermediary sets first the fee for the rival firms, and then the fee for the non rival firms. We derive the following proposition.

Proposition 2 *In the sequential switching game with rivalry in the first group, the optimum for the intermediary is to set* $A_1 = -\pi_1\left(\frac{N_1}{2} + 2, N_2\right)$ *and* $A_2 = \pi_2\left(N_1\right)$ *so as to attract all agents of both groups, provided that this scheme generate positive profits, i.e.,* $-N_1\pi_1\left(\frac{N_1}{2} + 2, N_2\right) + N_2\pi_2\left(N_1\right) > 0$.

Proof. The benefit function for the agents of group 1, who are moving first, is $\pi_1(n_1, n_2)$, while the benefit function for the agents of group 2, who are moving second, is $\pi_2(n_1)$. As for **stages 3 and 4** of the game, we can use the results of the previous section: all agents of group 2 make the same decision and the intermediary decides to attract them all as long as he has attracted at least half of the agents of group 1 beforehand; that is,

$$n_2^*(n_1) = \begin{cases} N_2 & \text{if } n_1 \geq N_1/2, \\ 0 & \text{otherwise.} \end{cases}$$

In **stage 2**, an equilibrium with $0 < n_1 < N_1$ agents switching occurs if and only if

$$\begin{cases} \pi_1(n_1, n_2^*(n_1)) - A_1 \geq \pi_1(N_1 - n_1 + 1, N_2 - n_2^*(n_1 - 1)), \\ \pi_1(N_1 - n_1, N_2 - n_2^*(n_1)) > \pi_1(n_1 + 1, n_2^*(n_1 + 1)) - A_1. \end{cases}$$

Using the values of $n_2^*(n_1 - 1)$, $n_2^*(n_1)$, and $n_2^*(n_1 + 1)$ given in Table 1, we can detail the latter two conditions as follows.

- For $0 < n_1 < \frac{N_1}{2} - 1$, the two inequalities become:

  $$-\pi_1(N_1 - n_1, N_2) < A_1 \leq -\pi_1(N_1 - n_1 + 1, N_2).$$

  As no agent of group 2 will switch afterwards, the intermediary has to pay a subsidy to the agents of group 1 he wants to attract; clearly, such an equilibrium induces losses for the intermediary.

- For $n_1 = \frac{N_1}{2} - 1$, the two inequalities become

  $$\pi_1\left(\frac{N_1}{2}, N_2\right) - \pi_1\left(\frac{N_1}{2} + 1, N_2\right) < A_1 \leq -\pi_1\left(\frac{N_1}{2} + 2, N_2\right),$$

  which are clearly incompatible as the left-hand side is positive. No such equilibrium is possible because the intermediary has to subsidize the $\left(\frac{N_1}{2} - 1\right)th$ agent to induce her to switch, but by doing do, he also attracts the $\frac{N_1}{2}th$ agent who is willing to pay a positive fee to interact with all agents of group 2 on the new platform.

- For $n_1 = \frac{N_1}{2}$, the two inequalities become

  $$\pi_1\left(\frac{N_1}{2} + 1, N_2\right) < A_1 \leq \pi_1\left(\frac{N_1}{2}, N_2\right) - \pi_1\left(\frac{N_1}{2} + 1, N_2\right)$$

  which supposes that $\pi_1\left(\frac{N_1}{2}, N_2\right) > 2\pi_1\left(\frac{N_1}{2} + 1, N_2\right)$.

- For $\frac{N_1}{2} < n_1 < N_1$, the two inequalities become

$$\pi_1 (n_1 + 1, N_2) < A_1 \leq \pi_1 (n_1, N_2).$$

As no agent of group 2 stays on the existing platform, the intermediary can charge a positive fee to the agents of group 1 he wants to attract. Such equilibria may then be profitable for the intermediary.

Applying the same logic, the equilibrium would involve no agent switching ($n_1 = 0$) if and only if $A_1 > -\pi_1 (N_1, N_2)$, and all agents switching ($n_1 = N_1$) if and only if $A_1 \leq \pi_1 (N_1, N_2)$.

We can now move to **stage 1** where the intermediary has to find the highest value of A_1 inducing a unique and profitable equilibrium in stage 2. To eliminate the unprofitable equilibrium with $n_1 = 0$, we need to have $A_1 \leq -\pi_1 (N_1, N_2)$. Such a subsidy also eliminates all equilibria with $\frac{N_1}{2} \leq n_1 < N_1$. The remaining equilibria are those with $0 < n_1 < \frac{N_1}{2} - 1$ and the one with $n_1 = N_1$. The former equilibria are clearly non profitable as a subsidy has to be paid to the n_1 agents of group 1 while no agent of group 2 will follow. To eliminate them, we take the most stringent condition, which is obtained for $n_1 = \frac{N_1}{2} - 2$ as π_1 decreases in its first argument; hence, we need to impose $A_1 \leq -\pi_1 \left(\frac{N_1}{2} + 2, N_2\right)$. Under this condition, the unique equilibrium is $n_1 = N_1$. This proves the proposition. ■

As the condition for positive profits cannot be ascertained with generic benefit functions, we use the profit functions derived in the following specific case.

Example 3 Successive vertical oligopoly. *Suppose that n_1 "sellers" of a homogeneous input compete à la Cournot and sell the input to n_2 "buyers", who transform it into independent final products (which implies that rivalry exists between the sellers but not between the buyers); suppose also that the intermediary attracts sellers first. Then, the intermediary finds a profitable way to enter*

Proof. In this setting, profits are computed as (see Appendix 7.1):

$$\pi_1 (n_1, n_2) = \frac{n_2}{2 (n_1 + 1)^2} \text{ and } \pi_2 (n_1) = \frac{n_1^2}{4 (n_1 + 1)^2}.$$

Hence,

$$-N_1\pi_1\left(\frac{N_1}{2}+2,N_2\right)+N_2\pi_2\left(N_1\right)=\frac{N_1N_2}{4}\frac{\left(N_1-2\right)^3+2N_1\left(5N_1+4\right)}{\left(N_1+6\right)^2\left(N_1+1\right)^2}>0,$$

meaning that the intermediary finds a profitable way to enter. ■

4.2 Sequential switching with rivalry in the second group

We now inverse the roles: agents of group 1, who are moving first, earn $\pi_1\left(n_2\right)$, and agents of group 2, who are moving second, earn $\pi_2\left(n_1,n_2\right)$. In **stage 4**, the rivalry between agents implies that any partition of group 2 can emerge at the equilibrium. An equilibrium with $0<n_2<N_2$ firms switching requires

$$\begin{cases}\pi_2\left(n_1,n_2\right)-A_2\geq\pi_2\left(N_1-n_1,N_2-n_2+1\right),\\ \pi_2\left(N_1-n_1,N_2-n_2\right)>\pi_2\left(n_1,n_2+1\right)-A_2.\end{cases}$$

Define

$$a_2\left(n_1,n_2\right)\equiv\pi_2\left(n_1,n_2\right)-\pi_2\left(N_1-n_1,N_2-n_2+1\right).$$

The two conditions for an equilibrium with $0<n_2<N_2$ can then be reexpressed as

$$a_2\left(n_1,n_2+1\right)<A_2\leq a_2\left(n_1,n_2\right).$$

Similarly, we have an equilibrium with $n_2=0$ if $A_2>a_2\left(n_1,1\right)$ and an equilibrium with $n_2=N_2$ if $A_2\leq a_2\left(n_1,N_2\right)$.

The function $a_2\left(n_1,n_2\right)$ measures the highest fee an agent of group 2 is willing to pay for joining the new platform on which n_1 agents of group 1 and n_2 agents of group 2 are active. As $a_2\left(n_1,n_2\right)$ decreases with n_2, it can be seen as as demand function for the new platform; as $a_2\left(n_1,n_2\right)$ increases with n_1, this demand function moves upward when more agents of group 1 are present on the new platform.

As $a_2\left(n_1,n_2\right)$ decreases with n_2, the $\left(n_2+1\right)$ above conditions define a sequence of adjacent intervals, meaning that any value of A_2 corresponds to a unique equilibrium. Hence, in **stage 3**, the intermediary's problem is equivalent to choosing the value of n_2 that maximizes its revenue: $n_2a_2\left(n_1,n_2\right)$. To proceed with the solution of the game, we impose some additional structure on the generic benefit functions. We make the following two assumptions:

Assumption 1. $n_2 a_2 (n_1, n_2)$ is a single-peaked function of n_2;

Assumption 2. the unique maximum of $n_2 a_2 (n_1, n_2)$, $n_2^* (n_1)$, is weakly increasing in n_1.

These two assumptions seem natural: Assumption 1 says that for each value of n_1, the intermediary determines a unique value of n_2 in stage 3, and Assumption 2 says that this value does not decrease if more agents of group 1 have been attracted beforehand.[13] As will be shown below, these two assumptions are fulfilled in the specific example of successive Cournot oligopolies. Naturally, $n_2^* (0) = 0$. Indeed, if no agent of group 1 has switched beforehand, the intermediary will have to pay subsidies to agents of group 2, which it cannot credibly do at this stage (as it always has the possibility to shut its activity down).

Before turning to stages 1 and 2, which are slightly more technical, we summarize our results in the next proposition.

Proposition 4 *In the sequential switching game with rivalry in the second group, the intermediary sets* $A_1 = \pi_1 (n_2^* (1)) - \pi_1 (N_2) \leq 0$ *and* $A_2 = \pi_2 (N_1, n_2^* (N_1))$, *where* $n_2^* (n_1) = \arg\max_{n_2} n_2 a_2 (n_1, n_2)$. *Thereby, it attracts all* N_1 *agents of group 1 and a number* $n_2^* (N_1)$ *agents of group 2. This scheme is profitable as long as*

$$N_1 (\pi_1 (n_2^* (1)) - \pi_1 (N_2)) + n_2^* (N_1) \pi_2 (N_1, n_2^* (N_1)) > 0.$$

Proof. The proof requires the analysis of stages 2 and 1. Moving to **stage 2**, we first show that there is *no equilibrium where group 1 is partitioned between the two platforms.* An equilibrium with $0 < n_1 < N_1$ would require:

$$\begin{cases} A_1 \leq \pi_1 (n_2^* (n_1)) - \pi_1 (N_2 - n_2^* (n_1 - 1)) \equiv \bar{A}_1, \\ A_1 > \pi_1 (n_2^* (n_1 + 1)) - \pi_1 (N_2 - n_2^* (n_1)) \equiv \underline{A}_1. \end{cases}$$

[13] Considering $a_2 (n_1, n_2)$ as continuous in its two arguments, Assumptions 1 and 2 are fulfilled if we have (i) $2 (\partial a_2 / \partial n_2) + n_2 (\partial^2 a_2 / \partial n_2^2) < 0$ and (i) $(\partial a_2 / \partial n_1) + n_2 (\partial^2 a_2 / \partial n_1 \partial n_2) > 0$.

For the two inequalities to be compatible, we need $\bar{A}_1 > \underline{A}_1$ or

$$\pi_1 (N_2 - n_2^* (n_1)) - \pi_1 (N_2 - n_2^* (n_1 - 1)) > \pi_1 (n_2^* (n_1 + 1)) - \pi_1 (n_2^* (n_1)) .$$

The right-hand side is non-negative as, from Assumption 2, $n_2^* (n_1 + 1) \geq n_2^* (n_1)$ and π_1 is an increasing function; by the same token, the left-hand side is non-positive. We thus have a contradiction, which proves our result.

It follows that the two potential equilibria at stage 2 are: $n_1 = 0$ if and only if $A_1 > \pi_1 (n_2^* (1)) - \pi_1 (N_2)$, and $n_1 = N_1$ if and only if $A_1 \leq \pi_1 (n_2^* (N_1)) - \pi_1 (N_2 - n_2^* (N_1 - 1))$.

In **stage 1**, we require unique implementation of the (potentially) profitable equilibrium, i.e., $n_1 = N_1$. Therefore, we need $A_1 \leq \pi_1 (n_2^* (N_1)) - \pi_1 (N_2 - n_2^* (N_1 - 1))$, so that $n_1^* = N_1$ is an equilibrium, *and* $A_1 \leq \pi_1 (n_2^* (1)) - \pi_1 (N_2)$, so that $n_1^* = 0$ is not an equilibrium. Using Assumption 2, we can establish that the latter condition is more stringent than the former. Indeed,

$$\pi_1 (n_2^* (1)) - \pi_1 (N_2) < \pi_1 (n_2^* (N_1)) - \pi_1 (N_2 - n_2^* (N_1 - 1)) \iff$$
$$\pi_1 (n_2^* (1)) - \pi_1 (n_2^* (N_1)) < \pi_1 (N_2) - \pi_1 (N_2 - n_2^* (N_1 - 1)) ,$$

where the left-hand side is non-positive and the right-hand side, non-negative.
■

As above, we examine whether the profitability condition is satisfied in our specific example.

Example 5 Successive vertical oligopoly (continued). *Suppose now that n_2 "sellers" of a homogeneous input compete à la Cournot and sell the input to n_1 "buyers", who transform it into independent final products; suppose also that the intermediary attracts buyers first. Then, the intermediary attracts all buyers and a single seller. Moreover, the intermediary, finds it profitable to enter only when the initial numbers of buyers and sellers in the industry are relatively low. Otherwise, it is too costly to subsidize the entire group of buyers given the fee that the intermediary will be able to extract from the monopoly seller.*

Proof. Profits are computed as

$$\pi_1 (n_2) = \frac{n_2^2}{4 (n_2 + 1)^2} \text{ and } \pi_2 (n_1, n_2) = \frac{n_1}{2 (n_2 + 1)^2}.$$

To apply the proposition, we need to compute $n_2^*(N_1)$ and $n_2^*(1)$ where $n_2^*(n_1)$ maximizes

$$n_2 a_2(n_1, n_2) = n_2 \frac{n_1}{2(n_2+1)^2} - n_2 \frac{N_1 - n_1}{2(N_2 - n_2 + 2)^2}.$$

For $n_2 \geq 1$, the two terms decrease with n_2. There are thus two possible optima: $n_2^*(n_1) \in \{0, 1\}$. For $n_1 = N_1$, it is obvious that $n_2^*(N_1) = 1$ is the optimum. For $n_1 = 1$, the optimum is $n_2^*(1) = 1$ if and only if $a_2(1,1) > 0$, i.e., iff $N_2^2 + 2N_2 + 5 \geq 4N_1$; otherwise the optimum is $n_2^*(1) = 0$. Therefore, we need to consider two cases.

- If $N_2^2 + 2N_2 + 5 \geq 4N_1$, the intermediary sets $A_1 = \pi_1(1) - \pi_1(N_2)$ and $A_2 = \pi_2(N_1, 1)$, which attracts all buyers and a single seller. The intermediary's profit is positive if and only if $N_2^2 - 6N_2 - 3 < 0$, which requires $N_2 \leq 6$ (and a necessary condition for this to be compatible with $N_2^2 + 2N_2 + 5 \geq 4N_1$ is $N_1 \leq 13$).
- If $N_2^2 + 2N_2 + 5 < 4N_1$, the intermediary sets $A_1 = -\pi_1(N_2)$ and $A_2 = \pi_2(N_1, 1)$, which attracts all buyers and a single seller. The intermediary's profit is positive if and only if $N_2^2 - 2N_2 - 1 < 0$, which requires $N_2 = 2$ (and a necessary condition for this to be compatible with $N_2^2 + 2N_2 + 5 < 4N_1$ is $N_1 \geq 4$).

It follows that the intermediary finds it profitable to enter only when the initial numbers of buyers and sellers in the industry are relatively low. Otherwise, it is too costly to subsidize the entire group of buyers given the fee that the intermediary will be able to extract from the monopoly seller. ■

4.3 Simultaneous switching game

The main result we obtain the simultaneous switching game is summarized in the following proposition.

Proposition 6 *In the simultaneous switching game with partial rivalry, the best possible option for the intermediary is to attract all agents of the non-rival group and either all or a strict subset of the agents of the rival group.*

In the former case, the intermediary divides the rival group and conquers the non-rival group; the opposite prevails in the latter case. Whether any of these options is profitable depends on the specificities of the two groups.

To prove this proposition, let group 1 be the rival group, so that the benefit functions are $\pi_1(n_1, n_2)$ and $\pi_2(n_1)$. In **stage 2**, the absence of rivalry in group 2 implies that all agents make the same decision, meaning that the only two possible equilibrium values for n_2 are 0 if $A_2 > \pi_2(n_1) - \pi_2(N_1 - n_1)$, and N_2 if $A_2 \leq \pi_2(n_1) - \pi_2(N_1 - n_1)$. As for group 1, all partitions can emerge at equilibrium. Agents of group 1 split between the two platforms if

$$\begin{cases} \pi_1(n_1, n_2) - A_1 \geq \pi_1(N_1 - n_1 + 1, N_2 - n_2), \\ \pi_1(N_1 - n_1, N_2 - n_2) > \pi_1(n_1 + 1, n_2) - A_1. \end{cases}$$

They all stay on the existing platform if $A_1 > \pi_1(1, n_2) - \pi_1(N_1, N_2 - n_2)$, or they all switch to the new platform if $A_1 \leq \pi_1(N_1, n_2) - \pi_1(1, N_2 - n_2)$. We thus have six possible types of equilibria, under the following sets of conditions, which are represented in Figure 2.

$$\begin{array}{llll} (0,0) \quad \text{if} \begin{cases} A_1 > -\pi_1(N_1, N_2) \\ A_2 > -\pi_2(N_1), \end{cases} & & (N_1, N_2) \quad \text{if} \begin{cases} A_1 \leq \pi_1(N_1, N_2) \\ A_2 \leq \pi_2(N_1), \end{cases} \\ (0, N_2) \quad \text{if} \begin{cases} A_1 > \pi_1(1, N_2) \\ A_2 \leq -\pi_2(N_1), \end{cases} & & (N_1, 0) \quad \text{if} \begin{cases} A_1 \leq -\pi_1(1, N_2) \\ A_2 > \pi_2(N_1), \end{cases} \end{array}$$

$$(n_1, 0) \quad \text{if} \begin{cases} -\pi_1(N_1 - n_1, N_2) < A_1 \leq -\pi_1(N_1 - n_1 + 1, N_2), \\ A_2 > \pi_2(n_1) - \pi_2(N_1 - n_1), \end{cases}$$

$$(n_1, N_2) \quad \text{if} \begin{cases} \pi_1(n_1 + 1, N_2) < A_1 \leq \pi_1(n_1, N_2), \\ A_2 \leq \pi_2(n_1) - \pi_2(N_1 - n_1). \end{cases}$$

Considering **stage 1**, we observe that in order to implement a unique and (potentially) profitable equilibrium, the intermediary must choose fees (A_1, A_2) in one of the two shaded areas of Figure 2: at least one fee must be positive and the intermediary excludes equilibria in which none of the firms that face the positive fee join the new platform. Let us consider these two

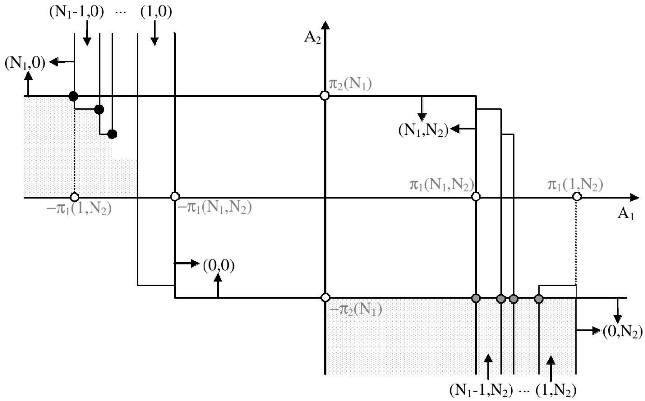


Figure 2: Simultaneous switching game with partial rivalry

areas in turn. In the *top-left area*, the equilibrium is (N_1, N_2), a subsidy is paid to the agents of group 1 ($A_1 < 0$) and a fee is charged to the agents of group 2 ($A_2 > 0$). The intermediary maximizes his profits by choosing, for a given A_i, the highest possible A_j within the limits of the area. The optimum is to be found among the black dots in Figure 2, which are defined by $A_1 = -\pi_1(x, N_2)$ and $A_2 = \pi_2(N_1 - x + 1) - \pi_2(x - 1)$, with $1 \leq x \leq N_1/2$. Note that x is smaller than or equal to $N_1/2$ to guarantee that the fee paid by firms of group 2 is positive. The subsidy given to group 1 is thus computed so as to eliminate unprofitable equilibria where n_1 agents of group 1 switched and are not followed by any agent of group 2. Hence, in this area, the intermediary solves:

$$\max_x \Pi_I^{TL} = -N_1\pi_1(x, N_2) + N_2(\pi_2(N_1 - x + 1) - \pi_2(x - 1)).$$

In the *bottom-right area*, equilibria are of the type (n_1, N_2) with $1 \leq n_1 \leq N_1$. Here, a fee is charged to the rival group ($A_1 > 0$), whereas the non-rival group is subsidized: $A_2 = -\pi_2(N_1)$, which corresponds to the lowest subsidy excluding the no-participation equilibrium. Again, profit maximization calls for selecting the highest possible A_j for a given A_i within the limits of the area. The intermediary chooses thus between the gray dots in Figure 2 and finds the optimum solving:

$$\max_{n_1} \Pi_I^{BR} = n_1 \pi_1 (n_1, N_2) - N_2 \pi_2 (N_1).$$

The generic benefit functions do not allow us to state more than the above proposition. To ascertain whether the intermediary can enter profitably, we return to our specific example. As we show below, none of the two options yields a positive profit, which stands in sharp contrast with the sequential switching games where there always existed room for profits.

Example 7 Successive vertical oligopoly (continued). *Suppose now that n_1 "sellers" of a homogeneous input compete à la Cournot and sell the input to n_2 "buyers", who transform it into independent final products. Then, the intermediary is unable to profitably launch his new platform (unless $N_1 = 2$).*

Proof. As before, profits are given by

$$\pi_1 (n_1, n_2) = \frac{n_2}{2 (n_1 + 1)^2} \text{ and } \pi_2 (n_1) = \frac{n_1^2}{4 (n_1 + 1)^2}.$$

The maximization problem in the bottom-right area writes as

$$\max_{n_1} \Pi_I^{BR} = n_1 \tfrac{N_2}{2(n_1+1)^2} - N_2 \tfrac{N_1^2}{4(N_1+1)^2}.$$

From the first-order and second-order conditions, it is easily seen that the optimal value is $n_1 = 1$. At $n_1 = 1$, we have

$$\Pi_I^{BR*} = -N_2 \tfrac{N_1^2 - 2N_1 - 1}{8(N_1+1)^2} < 0 \text{ for } N_1 > 2.$$

In the top-left area, the intermediary chooses the value of x (with $1 \leq x < N_1/2$) that maximizes

$$\Pi_I^b = -\tfrac{N_1 N_2}{2(x+1)^2} + N_2 \left(\tfrac{(N_1 - x + 1)^2}{4(N_1 - x + 2)^2} - \tfrac{(x-1)^2}{4x^2} \right).$$

Note that in the second term, $(N_1 - x + 1)^2 / (N_1 - x + 2)^2 < 1$. Therefore,

$$\Pi_I^b < -\tfrac{N_1 N_2}{2(x+1)^2} + N_2 \left(\tfrac{1}{4} - \tfrac{(x-1)^2}{4x^2} \right) = -\tfrac{1}{4} \tfrac{2x^2 (N_1 - 2x) + (2x+1)(x-1)^2}{(x+1)^2 x^2} < 0$$

Hence, $\Pi_I^b < 0$. ∎

4.4 Discussion

This section has emphasized two main results that are illustrated by the successive vertical oligopoly example. First, rivalry in one group does not necessarily ease the launch of a profitable new platform. In the previous section, we showed that if both groups were non rival, the intermediary could always launch a profitable platform. By contrast, we show in this section that in the successive vertical oligopoly example, rivalry on one side does not make the new platform profitable if the game is simultaneous, and if the game is sequential with few agents and with rival agents targeted in the second place.

Second, simultaneity in the game creates multiple equilibria. Some of these equilibria are non profitable for the intermediary whereas a good equilibrium could also occur for the same set of fees. By using sequentiality, the intermediary is free to set the fee for the second group after he has attracted some firms from the first group. This flexibility is sufficient to eliminate the non profitable equilibrium and to guarantee that the other equilibrium will emerge.

5 Full rivalry

We suppose now that rivalry prevails in both groups of agents. In each group agents are willing to differentiate from other agents. Thus, it is easy to attract some of them to the new platform that offers them a way to differentiate. However, it is difficult for the new platform to attract all agents of one group because of this desire to differentiate from rivals. Under partial rivalry the intermediary sets fees so as to attract all non rival agents. This makes the platform attractive for rival agents. Here, rivalry in both groups makes it difficult to attract all agents of one group. Hence, it is more costly to make the new platform attractive for one group if there are few agents of the other group. Thus, rivalry in both groups creates opposite forces: it eases the attraction of a few set of agents from both groups to the new platform, but it makes it difficult to attract large groups which makes the new platform less attractive. It is thus not clear whether the new platform will be easier to launch than under partial rivalry.

The generic benefit functions are thus $\pi_1(n_1, n_2)$ and $\pi_2(n_1, n_2)$. Because of the multiple sources of externalities, it is not surprising that the analysis becomes rather intricate and that almost no general insight can be drawn when using generic benefit functions. Let us briefly show why.

5.1 Generic benefit functions

In the *sequential switching game*, stages 3 and 4 are as in the partial rivalry case with rivalry in the second group. That is, **stage 4** defines the decreasing "demand function " $a_2(n_1, n_2)$ for the new platform, and the intermediary maximizes $n_2 a_2(n_1, n_2)$ in **stage 3**. This determines a unique maximum, $n_2^*(n_1)$, which is non-decreasing in n_1 (under Assumptions 1 and 2). Moving now to **stage 2**, an equilibrium with $0 < n_1 < N_1$ requires:

$$\begin{cases} \pi_1(n_1, n_2^*(n_1)) - A_1 \geq \pi_1(N_1 - n_1 + 1, N_2 - n_2^*(n_1 - 1)), \\ \pi_1(N_1 - n_1, N_2 - n_2^*(n_1)) > \pi_1(n_1 + 1, n_2^*(n_1 + 1)) - A_1. \end{cases}$$

Defining

$$a_1(n_1) \equiv \pi_1(n_1, n_2^*(n_1)) - \pi_1(N_1 - n_1 + 1, N_2 - n_2^*(n_1 - 1)), \qquad (2)$$

we can rewrite the previous two conditions as $a_1(n_1 + 1) < A_1 \leq a_1(n_1)$. The problem here is that there is no guarantee that these two inequalities define an open interval. Indeed, an increase in n_1 has two contrasting effects on $a_1(n_1)$. On the one hand, there is a direct negative effect stemming from the rivalry among agents of group 1: as n_1 increases, $\pi_1(n_1, .)$ decreases and $\pi_1(N_1 - n_1 + 1, .)$ increases, making a_1 decrease. On the other hand, there is a positive indirect effect through the increase in the number of agents of group 2 who will be attracted subsequently: as n_1 increases, so do $n_2^*(n_1)$ and $n_2^*(n_1 - 1)$, making $\pi_1(., n_2^*(n_1))$ increase, $\pi_1(., N_2 - n_2^*(n_1 - 1))$ decrease, and so a_1 increase (at least weakly). In other words, we do not know whether $a_1(n_1)$ corresponds to a decreasing demand function for the new platform and, hence, *we are unable to solve the first two stages of the game.*

In the *simultaneous switching game*, any partition of the two groups of firms can now emerge at equilibrium. There are thus nine possible equilibrium configurations. As we cannot balance the effect of the conflicting inter- and intra-group externalities, we cannot either delineate the areas in the (A_1, A_2) plane for which each possible equilibrium obtains; in particular, we cannot identify precisely the areas with multiple equilibria and, therefore, we cannot express the conditions for unique implementation. Worse, specific profit functions do not really help as it appears that there exist combinations of fees for which there are multiple equilibria with positive participation of both groups to the new platform. (to be extended with an example).

As a consequence, we restrict ourselves to examine the sequential switching game in the specific setting of a successive vertical oligopoly.

5.2 Successive vertical oligopoly

In this example, benefit functions are derived from the interaction between a set of sellers and a set of buyers on the market for an intermediate input. The difference with the previous examples is that we assume now that the buyers of the input transform it into a homogeneous final good, meaning that rivalry also exists among buyers. The successive Cournot oligopoly is analyzed in Appendix 7.1. We just reproduce here the equilibrium profits obtained by a typical seller and a typical buyer on a platform where n_s sellers and n_b buyers are active:[14]

$$\begin{aligned} \text{Seller's profit} \quad &: \quad \pi_s(n_s, n_b) = \frac{n_b}{(n_s+1)^2 (n_b+1)}, \\ \text{Buyer's profit} \quad &: \quad \pi_b(n_s, n_b) = \frac{n_s^2}{(n_s+1)^2 (n_b+1)^2}. \end{aligned}$$

It is easily checked that both functions exhibit positive inter-group externalities, as well as negative intra-group externalities.

Because the two functions differ, the order of moves for the two groups matter. We therefore contrast a 'buyers first' scenario and a 'sellers first' scenario. However, despite the asymmetry between the two groups, we will see that the equilibria of the two scenarios are qualitatively similar. A first similarity is observed at the last two stages of the game. As we mentioned

[14]Recall that we assume single-homing on the part of both sellers and buyers. In this specific example, one can think of the platforms as being two separate geographical markets or two virtual marketplaces with exclusive membership contracts. (complete if necessary)

above in the general analysis, stage 4 defines the decreasing "demand function ", $a_2(n_1, n_2)$, expressed by the second group for the new platform and in stage 3, the intermediary chooses n_2 so as to maximize

$$n_2 a_2(n_1, n_2) = n_2 \pi_2(n_1, n_2) - n_2 \pi_2(N_1 - n_1, N_2 - n_2 + 1). \qquad (3)$$

Here, whether sellers or buyers are the second group, *the intermediary never finds it optimal to attract more than a single firm.* To understand this result, we first note that in a Cournot market for a homogeneous product with linear costs (which is the case both on the sellers' and on the buyers' side), industry profits are maximized under monopoly (see Amir, 2003). We indeed check that $n_s \pi_s(n_s, n_b)$ decreases with n_s and $n_b \pi_b(n_s, n_b)$ decreases with n_b. Second, as more agents of one group move to the new platform, the fall back position of each agent (i.e., the profit this agent would achieve by unilaterally switching back to the existing platform) improves; this means that the total compensation the intermediary has to pay, $n_2 \pi_2(N_1 - n_1, N_2 - n_2 + 1)$, increases with n_2. In sum, the first term in expression (3) decreases with n_2, while the second increases with n_2; as the second term is preceded by a minus sign, the whole expression decreases with n_2 and this result is independent of the partition of the first group between the two platforms. Hence, for all $n_1 \geq 1$, $n_2 a_2(n_1, n_2)$ decreases with n_2, meaning that $n_2^*(n_1) \leq 1$. We still need to check whether $a_2(n_1, 1)$ is positive (because, otherwise, $n_2^*(n_1) = 1$ just minimizes the intermediary's losses). As $a_2(n_1, 1)$ is an increasing function of n_1, it suffices to show that $a_2(1, 1) \geq 0$. We check that, for $N_s, N_b \geq 3$, we have

$$\begin{aligned} a_{2s}(1,1) &= \pi_s(1,1) - \pi_s(N_s, N_b - 1) = \tfrac{\left(N_s^2+2N_s-7\right)(N_b-1)+(N_s+1)^2}{8N_b(N_s+1)^2} > 0, \\ a_{2b}(1,1) &= \pi_b(1,1) - \pi_b(N_s - 1, N_b) = \tfrac{(N_sN_b+5N_s-4)(N_s(N_b-3)+4)}{16N_s^2(N_b+1)^2} > 0. \end{aligned}$$

The intermediary's optimum at **stage 3** is thus given by:

$$n_2^*(n_1) = \begin{cases} 1, & \forall n_1 \geq 1 \\ 0 & \text{if } n_1 = 0, \end{cases}$$

with $n_2^*(n_1) = n_s^*(n_b)$ in the 'buyers' first' scenario, and $n_2^*(n_1) = n_b^*(n_s)$ in the 'sellers first' scenario.

We can now move to **stage 2** and rewrite expression (2), which measures the willingness to pay for agents of group 1 to join the new platform, as:

$$\begin{aligned} a_1(n_1) &= \pi_1(n_1, n_2^*(n_1)) - \pi_1(N_1 - n_1 + 1, N_2 - n_2^*(n_1 - 1)) \\ &= \begin{cases} \pi_1(1,1) - \pi_1(N_1, N_2) & \text{for } n_1 = 1, \\ \pi_1(n_1, 1) - \pi_1(N_1 - n_1 + 1, N_2 - 1) & \text{for } 2 \leq n_1 \leq N_1. \end{cases} \end{aligned}$$

The equilibrium involves $0 < n_1 < N_1$ firms switching to the new platform if and only if $a_1(n_1 + 1) < A_1 \leq a_1(n_1)$. Similarly, all firms switch at equilibrium if and only if $A_1 \leq a_1(N_1)$, and no firm switches if and only if $A_1 > a_1(1)$. For $2 \leq n_1 < N_1$, it is clear that $a_1(n_1) > a_1(n_1 + 1)$, so that the equilibria with n_1 ranging from 2 to N_1 are all possible and exclusive of one another. We still need to check whether $a_1(1) > a_1(2)$. This is so as long as

$$\pi_1(1,1) - \pi_1(N_1, N_2) > \pi_1(2,1) - \pi_1(N_1 - 1, N_2 - 1).$$

We show in Appendix 7.1 that this inequality is satisfied in the two scenarios. This implies that each value of A_1 determines a unique equilibrium in stage 2. Naturally, among all values of A_1 inducing the same equilibrium, say with n_1 firms switching, the intermediary will select the largest, i.e., $a_1(n_1)$.

The latter findings allow us to fully describe the intermediary's problem in **stage 1**:

$$\max_{n_1} \Pi_I(n_1) = n_1 a_1(n_1) + a_2(n_1, 1).$$

The first term in this profit function is the total fee (or subsidy) charged (or paid) to the firms of the first group, and the second term is the fee charged to the single firm of the second group.

We demonstrate in Appendix ??? that the following results hold in the two scenarios: (i) $\Pi_I(n_1)$ is single-peaked; this implies that there is a single value, n_1^*, that maximizes the intermediary's profit and that can be approximated by the value of n_1 such that $d\Pi_I(n_1)/dn_1 = 0$ (it is actually the integer that is the closest to this value); (ii) $\Pi_I(n_1^*) > 0$; (iii) n_1^* is comprised between 1 and $N_1/3$; (iv) n_1^* is non decreasing in N_1 and non increasing in N_2. Moreover, comparing the two scenarios, it appears that it is more profitable to attract buyers first if $N_b \geq N_s$ and to attract sellers first otherwise.[15]

Collecting these results, we can state:

Proposition 8 *In the sequential switching game with rival sellers and rival buyers, the intermediary's optimal strategy is to attract no more than one third of the firms of the first group, and to grant a monopoly to one firm of the second group. This strategy yields positive profits whichever group of firms is attracted first, but it is generally more profitable to attract the largest group first.*

5.3 Discussion

Under partial rivalry, the profitability of the new platform depends on the group that is targeted first. Under full rivalry, our example shows that the launch of the new platform is profitable whatever the group that is targeted first. The small size of the new platform (monopoly on one side and less than one third of firms on the other side) reduces the attractiveness of the platform. However, the small size of the platform also offers all agents an opportunity to differentiate from rivals. The specific example used in the proposition shows that the intermediary finds a profitable way to deal with this trade-off.

In the absence of rivalry, the new platform attracts all agents. Under rivalry on one side, the new platform does not attract all rival agents if rivals are targeted second. Under rivalry on both sides, our example shows that the platform attracts few agents of both groups. Hence, rivalry tends to reduce the size of the new platform. However, it has an ambiguous effect on the profitability of the new platform.

Under partial rivalry, the intermediary divides and conquers: he subsidizes rival agents (the sellers) if he seeks to attract all of them (in the sellers first strategy), whereas he subsidizes the non rival agents (the buyers) if he seeks to attract few rival agents (in the buyers first strategy). The first group that is attracted is subsidized and all members of that group switch

[15] This rule admits a few exceptions in very small industries (i.e., when $N_s \leq 6$ and $N_b \leq 4$).

to the new platform (divide strategy). The other group has then no other choice than to switch and to pay a positive fee (conquer strategy).

By contrast, all agents pay a positive fee in the specific example of full rivalry (this is proved in Appendix ???). The reason is that both groups wants to differentiate and are keen to join the new small platform. The first group accepts to pay positive fees because they face less than one third of the competing agents and the second group also agree to pay positive fees because the firm that switches benefits from a monopoly. The positive indirect network effect is not sufficient to counter the benefits of smaller competition.

Under full rivalry, the willingness to move to a market with a fewer number of agents is stronger on the side which counts the largest number of agents. It is thus easier for the intermediary to attract a large number of agents of this side. Because the intermediary must attract a relatively large number of buyers in the buyers first strategy, and a relatively large number of sellers in the sellers first strategy, he chooses the strategy for which it is easier to attract agents: buyers first if $N_b \geq N_s$, sellers first otherwise. With many firms easily attracted on board, the intermediary is able to extract a larger rent from the single firm that he will attract from the other side afterwards.

6 Conclusion

(to be written)

References

[1] Abiru, M., Nahata, B., Raychaudhuri, S., and Waterson, M. (1998). Equilibrium structures in vertical oligopoly. *Journal of Economic Behavior & Organisation* **37**: 463-480.

[2] Amir, R. (2003). Market structure, scale economies and industry performance. CORE Discussion Paper 2003/65. Université Catholique de Louvain, Belgium.

[3] Armstrong, M. (2004). Competition in two-sided markets. Mimeo, Nuffield College, Oxford.

[4] Belleflamme, P. and Toulemonde, E. (2004). Emergence and entry of B2B markets. CORE Discussion Paper 2004/78. Université Catholique de Louvain, Belgium.

[5] Brunn, P., Jensen, M., and Skovgaard, J. (2002). e-markets: Crafting a winning strategy. *European Management Journal* **20**: 286-298.

[6] Caillaud, B., and Jullien B. (2003). Chicken and egg: Competition among intermediation service providers. *RAND Journal of Economics* **34**: 521-552.

[7] Evans, D.J. (2003). The antitrust economics of multi-sided platform markets. *Yale Journal on Regulation* **20**: 325-381.

[8] Gabszewicz, J., and Wauthy, X. (2004). Two-sided markets and price competition with multi-homing. CORE Discussion paper 2004/30. Louvain-la-Neuve, Belgium.

[9] Hagiu, A. (2004). Optimal pricing and commitment in two-sided markets. Mimeo.

[10] Jullien, B. (2004). Two-sided markets and electronic intermediaries. Mimeo. CES ifo Economic Studies Conference on 'Understanding the Digital Economy: Facts and Theory'. Munich.

[11] Kaplan, S., and Sawhney, M. (2000). E-hubs: the new B2B markets. *Harvard Business Review* **78(3)**: 97-106

[12] Katz, M., and Shapiro, C. (1985). Network externalities, competition and compatibility. *American Economic Review* **75**: 424-440.

[13] Lucking-Reiley, D., and Spulber, D.F. (2001). Business-to-Business Electronic Commerce. *Journal of Economic Perspectives* **15**: 55-68.

[14] Milliou, C., and Petrakis, E. (2004). Business-to-business electronic markets: Joining a public or creating a private. *International Journal of Finance and Economics* **9**: 99-112.

[15] Nocke, V., Peitz, M., and Stahl, K. (2004). Platform ownership. Mimeo. University of Manheim, Germany.

[16] Ordanini, A., Micelli, S., and Di Maria, E. (2004). Failure and success of B-to-B exchange business models: A contingent analysis of their performance. *European Management Journal* **22**: 281-289.

[17] Popović, M. (2002). B2B e-markets. Mimeo European Commission's Electronic Commerce Team (Information Society Directorate General). Brussels.

[18] Rochet, J.-C., and Tirole, J. (2002). Cooperation among competitors: Some economics of payment card associations. *RAND Journal of Economics* **33**: 549-570.

[19] Rochet, J.-C., and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association* **1**: 990-1029.

[20] Rochet, J.-C., and Tirole, J. (2004). Two-sided markets: An overview. Mimeo. IDEI, Toulouse.

[21] Schmalensee, R. (2002). Payment systems and interchange fees. *Journal of Industrial Economics* **50**: 103-122.

[22] Soubeyran, A., and Weber, S. (2002). District Formation and Local Social Capital: A (Tacit) Co-opetition Approach. *Journal of Urban Economics* **52**: 65-92.

[23] Yoo, B., Choudhary, V., and Mukhopadhyay, T. (2003). Neutral vs. biased markets: A comparison of electronic B2B markets with different ownership structures. Mimeo. University of California, Irvine.

7 Appendix

7.1 Successive vertical oligopoly

We solve the game backwards, starting with buyers' decisions. Consider a platform with n_b buyers, each one ($i = 1, \dots, n_b$) producing one variety of the final product. The representative consumer has a quadratic surplus function:

$$U(q_1, q_2, \dots, q_{n_b}) = \sum_{i=1}^{n_b} q_i - \frac{1}{2}\left(\sum_{i=1}^{n_b} q_i^2 + \gamma \sum_{i=1}^{n_b}\sum_{j\neq i} q_i q_j\right) - \sum_{i=1}^{n_b} p_i q_i$$

where $0 \leq \gamma \leq 1$ indicates the strength of product substitutability. Maximizing consumer's surplus yields the linear inverse demand schedule $p_i = 1 - q_i - \gamma Q_{-i}$ (with $Q_{-i} = \sum_{j\neq i} q_j$) in the region of prices where quantities are positive. Suppose that the unit cost for a buyer is entirely given by the price w paid for the intermediate product. The first-order condition for profit maximization yields

$$w = 1 - 2q_i - \gamma Q_{-i}.$$

Summing on all firms and taking advantage of symmetry gives

$$w = 1 - \frac{2 - \gamma + \gamma n_b}{n_b} Q.$$

Because of the one-for-one transformation technology, the total quantity of the final product (Q) is equal to the total quantity of the intermediate product (X). The previous expression gives thus the inverse demand function for the sellers. So, seller j (with $j = 1, \dots, n_s$), whose marginal cost of production is assumed to be equal to zero, has the following first-order condition for profit maximization:

$$1 - 2\tfrac{2-\gamma+\gamma n_b}{n_b} x_j - \tfrac{2-\gamma+\gamma n_b}{n_b} X_{-j} = 0.$$

At the symmetric Cournot equilibrium, each seller produces a quantity

$$x = \frac{n_b}{(n_s + 1)(2 - \gamma + \gamma n_b)}.$$

The total quantity exchanged on the platform is thus equal to

$$Q(n_s, n_b) = X(n_s, n_b) = \frac{n_b n_s}{(n_s + 1)(2 - \gamma + \gamma n_b)}.$$

Using the latter expression, one computes the firms' profits at the equilibrium:

$$\begin{aligned} \text{for sellers, } \pi_s(n_s, n_b) &= \frac{n_b}{(n_s+1)^2 (2 - \gamma + \gamma n_b)}, \\ \text{for buyers, } \pi_b(n_s, n_b) &= \frac{n_s^2}{(n_s+1)^2 (2 - \gamma + \gamma n_b)^2}. \end{aligned}$$

In this model, there is always rivalry among sellers as they produce a homogeneous input. However, the degree of rivalry among buyers increases with the value of γ. We consider two cases.

- *No rivalry among buyers* ($\gamma = 0$):

$$\pi_s(n_s, n_b) = \frac{n_b}{2(n_s+1)^2} \text{ and } \pi_b(n_s, n_b) = \frac{n_s^2}{4(n_s+1)^2}.$$

- *(Maximum) rivalry among buyers* ($\gamma = 1$):

$$\pi_s(n_s, n_b) = \frac{n_b}{(n_s+1)^2(n_b+1)} \text{ and } \pi_b(n_s, n_b) = \frac{n_s^2}{(n_s+1)^2(n_b+1)^2}.$$

In the latter case, to complement the analysis of Section 5.2, we establish two useful results:

- $\pi_s(1,1) - \pi_s(N_s, N_b) > \pi_s(2,1) - \pi_s(N_s - 1, N_b - 1)$. Indeed,

$$\begin{aligned} &\pi_s(1,1) - \pi_s(N_s, N_b) - \pi_s(2,1) + \pi_s(N_s - 1, N_b - 1) \\ &= \tfrac{1}{8} - \tfrac{N_b}{(N_s+1)^2(N_b+1)} - \tfrac{1}{18} + \tfrac{N_b-1}{N_s^2 N_b} \\ &= \tfrac{5}{72} - \tfrac{1}{N_s^2 N_b(N_b+1)} + \tfrac{(1+2N_s)N_b}{(N_s+1)^2(N_b+1)N_s^2} > 0 \ \forall N_s, N_b \geq 3. \end{aligned}$$

- $\pi_b(1,1) - \pi_b(N_s, N_b) > \pi_b(1,2) - \pi_b(N_s - 1, N_b - 1)$. Indeed

$$\begin{aligned} &\pi_b(1,1) - \pi_b(N_s, N_b) - \pi_b(1,2) + \pi_b(N_s - 1, N_b - 1) \\ &= \tfrac{1}{16} - \tfrac{N_s^2}{(N_s+1)^2(N_b+1)^2} - \tfrac{1}{36} + \tfrac{(N_s-1)^2}{N_s^2 N_b^2} \\ &= \tfrac{5}{144} - \tfrac{2}{(N_s+1)^2 N_b^2} + \tfrac{1}{(N_s+1)^2 N_b^2 N_s^2} + \tfrac{(2N_b+1)N_s^2}{(N_s+1)^2(N_b+1)^2 N_b^2} > 0 \ \forall N_s, N_b \geq 3. \end{aligned}$$

Network Externalities and Interconnection Incentives

MARGIT A. VANBERG

Centre for Economic Research (ZEW), Mannheim

November 2005

Abstract

The majority of industrial organizations literature on network externalities looks at firm behavior under given market characteristics. The present paper instead asks the question whether the presence of network externalities can change market characteristics, specifically, whether an initially large market player can decline cooperation (interconnection) with competing network operators and thereby gain a dominant position when network externalities are significant. The paper comes to the conclusion that only when a network operator already has network specific market power due to the ownership of a monopolistic bottleneck network area, will network externalities enable the operator to increase his market dominance. In competitive markets or in contestable natural monopolies, however, network externalities will not lend network specific market power to an initially large operator. In these markets, the market process can be expected to solve the trade-off between ensuring cooperation between competing operators and at the same time safeguarding competition in product characteristics and quality of service.

Keywords: network externalities, interconnection, regulation
JEL classification: L15; L43

Address: Centre for European Economic Research (ZEW)
Research Group Information and Communication Technologies
P.O. Box 103443
D- 68034 Mannheim
Phone: +49/621/1235-351
Fax: +49/621/1235-333
E-mail: vanberg@zew.de

1 Introduction

The majority of industrial organizations literature on network externalities treats network externalities in an oligopoly context. The focus in this literature is to understand firm behavior under given market characteristics. The present paper, instead, analyses whether network externalities can, in fact, change market characteristics. What is the effect of network externalities on competition and cooperation in network industries? Can network externalities lend network specific market power to market players and thereby allow one or more firms to dominate the market?

The paper uses the disaggregated regulatory approach (see Knieps, 1997) to isolate the effects that network externalities have on market structure and the competitive process. The analysis comes to the conclusion that only when a network operator already has network specific market power due to the ownership of a monopolistic bottleneck network area, will network externalities enable the operator to increase his market dominance. In this case network externalities increase the need for regulation. Besides network access regulation to the monopolistic bottleneck, open standards need to be enforced between the monopolistic network area and markets related to this area by indirect network effects. In competitive markets or in contestable natural monopolies, however, network externalities will not lend network specific market power to an initially large operator. Market intervention is counterproductive in these cases.

The paper is organized as follows. Section 2 discusses the characteristics of network externalities. Section 3 introduces the disaggregated regulatory approach and discusses relevant parts of the literature on network externalities with a view to the question of how the presence of network externalities influences market processes. Conclusions are presented in section 4.

2 Defining Network Externalities

Network externalities are a special kind of external effect. External effects are present whenever the production or consumption of a good or service results in costs or benefits experienced by third parties without these parties receiving compensation for the costs incurred respectively paying for the benefits received. The lack of a price attached to the external effect prevents market processes from leading to an efficient production or consumption level of these goods.

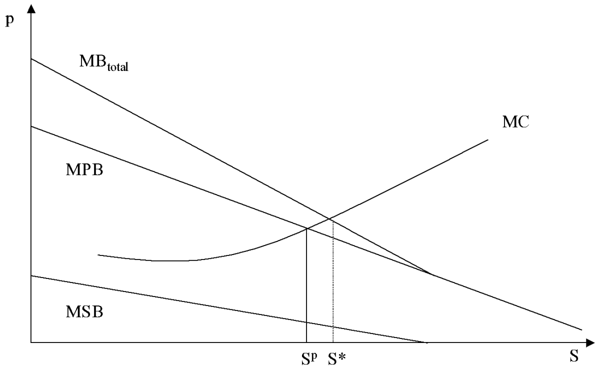
When the benefit received from consuming a good or service is positively related to the number of other purchasers of the good, then this good exhibits positive network externalities. Formally, when the utility $U_{i,j}$ that individual i receives from consuming a particular network technology j is dependent not only on the product characteristics of product j, T, but also on the number of other users of this technology, S, then this utility function reflects network externalities:

$$U_{i,j}(S, T) < U_{i,j}(S', T); \qquad \text{for } S < S'$$

The utility derived from the consumption of a network good can be decomposed into two components (Blankart and Knieps, 1992: 80) the "technology effect", measuring the utility derived from the product characteristics of the network good and the "network effect" measuring the utility derived from the number of other purchasers of the same good. When the network effect is of high importance, as for instance in communication networks, (think, for instance, of fax machines), then the utility of being the only person to consume this communications technology may correspond to $U_{i,j}(1, T) = 0$. On the other hand, when the technology effect is of particular importance, then a consumer can derive a higher utility from consuming the most preferred technology (with characteristics T_1), even when there are less users of this technology compared to the second placed technology (with characteristics T_2): $U_{i,1}(S, T_1) > U_{i,2}(S', T_2)$, for $S < S'$.

The term externality is applied for utility functions exhibiting these properties because the demand effect which results from an increase in S is not internalized through the price system of the market. Existing users do not reveal the extra benefit they obtain from additional users by, for instance, compensating newcomers for the benefit they bring to the installed base. Rather, new consumers base their decision for or against purchasing the technology only on their own preferences for the product characteristics and on their knowledge of the existing and perhaps the expected size of the user base.

Figure 2.1: Unexploited gains from further network participation



While a static analysis is inappropriate to study external effects in the context of dynamic industries the simple diagram depicted in Figure 2.1 serves well for the purpose of illustrating that in the presence of network externalities the equilibrium network size will generally be smaller than the socially optimal network size such that the competitive equilibrium is likely to reflect an inefficient welfare loss. Figure 2.1 shows the marginal cost function (MC), the marginal private benefit (MPB) and the marginal social benefit (MSB) functions for a network good. The total benefit function (MB_{total}) exceeds the private benefit function because it includes the benefits which accrue to other users of the technology when an additional user joins the network. In the social optimum, the level of network participation would be S*, in which the sum of the private and social benefits from further network participation (MB_{total}) are equal to the marginal cost incurred by the marginal consumer. The private equilibrium S^p is smaller then S*, however, because a user will purchase the technology only when the private benefits received exceed the private marginal costs. The private consumption decision leads to too little network participation compared to the social optimum. Only coordination between the users of the technology can internalize the external benefits and increase network participation to S*. Existing consumers would, for instance, have to agree to subsidize the marginal consumer. The question of how and when users will coordinate their consumption decisions in the presence of network externalities will be taken up again in sections 3.3 and 3.4.

2.1 Direct and indirect network externalities

Before analyzing the consequence for the functioning of market mechanisms from the presence of network externalities, it is worthwhile to introduce a few more definitions. In the basic definition of network externalities given above, it was assumed that the network effect is a function only of the number of purchasers of the same good. The literature on network externalities distinguishes more precisely between direct network externalities and indirect network externalities (Katz and Shapiro, 1985: 424). The term direct network externalities refers to the immediate effect the size of a network has on the utility received from participation in that same network. This is the effect introduced above. The term indirect network externalities refers to the effect the size of a network can have on markets for complementary goods or services. When the number of subscribers of a network increases, this can positively affect the variety of complementary goods and services offered on complementary markets. A common illustration of indirect network externalities is the interrelationship between the diffusion of compact disc players and the variety of CDs offered. Early adopters of compact disc players experienced a utility increase from the later broader diffusion of CD players, because only then did the variety of CDs offered reflect the full spectrum of music styles. In general, it can be said that indirect network externalities often occur on markets interrelated by a so-called hardware/software relationship.

2.2 Pecuniary network externalities

A further distinction made in the literature on externalities is the one between pecuniary network externalities and technological network externalities. As will be shown below, pecuniary network externalities are demand effects that are transmitted through the price system, whereas technological network externalities impose benefits (or costs) outside of market mechanisms. Since pecuniary externalities are common in network industries, it is worthwhile to consider this point at length.

Pecuniary network externalities can, for instance, result from an outward shift in demand in a decreasing costs industry. When the demand shift allows the producer to reach a lower level of production costs both new consumers and existing consumers benefit from a lower price level. Formally the utility increase to existing consumers could, of course, also be illustrated by the utility function $U_{i,j}(S,T)$ introduced at the beginning of this chapter. However, the difference to "real" external effects is that in the case of pecuniary externalities the utility increase is only indirectly related to the number of network participants. The direct relationship is between the fall in prices, induced by increased participation, and the resulting increase in consumers' surplus. Existing consumers appropriate the increase in consumer rents as a result of a functioning price system. So, in fact, with pecuniary externalities there is no external social benefit when an additional user joins the network. Obviously, it is inappropriate to use the term 'externality' in the context of pecuniary effects, since they are transmitted through the price mechanism of the market.

Liebowitz and Margolis (1994: 137) find that the traditional differentiation between technological and pecuniary externalities is not consistently applied in more recent work on network effects. The large fixed costs associated with setting up a network infrastructure make economies of scale common in network industries. Networks therefore regularly feature pecuniary effects. Because communications networks are also prominent examples of markets characterized by technological network externalities, the distinction between these effects is not always evident.[1] In a policy context it is however important to distinguish the two effects. Pecuniary externalities should not attract policy intervention because they are a part of a functioning price mechanism. Technological externalities may, however, justify policy directed at internalizing the externality. Section 3 will go deeper into the analysis of when policy intervention is called for and when the market can be expected to generate alternative institutions to internalize external effects, such as, for instance, private standardization committees (see Blankart and Knieps, 1993 and Knieps, 1994).

2.3 Pareto-relevance of network externalities

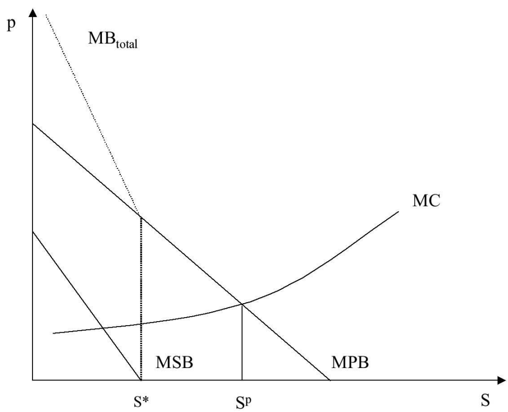
It is important to realize that not all correctly identified technological network externalities result in too little network participation, even when no deliberate internalization of the external effect is given. External effects that are inframarginal, in the sense that they would not change the equilibrium outcome even if internalized into the market mechanism, do not lead to a pareto-inferior equilibrium (Liebowitz and Margolis [1994: 140]). The scenario of inframarginal externalities is a realistic one because it is reasonable to assume that while many networks profit from increasing membership up to a critical size, further growth beyond that point would yield no

[1] It may be that the confusion Liebowitz and Margolis (1994) identify in the context of network industries with respect to the differentiation of pecuniary and technological externalities is explained by the following: Network industries, at least in part, are often characterized by economies of scale in the relevant output region. These network areas, as natural monopolies, lend market power to the incumbent firm, such that the market process cannot be relied upon to bring about the socially optimal outcome. The close relationship between economies of scale and pecuniary effects may explain why pecuniary externalities are sometimes classified as cases in which market mechanisms lead to unwanted outcomes.

further external benefit. Rather, when the critical size of the network is small relative to the number of potential members, several networks can coexist and compete for membership. Therefore, whenever network effects are inframarginal, policy interventions in the market would be misguided.

Figure 2.2 is a graphic illustration of inframarginal network externalities in a static environment. The marginal social benefit from further network participation is zero once network size has reached S*. Because the private equilibrium S^p, where marginal private benefits equal marginal private costs, is to the right of S*, the socially optimal equilibrium coincides with the private equilibrium.

Figure 2.2: Inframarginal network externalities



If one considers policy intervention a viable means of correcting market outcomes which are thought to be inefficient due to externalities, then they can only be justified when the external effects to be corrected are pareto-relevant in equilibrium. If the external effects are inframarginal, intervention will create deviations from the efficient outcome, rather than help to eliminate them. Of course, determining the pareto-relevance of network externalities presupposes that the cost and benefit schedules are known. Because such knowledge is generally not available to policy makers, there is always a great risk of false positives (and false negatives) in these policy decisions.

2.4 Network externalities, product variety and innovation[2]

As was shown in Figure 2.1, internalizing pareto-relevant network externalities requires that consumers gather on one uniform technology as long as the marginal social benefit from further network participation continues to be positive. However, when consumers derive utility not only from network size but also from the product characteristics (the technology effect), internalizing network externalities in this way may not correspond to the socially optimal solution. When consumers' tastes for product characteristics differ, there may be a conflict between internalizing network externalities and catering to individual preferences. Consumers with heterogeneous tastes will not be able to agree on a single preferred set of product characteristics. They may prefer to consume a variety of products offering less network benefits to compromising on product characteristics in order to maximize the utility received from belonging to one network.

This problem is similar to the trade-off between optimal firm size and product variety in the context of production economies of scale. With falling average costs over the relevant output region, exhausting production economies of scale is possible only at the expense of less product diversity. When consumers have heterogeneous tastes, it is however not necessarily in society's interest to decrease product diversity in order to decrease production costs. There is an extensive body of literature devoted to finding the socially optimal solution to this conflict [See for instance Lancaster (1975), Spence (1976), Dixit and Stiglitz (1977)]. A general result of this literature is that when consumers prefer product variety, then higher costs of producing more variety do not necessarily reflect an inefficiency in the market. Rather, the increase in utility from consuming a product close to the individual preferences can compensate for the additional resources devoted to production in this market. Transferring this result to the trade-off between network externalities and product variety indicates that not to reach the optimal network size in a market featuring pareto-relevant network externalities may not indicate an inefficiency when this is a result of more product variety in the market.

---

2 See Blankart and Knieps (1994) for a more elaborate discussion of the trade-off between externalities, variety and innovation.

There is a further reason not to prefer the exhaustion of network externalities by requiring all consumers to choose one network. Only an environment which offers diversity will bring about product innovations. Such innovations are important in dynamic markets with continual technological progress. Consumers participate in the search for better technologies by trying out new products. Markets featuring network externalities are at a disadvantage with respect to this competition of new ideas because new entrants offering newer technologies have no established customer base and can therefore not offer the same benefits from the network effect as established firms. Consumers may be reluctant to try out the new products which have a disadvantage in the network effect and offer only uncertain benefits with respect to their new product characteristics.

It is not likely that decentralized decision-making by consumers will solve the trade-off between externalities, variety and innovation. A single consumer cannot choose the technology he or she would prefer in equilibrium without knowing how the other consumers in the market will decide in different constellations. Because this information is impossible to come by, a consumer can always take a wrong decision. Because of this, the question arises whether policy intervention can help resolve the conflict between network size, variety and innovation. However, also policy makers would need to know the individual consumers' preferences and would have to aggregate them in order to approximate a socially optimal policy. Strictly speaking, policy makers would furthermore have to take into account the development of future technologies to solve not only the conflict between network size and variety but also take into account new innovations. Given that this information cannot be obtained, it is impossible that policy intervention can bring about a socially optimal outcome (Blankart and Knieps, 1994: 458). Therefore, neither free market mechanisms nor public policy is optimal. This discussion is taken up again in section 3.4.

2.5 Summary

The purpose of this section was to draw up the deviations from perfect competition to be expected in markets exhibiting network externalities but to also heighten the awareness for the pitfalls of calling for market intervention whenever externalities are suspected. For one thing, it is possible that the externalities are not technological or even if they are, that they are not relevant in market equilibrium. Furthermore, it was argued that even when network externalities are pareto-relevant in equilibrium, there is a trade-off between network size, variety and search for new technologies, such that internalizing the externalities is not always the optimal solution. Resolving the trade-off between these three goals is far from a simple task. Both non-intervention market outcomes and policy-induced outcomes will not be able to solve the trade-off optimally. Section 3 is devoted to analyzing the functioning of market processes in the presence of network externalities. Based on the findings of this analysis an evaluation of the market outcomes and the possibilities for policy measures is possible.

3 Market processes in the presence of network externalities

The implementation of policy measures geared to internalizing external effects, for example the implementation of a Pigouvian tax, was traditionally considered a viable means of restoring efficient production in the presence of external effects.[3] It was already argued above that such intervention is difficult, particularly in dynamic markets, in which the trade-off between externalities, variety and innovation plays a role. Taking the limitations to policy intervention into account, it is obvious that market outcomes cannot and should not be measured against a hypothetical standard of a social optimum which is not realizable.[4] Rather, the standard should be an outcome which can realistically be expected to emerge from the political process. [5] Given such alternatives, Hayek (1945), for instance, emphasizes that the competitive process is the most efficient means of using the individual knowledge dispersed among the members of society.[6] In his view, the market spontaneously generates institutions in which decentralized choices of consumers can be directed to pursue a common goal and it is much better at doing this than any centralized planning committee. The following sections will analyze whether such a belief in the ability of the market to coordinate decentralized decision-makers is warranted also in the presence of network externalities or whether it is rather to be expected that market processes are inhibited to such a significant degree that even given all its limitations, public policy can be expected to improve on the market outcome.

---

3 See Baumol and Oates (1988) for an overview of the policy instruments which are applied to internalize external effects in the context of environmental management.

4 See also Demsetz' argumentation that policy considerations should not be based on a nirvana approach (Demsetz, 1969).

5 Besides the fact that optimal policy intervention is hindered by missing information, research in political economy further warns that policy makers also lack the necessary knowledge to predict correctly the consequences of their actions [notion of "division of knowledge", Hayek (1945)] and are not oriented on maximizing public welfare [public choice theory, Buchanan (1987)].

6 Austrian economics in general rejects the concept of "social welfare" as a criterion against which market outcomes can be measured. From the point of view of Austrian economics "coordination efficiency", by which is meant the ability of the market process to coordinate the actions of the individual members of society, is an appropriate measure by which the functioning of the market can be evaluated. This concept reveals the relative importance of dynamic processes in Austrian economics as compared to the evaluation of transient equilibria. (See Schmidtchen, 1990, especially at 136-7.)

3.1 A disaggregated regulatory approach to market structure and network externalities

There is a fairly broad body of theoretical economics literature on network externalities and their effect on market equilibrium.[7] Typical research questions in this literature are: What influences consumers' decision on which network to join? How do firms decide whether to cooperate or not? How do the private and the social incentives for compatibility between competing technologies compare?[8] Under what circumstances can a market tip in favor of inferior products or services and become "locked-in"?[9] How should standards be set, i.e. is it optimal to leave standardization to market forces, should industry groups cooperate, or is it optimal to mandate standards?[10]

The following discussion of the literature is structured along the assumptions made therein about market structure. The categories of "monopolistic bottleneck market", "contestable market" and "competitive market" used below are taken from the disaggregated approach to market power regulation (see Knieps, 1997). This approach specializes in the localization of market power in network industries. It uses the concept of network-specific market power to characterize network areas which justify regulatory intervention. A firm has network-specific market power when there is no active competition and when no potential competitor can enter the market such that the incumbent firm can realize supra-competitive profits over an extended time period.

In order to prove the existence of network-specific market power it is therefore essential to understand the entry-conditions in a particular market. The disaggregated regulatory approach uses Stigler's definition of barriers to entry to narrow down the entry conditions which hinder the dissipation of supra-competitive profits even over a longer time period. Stigler defines: "A barrier to entry [...] as a cost of producing (at some or every rate of output) which must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry" (Stigler, 1968: 67). This definition interprets only real asymmetric cost advantages of incumbent firms as barriers to entry, but not, for instance, the capital requirements which entrants face in any industry. This is so because the capital requirements are the same for incumbents as for entrants. From the point of

---

7 Surveys of the literature on network externalities are for instance given in Farrell and Saloner (1987), Gilbert (1992) and Holler (1996).

8 For the first three questions see the seminal papers on network externalities in an oligopoly context by Farrell and Saloner (1985, 1986) and Katz and Shapiro (1985, 1986).

9 See, for instance, David (1985) and Arthur (1989).

view of the disaggregated regulatory approach, wider definitions of entry-barriers than Stigler's definition are not suited to localizing network-specific market power. Definitions which treat mere hindrances to immediate market entry, as entry barriers, will not localize stable network specific power. For instance, the classical work on entry barriers by Bain includes product differentiation advantages of incumbents, such as advertising expenditures, among the barriers to entry. It can be argued, however, that as long as the costs of differentiating one's service or product are the same for both incumbents and newcomers, product differentiation activities can be undertaken by either firm, such that incumbents have no strategic advantage. Advertising outlays will not deter market entry in the long run, when supra-competitive profits continue to be made in the industry over a substantial time period. Along the same lines, even economies of scale and scope in the relevant region of market demand cannot be considered barriers to entry in the long run, as long as both the incumbent and the entrant have access to the same production technologies and therefore to the same cost functions. This is why not even an incumbent in a contestable natural monopoly is seen as having network specific market power.

This paper uses the market segmentation of the disaggregated regulatory approach as a backdrop for the discussion of the effect network externalities have on market outcomes. This allows us to disregard the influence of cost characteristics on market structure as these are already accounted for in the market segmentation of the disaggregated approach. The focus can therefore turn to identifying the influence that network effects have on market structure and on the competitive process.

The disaggregated regulatory approach differentiates between network areas which are characterized as monopolistic bottlenecks, network areas which are considered contestable natural monopolies, and network areas in which active competition is possible. The combinations of these cases with network externalities, as they are analyzed in the following subsections, are illustrated in Table 3.1.[11]

---

[10] See, for instance, Farrell and Saloner (1988) and Gandal (2002).

[11] Blankart and Knieps (1995) organize their analysis of the need for ONP (open network provision) regulation in networks exhibiting both network externalities and economies of scale in a very similar way. They differentiate the three different forms of the cost function mentioned above on both sides of the market (supply-side and demand-side). Their analysis therefore covers nine different scenarios. The analysis here presupposes competitive conditions on the demand-side of the market and therefore differentiates only a subset of three of the scenarios treated by Blankart and Knieps.

Table 3.1: Combinations of Cost and Demand Characteristics in Network Industries

Cost-side / Demand side	Monopolistic bottleneck	Contestable natural monopoly	Active competition
network externalities	Case 1	Case 2	Case 3

The majority of industrial organizations literature on network externalities is embedded in an oligopoly context.[12] The focus of this literature is to understand firm behavior under given market conditions. This chapter, in contrast, seeks to answer the question whether network externalities can be the source of market power, i.e. change market structure. The disaggregated regulatory approach places the entry conditions of a market at the basis of every analysis of market power. Since the following discussion of the literature is based on this theory, there is no need to treat the oligopoly case separately.[13] A market either features entry barriers which lend market power to active firms (Case 1) or a market is characterized by free entry such that active firms have no market power (Cases 2 and 3). The disaggregated theory predicts that an oligopoly market of few active firms with significant barriers to entry will lead to market outcomes comparable to those of a monopolistic bottleneck (Case 1), whereas a market with few active firms but no barriers to entry will lead to market outcomes comparable to competition (Case 3).[14]

A major drawback of the literature on network externalities in an oligopoly context is that the standard oligopoly assumption of a fixed number of active firms enters into the analysis elements of market power which can later not be differentiated from market imperfections resulting from network externalities. The approach of this chapter will therefore be to differentiate clearly between Case 1 which contains pre-defined market power due to the cost-characteristics of the market and the cases 2 and 3 which are not characterized by any entry restrictions resulting from

[12] See, for instance, the seminal articles in this branch of literature by Katz and Shapiro (1985) and Farrell and Saloner (1985).

[13] See also Blankart and Knieps (1995: Footnote 2).

[14] A competitive market with few active firms but no barriers to entry is, for instance, described in the theory of monopolistic competition (Chamberlin, 1950). This theory provides a broader understanding of competition than, for instance, the theory of perfect competition. Most importantly, it is compatible with more realistic assumptions on market characteristics allowing, for instance, fixed costs of production, product differentiation, advertising and some degree of price setting liberties on the part of the firms. Due to the assumption of free market entry, any profits "above those necessary to maintain capital and business ability in the field" (Chamberlin, 1950: 169), will attract market entry until no extra profits are being earned. The demand curve shifts inward as new firms enter the market until it is tangent

production costs. In these two cases suppliers are disciplined either by potential or by actual competition. The number of firms in these cases is endogenously determined through the interplay of supply and demand conditions. If market power can be located in these markets once network externalities are added, then this is a stronger indication that network externalities cause market power, than when market power is ascertained in one of the oligopoly models mentioned above.

3.2 Case 1: Network externalities in a monopolistic bottleneck

The first case to be considered here is a monopolistic bottleneck with network externalities. Non-contestable natural monopolies are characterized by cost subadditivity in the relevant output region paired with sunk costs of production.[15] The disaggregated regulatory approach argues that monopolistic bottlenecks, i.e. network areas showing both of these characteristics, lend network-specific market power to the owner of the monopolistic bottleneck and therefore justify sector-specific regulation.

3.2.1 Direct network externalities in a monopolistic bottleneck area

When network externalities are added to a monopolistic bottleneck network area, the monopolists' profit maximization behavior will generally lead to a market outcome that is not compatible with exhausting the direct network externalities present in the market. Farrell and Saloner (1992:13), for instance, argue that when the monopolist cannot price discriminate, he will not be able to appropriate the utility increase existing consumers experience with a higher subscription rate. Rather, a monopolist constricted to linear prices will continue to maximize profits by maximizing the willingness to pay of the marginal consumer. The monopolist's price will take into account the marginal consumers' valuation of the network externalities, but not the valuation of the increased network size by all inframarginal users. Compared to a competitive market outcome, the monopoly price (above marginal cost) will aggravate the problem of too little network participation (Katz and Shaprio, 1994: 101).[16] The market equilibrium in a monopoly will be to the left even of the private equilibrium S^p in Figure 2.1.

---

to the average cost curve. The number of firms active in the industry is determined by the condition that profits are equal to zero. There is no market power in equilibrium.

15 On the concept of cost subadditivity see Sharkey (1982: Chapter 4). In the single product case economies of scale over the relevant output region are a sufficient condition for subadditivity. In the multi-product case subadditivity requires both economies of scale for every single output separately as well as "economies of joint production" for all outputs over the relevant output region.

16 In a slightly different partial equilibrium oligopoly model Katz and Shapiro (1985) argue that under the assumption that consumer expectations on network size be fulfilled in equilibrium, a monopoly context

3.2.2 Indirect network externalities in a monopolistic bottleneck area

Monopolists have an incentive to transfer market power from the monopolized market to vertically and horizontally related markets. Indirect network externalities, which affect markets complementary to the monopolistic bottleneck area, can provide the monopolist with the necessary prerequisites to pursue such strategies. For instance, when complementary markets are dependent on a common standard with the monopolistic bottleneck, the owner of the monopolistic bottleneck could refuse non-discriminatory open standards to competitors. As an example consider markets characterized by a hardware/software relationship. If the hardware market were a monopolistic bottleneck, producers of related software would depend on open standards, which the monopolist could refuse in favor of an integrated or a cooperating software producer.

3.2.3 Public policy for monopolistic bottleneck areas featuring network externalities

This discussion shows that the presence of network externalities in a non-contestable natural monopoly will not reverse the fact that the owner of a monopolistic bottleneck has network-specific market power. Rather, network externalities offer the potential for further strategic actions to an uncontested monopolist. Therefore, network externalities increase the need for regulation in the monopolistic bottleneck area. In addition to open access regulation of the network elements, there is a need to pay attention to complementary markets and compatibility issues. Open standards need to be enforced such that competitors on vertically related markets benefit equally from indirect network externalities that emanate from the monopolized market. For example, product compatibility between the monopolist's products and the products of the competition should be ensured, such that an increase in the monopolist's output initiates demand effects not only for the monopolist's complementary products but also for the products of independent competitors.

The question remains, which standards are to be used by the bottleneck owner. Allowing the monopolist to decide on standards may already be discriminating towards competitors. Blankart and Knieps (1995: 289ff.) discuss the practical design of the regulation of standards. They suggest that, when it is possible, a contest for best standard can be held before monopoly is awarded. This is, however, feasible only prior to product introduction. Very often the need for a new standard

---

will lead to consumers expecting the network size to be smaller compared to a competitive market. Automatically consumers therefore adjust their willingness to pay for the network service downward in a monopoly context. This decreases the willingness to pay and ultimately reduces even further the equilibrium network size.

evolves in markets of established products when a new application is introduced. In these cases the regulation of open standards must be designed carefully. First of all, both the monopolist and its competitors will have legitimate interest in influencing the standard because they have the best knowledge of the technical and qualitative requirements of their applications. Secondly, governments may be prone to promote standards which serve their own interests rather then societies interests. The standard-setting process has to take these aspects into consideration. This discussion is taken up again in section 4.

3.3 Case 2: Network externalities in a contestable natural monopoly

Natural monopolies with no significant irreversible costs of production, are contestable natural monopolies. A contestable natural monopoly can support only one active firm. However, the monopolist does not possess network specific market power because in a contestable market potential competitors have access to the same cost function as the active firm and market entry does not require investing into irreversible assets. Under these conditions, a potential competitor will enter the market and undersell the monopolist should he be charging super-competitive rates.[17] Therefore it was argued that the possibility of market entry by potential competitors disciplines the monopoly supplier.

The effectiveness of potential competition hinges on the readiness of consumers to switch to the new entrant should he supply a superior product. The question now becomes whether the presence of network externalities in a market can change the fact that a superior product can establish itself in competition with a contestable monopolist. Can network externalities prohibit consumers from switching to a new entrants' product even when the cost characteristics of the market allow free entry and exit? Can network externalities introduce barriers to entry as defined by Stigler (1968: 67)?

Some economists have argued that network externalities can hinder new entrants from establishing themselves in a market even when they have an objectively superior product to sell. This assumption of so-called "inefficient lock-in" is seen to result from the fact that a product featuring network externalities has to gain a critical mass of users in order to become viable. When network externalities are strong, a new product will fail to be of any noteworthy use to consumers until a required minimum of other users (the critical mass) is given. An obvious example is a telephone network. In a world without interconnection, a user will not switch to a new telephone network featuring better technology at lower prices as long as there are no subscribers on that network to communicate with. The theory predicts, that if all consumers postpone purchase of a product with network externalities until the critical mass is reached, then new entrants will not be able to establish themselves. A cheaper product or a product of better quality would not be sufficient to gain a user base in the face of strong network externalities which guide users to previously established networks.

It is important to understand that the problem of reaching a critical mass does not compare to ordinary switching costs. Switching costs are a factor in many markets that do not feature network externalities. The difference between switching costs and the problem of critical mass is that in the case of ordinary switching costs a consumer can weigh her private costs and benefits from switching to a new product and make an informed choice. For instance, a consumer may consider it costly to learn to use a new mobile telephone of an unknown manufacturer, because of an unfamiliar menu navigation. She can compare these costs to the benefit received from the new phone in terms of new features, better design, lower price etc. Liebowitz would argue that such switching costs "are real costs, and if the new product is not sufficiently better to outweigh those costs, then it is efficient for society to stick with the old" (Liebowitz, 2002: 36). Ordinary switching costs do not cause market failure. Rather, they deliver important information in the market process. Sufficiently superior products will overcome switching costs of this kind and there are plenty of examples in real markets as proof of this.

What is different for consumers deliberating to switch products in a market exhibiting network externalities is the fact that the individual consumer does not have all the information needed to compare the relevant costs and benefits of switching. The individuals' costs and benefits from switching to the new product always depend on the product choice made by other consumers. For a single consumer to make an informed product choice all consumers would need to communicate their interests ahead of time.

---

[17] When the entrant has not reached the scale economies of the incumbent, his actual production costs will exceed the incumbent's costs. However, because the entrant expects to replace the incumbent, he is willing to sell at prices equivalent to the long run minimum average costs.

Given these difficulties, the economics profession has produced different viewpoints on the question whether network externalities can substantially reduce the threat of potential competition and therefore lend market power to an otherwise contestable monopolist. While, for instance, Arthur (1989) and David (1985) argue that market equilibrium will be influenced to a significant degree by chance events, Katz and Shapiro (1994: 112) argue that "the market efficiency is unclear, once recognition is given to the many private institutions that arise to achieve coordination and internalize externalities. [...] there are many possible responses of systems markets to these problems that involve no government intervention whatsoever." The following three sections illustrate the aspects which have been shown to be important for achieving market coordination in the presence of network externalities.

3.3.1 The role of information in overcoming the critical mass problem

In a contestable natural monopoly the choice between technologies can be reduced to a two-dimensional choice between staying with the incumbent's technology or switching to a new entrant. Farrell and Saloner (1985) model a very similar market. Here n firms decide between a status quo standard T_1 and a new standard T_2. By assumption, all firms agree that the new standard is superior, but because of network externalities, switching is only beneficial if all other firms switch. In equilibrium, when all firms have complete information on the other firms' preferences, then all firms will switch to the new standard.[18] Only when the preferences of the other firms are not common knowledge, then only those firms with a very high preference for the new product will switch early and it is not clear, whether all other firms will follow, even when the new standard is preferred by all consumers. Farrell and Saloner call this phenomenon "excess inertia". When consumers have different preferences with respect to the two products it is also possible that "excess momentum" occurs. Those that favor the newer product switch early and then may regret this when they are not followed by the users preferring the older product.

The Farrell and Saloner model focuses on the importance of information. In their model consumers need to be able to communicate their preferences in order to co-ordinate a switch that leads to a superior market outcome. Otherwise either excess inertia or excess momentum can leave some

[18] This outcome results from the fact that each individual i $\in$ (1,2,...n) prefers to switch to the new technology when all individuals 1, ... i-1 have already switched and when he believes the rest will switch when he switches. The individual believes that i+1, ...n will switch when he switches because he knows that they have the same preferences as he does. Therefore individual i will switch and all others will follow (see Farrell and Saloner, 1985: 73).

users stranded with a less-preferred technology. Central to the question whether network externalities impede the efficient market process is therefore the likelihood of consumers succeeding in reducing the information problem by communicating their preferences.

3.3.2 The role of consumer expectations in overcoming the critical mass problem

The likelihood that consumers can spontaneously organize to systematically reveal to each other their product preferences depends to a large degree on the extent of the market. A small user group is more likely to achieve coordination than a market with many anonymous buyers. The later case is modeled by Arthur (1989), who assumes a situation in which consumers have no further information on the prospects of a product except for the number of previous customers of the product. They decide which technology to use based only on the technological characteristics and on the number of existing users. The process of technology adoption is then stochastically determined by a chance sequence in which consumers make their technology choice. Arthur uses this model to show how historical events, even if insignificant at the time, play a significant role in the long-run adoption of products and technologies.[19] Early events that, by chance, favor one technology will grant this technology a significant installed base such that when a superior product arrives on the scene at a later time, even those consumers that would prefer this newer product will buy the inferior existing product as it offers the larger network benefit. Thus, according to this model, society can experience inefficient lock-in[20].

Two assumptions are critical for Arthur to arrive at his prediction. First, consumers' preferences for product characteristics (the technology effect) are exogenous to the model. They cannot be altered by making available newer information or by changing product characteristics. Second, the

---

[19] David (1985: 332) even speaks of "historical accidents" in this context.

[20] The best-known exponents of the theory of inefficient lock-in are David (1985) and Arthur (1989). See also Cowan (1991). Liebowitz and Margolis (1995) offer a very interesting critique of the literature on path-dependency and lock-in by differentiating three different forms of path dependency of which only one leads to market failure in the sense that policy interventions can be shown to improve on the market outcome. This case presupposes that "there exists or existed some feasible arrangement for recognizing and achieving a preferred outcome, but that outcome is not obtained" (pg. 207). The authors argue that this third case is based on restrictive and implausible assumptions, especially because it is assumed that at least one person must know of the better solution, but this person does not turn this knowledge into profit by coordinating buyers to the superior outcome (pg. 217). Nevertheless, in academic discussions the policy implications of this case are applied to the two other more realistic cases of path-dependency which assume that "the existence or superiority of alternative paths are not known at the time that initial decisions are made" (pg. 211). Liebowitz and Margolis argue that in these cases, there is not real inefficiency in the market outcome as there is no knowledge of the superior outcome in the economy.

consumers' utility is endogenously determined only by the number of existing users of a product (the network effect). Expectations of future consumption levels are not considered. As a result, only historical chance events influence the consumer decision which network to join.

These random events can, however, be dominated by the influence of strategic actions taken by market participants, once consumer expectations regarding future network size and changes in technology are taken into account. While it is true, that consumers will often lack the organization necessary to reveal their preferences to each other, a new entrant, with a possibly superior product, has opportunities to disclose relevant information on his products' chances and to influence consumer buying decisions. And when he believes in the superiority of his product he has an incentive to do so. As Liebowitz and Margolis (1990: 4) suggest: "An owner with the prospect of appropriating substantial benefits from a new standard would have an incentive to share some of the costs of switching to a new standard. This incentive gives rise to a variety of internalizing tactics." These tactics are discussed in the following section.

3.3.3 The role of sponsorship in overcoming the critical mass problem

There are several papers in the literature which focus on these abilities of firms to "sponsor" their products in order to overcome the critical mass problem. Katz and Shapiro (1986), for instance, analyze technology adoption given that a new entrant is willing to incur short-run losses in order to establish his technology. This model is interesting in the context of the contestable monopoly case discussed here, because it considers two competing technologies available in two time periods. Period one consumers take into account the probable consumption decision of period two consumers when making their technology choice. Period two consumers make their technology choice dependent on the technology chosen in period one. Preferences of consumers are assumed to be homogeneous and consumers are assumed to have rational expectations. In the first period, consumers unanimously choose the technology yielding the highest consumer surplus at that time. In the second period, consumers compare the benefit from the combined technology and network effects of both products. The second period is thus comparable to the contestable monopoly case in which a newer technology competes with an incumbent technology.

Then, however, these more common cases lack the strong policy implications of the case in which a superior outcome is known to be feasible.

The model assumes that the newer technology will win the market in the second period only when the extent of the price/quality superiority over the previous technology is worth more to consumers than the installed base advantage the existing technology offers. The paper thus emphasizes the possibilities an entrant has to influence the price/quality superiority of his product. It is argued that the entrant can differentiate his product so as to offer more benefits from the technology effect of his product. When consumers' expectations of future network size factor into their benefit function, the new entrant also has the possibility to convince consumers that his product will offer more benefits from the network effect in the future: early product announcement and a penetration price strategy can, for instance, induce some first period consumers to wait until the second period to make their purchase decision, thereby decreasing the installed base advantage of the first technology. Furthermore, this may convince new users that the entrant will pull sufficient demand to be able to offer similar network benefits as the incumbent.[21]

Katz and Shapiro (1986: 832) show that when neither of the two technologies is sponsored then there is a tendency in the market to standardize on the initially superior network, even when this is not the socially optimal outcome in later periods. This confirms Arthur's (1989) result. However, when both technologies are sponsored there is a tendency that the market will standardize on the technology which has lower costs in the second period (Katz and Shapiro, 1986: 838). Interpreting lower costs as a signal for a more advanced technology, this model shows that sponsoring can be an efficient means for a new entrant with a superior product to overcome critical mass restrictions.

There are multiple strategies an entrant can follow to sponsor his product.[22] Whenever consumer expectations of future market shares play a role, advertising can, for instance, increase a product's reputation and influence the expected sales of a product. This will in turn increase the realized demand share in the present. The entrant can also make the price of his service dependent on network size. As long as the network is small, consumers pay a lower price. This pricing scheme acts as insurance for consumers. Only when network size increases will membership become more expensive. The risk of being stranded on the "wrong network" is thereby reduced. A further pricing strategy is to subsidize a basic service and thereby bind more consumers to a technology, while adding mark-ups to advanced services and applications in order to cover the losses obtained on the basic service. The entrant can also use subscription in order to build a customer base before

[21] In a related paper Farrell and Saloner (1986) analyze the welfare effects of product pre-announcement and strategic pricing in dynamic markets with network externalities.

[22] See also Katz and Shapiro (1994: 102).

actually introducing his product. Consumers that subscribe in advance commit to switching to the new product once an appointed number of subscribers is reached.

3.3.4 Public Policy for contestable markets featuring network externalities

The discussion of the various ways in which information dissemination, consumer expectations and sponsorship can influence the markets abilities to achieve coordination even in the presence of network externalities shows, that the theory of inefficient lock-in may throw an overly pessimistic view on the problem of network externalities in contestable markets. For the purposes of taking policy measures in such markets it would be dangerous to conclude that a firm which has been able to dominate a market has market power. It may well be that no superior product has as yet been offered by potential market entrants. From their studies of real-world examples Liebowitz and Margolis conclude that the dominance of a particular product can often be explained by superior quality and/or price advantages offered to consumers.[23] Using Stigler's definition of entry barriers they conclude: "...it is unclear how network effects, economies of scale, or any of the other factors at work favor the incumbent relative to the challenger. The incumbent had to coordinate consumers to adopt his product, whether it was first in the market or replaced a previous incumbent. In either case, getting consumers to come on board is a cost imposed on all market entrants, late or early, and does not necessarily favor early firms" (Liebowitz and Margolis, 2001: 164-165). Therefore, network externalities alone do not lend market power to a contestable natural monopolist which would justify policy interventions into the market process. Rather, market institutions are creative in finding ways by which consumer preferences are revealed such that network externalities can be internalized within regular market operations. Policy makers would have difficulty improving on market outcomes.

[23] Liebowitz and Margolis have put the hypothesis of inefficient lock-in to an empirical test by analyzing real-world markets commonly associated with this phenomenon. In "The Fable of the Keys" (Liebowitz and Margolis, 1990) investigate the history of the Qwerty keyboard, which has long served as the standard example for an inefficient lock-in, because it is argued to be inferior in terms of typing efficiency when compared to keyboards developed subsequently. In Liebowitz and Margolis (2001) they analyze the hypothesis of inefficient lock-in in a competitive context. They examine the claim that network effects in the software market are responsible for Microsoft's success in particular product groups. In the antitrust case against Microsoft it was argued that Microsoft dominates particular software markets only because the software is part of the Microsoft operating system and not because of product quality and price advantages. Liebowitz and Margolis use software reviews from leading computer magazines to compare the product rankings according to expert opinions with measures of market share. They find that the market success of, for instance, the spreadsheet program Excel and Microsoft's Internet browser Explorer can also be explained by a quality lead over the competing product as well as competitive price setting.

With respect to the allocative efficiency of the market outcome to be expected in an unregulated contestable market featuring network externalities, it can be said that since the cost conditions of the market favor only one active firm, there is no threat of network islands developing. Furthermore, should the operator try to enforce a price above the competitive level this would invite market entry. Only the competitive price protects the operator from replacement by a potential competitor. This competitive price level leads to a network size which internalizes the network externalities to a large degree (comparable to equilibrium S^p in Figure 2.1.).

3.4 Case 3: Network externalities in a competitive market

The last case to be discussed here is the competitive case.[24] In a competitive market featuring substantial network externalities, firms will compete not only in price and product characteristics, but also in the dimension of network size. The stronger the network externalities, the more important will be the "network effect" a product offers compared to its "technology effect" (see section 2 above).

The important difference to case 2, in which network externalities are analyzed in the context of a contestable natural monopoly, is that in a competitive market users will have a choice between several operating networks. The trade-off between standardization and variety, mentioned in section 2.4, takes on its full meaning only in this context. An increased variety of technologies available for consumption makes it more difficult for consumers to resolve the trade-off between choosing the technology with the most preferred product characteristics and benefiting from being a part of a network with a large number of users. Consumers can be expected to have different preferences with respect to their preferred technological and qualitative product characteristics and also different valuations of the network effect. It is likely that in this scenario consumers will be better off when a variety of network islands caters to particular consumer tastes, rather then when one large network offers a compromise between the demanded product characteristics (see for instance a model by Shy, 2001: 27-35). However, the more important the network effect, the more willingly will consumers give up specialized technologies and concentrate on the network offering the largest user base. The aim of this section is to understand how the competitive market process will solve the trade-off between taking advantage of network externalities (implicating a small number of active firms) and offering product variety and competition in the market (implicating a larger number of active firms).

As was discussed in section 2.4, a trade-off between scale and variety is also a common problem in markets featuring supply-side economies of scale. Here the trade-off between scale and variety results from the fact that the cost efficiencies of larger scale production can be exhausted only at the expense of a decrease in product variety. Multi-firm production (when operating below minimum efficient scale) entails a duplication of fixed costs, such that variety comes at the price of higher average production costs. With network externalities in a market, the requirements on size are on the demand-side rather than on the supply-side of the market. This leads to an important difference in the inefficiencies which can result from multi-firm production. The problem with network externalities is not that multi-firm production increases average costs, but rather that users will not be able to benefit from the positive network externalities when split on different network islands. Unlike in the case of economies of scale, there is a solution to this problem of the type "you can have your cake and eat it too", i.e. which allows the consumer to profit both from product variety and the positive network externalities. This solution presupposes that rival firms will cooperate by making their networks compatible. Compatibility between self-contained network islands allows users to communicate and share network benefits while competition is preserved in product dimensions not affected by the cooperation.[25] With full compatibility the consumer will base his product choice on technological characteristics and price. Network reach will be the same for all products. The main question therefore is what incentives firms in a competitive market have to make their networks compatible.

3.4.1 Firm incentives for compatibility

When network externalities are strong and the consumer valuation of compatibility accordingly high, the question is whether firms are likely to voluntarily engage in cooperation but salvage competition in product characteristics or whether larger firms will have an incentive to decline

[24] The reference scenario of competition in this case is not the textbook model of a perfectly competitive market but rather an understanding of competition which allows product differentiation and therefore concedes a degree of price setting competencies on the part of the firm. See also footnote 14.

[25] Put differently, supply-side scale economies determine the cost-efficient physical network size, while user externalities determine the efficient number of compatible network users. Since physical network size cannot be shared between independent network owners, supply-side economies of scale lead to higher firm concentration in a market. User externalities do not have this inevitable effect, since the "user size" of a network can be shared through cooperation.

cooperation in the hopes of collecting supra-competitive profits from offering the largest network. To answer this question it is useful to consider the costs and benefits of cooperation from the perspective of the firm. Most substantial on the cost side is that, by making its network compatible, a firm gives up network size as a differentiating product characteristic in its competition with rival firms. This aspect is especially costly to firms with initially large networks. Furthermore, cooperating firms have to invest more in product differentiation with respect to remaining differentiating characteristics. Lastly, there are transaction costs involved in coordinating the cooperation of firms with conflicting interests.

On the other hand, there are positive aspects to compatibility which may outweigh these costs. The most substantial benefit derived from cooperation from the point of view of the firm is the possibility to benefit from a significant demand shift generated by the increase in overall network size. When total demand for network services rises due to several networks becoming compatible, then even initially large networks can experience a demand increase for their network services. Economides (1996), for instance, shows that even a monopolist, can profit from inviting entrants, even subsidizing their market entry, and making his services compatible to theirs. This counterintuitive result is explained by the fact that when consumer demand is a function of the expected network size and when network externalities are sufficiently strong, then the demand effect from a larger network pushes the demand curve outwards to such an extent that the higher market wide sales allow the monopolist to sell so much more of his product compared to the monopoly output that even at a lower competitive price he can increase his profit.[26] The monopolist can realize this output level and this price only in a competitive setting. A monopolist cannot credibly commit to producing a competitive output/price combination because, given a level of demand, it is always in the monopolists' interest to restrict output and raise prices. Since consumers expect the monopolist to offer network services only at higher prices, they expect a smaller network in a monopoly and their demand at any given price will be correspondingly lower in a monopoly than in a competitive market.

A further benefit to cooperation is that it eliminates competition to establish the largest network and therefore frees resources which can be put to other uses. These can, for instance, be invested into competing in other product characteristics. The more important technological differentiation is

[26] The elasticity of demand obviously plays an important role in this context. The more elastic the demand, the more pronounced will be the demand expansion induced by more rivalry in the market.

to consumers, the more important it will be to invest into product innovations. This is likely to be of greater significance in dynamic markets.

3.4.2 A model of network externalities in Cournot competition

To my knowledge there exists no model of network externalities in a competitive market. The literature on compatibility choice mostly uses Cournot or Bertrand competition, assuming a fixed number of firms in the market. The section discusses why the results of these models cannot be transferred to a free-entry market environment. It looks at the assumptions of the common oligopoly models in the literature on network externalities and on their influence on the results derived therein.

The seminal paper on network externalities in oligopolistic competition is Katz and Shapiro (1985). The authors analyze equilibrium outcomes and firm incentives for compatibility. The model defines a consumers' utility as:

$U_{i,j}(S,T) = r_i + v(S_j^e)$ for consumers $i = 1,2,...,m$ and technologies $j = 1,2,...,n$.

Parameter r_i stands for the basic willingness to pay for network services (this is the willingness to pay which results from the "technology effect"). As technologies are assumed to be homogeneous in all characteristics but network size, r does not differ across technologies j, but only across consumers i, assumed to have different basic valuations of network services. The parameter r is assumed to be uniformly distributed between minus infinity and A with density one and A is assumed to be a positive value.[27]

The consumers' valuation of the network externality is expressed in the externality function $v(S_j^e)$, where S_j^e is the sum of the network users expected to be compatible with technology j. All consumers have the same valuation of the network effect. The model assumes $v'(S_j^e) > 0$, $v''(S_j^e) < 0$ and $\lim v'(S_j^e) = 0$ as $S_j^e \to \infty$. Users counted among S_j^e are either

[27] These assumptions on the distribution of r ensure a linear demand curve for technology j.

subscribers of technology j or subscribers of networks compatible with technology j. When x_k^e stands for the expected network size of technology k, S_j^e can be expressed as:

$S_j^e = \sum_{k=1}^{m} x_k^e$, where $k = 1,2,...,m$ are the networks compatible with j.

The static, one-period model is solved in three steps. In a first step, consumers build expectations about the future network size associated with each technology. In a second step, firms choose their output given the consumer expectations of network size and assuming the output of rival firms as given (Cournot competition).[28] This output game determines the prices that firms charge in equilibrium. In a third step, consumers make their purchase decision by maximizing their utility. When the price of a technology is given by p_j, a consumer chooses the technology for which utility is maximized (and is equal to or greater than the utility realized with the next best alternative resource utilization). Formally, the consumer's decision function is:

$$\max_j [U_{i,j}(S,T) - p_j] \geq U_{\min}$$

Assuming that $U_{\min}$ is equal to zero, the consumer maximizes

$$U_{i,j}(S,T) - p_j = r_i + v(S_j^e) - p_j . \qquad (1)$$

If this expression is negative for all available technologies j the consumer will join none of the available networks.

Since the consumer's basic willingness to pay r_i is the same for every available technology j, the consumer will choose that technology for which the difference, $[v(S_j^e) - p_j]$, is maximized, respectively that technology for which the price adjusted for the network effect offered by a particular technology, $[p_j - v(S_j^e)]$, is minimized. Katz and Shapiro call this price the "hedonic price" of technology j (Katz and Shapiro, 1985: 427).

[28] The model assumes constant marginal costs.

Since all consumers are assumed to have the same evaluation of the network externality, all consumers will choose that technology which is available at the lowest hedonic price. Therefore, the equilibrium will only feature more than one firm with positive output, if these firms offer their technology at the same hedonic price.

The equilibrium concept employed in the model is that of fulfilled expectations Cournot equilibrium (FECE). This concept assumes that all firms maximize their profits by choosing their output. Consumer expectations and the output level of rival firms are given. Production costs are assumed to be the same for all firms; all costs are non-variable. In equilibrium expected sales must be equal to actual sales.

Assuming that fixed costs are zero and redefining r_i as the willingness to pay which exceeds marginal costs, allows the authors to set up a profit function, which, differentiated with respect to output, yields output equations for all firms. Solved simultaneously, these equations yield the following equilibrium output:

$$x_j^* = [A + n \cdot v(S_j^e) - \sum_{k \neq j} S_k^e]/(n+1)$$

Since the number of active firms n is exogenous to the model, the model leads to the conclusion that those networks which consumers expect to be dominant (high value of $v(S_j^e)$), for instance because of their initial network size or because of their reputation, will in fact be dominant in equilibrium (high x_j^*). Therefore, within their framework, the authors conclude that network externalities lend market-power to initially large firms.

This equilibrium outcome is a result of the specifications of the model. Particularly, the model does not include a positive value to product variety. For consumers to value variety, the parameter r, reflecting the willingness to pay for the technological characteristics, would have to vary in technologies j. If this were the case, then not all consumers would prefer the same network, even if their valuation of the network externality were the same. In this case, consumer expectations of network size would not dominate the consumer's choice of network. Furthermore, the assumption that all consumers have the same valuation of the network effect is also unrealistic but very restrictive with respect to the results of the model. Lifting these assumptions would allow different combinations of price, technological characteristics and network size to exist simultaneously in the market. The result that networks which have an initially large user-base will come to dominate a market would no longer hold. More realistic market conditions (heterogeneous consumers, product differentiation and free market entry) are likely to lead to more firms in equilibrium.

Within their model Katz and Shapiro do analyze how their result would change if they lifted the assumption of a fixed number of firms and assumed instead that the number of firms increases indefinitely. The equilibrium output then converges on:

$$\lim_{n\to\infty} x_j^* = v(S_j^e) - \lim_{n\to\infty}[\sum_{k\neq j} v(S_k^e)/(n+1)]$$

In words, as the number of firms grows, firm j's equilibrium output will be the larger, the larger the expected network size of firm j as compared to the network sizes of its competitors. When firms are expected to have the same network size in equilibrium, either because of full compatibility in the market or because firms are grouped in network islands of equal size, then the equilibrium output of each single firm will tend towards zero (because the second term of the above equation, $\lim_{n\to\infty}[\sum_{k\neq j} v(S_k^e)/(n+1)]$, will tend towards $v(S_k^e)$ and $v(S_k^e)$ is equal to $v(S_j^e)$ when networks are symmetric). Using this result, Katz and Shapiro (1985: 429) show that when there is full compatibility between all active firms then there is a unique symmetric equilibrium which converges on the perfectly competitive equilibrium when the number of firms becomes increasingly large.

This competitive case analyzed by Katz and Shapiro is trivial. Since the model does not allow competition by product differentiation the firms are homogeneous whenever there is full compatibility, since compatibility takes away the last differentiating aspect, namely network size. Naturally a market with an infinite number of homogeneous firms will converge on the perfectly competitive equilibrium. A realistic model of competition which allows for consumer valuation of product differentiation in technological characteristics as well as in network size will necessarily not predict perfect competition in a competitive market featuring network externalities. This is, as was already stated above, not the criterion against which market performance should be measured.

Katz and Shapiro (1985) also analyze the firms' choice to make their technologies compatible by comparing the difference in realized profit from the non-compatibility case to the compatibility case. They argue that only when the expected profit with compatibility is larger than the profit in the incompatibility case, will an individual firm have an incentive to cooperate. The authors show that within their framework not all suppliers necessarily benefit from increased compatibility. Even when overall output rises, initially large networks may have a lower output in a compatible equilibrium. Therefore, when firm size is highly asymmetric, firms may not voluntarily agree on a common standard. Rather, initially larger firms will tend to favor incompatibility.

When the authors enter the possibility that firms make side-payments among one another, it becomes sufficient that average profits increase through cooperation for firms to voluntarily agree on compatibility. Those firms which gain disproportionately from compatibility can compensate the losses of those firms that would otherwise prefer not to cooperate. Because the condition that average profits rise for compatibility to be a viable outcome is less restrictive than the condition that profits increase individually, side payments increase the likelihood that firms reach a mutually beneficial compatibility agreement, even when they are of substantially different initial size.[29]

3.4.3 Public policy for competitive markets featuring network externalities

The above discussion shows that cooperation is likely in a competitive market with substantial network externalities, especially when consumer tastes vary and firms have the possibility to differentiate their services not only in network size, but also along technological and qualitative characteristics. A larger firm can hope to benefit from an overall demand increase resulting from increased network benefits and hold on to its previous customers by offering superior product characteristics. Product differentiation can therefore strengthen the incentives to cooperate on compatibility issues. And as was argued by Katz and Shapiro, allowing for side-payments further

[29] Compensation here refers strictly to payments made in order to induce the other party to agree on a common standard. These payments have to be differentiated clearly from payments which are compensation for costs incurred by the other party as a consequence of the compatibility. For example, if one network offers another network free licenses for its software products in order to induce it to agree to interconnect this can be considered a side-payment. When, however, payments flow in direct relation to the costs of achieving compatibility, such as transportation compensation for the costs incurred in terminating one another's traffic, then these payments do not reflect side-payments.

increases the opportunity for market participants to internalize the network externalities. There are even cases in which side-payments may flow from the initially larger network to smaller competitors. These analysis show that payments made to induce competitors to become compatible can be interpreted as compensating a firm for a kind of property right it has to the positive network externalities associated with its network size.[30] Such payment should not be interpreted as reflecting market power. Rather, they redistribute the gains from network compatibility in a way that all firms are better off.

Up until this point we have regarded the problem of compatibility only from the point of view of the firms. It is possible, however, that the social incentives for compatibility are not congruent to the firm incentives for compatibility. First, there can be too little cooperation among firms when cooperation is in the interest of society but firms cannot appropriate enough of the benefits from cooperation to exceed their costs. Second, there can be too much cooperation when firms cooperate even though society's costs to cooperation (see the following section) exceed the benefits. Is this a reason to assume that there is, after all, a responsibility to bring about the correct level of compatibility through government intervention? Whether the move to compatibility is efficient from a social welfare point of view depends on how the costs of compatibility relate to the benefits derived from the internalized network externalities (see Katz and Shapiro, 1985: 438). Because both the benefits and the costs are impossible to measure, there is no reason to assume that policy makers can bring about a closer congruence between the private and social incentives for compatibility than will emerge from an unaided market process. The information requirements for determining the socially optimal level of compatibility are impossible to fulfill. Therefore, the need for government intervention into competitive markets featuring network externalities is not substantiated by the present discussion of market processes in the presence of network externalities.

[30] There are strong parallels here to the Coase Theorem, which postulates that as long as property rights are well-defined, private negotiations can lead to efficient market outcomes. See also Katz and Shapiro (1986: 825).

4 Standards as a prerequisite for cooperation

So far we have discussed the compatibility decision in the abstract. Taking a more practical view, technological requirements will have to be fulfilled before network services become compatible across network boundaries. How do firms decide on a standard to which all compatible networks have to conform?

One can differentiate between three general mechanisms by which agreement between firms is reached. The first is standardization by multilateral agreement. For this the firms convene in standardization committees charged with reaching a consensus on a particular product standard. The second mechanism is a unilateral predetermination of a particular standard by a leading firm in the industry. The remaining firms either adopt this standard or employ adapters or gateway technologies in order to make their products compatible to the standard set by the industry leader. Lastly, an industry standard can be imposed by governmental decree.

4.1 Voluntary standardization

When governmental intervention is not deemed appropriate (as was the case in the above discussed contestable markets and competitive markets), then either the first or the second of these mechanisms will be used to fix the industry standard. What are the relative advantages of either multilateral or unilateral standardization?

Multilateral standardization aspires to a consensual decision by all affected parties. This makes a multilateral standardization process very political and drawn-out. The advantage of such multilateral agreements, on the other hand, is that they are more likely to come to a sustainable solution. Unilateral choice of a standard, in comparison, is most likely quicker. The standard may however not be adopted readily, should conflicts arise between several players wanting to take on the role of market leader. In the end, incompatible standards may prevail.

Farrell and Saloner (1988) compare coordination by explicit communication in standardization committees and standardization by unilateral predetermination of a standard. They assume that everyone prefers standardization over incompatibility, regardless of the standard that is finally agreed upon. In finitely repeated games committees are shown to dominate unilateral standardization because they are more likely to achieve coordination.[31] Only a hybrid system, in

[31] There is no difference in infinitely repeated games.

which both standardization by committees and standardization by unilateral action are possible, is more efficient than standardization by multilateral agreement alone. This is so because the threat of pre-emptive action by the industry leader entices committees to work more effectively such that agreement is reached sooner.

In conclusion, the above discussion showed that when markets are considered competitive, then private solutions to the standardization problem can be expected. In these cases standardization committees often take on the important task of information dissemination in the market. Public policy should therefore allow and even encourage standardization committees. This is in contrast to traditional competition policy, which is skeptical towards joint decisions taken by members of the same industry, suspecting collusive and anti-competitive behavior. In markets featuring network externalities competition policy must, however, respect the difference between welfare-enhancing standardization and anti-competitive collusion.

4.2 Standardization by governmental decree

In section 3.2 it was argued that government intervention is deemed appropriate when standardization of vertically related product markets is desired, and when one of these markets is a monopolistic bottleneck. The first alternative for government control of market power is to impose government designed standards. What can be said about the efficiency of standardization by governmental decree?

A fundamental problem with standardization by government is the fact that the politicians charged with the standardization do not necessarily have the best interest of consumers and producers in the market in mind. Governments generally delegate standardization tasks to bureaucracies that are closer to individual industries. As Blankart and Knieps (1993: 46) put it, there is no "...democratic link between those who define and enforce standards and those who are subject to standards." Bureaucrats, other then members of government, do not depend on being re-elected every term. Rather, they can pursue goals such as maximizing their influence and budget, which may not be related to optimally serving the interests of the general electorate. Rather, bureaucrats are likely to be especially susceptible to rent-seeking such that the firm willing and able to invest into rent-seeking behavior, i.e. the firm with market power, can use the bureaucrats to pursue its own agenda. Furthermore, for reasons of self-legitimisation, the bureaucrats have an interest to extend their realm of action beyond the market areas affected by market power.

Because of these problems with governmental standard-setting it is important that rules are defined which limit the realm in which public standard setters are allowed to become active. If governmental standards are deemed necessary, there should be limitations concerning how deep standardization by bureaucrats may go. Blankart and Knieps (1993:40-44) suggest that network services are technological systems of many interrelated components which form a technological hierarchy. Standardization can be applied to any subset of these components. Since standards built on one another in the same way that technological components are part of a logical structure, one can speak of various degrees of standardization where the standardization of basic technological functions is a prerequisite for a "deeper standardization", involving also the specialized functions higher up in the technological hierarchy. With this view, standardization becomes a gradual process. From the point of view of affected firms there is a diminishing marginal return to standardization. Adopting a basic standard offers higher marginal benefits then joining a higher-level standard because basic standards have a wider dissemination. Using a compatible basic standard is a prerequisite for many interactions with horizontally/vertically related firms whereas compatible higher-level standards, specialized to specific applications, are required only seldomly and only in interactions with specific partners.

Because the positive network effects of adopting a uniform basic standard are clearly more important than the network effects of a uniform standard for advanced applications, Blankart and Knieps (1992:84) argue that government regulation should be confined to basic standards. To counteract the tendency of bureaucracies to tend to interfere too much rather then too little, the standardization of advanced applications should, however, not be allowed. Of course, the practical implementation of this rule hinges on the assumption that basic functions can be identified. Blankart and Knieps (1994: 459) argue that with regard to network services, basic functions are often tied to the infrastructure facilities of the network services whereas applied functions are tied to the applications offered on the network. A further indication as to whether a function can be counted among the basic functions is the expected dissemination of the standard within the industry. A standard with a wide dissemination regulates basic functions applied by many market participants. Standards with limited dissemination are or particular functions not used widely.[32]

[32] In a later paper Knieps (1995: 294) suggests an even more restrictive alternative. Here he argues that governmental activities can be limited to the regulation of access to monopolistic bottlenecks alone. In his view, reducing the market power of the bottleneck owner by price and quality control should suffice

5 Conclusions

This review of the literature on network externalities was by far not comprehensive. It disregards, for instance, the related branch of literature which focuses on network interconnection charges and competition in the telecommunications industry. Seminal articles in this tradition are Armstrong (1998 and 2002), as well as Laffont, Rey and Tirole (1998a and 1998b).[33] These papers focus on interconnection charges in the telecommunications industry and the question whether the regulation of these charges needs to prevent network operators from utilizing the interconnection charge as an instrument of collusion. Although the general results presented in these papers have important implications for the question discussed in the present analysis, it is beyond the realms of the present paper to go into their specific focus.

The papers which were evaluated for the present analysis are part of a tradition which takes a more general and abstract view on network externalities. The chosen papers show the representative themes which play a role in the analysis of how market processes are affected by network externalities. It was shown that the models presented in this chapter arrived at their specific results using restrictive assumptions. Based on these models it is not possible to reach universal conclusions on the changes to be expected in the competitive process when network externalities are present. Already early on in the academic discourse of compatibility choice in free market processes Besen and Johnson (1986: 18) concluded that the applicability of the theoretical models on network externalities to real world markets is limited. This evaluation is still appropriate today, even when the models have since been refined. Nevertheless, the analysis presented in this paper provide useful insights into the mechanisms that influence the ability of market processes to solve the trade-off between user externalities, product variety and search for new technologies. Farrell and Saloner (1985), for instance, emphasize the role of access to information on consumer preferences. Katz and Shapiro (1994) emphasize what firms can do to motivate consumers to reveal their preferences. The preparedness to sponsor a product (Katz and Shapiro, 1986) as well as to make side-payments (Katz and Shapiro, 1985) was also shown to have substantial influence on the market outcome.

---

to eliminate the monopolists' incentive to prevent coordinated standardization activities in multilateral committees.

The theories indicate that market processes very often generate novel ways for coordinating users and firms – especially in markets with heterogeneous consumers and product differentiation. Keeping in mind the various impediments to governmental institutions imposing efficient policy (i.e. Buchanan, 1987 or Blankart and Knieps, 1993: 46) the discussion suggests that in many realistic market environments the spontaneous market order has a comparative advantage in internalizing network externalities as compared to administered policy. To support regulation as a means of internalizing network externalities requires strong evidence that the many ways by which market participants can solve the information problem are not functional under particular circumstances, as for instance in the case of an uncontestable natural monopoly. In the remaining cases, government imposed standards cannot outperform standardization committees because a regulator does not have nearly enough information on the preferences of consumers and firms and also cannot foresee the technological developments in the market. Furthermore, even if the regulator had as much information as the sum of the members of a standardization committee, the regulator does not have the instruments at hand which allow a market-oriented aggregation of the preferences (such as side-payments resorted to by firms). Instead, the regulator is likely to be the target of rent-seeking behavior by firms with high stakes in the standardization decision. Therefore, the only role for government is to apply general competition policy in a way that allows beneficial standardization committees in which the members of one industry convene.

---

[33] These models in their most basic form assume that interconnection is given and that network externalities do not play a role. Only when price-discrimination between on-net and off-net calls are analysed do network size and network externalities play a role.

References

Armstrong, M. (2002), The Theory of Access Pricing and Interconnection, in: Cave, M., S. Majumdar and I. Vogelsang (Eds.), *Handbook of Telecommunications Economics: Vol. I*, Amsterdam.

Armstrong, M. (1998), Network Interconnection in Telecommunications, *The Economic Journal*, 108, 545-564.

Arthur, W. B. (1989), Competing Technologies, Increasing Returns, and Lock-In by Historical Events, *The Economic Journal* 99, 116-131.

Baumol, W. J. and W. E. Oates (1988), *The Theory of Environmental Policy*, Cambridge.

Besen, M. S. and L.L. Johnson (1986), *Compatibility Standards, Competition, and Innovation in the Broadcasting Industry*, Santa Monica, CA.

Blankart, Ch. and G. Knieps (1992), Netzökonomik, *Jahrbücher für neue politische Ökonomie*, 73-87.

Blankart, Ch. and G. Knieps (1993), State and Standards*, Public Choice* 77, 39-52.

Blankart, Ch. and G. Knieps (1994), Kommunikationsgüter ökonomisch betrachtet, *Homo oeconomicus* 11, 449-463.

Blankart, Ch. and G. Knieps (1995), Marked-oriented Open Network Provision*, Information Economics and Policy* 7, 283-296.

Buchanan, J. M. (1987), The Constitution of Economic Policy, *American Economic Review* 77, 243-250.

Chamberlin, E. H. (1950), *The Theory of Monopolistic Competition*, 7th ed., Cambridge, MA.

Cowan, R. (1991), Tortoises and Hares: Choice Among Technologies of Unknown Merit, *The Economic Journal* 101, 801-814.

David, P. A. (1985), Clio and the Economics of QWERTY, *American Economic Association Papers and Proceedings* 75, 332-337.

Demsetz, H. (1969), Information and Efficiency: Another Viewpoint, *Journal of Law and Economics* 12, 1-22.

Dixit, A. K. and J. E. Stiglitz (1977), Monopolistic Competition and Optimum Product Diversity, *American Economic Review* 67, 297-308.

Economides, N. (1996), Network externalities, complementarities, and invitations to enter, *European Journal of Political Economy* 12, 211-233.

Farrell, J. and G. Saloner (1985), Standardization, Compatibility, and Innovation, *Rand Journal of Economics* 16, 70-83.

Farrell, J. and G. Saloner (1986), Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation, *American Economic Review* 76, 940-955.

Farrell, J. and G. Saloner (1987), Competition, Compatibility and Standards: The Economics of Horses, Penguins and Lemmings, in: Gabel, L. (Ed.), *Product Standardization and Competitive Strategy*, Fontainebleau.

Farrell, J. and G. Saloner (1988), Coordination Through Committees and Markets, *RAND Journal of Economics* 19, 235-252.

Farrell, J. and G. Saloner (1992), Converters, Compatibility, and the Control of Interfaces, *Journal of Industrial Economics* 40, 9-35.

Gandal, N. (2002), Compatibility, Standardization, and Network Effects: Some Policy Implications, *Oxford Review of Economic Policy* 18, 80-91.

Gilbert, R. J. (1992), Symposium on Compatibility, *Journal of Industrial Economics* 40, 1-8.

Hayek, F. A. (1945), The Use of Knowledge in Society, *American Economic Review* 35, 519-530.

Holler, M. (1996), Die Modelierung von Netzeffekten und Ansätze industriepolitischer Steuerung, in: Boettcher,E; Herder-Donereich; Schenk, K.-E; Schmidtchen,D. (Eds.), *Jahrbuch für Neue Politische Ökonomie*, Tübingen.

Katz, M. and C. Shapiro (1985), Network Externalities, Competition, and Compatibility, *American Economic Review* 75, 424-440.

Katz, M. and C. Shapiro (1986), Technology Adoption in the Presence of Network Externalities, *Journal of Political Economy* 94, 822-841.

Katz, M. and C. Shapiro (1994), Systems Competition and Network Effects, *Journal of Economic Perspectives* 8, 93-115.

Knieps, G. (1994), Standards und die Grenzen der unsichtbaren Hand, *ORDO* 45, 51-62.

Knieps, G. (1995), Standardization: The Evolution of Institutions versus Government Intervention, in Gerken, L. (Ed.), *Competition AmongInstitutions*, Houndmills.

Knieps, G. (1997), Phasing out Sector-Specific Regulation in Competitive Telecommunications, *Kyklos*, 50, 325-339.

Laffont, J.-J., P. Rey and J. Tirole (1998a), Network Competition: I. Overview and Nondiscriminatory Pricing, *Rand Journal of Economics* 29, 1-37.

Laffont, J.-J., P. Rey and J. Tirole (1998b), Network Competition: II. Price Discrimination, *Rand Journal of Economics* 29, 38-56.

Lancaster, K. (1975), Socially Optimal Product Differentiation, *American Economic Review* 65, 567-585.

Liebowitz, S. (2002), *Re-Thinking the Network Economy: The True Forces That Drive the Digital Marketplace*, New York.

Liebowitz, S. and S. Margolis (1990), The Fable of the Keys, *Journal of Law & Economics* 33, 1-25.

Liebowitz, S. and S. Margolis (1994), Network Externality: An Uncommon Tragedy*, Journal of Economic Perspectives* 8, 133-150.

Liebowitz, S. and S. Margolis (1995), Path Dependence, Lock-In, and History, *Journal of Law, Economics, & Organization* 11, 205-226.

Liebowitz, S. and S. Margolis (2001), Network Effects and the Microsoft Case, in: Ellig, J. (Ed.), *Dynamic Competition and Public Policy: Technology, Innovation, and Antitrust Issues*, Cambridge.

Schmidtchen, D. (1990), Neoclassical and Austrian Theory of Economic Policy: Differences Constitutional Policies, in: Bosch, A., P. Koslowski and R. Veit (Eds.), *General Equilibrium or Market Process*, Tübingen.

Sharkey, W. (1982), *The Theory of Natural Monopoly*, Cambridge.

Shy, O. (2001), *The Economics of Network Industries*, Cambridge.

Spence, M. (1976), Product Differentiation and Welfare, *American Economic Review* 66, 407-414.

Stigler, G. (1968), *The Organization of Industry*, Homewood, Ill.

UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO DI ECONOMIA

THE LONG-RUN IMPACT OF ICT

Francesco Venturini

QUADERNO DI RICERCA n. 254

Abstract

Using some new techniques of panel cointegration analysis, this paper describes the long-run impact of digital capital on the aggregate performance of the US and EU-15 member countries.

ICT is found to significantly impact on output levels without substantial cross-country variation when one adopts the dynamic extension of panel OLS (PDOLS). In this case, however, the long-run elasticity of factor inputs does not differ from the one estimated in the short-run.

The time-series version of seemingly unrelated regression (DSUR) provides more plausible findings, showing a significant cross-countries heterogeneity. The effect of ICT on growth appears relevant - and higher than emerging from short-differences - for most economies but not for the EU largest countries.

JEL Class.: C33, E20, O47.

Keywords: ICT, economic growth, panel cointegration analysis.

Indirizzo: venturini@dea.unian.it. Polytechnic University of Marche, Ancona (Italy).

The Long-Run Impact of ICT*

Francesco Venturini

1 Introduction

In a recent paper Jorgenson and Vu (2005) have shown that capital input accumulation has been the primary source of the world's output growth between the end of the twentieth century and the beginning of the current one. Since 1995 the acceleration in the growth rate of output and labour productivity can be traced for a large fraction to the advances in Information and Communication Technology (ICT).
The impressive improvement in the price-performance ratio of microelectronic components has fuelled the rise in technical efficiency of ICT producing industries and the rapid adoption of computers, software and communication equipment by firms and households, as a consequence of price decline.
The growth impact of ICT has been particularly sizeable in the US as well as in some other countries of OECD area (Finland, Korea and Australia). Also, it has been substantial even in such developing economies as China and India. Instead, aside from few episodes, Europe seems to have lost momentum; for this reason, recently, the EU institutions have renewed with great emphasis their medium-term initiative towards the construction of a common information-based economic space (*i2010*; see EC (2005)).

Thus far the relevance of ICT for national performance has been detected mainly through growth accounts, decomposing output growth into the income share-weighted rise of various factor inputs. On the other hand, econometric literature has focussed principally on short-run effects of high-tech equipment, sometimes comparing countries on the basis of industry data. After a decade from the advent of the often-called information age and, not secondarily, a quarter of century from the first wave of investments in office machinery, it seems useful to investigate the growth effects of ICT from a

* *The author wishes to thank Giulio Palomba for his precious suggestions. Dale Jorgenson, Khuong Vu and Peter Pedroni are gratefully acknowledged for providing some data and codes used in the econometric analysis. Any error is solely responsibility of the author. Comments are welcomed.*

long-run perspective, through a panel cointegration analysis. Given the nature of general purpose technology, the productive impact of digital capital is likely to fully materialize only in a long-term horizon, especially at the most aggregate level of analysis. Accordingly, by applying the usual methods of estimation on short-differences, there is the risk that a part of its contribution remains neglected.

This paper aims at gauging the impact of high-tech equipment on GDP levels for the US and the EU-15 countries over the period 1980-2004. It employs two newly available procedures which represent the dynamic extension of panel ordinary least squares and seemingly unrelated regression (PDOLS and DSUR). However, the latter will be shown to be more suited for our international comparison on ICT, as able to account more powerfully for cross-country dependence, yielding non-trivial differences in results.

The remainder is organized as follows. Section 2 surveys the empirical literature on the role of ICT capital in aggregate performance. Section 3 presents the analytical framework at the basis of the work; it describes the properties of PDOLS and DSUR but, at the same time, discusses the disadvantages related to their application to our data. A short statistical analysis is reported in section 4 where output growth is decomposed into factor inputs' contributions.

Section 5 lays out the econometric findings. Initially, it analyses the trend stationary properties of series (par. 5.1); then, it quantifies the long-run impact of ICT by estimating an aggregation production function (par. 5.2).

Results do vary sizeably in relation to the technique utilised. ICT is found to indistinctly affect output within PDOLS regression, after controlling for country-specific heterogeneity. In this case, however, long-run elasticity does not differ from the one estimated in the short-run.

By employing DSUR there is confirmation that ICT matters for growth, but not uniformly within Europe. In line with the sound of growth accounts studies and previous econometric literature, it is documented that the largest EU economies (France, Germany Italy and UK) have not benefited significantly from digital capital.

Moreover, there emerges a wide discrepancy between the long-run results and the ones obtained on short differences by static SUR, stressing the importance of adopting panel cointegration techniques to study the growth effects of ICT. Finally, section 6 concludes.

2 ICT and economic growth

The impact of ICT investment on economic performance has been scrutinized from more than one perspective. There is a large international evidence that computer use exerts a positive effect on firms' productivity. Dedrick *et al.* (2003) and Pilat (2004) survey this large body of studies concluding that, to be profitable, IT equipment requires complementary investments in communication equipment and software, together with some other collaborating inputs such as human capital and organizational factors. ICT capital acts as enabler of further innovations in many business activities, with a clear advantage for companies undertaking R&D projects.
Yet, as pointed out by Pilat (2004), the benefits gained at the firm-level may be not appearing in aggregate statistics as the poor performance of less productive businesses may obscure the growth of the most innovative ones. This aspect is more accentuate in presence of strong market rigidities that prevent successful firms from emerging, reducing the incentives to high-tech investment. According to Bassanini and Scarpetta (2002), such institutional factors make Europe a scarcely dynamic economic space, in part explaining why ICT has contributed to economic growth to a smaller extent with respect to the US[1].

The prominence of information technology for the US aggregate economy was initially stressed by Jorgenson and Stiroh (2000) and Oliner and Sichel (2000). At the beginning, however, it was believed that US productivity gains were confined within IT production sectors (Gordon (2000))[2]. Now, instead, there is a large consensus on the pervasiveness of growth effects of ICT. The formidable fall in semiconductors' price has fuelled both TFP growth of ICT-producing sectors and high-tech capital deepening in the rest of the economy, accounting for the full one-percent acceleration in labour productivity occurred after 1995 (Jorgenson *et al.* (2003))[3].
Either a lower ICT specialization or a smaller usage of innovative equipment are considered the main determinants of the slower growth experienced by Europe (Timmer *et al.* (2003)).
Nevertheless, a low degree of high-tech specialisation is not necessarily bad for growth as stressed by the performance of Australia, Canada and Mexico

[1] See also Daveri (2004).

[2] Investigating the localization of productivity acceleration within the US, Daveri and Mascotto (2002) find that states where labour productivity grew at a faster rate present an IT specialisation above the national mean. For the EU-15 member states, the evidence provided by O'Mahony and van Ark (2003) goes towards the same direction.

[3] Cette *et al.* (2005) postulate that if the decline in relative prices did persist, the potential output growth in the US could be enhanced by over two percent points per year.

(Pilat and Wolfl (2004)). In terms of welfare, then, relevant benefits from technical advances in ICT production also accrue to using countries, as a consequence of price decline (Bayoumi and Haacker (2002)).

Searching for the industry sources of US resurgence, Stiroh (2002b) and Nordhaus (2005) observe that it originated entirely from those sectors that produce and intensively use digital capital. Relative to the US, O'Mahony and van Ark (2003) point out that the EU is severely lagging in some ICT using service sectors like finance, wholesale and detail trade, where new asset types facilitated radical business re-organisations in the last decade (McGuckin *et al.* (2004)).

The econometric evidence on the nexus between high-tech capital deepening and industry labour productivity growth is mixed. According to Stiroh (2002a), US manufacturing sectors have not taken a particular advantage from high-tech assets, apart from ICT producing firms. On the other hand, O'Mahony and Vecchi (2005) find a positive effect for all US market industries, but not for the UK. For the latter country, however, a favorable indication comes from Oulton and Srinivasan (2005).

At the highest level of aggregation, the scarcity of econometric studies is attributable to the limited availability of comparable statistics. Moreover, evidence does vary depending on the nature of the data employed, timeframe, country coverage and estimation technique.

Relying upon a private source (until 1993), Dewan and Kraemer (2000) find a significant contribution to output from ICT only for developed countries. Park and Shin (2004) 'update' that kind of study to a more recent period (1992-2000), employing the World Development Indicators by the World Bank. A positive effect can be identified either for richer or less industrialised nations, proportionally to the relative level of IT capital. In addition, there emerges an indirect effect of ICT on productivity growth.

A less orthodox attempt of assessing the existence of ICT knowledge spillovers is carried out by Dutta and Otsuka (2004), using patents application data for a small group of nations. Given that the flow of new knowledge (applications) is strongly and positively correlated to the stock of patents applied by high-tech industries, ICT patents are used as a proxy of knowledge within an output production function framework. Nevertheless, in contrast to the prescriptions of new growth theories, GDP is not affected by knowledge input, perhaps due to the short time span considered in this study[4].

Similar in the spirit is the work by Becchetti and Adriani (2005) who regard

[4]Likewise, Aiginger and Falk (2005) observe that the technological specialisation, measured by the share of high-tech exports and EPO ICT patent application, does not add much information to R&D intensity as a source of growth for OECD countries.

ICT as enabler of knowledge diffusion in a growth regression *à la* Mankiw-Romer-Weil. In this respect, the uptake of new technologies emerges as a crucial additional factor to explaining income differentials across countries.

Fuss and Waverman (2005) evaluate instead whether ICT engender networking effects, building a system of simultaneous equations where TFP is supposed to depend on the penetration rate (and digitalisation) of telecom infrastructures. The underlying idea is that advanced communication equipment puts in connection the stock of computers, fuelling the social value of ICT capital over private return. Disentangling TFP into various sources (scale economies, time trend and spillovers), Fuss and Waverman (2005) document that ICT externality was the main contributor of productivity in the late 1990s for most OECD countries. Following this line of discussion, it is reasonable to believe that productivity spillover might be even higher than estimated in this paper as it does not consider the benefits stemming from the connection with IT capital of trading partner countries.
To a broad extent, this is the goal pursued by Lee and Guo (2004). They explicit foreign high-tech investment as a determinant of national TFP growth, reporting a robust evidence in favour of spillovers going from richer to less developed countries. According to Gholami *et al.* (2005), the relation between ICT and international trade is double-sense. On one hand, digital capital boosts foreign direct investment from more industrialized economies since it facilitates the access to information on new markets and co-ordination with headquarters. On the other hand, the inflow of FDI also favours the dissemination of new technologies in less developed countries.

3 Analytic framework

After 1995 there has been a valuable research effort to assess the fraction of output growth traceable to the deployment of digital capital. Aside from few exceptions, econometric studies have adopted techniques more suited for the short-run, being based on first-differenced variables. This has advantage of working with stationary series and using traditional inference to test the robustness of results. As known, however, it happens at the cost of losing some useful information when there exists a long-run stationary relation between dependent variable and regressors (cointegration).

Real investment in high-tech equipment has accelerated enormously in the last decade, growing annually at double-digit rates. Although, it should not be forgotten that the installed capacity was no-negligible already in the first half of the 1990s. At least partially, current earnings may originate from the past (Gordon (2004)). Given the nature of general purpose technolo-

gies, ICT needs a long time to yield returns since its introduction is usually accompanied by business re-organizations, complementary investment and, more generally, adjustment costs. This is the main explanation put forward by Brynjolfsson and Hitt (2003) and Oulton and Srinivasan (2005) to justify why high-tech capital exhibits a larger coefficient when estimated on long differences in comparison to annual growth rates.
Moreover, it is likely that both direct effects and productivity spillovers of ICT become more apparent in the long-run at an economy-wide level, in light of those compensation effects working between firms and/or industries. In this connection, looking at a long-term horizon, O'Mahony and Vecchi (2005) find a significantly positive impact of ICT on output growth, which is also compatible with the presence of spillovers, being well above the income share.

This paper seeks to measure the long-run elasticity of digital capital across a moderate panel of countries, by estimating an output production function through two new techniques of cointegration analysis. The former is the dynamic version of panel OLS estimator (PDOLS; Mark and Sul (2003)). Recently, it has been used by Bottazzi and Peri (2006) to study the impact on national stock of knowledge (patents) of R&D expenditure and the patenting ability of trading partner countries. The latter estimator has been proposed by Mark *et al.* (2005) and Moon and Perron (2005), consisting in the time series extension of seemingly unrelated regression (DSUR).

Assuming a log-linear Cobb-Douglas production function, GDP (Y) can be expressed as dependent on hours worked (H), traditional capital and ICT assets (K_N and K_{ICT})[5]:

$$y_{it} = \underline{\alpha}_i' \underline{d}_{it} + \underline{\beta}_i' \underline{x}_{i,t} + u_{it}$$

$$\ln Y_{it} = \alpha_{0i} + \alpha_{1i} t + \beta_{1i} \ln H_{it} + \beta_{2i} \ln K_{N,it} + \beta_{3i} \ln K_{ICT,it} + u_{it}; \qquad (1)$$

i=1,...,N denotes the cross-sectional units (N=16), t=1,...,T time dimension (T=25) whilst $\underline{d}_{it}$ is the vector of deterministic components. This includes an individual intercept, a time trend and, also, common time dummies when running PDOLS.
It is evident that equilibrium error u_{it} is affected by the endogeneity with own-equation regressors and cross-dependence with the disturbances of other

[5] The hypothesis that a Cobb-Douglas specification is valid for the entire sample may be debatable. For the US, Antràs (2004) reports evidence against this assumption while, for instance, it seems to hold for the Finnish case only on a very long time horizon (Jalava *et al.* (2005)). However, for a large panel of countries, Kumbhakar and Wang (2005) obtain with a Cobb-Douglas factor inputs' elasticities very close to the ones resulting from a translog specification, once controlled for heterogeneity (fixed effects).

equations. If first-differenced variables are stationary, they can be modelled as correlated random walks ($\Delta x_{it} = e_{it}$) and the long-run covariance matrix of equations' system error (u_{it}, e_{it}) can be represented as follows:

$$\Omega = \begin{pmatrix} \Omega_{u'u} & \Omega_{u'e} \\ \Omega_{e'u} & \Omega_{e'e} \end{pmatrix}.$$

The covariance matrix of u_{it} shows the degree of correlation among equations ($\Omega_{u'u}$); non-null values of off-diagonal parameters determine the inefficiency of least squares estimator, imposing the adoption of seemingly unrelated regression. $\Omega_{e'e}$ regulates instead the cross-equation dependence of regressors, explaining why some efficiency gain can be obtained by a system estimator relative to the ones based on single equations. Finally, $\Omega_{e'u}$ models the endogeneity between error term and regressors that, within each equation, is source of bias for static estimation techniques.

In order to purge the equilibrium errors u_{it} from the effect of reverse causality, Mark *et al.* (2005) suggest of including into equation (1) p lags and leads of first-differenced regressors of own-equation (Δx_{it}). The insertion of the ones of other panel units (Δx_{jt}) is designed to remove cross-section dependence.

Operatively, PDOLS and DSUR can be obtained through a feasible two stages procedure. As a first step, any individual dependent variable (y_{it}) and each element of regressors' matrix ($\hat{\mathbf{x}}_{it}$) are regressed on the vector of p lags and leads (v_{it}). Then, after stacking residuals (denoted by a hat) of auxiliary regressions into a system, the cointegration vector can be computed by means either of the estimator of least squares or seemingly unrelated regression[6] :

$$\underline{\hat{\beta}}_{pdols} = [\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \hat{\mathbf{x}}_t^{'}]^{-1} [\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \hat{\underline{y}}_t]$$

$$\underline{\hat{\beta}}_{dsur} = [\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \Omega_{u'u}^{-1} \hat{\mathbf{x}}_t^{'}]^{-1} [\sum_{t=p+1}^{T-p} \hat{\mathbf{x}}_t \Omega_{u'u}^{-1} \hat{\underline{y}}_t].$$

$\underline{\hat{\beta}}_{dsur}$ is thus shaped as computed under the assumption of cross-sectional homogeneity of cointegrating vector ($\beta_1 = ... = \beta_N$; *restricted* DSUR) which seems plausible when only few annual observations are available (Mark and Sul (2003); p. 4). Accordingly, individual heterogeneity is confined to short-run dynamics.

[6]The effect of truncation is not discussed for sake of brevity. $\hat{\mathbf{x}}_t = (\underline{x}_{1,t}, ..., \underline{x}_{N,t})$ is a k x N matrix.

In addition, the Wald test aimed at assessing the similarity among single-equation cointegration vectors cannot be used when time dimension is limited (less than 300) as leads to reject the null hypothesis of homogeneity, being systematically oversized.

The family of PDOLS and DSUR estimators represents a valid alternative to error correction models (ECM) where short- and long-run parameters are estimated jointly. The latter framework has been adopted by Guellec and van Pottelsberghe (2004) to study the impact of R&D expenditure on TFP - estimated through three-stage least squares and static SUR- for a panel of countries similar to ours.

To account for heterogeneity in a ECM setting, Pesaran *et al.* (1999) propose a maximum likelihood approach that imposes common long-run parameters while allowing a different (short-run) dynamics among the panel units (pooled mean group; PMG). The logic underlying such a procedure is close to one employed in this paper but, computationally, requires a larger amount of cross-sectional observations. PMG estimator has been employed by O'Mahony and Vecchi (2005) and, also, by McMahon *et al.* (2005) to assess the effect of ICT on global investment cycle. Nevertheless, given the short size of panel, the latter work utilises a modified version of PMG based on SUR.

Numerous alternative procedures of panel cointegration analysis have been developed recently. For instance, Larsson *et al.* (2001) extend to multiple equations the Johansen's method of rank cointegration. Instead, Pedroni (2000) suggests a version of fully modified OLS estimator (FM-OLS) corrected for long-run endogeneity. This body of literature is accurately surveyed by Breitung and Pesaran (2005).

4 Data characteristics and some descriptive results

Our study employs the GGDC Total Economy Growth Accounting database that has been developed at the University of Groningen (Netherlands) by Bart van Ark and his scholars[7]. It includes all the European Union member states before the enlargement (EU-15) and refers to the period 1980-2004. As a measure of output it considers GDP net of rentals paid for residential buildings in order to avoid any distortion related to large cross-country differences

[7]http://www.ggdc.net/dseries/growth-accounting.html, release June 2005. See van Ark *et al.* (2002) and Timmer *et al.* (2003) for details and the underlying growth accounting methodology.

in their measurement. Hours worked (in million thousands) are adopted as a proxy of labour input and, as a result, the contribution of labour quality is embedded into the residual (TFP).
Capital is disentangled into three kinds of ICT asset (computer and office machinery, communication equipment and software) and three non-ICT related types (non-IT equipment, transport equipment and non-residential buildings). For high-tech investment, a qualitative adjustment based on the US hedonic deflators is made to guarantee a homogeneous treatment of technical characteristics, especially for computers (price harmonisation).

All monetary variables have been converted from national currencies into US constant dollars of 2002 but are not transformed into a PPP base, given the lack of relative prices for ICT.
A vigorous debate has developed around the validity of purchasing power parity hypothesis, that is whether exchange rates compensate disparities in relative prices over a long-term horizon (see Sarno and Taylor (2002)). Evidence does vary according to sample coverage and data frequency; the usage of monthly price indexes may lead to refute such an assumption, as demonstrated, among others, by Moon and Perron (2005) adopting DSUR. Using highly frequency series Coakley *et al.* (2005) have shown that the parity hypothesis cannot be accepted when built on consumption price index while the opposite happens if based on producer price indexes.
Annual data instead usually provide favourable evidence (Pedroni (2004)). This is also confirmed for the group of EU-15 member countries by trend stationary of PPP index (OECD (2005))[8]. Therefore, as common in cross-country regressions focused on the long-run, the bias provoked by the missing conversion of monetary series into international dollars should be minimal[9].

As for descriptive analysis, **Table 1** reports the decomposition of output growth into factor inputs' contribution. If one considers Europe and United States as a whole (section A), it is possible to notice that the share-weighted growth of ICT capital has been as high as the one of traditional assets over the period 1980-2004. In average, it amounts to 0.6%-points per year, reflecting an annual average growth of 14% and an income share of 0.04. ICT accounts for one seventh of all capital services.
Looking at various time intervals, it is striking the acceleration in TFP growth occurred after 2001. In the last years, labour contribution zeroed while the value of ICT capital services more then halved with respect to 1995-2001.

[8]Trend stationary is displayed in Table A.1 of Appendix. Panel unit roots tests utilised are described in section 5.1.

[9]It should be reminded that the log-levels specification employed in this work includes a time trend whilst short-run variations are eliminated by the first step of regression.

Taking into account the EU-15 and US distinctly (section B), it becomes evident that productivity exploded only in the US between 2001 and 2004, while it showed an inexorable downward trend in Europe since the mid-1990s. In the latter period the EU-15 has not been able to technologically catch up the US; the growth contribution of high-tech equipment decreased to 0.2% per year from 0.6 of 1995-2001 and, in practice, the delay remained almost unchanged.

For the aim of the work it is particularly useful to focus on country-specific data. **Table 2** shows the wide heterogeneity within Europe in the contribution of ICT which ranges from 0.22 percentage points of Greece to 0.70-0.74 of Belgium and Luxembourg. The aggregate contribution of ICT hides a substantially different dynamics in high-tech expenditure (office machinery, communication equipment and software). A large fraction of the EU lag can be ascribed to the small contribution of computers and software, especially in the major continental economies. Figures comparable to the US are exhibited only by Belgium and Luxembourg for computers, Scandinavian countries for software and, finally, Finland and Italy for communication equipment.

With regard to the other sources of growth, TFP arises as the main driver of output for most countries except for Austria, Greece, Italy, Luxembourg and Spain. It is remarkable the reduction in the growth contribution of labour quantity (hours worked); relevant values are shown only from those countries that liberalised more intensively the labour market during the 1990s (Ireland, Luxembourg, Netherlands and Spain). Likely, as illustrated by Jorgenson and Vu (2005), human capital has been another key factor for development but this does not appear in our data as incorporated into Solow's residual.

Finally, it should be pointed out that traditional capital accounted for a sizeable part of GDP growth either in most catching-up countries or such advanced economies as France and Italy. In the latter group, it has been the driver of economic growth, showing a contribution largely superior to high-tech capital, contrarily to what happens for Germany.

5 Econometric results

5.1 Panel unit roots and cointegration analysis

This section is devoted to demonstrate that the macro-economies series employed in the estimation are trend stationary and there exists a relation of cointegration among them.

As a first step, we need to show that log-levels variables are not stationary

but they do if considered in first differences. To check the integration degree of series we employ the t-bar statistics developed by Im, Pesaran and Shin (2003) (in so forth IPS) that consists in an average of ADF tests carried out on each country equation.
IPS test assumes a null hypothesis of non-stationary ($\beta_i = 0$ for all i). It diverges to a negative infinite under the alternative one, allowing for heterogeneity in short-run dynamics ($\beta_i < 0$ for some i). Contemporaneous interdependence is removed by subtracting out the cross-sectional mean (time demeaning), that is equivalent to working with common time dummies:

$$\Delta y_{it} = \underline{\alpha}_i^{'} \underline{d}_{it} + \beta_i y_{it-1} + \sum_{p=1}^{T} \gamma_{ip} \Delta y_{it-p} + u_{it}.$$

Im *et al.* (2003) have demonstrated that t-bar is more powerful than the previous generation of unit roots tests, based on the alternative hypothesis of homogeneity ($\beta_i < 0$ for all units), as the one proposed by Levin and Lin (1993) (hereinafter LL).

Table 3 reports both kinds of tests where the optimal number of ADF lags is chosen by a step-wise procedure minimizing the Akaike information criterion. Along with country-specific intercepts, a time trend is included in log-levels specification but not with annual growth rates. The acceptance of the null hypothesis in the first regression (levels) and, contemporaneously, the rejection in the second one (growth rates) mean that our series are trend stationary.
Unequivocally, this is the conclusion indicated by inference based on IPS test. Moreover, the log-levels specification makes apparent the relatively large power of such a test that, in contrast to LL, points to non-stationary of both capital series. This outcome signals the presence of a considerable degree of cross-sectional heterogeneity for these two variables.
According to Pesaran (2005), IPS test is not fully affordable with high levels of cross-sectional dependence, as only partly removed by time demeaning. In this case, the coefficient of lagged level is downward biased and the null hypothesis of non-stationary is over-rejected by t-bar[10]. However, as the

[10] In alternative, Pesaran (2005) proposes a test consisting in a average of t-ratio statistics carried out on Dickey-Fuller regressions ($CIPS = (1/n)\sum_i CADF_i$) which are augmented either with the lagged level or the growth rate of cross-sectional mean of the dependent variable ($\overline{y}_{t-1}$ and $\Delta\overline{y}_{t-p}$). When residuals are serially correlated the equation can be expressed as follows:

$$\Delta y_{it} = \underline{\alpha}_i^{'} \underline{d}_{it} + \beta_i y_{it-1} + \phi_i \overline{y}_{t-1} + \sum_{p=0}^{T} \gamma_p \Delta \overline{y}_{t-p} + \sum_{p=1}^{T} \gamma_{ip} \Delta y_{it-p} + u_{it}.$$

values of test statistics for log-levels in **Table 3** show, this does not hold for our case and IPS test works reasonably well.

Next we have to verify whether macro-economic series are cointegrated, that is there exists a long-run production function. In so doing, we rely upon the ADF-type statistics proposed by Pedroni (1999) and (2004); they belong to a set of seven tests built on the residuals of least squares regression of the potentially cointegrated relation (eq. 1). All these tests are shown to be robust to double-sense causality. They can be distinguished into two types (panel and group mean tests), both sharing the null hypothesis of no cointegration but diverging for the alternative one. Being computed on pooled annual data, panel tests assume a common cointegrating vector while the group mean tests, which consist in between-averages of the individual statistics, admit heterogeneity in parameters.
As analogous to the augmented Dickey-Fuller approach, panel ADF statistics can be regarded as the closest to the unit roots test proposed by Levin and Lin (1993), whilst the group mean ADF statistics to the test devised by Im *et al.* (2003). The latter should be preferred, especially in short panels, otherwise the null hypothesis of no cointegration might be accepted even though valid for few units.
Nevertheless, as **Table 3** illustrates, this possibility is excluded for our analysis; the null hypothesis is always rejected when the output production function is specified with time trend (along with time dummies). This indicates the existence of a cointegration relation among output and factor inputs[11].

5.2 Estimation of direct effect of ICT on GDP

Panel dynamic ordinary least squares estimation (PDOLS)

The estimation of long-run output production function is carried out initially by panel DOLS[12]. The first section of **Table 4** considers the EU-15 countries and the US as a whole (columns I-IV). The second part instead leaves out

Along with IPS test, this statistics has been employed to check the trend stationary of PPP index for the EU-15 member countries displayed in Table A.1 of Appendix.

[11] Both statistics are one sided tests which are distributed as normal standard, diverging to a negative infinite under the alternative hypothesis (cointegration). Panel statistics are not weighted for the long-run variance as outperforming the weighted tests in small sample. See Pedroni (2004), note 4.

[12] Panel cointegration estimation is carried out with Gauss codes made available by Donggyu Sul. Both in PDOLS and DSUR, the maximum value for the step-down procedure selecting the optimal number of lags (and leads) of first-differenced regressors is fixed to one; standard errors are corrected with pre-weighting method (see Mark *et al.* (2005); p. 802).

Finland and Sweden as, how it will be discussed later, they are source of a severe noise for the estimation (col. V-VI).

For sake of completeness, columns I-II also display the estimates obtained without trend, even though in the following attention will be paid only to trend stationary case, based on the inclusion of this deterministic component. It should be noticed from the first two columns that the elasticity of factor inputs are slightly different from the income shares reported in section 3. Labour and traditional capital are estimated a somewhat smaller (about a fifth) contrarily to ICT whose coefficient is three times higher[13]. In line with a large international evidence, there is proof of slightly decreasing returns (around 0.90).

The inclusion of time trend reduces remarkably the size of coefficients. Since macro-economic series grow uniformly over time, a large fraction of their variance can be explained deterministically, lowering R-squared (col. III-IV).

Note that traditional assets are no longer significant. Instead, the coefficient of hours worked passes from 0.52 to 0.32 whilst ICT capital from 0.14 to 0.09 when not controlling for the effect of contemporaneous shocks. Introducing time dummies reduces only the output elasticity to labour, given the more pronounced cyclical nature of this factor input.

The insignificance of low-tech capital might be interpreted in two alternative ways. Once controlled for time trend, its dynamics may be characterised by a low variance so to be uninformative for explaining output levels.

On the other hand, it may depend on some noise in data. Indeed, looking at de-trended values, it is possible to see a clear break for Finland and Sweden around 1991-92[14]. As discussed by Daveri and Silva (2004), Finland started a transition from a semi-planned to an open-market economy in those years, dismissing a large part of its inefficient over-installed capacity (capital shedding). A similar story seems to hold for Sweden as well; Lindbeck (2000) argues that the severe recession of the early 1990s forced Swedish firms to a more intensive utilisation of capital. This effect overlapped to a deceleration in real investment started in the mid-80s, leading to a downsized process of capital accumulation.

As a consequence, we have re-estimated the output production function excluding such countries (col. V). Expectedly, traditional capital becomes sig-

[13]Table 1 shows the annual average of pooled income shares (0.69 for labour, 0.26 for traditional capital and 0.04 for ICT equipment). Being time invariant, these values coincide with the cross-sectional average of time-series means.

[14]Unified Germany does not present any relevant change in series, as macro-economic aggregates have been estimated backwardly from post-unification levels by using the annual growth rates of West Germany (see Timmer *et al.* (2003)).

nificant at a 1% level, showing a coefficient of nearly 0.19. Moreover, in comparison to column IV, there is a fall in the elasticity of labour and ICT capital. It should be also emphasized that restricting the focus on the European member states does not modify results at all.

At this point it seems useful to compare the long-run elasticity of factor inputs with the short-run values which can be estimated by applying usual techniques to first-differenced (stationary) variables.

Table 5 shows the results for the full sample (US and EU-15) either over the whole timeframe or for the post-1992 period, thus to avoid any possible distortion due to the break in Swedish and Finnish series. Similarly to PDOLS, the model is also re-estimated over 1980-2004 dropping out these countries (col. VI-IX). In order to guarantee some minimal heterogeneity among panel units, all specifications include country-specific intercepts, allowing for time dummies but not time trend.

Despite the well-known problems of reverse causality (endogeneity), panel static OLS regression yields coefficients not dissimilar from such instrumental variables procedures as IV-2SLS and the one-step difference GMM estimator (Arellano and Bond (1991))[15]; a consistent finding is obtained by Rincon and Vecchi (2004) in a firm-level analysis on ICT. Diff-GMM relies upon valid instruments as illustrated by the p-value of Hansen test and seems appropriate for our model in light of Arellano-Bond tests on serial correlation of residuals. Although, by removing fixed-effects, it loses some useful information in coefficients' estimation in comparison with IV-2SLS.

For the whole period and full sample (col. I-III), there is confirmation of the irrelevance of low-tech capital for GDP growth, in line with previous outcomes. On the other hand, the coefficient of hours worked is larger than in the long-run while ICT shows a smaller elasticity; at the best, the latter is estimated a tenth of percentage point lower than in the PDOLS regression (0.085 with IV-2SLS against 0.094).

Restricting the analysis to the most recent years (1992-2004) improves the estimation of traditional capital only by running IV-2SLS. By contrast, labour contribution lessens, probably because of the decline in hours worked per

[15] All short-run regressions utilise standard errors robust to heteroskedasticity. To reduce as much as possible the asynchrony of business cycle, IV-2SLS employs as instruments four-year lagged variables for the estimation over 1980-2004 and two-year lags for 1992-2004. Alternative procedures have provided findings qualitatively inferior the ones reported in Table 5. To limit the larger variability of small economies' observations, we run a weighted least squares regression but results do nearly coincide with OLS; instead, admitting AR(1) errors reduces of about a fifth the size of least squares coefficients. Finally, given the scarce persistency of first differenced variables, sys-GMM performed more poorly than diff-GMM (Blundell and Bond (1998)). All results cited hereinafter but that are not reported in the main text are available on request from the author.

employed occurred in Europe in the last fifteen years (McGuckin and van Ark (2005)).
More interesting outcomes emerge when Finland and Sweden are maintained out from the first-differences regression (col. VI-IX). Adopting IV-2SLS low-tech capital is significant at a 10% level and, most importantly, factor inputs' elasticities become close to the long-run ones reported in the last two columns of **Table 4**.
Therefore, no relevant gain seems to come from the usage of cointegration techniques in the estimation of output production; this finding leaves open some doubt on the validity of PDOLS estimator for our type of analysis and, accordingly, now we turn to estimate dynamic SUR.

Dynamic seemingly unrelated estimation (DSUR)

Thus far cross-countries correlation has been shaped through common time dummies but PDOLS findings display only small variation when they are introduced in the estimation.
Modelling cross-sectional dependence in this way may be too restrictive and relevant efficiency gains can be obtained by means of seemingly unrelated regression. The dynamic SUR estimator devised by Mark *et al.* (2005) computes one cointegrating vector for each single equation, verifying whether they can be considered statistically identical within the system through a Wald test. Unfortunately, this test is systematically oversized when the time dimension is no sufficiently long and, given T, this problem rises with the size of panel.
Furthermore, to implement DSUR there is need to calculate a larger amount of parameters with respect to panel DOLS. As a result, with relatively short time series available, one has to divide the cross-sectional units into various sub-groups[16], even though it lowers the estimates' precision in comparison with a full system regression.
Therefore, output production function is estimated by means of DSUR over different groups under the hypothesis of a homogeneous cointegration vector for each of them. Breaking sample has nevertheless the disadvantage of producing results that may differ among groups because of their composition, rather than the underlying production technology. As a consequence, the robustness of such differences is checked by a version of Chow test robust to heteroskedasticity (Wald test).
We assume a positive relation between scale and production technology and

[16]Mark *et al.* (2005) stress that cross-sectional units should be at least a twelfth of the time dimension to make unrestricted DSUR feasible and with all desirable properties.

divide the sample into three groups: big, medium and small countries[17]. This classification is robust to various alternatives and, as it will be evident later, leaves unchanged the final results of analysis. Preliminarily, we attempted to cluster countries on the basis of their interdependence - measured by the import share on GDP - but it did not provide convincing findings.

Table 6 presents either DSUR or PDOLS results to facilitate the comparison between these estimators and see how the estimation improves when cross-sectional dependence is more powerfully controlled for[18].

First of all, it should be noticed that the parameters of sub-regressions are always statistically different, as shown by the values of Wald test. The unique exception is given by the inclusion of the US in DSUR estimation where there emerges a similarity between large and small countries ($\chi^2(3) = 0.06$). This contrasts with the idea that, if some link exists between technology and scale of production, a larger similarity should be expected between large and medium-sized economies or, alternatively, between the ones localised in the middle and in the right tail of the distribution. However, we can anticipate, this anomaly will disappear when Finland and Sweden are excluded from the analysis.

PDOLS results present some remarkable points. First, labour elasticity is higher in the largest countries by a three times factor relative to the other states (around 0.62-0.67 against 0.21-0.25). Second, low-tech capital is uninformative for explaining output levels only for the first group of countries while, not surprisingly, it enters with a wrong sign for the medium-sized ones as including Sweden. Note that the same happens in DSUR for the smallest members of the EU due to the presence of Finland. Last but not least, digital capital exhibits a positive and significant coefficient only for the group of small countries (0.094) that is made up by some notorious ICT-intensive users like Denmark, Finland, Ireland and Luxembourg[19].

[17]Big countries: US, Germany, France, UK and Italy. Intermediate countries: Spain, Netherlands, Belgium, Sweden and Austria. Small countries: Greece, Portugal, Denmark, Finland, Ireland and Luxembourg.

[18]In a single-equation regression, PDOLS and DSUR coincide with the estimator of dynamic ordinary least squares of Saikkonen (1999); these results are presented in Table A.2 of Appendix. Given the wide discrepancy in estimates, the similarity of coefficients between one-equation and system regressions is always refused; for simplicity, the corresponding values of Wald test are not reported in Table 6 and 7.

[19]To check the robustness of this finding to sample composition, we have re-estimated the model separately for ICT-intensively users on the one hand and less technological advanced countries on the other (Greece and Portugal). The null hypothesis of homogeneity in the cointegration vector cannot be rejected, meaning that the similarity in the contribution of labour and traditional capital is stronger than heterogeneity appearing in the deployment of Information Technology.

As apparent from the right section of the table, relevant efficiency gains stem from using dynamic seemingly unrelated regression in place of PDOLS. Also, coefficients show some change. The growth effect of hours worked is now slightly smaller for the entire group of big countries (col. V), but it rises when the US are left out (col. VI). A comparable coefficient is presented by the smallest economies but not by medium-sized ones, which instead exhibit a rather downsized elasticity.
Traditional capital is significant at a 10% level considering together all the largest economies, showing a coefficient of 0.074. Restricting the focus on the EU states the coefficient jumps up to nearly 0.19 and reaches the highest level of significance. Again, the disappointing performance of the other two groups can be widely attributed to Sweden and Finland.
It is surely in the estimation of ICT impact that DSUR outperforms panel dynamic OLS. Digital capital emerges now as a driver of growth for most countries except for the EU big states. Its coefficient ranges from 0.054 for the intermediate group to 0.171 of smaller economies whilst, evidently, the value reported in column V (0.124) reflects the inclusion of the US[20].
This result is consistent with growth accounts evidence. In Europe the major continental economies have not been able to exploit the growth potential of high-tech equipment because of more general problems of competitiveness. Despite relatively higher investment rates in technologically advanced capital, also the performance of the United Kingdom is downsized if compared to the US, confirming the findings reported by Basu *et al.* (2004) and O'Mahony and Vecchi (2005).
The large improvement in estimating the long-run impact of ICT suggests that cross-country correlation may be stronger for this factor input than for labour and traditional assets. It is likely to mirror the technique adopted to deflate national investment, relying upon US hedonic prices (Schreyer (2002)). This aspect makes DSUR the optimal procedure to assess the growth effects of ICT across countries.

As in the foregoing analysis, now we turn to assess how the inclusion of Sweden and Finland affects the results. There is an important distinction between PDOLS and DSUR; the former estimation does no longer exhibit significant differences among various groups and, thus, all countries can be pooled. Note, however, that this specification has been already estimated and presented above in **Table 4** (col. V-VI).
One the other hand, by running DSUR significant divergencies persist be-

[20] It is interesting to underline that factor inputs' elasticity are aligned to income shares only for France, Germany, Italy and UK.

tween the major EU countries and the rest of the sample[21]. Therefore, we have re-run dynamic seemingly unrelated regression joining together small and medium-sized countries. Outcomes are summarized in the left section of **Table 7** which, for comparative aims, reports once again the values relative to the top economies.
A rather clear picture emerges now on the role played by various factor inputs in the European development over the last quarter of century.
The largest economies exhibit a growth pattern dominated by labour input. Instead, the deployment of low-tech capital has been substantially homogeneous among countries whilst the ICT contribution does vary remarkably (0.041 vs 0.117). Indirectly, **Table 7** provides confirmation of the US leadership in terms of high-tech capital utilisation, given that the coefficient reported in the first column exceeds the one of EU countries.

Finally, to compare long- and short-run elasticities we have also estimated static version of seemingly unrelated regression (Zellner (1962)) on first-differenced variables. Albeit the well-known problems of endogeneity between regressors and disturbances, one can observe that the coefficients of dynamic regression are always larger[22], further witnessing the importance of using DSUR for this kind of study.

Sensitivity analysis of parameters

One of the main outcomes of the preceding analysis is that the size of ICT elasticity is always above the income share when significant. Several arguments have been advanced in productivity literature to explain a similar result, all pointing to the inadequacy of neoclassic assumptions at the basis of income shares' calculation (perfect competition and full exhaustion of output).
Our evidence excludes increasing returns and, also, error measurements given the usage of hedonic deflators for high-tech expenditure. Furthermore, it is difficult to believe that the market of ICT assets behaves less competitively than the one of traditional equipment. Thus, only two explanations remain valid for our findings (see Stiroh (2002a)).
On one hand, the size of ICT coefficient might reflect the presence of external effects caused by networking or productivity spillovers that push up the social value of digital capital over private return.
On the other hand, it might be upward biased because of the omission of

[21]Intermediate regressions are reported in Table A.3 of Appendix.

[22]It makes exception labour contribution in the smallest economies. As above discussed, the strong cyclical nature of hours worked makes this factor input extremely sensitive to double-sense causality.

labour quality in output production function. Increasing levels of firms' spending in IT have been accompanied over time by complementary investments in human capital which, by nature, tend to further enhance the returns of IT equipment. The mutual self-enforcing effect between these factors is stressed by Brynjolfsson and Hitt (2003) who underline the difficulty of disentangling the two components in absence of accurate data. Sometimes, however, ICT elasticity shows small variation between using quality-adjusted or raw labour data (O'Mahony and Vecchi (2005)).

In the following we check the sensitivity of ICT coefficient on a group of countries (France, Germany, Netherlands, UK and US) for which labour quality series are available from a consistent source, even though for a shorter time span (1980-2000)[23].

Table 8 reports DSUR results. For comparative aims, the first two columns also present findings relative to 1980-2004 based on hours worked. The coefficient of ICT changes remarkably depending on the inclusion of the US; it amounts to 0.092 for the EU-4 countries, rising to 0.140 when the US are comprised. In light of the results of the previous section, the significance of digital capital for the EU countries is clearly attributable to the good performance of the Netherlands.

The second part of **Table 8** highlights the scarce covariation between ICT and human capital as the contribution of former factor remains stable albeit the quality adjustment of labour. By contrast, there is a fall in the coefficient of traditional assets and, as expected, an increase in labour elasticity[24].

Although this finding cannot be generalized as reflecting the subgroup composition, it leads to believe that the upward bias of ICT coefficient may be small for the overall sample as well (EU-15 and US). Indirectly, this supports our previous results.

6 Concluding remarks

Despite a global convergence in the uptake of digital technologies occurred after 1995, this work has shown that non-negligible differences persist within

[23]GGDC Industry Growth Accounting database. The growth rate of labour quality is calculated at an industry level and then aggregated for the overall economy. The relative level of labour quality among countries has been fixed to the values reported by Jorgenson and Vu (2005) using 1995 as benchmark year. Finally, a synthetic measure of labour input has been computed multiplying hours worked by the index of human capital.

[24]When the focus is restricted on the EU-4 countries, the coefficient of ICT does modify marginally. A long-difference regression for the period 1989-95 and 1995-2003 carried out on all EU-15 countries (less Luxembourg) and US, exploiting labour quality data built by Jorgenson and Vu (2005), qualitatively support the previous evidence.

Europe in the growth effects of ICT.

In line with the large body of growth accounts literature, the leading countries of the EU are found to sensibly lag behind the US in terms of productive impact of new asset types. On the other hand, there is a core of small dynamic economies in Europe whose growth pattern has been positively influenced by the deployment of Information Technology.

The disappointing technological performance of the major continental states is usually ascribed to the structural weakness of their economies, become apparent in the last decade. A low specialization on innovative productions and a rigid regulation of internal markets are the roots of the fall of competitiveness. These factors have lessened the incentives to high-tech investment, depressing the global returns of ICT; it is known that digital capital makes firms more flexible but, at the same time, needs a dynamic environment to yield efficiency gains.

In this connection, the renewed commitment of the EU institutions for creating a common digital platform (i-2010) and sustaining ICT usage goes towards the right direction. Yet, these interventions are unlikely to stimulate productivity until stronger policies for competition and innovation are not pursued by national institutions.

Another remarkable feature of the paper can be identified in the usage of panel cointegration techniques to estimate the sources of growth. At an economy-wide level of analysis, this is indispensable to assess the contribution of ICT whose nature of general purpose technology confines relevant productivity gains to the long-run. In this respect, dynamic seemingly unrelated regression arises as the most suited procedure of estimation, allowing to identify the wide discrepancies existing within Europe and between the EU and US.

References

AIGINGER, K. AND FALK, M. (2005). *Explaining Differences in Economic Growth among OECD countries.* Empirica, 32: 19–43.

ANTRÀS, P. (2004). *Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution.* Contributions to Macroeconomics, 4(1).

ARELLANO, M. AND BOND, S. (1991). *Some Tsts of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.* Review of Economic Studies, 58: 277–97.

BASSANINI, A. AND SCARPETTA, S. (2002). *Growth, Technological Change, and ICT diffusion: Recent Evidence from OECD Countries.* Oxford Review of Economic Policy, 18(3): 324–44.

BASU, S., FERNALD, J., OULTON, N. AND SRINIVASAN, S. (2004). *The Case of the Missing Productivity Growth: Or, Does Information Technology Explain why Productivity Accelerated in the United States but not in the United Kingdom?* In *NBER Macroeconomics Annual 2003* (edited by GERTLER, M. AND ROGOFF, K.), pp. 9–63. The MIT Press, Cambridge, MA.

BAYOUMI, T. AND HAACKER, M. (2002). *It's Not What You Make, It's How You Use IT: Measuring the Welfare Benefits of the IT Revolution across Countries.* IMF Working Paper No. 117.

BECCHETTI, L. AND ADRIANI, F. (2005). *Does the Digital Divide Matter? the Role of Information and Communication Technology in CrossCountry Level and Growth Estimates.* Economics of Innovation and New Technology, 14(6): 435–53.

BLUNDELL, R. AND BOND, S. (1998). *Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.* Journal of Econometrics, 87: 115–43.

BOTTAZZI, L. AND PERI, G. (2006). *The International Dynamics of R&D and Innovation in the Short and Long Run.* Economic Journal, Forthcoming.

BREITUNG, J. AND PESARAN, M. (2005). *Unit Roots and Cointegration in Panels.* Cambridge University Working Papers.

BRYNJOLFSSON, E. AND HITT, L. M. (2003). *Computing Productivity: Firm-Level Evidence.* Review of Economics and Statistics, 85(4): 793–808.

CETTE, G., MAIRESSE, J. AND KOCOGLU, Y. (2005). *ICT Diffusion and Potential Output Growth.* Economic Letters, 87(2): 231–4.

COAKLEY, J., KELLARD, N. AND SNAITH, S. (2005). *The PPP Debate: Price Matters!* Economics Letters, 88(2): 209–13.

DAVERI, F. (2004). *Why is There a Productivity Problem in Europe?* CEPS Working Document No. 205.

DAVERI, F. AND MASCOTTO, A. (2002). *The I.T. Revolution across the U.S. States.* IGIER Working Paper No. 226.

DAVERI, F. AND SILVA, O. (2004). *Not only Nokia: What Finland Tells Us about New Economy Growth.* Economic Policy, 19(38): 119–63.

DEDRICK, J., GURBAXANI, V. AND KRAEMER, K. L. (2003). *Information Technology and Economic Performance: a Critical Review of the Empirical Evidence.* ACM Computing Services, 35(1): 1–28.

DEWAN, S. AND KRAEMER, K. L. (2000). *Information Technology and Productivity: Evidence from Country-Level Data.* Management Science, 46(4): 548–76.

DUTTA, D. AND OTSUKA, K. (2004). *Knowledge Spillover from Information and Communication Technology: A Comparative Study of Australia, Japan, South Korea and Taiwan.* University of Sydney, School of Economics and Political Science, Working Paper ECON2004-8.

EC (2005). *i2010 - A European Information Society for Growth and Employment.*

FUSS, M. AND WAVERMAN, L. (2005). *The Networked Computer: the Contribution of Computing & Telecommunications to Economic Growth Productivity OR Why is there no New Economy in Old Europe. A Production Function Approach.* London Business School, Digital Transformations Working Paper No. 1.

GHOLAMI, R., LEE, T.-Y. AND HESHMATI, A. (2005). *The Causal Relationship between ICT and FDI.* UNU-WIDER Research Paper 2005/26.

GORDON, R. J. (2000). *Does the New Economy Measure up to the Great Inventions of the Past?* Journal of Economic Perspectives, 14(4): 49–77.

— (2004). *Five Puzzles in the Behaviour of Productivity, Investment, and Innovation.* NBER Working Papers n. 10660.

Guellec, D. and van Pottelsberghe, B. (2004). *From R&D to Productivity Growth: Do the Institutional Settings and the Source of Funds of R&D matter?* Oxford Bulletin of Economics and Statistics, 66(3): 353–78.

Im, K. Y., Pesaran, M. H. and Shin, Y. (2003). *Testing for Unit Roots in Heterogeneous Panels.* Journal of Econometrics, 115: 53–74.

Jalava, J., Poiola, M., Ripatti., A. and Vilmunen, J. (2005). *Biased Technical Change and Capital-Labour Substitution in Finland, 1902-2003.* HECER Discussion Paper No. 56, Helsinki.

Jorgenson, D. and Vu, K. (2005). *Information Technology and the World Economy.* Scandinavian Journal of Economics, 107(4): 631–50.

Jorgenson, D. W., Ho, M. S. and Stiroh, K. J. (2003). *Lessons from the U.S. Growth Resurgence.* Journal of Policy Modeling, 25(5): 453–70.

Jorgenson, D. W. and Stiroh, K. J. (2000). *Raising the Speed Limit: U.S. Economic Growth in the Information Age.* Brookings Papers on Economic Activity, 1: 125–211.

Kumbhakar, S. and Wang, H.-J. (2005). *Estimation of Growth Convergence using a Stochastic Production Frontier Approach.* Economics Letters, 88(3): 300–5.

Larsson, R., Lyhagen, J. and Lothgren, M. (2001). *Likelihood-Based Cointegration Tests in Heterogeneous Panels.* Econometrics Journal, 4: 109–42.

Lee, T. S.-Y. and Guo, X. (2004). *Information and Communications Technology (ICT) and Spillover: a Panel Analysis.* Paper presented at Econometric Society Far Eastern Meeting 2004.

Levin, A. and Lin, C. (1993). *Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties.* Unpublished manuscript, University of California, San Diego.

Lindbeck, A. (2000). *Swedish Economic Growth in an International Perspective.* Swedish Economic Policy Review, 71(1): 7–37.

MARK, N. AND SUL, D. (2003). *Cointegration Vector Estimation by Panel DOLS and Long-Run Money Demand.* Oxford Bulletin of Economics and Statistics, 65(5): 655–80.

MARK, N. C., OGAKI, M. AND SUL, D. (2005). *Dynamic Seemingly Unrelated Cointegrating Regression.* Review of Economic Studies, 72: 797–820.

MCGUCKIN, R. AND VAN ARK, B. (2005). *Productivity and Participation: an International Comparison.* Groningen Growth and Development Centre Research Memorandum GD-78.

MCGUCKIN, R. H., SPIEGELMAN, M. AND VAN ARK, B. (2004). *The U.S. Advantage in Retail and Wholesale Trade Performance: How Can Europe Catch up?* Paper presented at EPKE Final Conference on Information Technology, Productivity and Growth, 28-29 October 2004, London.

MCMAHON, M., STERNE, G. AND THOMPSON, J. (2005). *The role of ICT in the Global Investment Cycle.* Bank of England Working Paper No. 257.

MOON, H. AND PERRON, B. (2005). *Efficient Estimation of the SUR Cointegration Regression Model and Testing for Purchasing Power Parity.* Econometric Reviews, forthcoming.

NORDHAUS, W. (2005). *The Sources of the Productivity Rebound and and the Manufacturing Employment Puzzle.* NBER Working Paper No. 11354.

OECD (2005). *Annual National Accounts - volume I - Exchange Rate, PPPS, and Population.* No. 01. Paris.

OLINER, S. D. AND SICHEL, D. E. (2000). *The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?* Journal of Economic Perspectives, 14(4): 3–22.

O'MAHONY, M. AND VAN ARK, B. (2003). *EU Productivity and Competitiveness: an Industry Perspective. Can Europe Resume the Catching-up Process?* DG Enterprise, European Commission.

O'MAHONY, M. AND VECCHI, M. (2005). *Quantifying the Impact of ICT Capital on Output Orowth: A Heterogenous Dynamic Panel Approach.* Economica, 72: 615–33.

OULTON, N. AND SRINIVASAN, S. (2005). *Productivity Growth and the Role of ICT in the United Kingdom: an Industry View, 1970-2000.* CEP Discussion Paper No. 681.

PARK, J. AND SHIN, S. (2004). *The Maturity and Externality Effects of Information Technology Investments on National Productivity Growth.* University of Rhode IIsland, College of Business Admnistration, Working Paper Series 2004-05 No. 2.

PEDRONI, P. (1999). *Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors.* Oxford Bulletin of Economics and Statistics, 61: 653–70.

— (2000). *Fully Modified OLS for Heterogenous Cointegrated Panels.* In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics* (edited by BALTAGI, B.), ch. 15, pp. 93–130. JAI Press., Amsterdam.

— (2004). *Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis.* Econometric Theory, 20: 507–625.

PESARAN, H., SHIN, Y. AND SMITH, R. (1999). *Pooled Mean Group Estimation of Dynamic Heterogeneous Panels.* Journal of the American Statistical Association, 94: 621–4.

PESARAN, M. H. (2005). *A Simple Panel Unit Test in the Presence of Cross Section Dependence.* Cambridge University, DAE Working Paper No. 0346, revised January 2005.

PILAT, D. (2004). *The Economic Impacts of ICT - What Have We Learned Thus Far?* Paper presented at the 4th ZEW Conference on the Economics of Information and Communication Technologies, Mannheim, July 2-3, 2004.

PILAT, D. AND WOLFL, A. (2004). *ICT Production and ICT Use: What Role in Aggregate Productivity Growth?* In *The Economic Impact of ICT Measurement, Evidence and Implications*, ch. 5, pp. 85–104. OECD, Paris.

RINCON, A. AND VECCHI, M. (2004). *The Dynamic Impact of ICT Spillovers on Companies' Productivity Performance.* NIESR Discussion Paper No. 245.

SAIKKONEN, P. (1999). *Asymptotically Efficient Estimation of Cointegration Regressions.* Econometric Theory, 7: 1–21.

SARNO, L. AND TAYLOR, M. (2002). *Purchasing Power Parity and the Real Exchange Rate.* International Monetary Fund, Staff Papers, 49(1): 65–105.

SCHREYER, P. (2002). *Computer Price Index and International Growth and Productivity Comparisons.* Review of Income and Wealth, 48(1): 15–31.

STIROH, K. J. (2002a). *Are ICT Spillovers Driving the New Economy?* Review of Income and Wealth, 48(1): 33–57.

— (2002b). *Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say?* American Economic Review, 92(5): 1559–76.

TIMMER, M., YPMA, G. AND VAN ARK, B. (2003). *IT in the European Union: Driving Productivity Divergence?* Groningen Growth and Development Centre Research Memorandum GD-67.

VAN ARK, B., MELKA, J., MULDER, N., TIMMER, M. AND YPMA, G. (2002). *ICT Investments and Growth Accounts for the European Union 1980-2000.* Groningen Growth and Development Centre Research Memorandum GD-56. (revised March 2003).

ZELLNER, A. (1962). *An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias.* Journal of the American Statistical Association, 57: 348–68.

Table 1: **Overall growth contribution of ICT capital in Europe and the US** (1980-2004)
annual average growth rates (% points)

Table 1.A — OVERALL SAMPLE

	GDP growth	TFP	Hours worked	Non-ICT capital	ICT capital
		contributions:			
1980-2004	2.7	0.9	0.5	0.6	0.6
1980-85	2.5	0.9	0.3	0.6	0.8
1985-90	3.3	0.9	1.1	0.8	0.5
1990-95	2.0	0.9	0.1	0.6	0.4
1995-2001	3.1	0.8	0.8	0.6	0.8
2001-04	2.2	1.4	0.0	0.4	0.3

Decomposition of factor inputs contributions (1980-2004)

	Hours worked	Non-ICT capital	ICT capital
Contri-bution:	0.51	0.61	0.62
Income share	0.69	0.26	0.04
Growth rate	0.74	2.33	14.0

Table 1.B

	EUROPEAN UNION-15					UNITED STATES				
		contributions:					*contributions:*			
	GDP growth	**TFP**	**Hours worked**	**Non-ICT capital**	**ICT capital**	**GDP**	**TFP**	**Hours worked**	**Non-ICT capital**	**ICT capital**
1980-2004	2.1	1.0	0.1	0.7	0.4	3.2	0.9	0.9	0.6	0.8
1980-85	1.5	1.2	-0.7	0.6	0.4	3.2	0.6	1.1	0.6	1.0
1985-90	3.2	1.1	0.8	0.9	0.5	3.3	0.6	1.4	0.6	0.6
1990-95	1.6	1.2	-0.6	0.7	0.3	2.5	0.5	0.9	0.4	0.6
1995-2001	2.5	0.7	0.6	0.6	0.6	3.6	0.9	1.0	0.6	1.0
2001-04	1.4	0.5	0.1	0.5	0.2	3.0	2.3	-0.1	0.4	0.5

Source: Own elaboration on data from Timmer et. al (2003), updated June 2005. Contributions are share-weighted growth rates.

Table 2: **Sources of national GDP growth** (1980-2004)
annual average growth rates (% points)

	GDP	TFP	Hours worked	Non-ICT capital	ICT capital
Austria	2.09	0.73	0.14	0.79	0.43
Belgium	1.94	0.99	-0.05	0.31	0.70
Denmark	1.92	0.81	-0.18	0.64	0.66
Finland	2.35	1.82	-0.28	0.34	0.48
France	1.98	0.95	-0.21	0.95	0.29
Germany	1.75	1.39	-0.43	0.37	0.42
Greece	1.91	0.49	0.58	0.62	0.22
Ireland	5.33	3.01	0.62	1.38	0.32
Italy	1.71	0.42	0.15	0.76	0.38
Luxembourg	4.69	1.27	1.41	1.26	0.74
Netherlands	2.22	0.73	0.58	0.47	0.45
Portugal	2.49	1.14	0.27	0.76	0.33
Spain	2.81	0.84	0.67	0.97	0.33
Sweden	2.19	0.96	0.14	0.49	0.60
United Kingdom	2.57	1.29	0.14	0.61	0.52
EU-15	2.12	0.96	0.06	0.66	0.44
United States	3.16	0.89	0.93	0.55	0.78

Source: Own elaboration on data from Timmer et. al (2003), updated June 2005. Contributions are share-weighted growth rates.

Table 3: **Panel unit roots and cointegration analysis**

	UNIT ROOTS *Levels* (incl. time dummies and trend)			
	GDP	Hours worked	Non-ICT capital	ICT capital
Levin-Lin (1993)	-0.33	-0.11	-1.57	-3.09**
IPS (2003)	-1.06	-0.79	-1.20	-1.45
	First differences (incl. time dummies)			
Levin-Lin (1993)	-2.45**	-5.45**	-1.46	-2.18**
IPS (2003)	-4.29**	-7.20**	-3.98**	-5.66**
	PANEL COINTEGRATION *Levels* (incl. time dummies and trend)			
	Panel ADF		Group ADF	
Pedroni (1999)	-2.20**		-2.47**	

Notes: All statistics are distributed as standard normal, diverging to an infinite negative under the alternative hypothesis; a step down procedure is employed to select ADF lags for each equation. Variables are cross-sectionally de-meaned; time trend is admitted only in log-levels specifications.
Unit roots tests assume the null hypothesis of non-stationary; under the alternative hypothesis, Levin and Lin's test admits a common coefficient for the lagged dependent variable while IPS a non-homogenous one for at least a positive fraction of individuals.
Pedroni's tests assume the null hypothesis of no cointegration; the alternative hypothesis is of a homogeneous cointegration vector for panel ADF t-statistics and of heterogeneity for group ADF t-statistics; tests are not weighted for the long-run variance.
** significant at a 5% level.

Table 4: **Long-run estimation of aggregate production function** (1980-2004), *levels*

	PDOLS					
	US and EU-15 (full sample)				**US and EU-13**[a]	**EU-13**[a]
	I	II	III	IV	V	VI
Hours worked	0.522***	0.505***	0.321***	0.293***	0.261***	0.262***
	(0.079)	(0.079)	(0.031)	(0.05)	(0.032)	(0.033)
Non-ICT capital	0.235**	0.236***	0.021	0.028	0.188***	0.184***
	(0.096)	(0.087)	(0.043)	(0.043)	(0.058)	(0.058)
ICT capital	0.138***	0.138***	0.092***	0.094***	0.084***	0.085***
	(0.018)	(0.016)	(0.010)	(0.011)	(0.011)	(0.011)
Intercept	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Time dummies	*no*	*yes*	*no*	*yes*	*yes*	*yes*
Trend	*no*	*no*	*yes*	*yes*	*yes*	*yes*
Obs. (N*T)	400	400	400	400	350	350
Adj. R-squared	0.93	0.92	0.55	0.53	0.70	0.69

Notes: GDP is the dependent variable; all variables are in log-levels. Standard errors are parametrically corrected with pre-weighting method. The maximum lag in the step-down procedure selecting the number of leads (and lags) is fixed to 1.
[a] EU-13 excludes Finland and Sweden.
*, **, *** significant at 10, 5 and 1% levels.

Table 5: **Short-run estimation of aggregate production function**, *first differences*

	US and EU-15					US and EU-13[a]		EU-13[a]	
	1980-2004			1992-2004		1980-2004		1980-2004	
	OLS	**IV-2SLS** (4 lags)	**diff-GMM**	**IV-2SLS** (2 lags)	**diff-GMM**	**IV-2SLS** (4 lags)	**diff-GMM**	**IV-2SLS** (4 lags)	**diff-GMM**
	I	II	III	IV	V	VI	VII	VIII	IX
Hours worked	0.331***	0.378***	0.327***	0.148***	0.201***	0.268***	0.231***	0.260***	0.196***
	(0.066)	(0.073)	(0.087)	(0.055)	(0.051)	(0.063)	(0.066)	(0.066)	(0.060)
Non-ICT capital	0.116	0.073	0.108	0.285**	0.256	0.194*	0.205*	0.202*	0.215
	(0.087)	(0.090)	(0.097)	(0.127)	(0.158)	(0.112)	(0.114)	(0.114)	(0.121)
ICT capital	0.078***	0.085***	0.076***	0.082**	0.068**	0.087***	0.079***	0.084**	0.075***
	(0.030)	(0.031)	(0.018)	(0.033)	(0.032)	(0.033)	(0.018)	(0.034)	(0.019)
Obs. (N*T)	400	320	368	160	176	160	322	160	299
Adj. R-squared	0.62	0.65	0.61	0.83	0.68	0.67	0.65	0.67	0.63
F-test of no significant fixed effects (P-value)	3.15 (0.01)	4.72 (0.00)		6.59 (0.00)		5.48 (0.00)		5.30 (0.00)	
Hansen test p-value			1.00		1.00		1.00		1.00
AR(1) p-value	0.00	0.00	0.02	0.03	0.01	0.00	0.03	0.01	0.04
AR(2) p-value	0.18	0.17	0.33	0.54	0.27	0.41	0.46	0.30	0.67

Notes: GDP is the dependent variable; all variables are in first differences (annual growth rates). Standard errors robust to heteroskedasticity are in brackets. Time dummies and country-specific intercepts are included but not reported.
F-test checks the null hypothesis of insignificant fixed effects. The P-value of Hansen test for instruments' over-identification and Arellano-Bond tests of no serial correlation in residuals is reported on the bottom.
[a] EU-13 excludes Finland and Sweden.
*, **, *** significant at 10, 5 and 1% levels.

Table 6: **Long-run estimation of production function by groups: a comparison between PDOLS and DSUR** (1980-2004)

	PDOLS				DSUR			
	Big Countries	EU Big Countries	Medium Countries	Small Countries	Big Countries	EU Big Countries	Medium Countries	Small Countries
	I	II	III	IV	V	VI	VII	VIII
Hours worked	0.624***	0.670***	0.249***	0.209***	0.518***	0.710***	0.165***	0.653***
	(0.137)	(0.149)	(0.045)	(0.113)	(0.041)	(0.060)	(0.012)	(0.012)
Non-ICT capital	0.150	0.072	-0.280***	0.047***	0.074*	0.186***	0.028	-0.076***
	(0.107)	(0.109)	(0.093)	(0.045)	(0.042)	(0.054)	(0.033)	(0.029)
ICT capital	0.030	-0.007	0.008	0.094***	0.124***	0.041	0.054***	0.171***
	(0.053)	(0.059)	(0.033)	(0.017)	(0.014)	(0.029)	(0.006)	(0.007)
Obs. (N*T)	125	100	125	150	125	100	125	150
Adj. R-squared	0.88	0.82	0.24	0.46	0.81	0.95	0.33	0.76
Wald test of parameters' homogeneity among groups								
	PDOLS				DSUR			
	Big	EU Big	Medium	Small	Big	EU Big	Medium	Small
Big	0				0			
EU big		0				0		
Medium	30.1***	17.5***	0		8.43**	28.2***	0	
Small	17.4***	10.0**	10.0**	0	0.06	13.3**	15.3***	0

Notes: GDP is the dependent variable. All estimates include country-specific intercepts (fixed effects) and time trend; PDOLS also considers time dummies. Standard errors are parametrically corrected with pre-weighting method. The maximum lag in the step-down procedure selecting the number of leads (and lags) is fixed to 1. The Wald test checks the null hypothesis of no significant difference in cointegration vector among groups.
Big Countries: US, Germany, France, UK and Italy; *Intermediate Countries*: Spain, Netherlands, Belgium, Sweden and Austria; *Small Countries*: Greece, Portugal, Denmark, Finland, Ireland and Luxembourg.
*, **, *** significant at 10, 5 and 1% levels.

Table 7: **Long-run vs short-run estimation: a comparison between DSUR and SUR** (1980-2004)

	Long-run estimation (*levels*)			Short-run estimation (*first differences*)		
	DSUR			SUR		
	Big Countries	EU Big Countries	Small & Medium Countries[a]	Big Countries	EU Big Countries	Small & Medium Countries[a]
	I	II	III	IV	V	VI
Hours worked	0.518***	0.710***	0.121***	0.354***	0.229***	0.214***
	(0.041)	(0.060)	(0.001)	(0.096)	(0.111)	(0.054)
Non-ICT capital	0.074*	0.186***	0.193***	0.046	-0.046	0.206**
	(0.042)	(0.054)	(0.002)	(0.213)	(0.256)	(0.093)
ICT capital	0.124***	0.041	0.117***	0.038	0.003	0.082***
	(0.014)	(0.029)	(0.001)	(0.060)	(0.066)	(0.022)
Obs. (N*T)	125	100	225	120	96	216
Adj.R -squared	0.81	0.95	0.63	0.60	0.59	0.67
F-test of no significant fixed effects (P-value)				2.24 (0.07)	1.62 (0.19)	8.37 (0.00)

Notes: GDP is the dependent variable. All estimates include country-specific intercepts (fixed effects); time trend is comprised in log-levels specification but not in first differences; the reverse holds for time dummies. In DSUR, standard errors are parametrically corrected with pre-weighting method. The maximum lag in the step-down procedure selecting the number of leads (and lags) is fixed to 1.

[a] excludes Finland and Sweden.

*, **, *** significant at 10, 5 and 1% levels.

Table 8: **Parameters' sensitivity to the inclusion of labour quality**, EU-4 vs US (DSUR)

	1980-2004		1980-2000	
	EU-4 and US *(hours worked)*	**EU-4** *(hours worked)*	**EU-4 and US** *(hours worked)*	**EU-4 and US** *(quality-adjusted labour)*
	I	II	III	IV
Labour	0.308***	0.532***	0.020	0.121***
	(0.035)	(0.033)	(0.022)	(0.008)
Non-ICT capital	0.111***	0.086	0.311***	0.233***
	(0.040)	(0.059)	(0.038)	(0.030)
ICT capital	0.140***	0.092***	0.186***	0.183***
	(0.011)	(0.019)	(0.009)	(0.003)
Obs. (N*T)	125	100	105	105
Adj. R-squared	0.70	0.80	0.72	0.76

Notes: GDP is the dependent variable. All estimates include country-specific intercepts (fixed effects) and time trend. Standard errors (in brackets) are parametrically corrected with pre-weighting method. The maximum lag in the step-down procedure selecting the number of leads (and lags) is fixed to 1.
EU-4: France, Germany, Netherlands and UK.
*, **, *** significant at 10, 5 and 1% levels.

APPENDIX

Table A.1: **Unit roots tests for PPP index of EU-15 countries** (1970-2002), annual values

	IPS (Im, Pesaran and Shin, 2003)		**CIPS** (Pesaran, 2005)	
	test statistics	cv 5%	test statistics	cv 5%
Levels (with trend)	-1.50	-2.85	-2.14	-2.76
First differences	-3.49**	-1.85	-2.36**	-2.25

Sources: OECD (2004); United States are the numeraire country.
Time trend is included in log-levels specification but not in first-differences; values are not standardised. IPS test utilises cross-sectionally de-meaned data.
** significant at a 5% level.

Table A.2: **Dynamic OLS estimation for single equation**, (1980-2004) (Saikkonen, 1991)

	Hours worked		Non-ICT capital		ICT capital	
	Coef.	*SE*	*Coef.*	*SE*	*Coef.*	*SE*
Austria	-0.269	0.328	1.224**	0.561	-0.039	0.057
Belgium	-1.167***	0.237	-0.919***	0.156	-0.140***	0.026
Denmark	-0.607***	0.099	-0.123	0.103	-0.176***	0.025
Finland	0.024	0.473	-0.856***	0.217	0.680***	0.160
France	-0.849***	0.272	0.243***	0.055	-0.287***	0.075
Germany	-0.250	0.362	1.010***	0.052	0.073***	0.030
Greece	-0.647***	0.162	0.721***	0.128	0.060	0.055
Ireland	1.221***	0.153	-0.111	0.200	0.021	0.016
Italy	-0.727***	0.062	0.904***	0.127	0.067***	0.021
Luxembourg	-1.275***	0.307	2.020***	0.351	0.638***	0.090
Netherlands	0.627***	0.022	2.154***	0.094	0.166***	0.012
Portugal	1.154***	0.131	0.128***	0.038	-0.160***	0.037
Spain	0.391***	0.023	0.420***	0.064	0.192***	0.018
Sweden	1.906***	0.377	-0.547***	0.042	-0.094	0.076
United Kingdom	1.219***	0.083	0.382***	0.077	-0.014	0.015
United States	-0.378***	0.062	2.529***	0.186	0.005	0.017

Notes: GDP is dependent variable; all variables are expressed in cross-sectionally demeaned log-levels. All equations include country-specific intercepts (fixed effects) and time trend. Standard errors in brackets are parametrically corrected with pre-weighting method. The maximum lag in the step-down procedure selecting the number of leads (and lags) is fixed to 1.
*, **, *** significant at 10, 5 and 1% levels.

Table A.3.1: **PDOLS and DSUR estimation by groups excluding Finland and Sweden** (1980-2004)

	PDOLS				DSUR			
	Big Countries	EU Big Countries	Medium Countries[a]	Small Countries[b]	Big Countries	EU Big Countries	Medium Countries[a]	Small Countries[b]
	I	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*
Hours worked	0.624***	0.670***	0.217***	0.080	0.518***	0.710***	0.199***	0.147***
	(0.137)	(0.149)	(0.036)	(0.100)	(0.041)	(0.060)	(0.016)	(0.069)
Non-ICT capital	0.150	0.072	0.426***	0.242***	0.074*	0.186***	0.068	0.169***
	(0.107)	(0.109)	(0.131)	(0.070)	(0.042)	(0.054)	(0.053)	(0.053)
ICT capital	0.030	-0.007	0.063***	0.094***	0.124***	0.041	0.078***	0.180***
	(0.053)	(0.059)	(0.029)	(0.016)	(0.014)	(0.029)	(0.010)	(0.015)
Obs. (N*T)	125	100	100	125	125	100	100	125
Adj. R-squared	0.88	0.82	0.88	0.61	0.81	0.95	0.46	0.64

Wald test of parameters' homogeneity among groups

	PDOLS				DSUR			
	Big	EU Big	Medium	Small	Big	EU Big	Medium	Small
Big	0				0			
EU big		0				0		
Medium	0.04	0.99	0		5.34	21.5***	0	
Small	10.8**	5.22	15.9***	0	2.81	22.4***	2.21	0

Notes: GDP is dependent variable; all variables are expressed in log-levels. All estimates include country-specific intercepts (fixed effects) and time trend; PDOLS also considers time dummies. Standard errors in brackets are parametrically corrected with pre-weighting method. The maximum lag in the step-down procedure selecting the number of leads (and lags) is fixed to 1. The Wald test checks the null hypothesis of no significant difference in cointegration vector among groups.

[a] excludes Sweden; [b] excludes Finland.

*, **, *** significant at 10, 5 and 1% levels.

Table A.3.2: **PDOLS and DSUR estimation by groups excluding Finland and Sweden** (1980-2004)

	PDOLS			DSUR		
	Big & Medium Countries [a]	EU Big & Medium[a] Countries	Small Countries[b]	Big Countries	EU Big Countries	Small & Medium Countries[a,b]
	I	II	III	IV	V	VI
Hours worked	0.277***	0.276	0.080	0.518***	0.710***	0.121***
	(0.041)	(0.044)	(0.100)	(0.041)	(0.060)	(0.001)
Non-ICT capital	0.149	0.107	0.242***	0.074*	0.186***	0.193***
	(0.095)	(0.101)	(0.070)	(0.042)	(0.054)	(0.002)
ICT capital	0.052*	0.044	0.094***	0.124***	0.041	0.117***
	(0.029)	(0.031)	(0.016)	(0.014)	(0.029)	(0.001)
Obs. (N*T)	225	200	125	125	100	225
Adj. R-squared	0.64	0.58	0.60	0.81	0.95	0.63

Wald test of parameters' homogeneity among groups

	PDOLS			DSUR			
	Big & Medium[a]	EU Big & Medium[a]	Small[b]		Big	EU Big	Medium & Small [a,b]
Big & Medium	0			Big	0		
EU Big & Medium		0		EU Big		0	
Small	0.09	0.15	0	Medium & Small	6.16	38.6***	0

Notes: GDP is dependent variable; all variables are expressed in log-levels. All estimates include country-specific intercepts (fixed effects) and time trend; PDOLS also considers time dummies. Standard errors in brackets are parametrically corrected with pre-weighting method. The maximum lag in the step-down procedure selecting the number of leads (and lags) is fixed to 1. The Wald test checks the null hypothesis of no significant difference in cointegration vector among groups.

[a] excludes Sweden; [b] excludes Finland.

*, **, *** significant at 10, 5 and 1% levels.

Piracy on the Silver Screen

January 18, 2005

Preliminary Draft – Do not cite without the authors' permission

Rafael Rob
University of Pennsylvania

Joel Waldfogel
The Wharton School
University of Pennsylvania and NBER

Abstract

New information technology has reduced marginal production and distribution costs of information goods to negligible levels and promises to revolutionize many industries. Unpaid copies of digital products can be as good as paid first-generation copies, and their availability can undermine the ability of sellers to cover first-copy costs. As a result, unpaid distribution has emerged as a major issue facing the music and movie industries in the past few years. Using survey data on movie consumption by about 500 University of Pennsylvania college students, we ask whether unpaid consumption of movies displaces paid consumption. Employing a variety of cross-sectional and longitudinal empirical approaches, we find large and statistically significant evidence of displacement. In what we view as the most appropriate empirical specifications, we find that unpaid first consumption reduces paid consumption by about 1 unit. Unpaid second consumption has a smaller effect, about 0.20 units. These estimates indicate that unpaid consumption, which makes up 5.2 percent of movie viewing in our sample, reduced paid consumption in our sample by 3.5 percent.

Waldfogel thanks the Mack Center and the Wharton Electronic Business Initiative for financial support. Sarah Waldfogel provided excellent assistance with data input.

New information technology, which has reduced marginal production and distribution costs of information goods to negligible levels, promises to revolutionize many industries. Zero marginal costs can allow higher profits and lower prices. Moreover, costless distribution allows sellers to use innovative marketing strategies such as free samples with partial functionality (Shapiro and Varian, 1999). But once the genie of negligible marginal costs is out of the bottle, the same technological changes that empower sellers can also enable unpaid distribution and redistribution. With digital goods, unpaid copies can be as good as paid first-generation goods, and the willingness and ability of consumers to share copies can undermine the ability of sellers to cover first-copy costs. Unless unpaid distribution stimulates demand for complementary goods which can continue to generate revenue, the information industries may ultimately suffer.

Over the past five years, unpaid distribution (variously known as "file-sharing" and "piracy") has emerged as a major issue facing producers of information products such as music and movies. After revenue growth for many years, US music industry revenue has declined from 2000 to 2003 and was flat in 2004, and many in the industry blame file-sharing [1]. The motion picture industry is similarly concerned about piracy, although overall Hollywood revenue continues to rise even if theatrical box office receipts have been flat over the past few years.[2]

While it is perhaps surprising at first blush, theory gives no unambiguous guidance on the effect of unpaid consumption on sales. On one side, unpaid copies are

[1] See the RIAA's "AntiPiracy" at http://www.riaa.com/issues/piracy/default.asp as well as the RIAA's "2004 Yearend Statistics" at http://www.riaa.com/news/newsletter/pdf/2004yearEndStats.pdf (both accessed December 15, 2005).
[2] The Motion Picture Association of America estimates $3.5 billion in losses to various forms of piracy in 2002. See the MPAA's "2003 Piracy Fact Sheet Overview" at http://www.mpaa.org/PiracyFactSheets/PiracyFactSheetOverview.pdf (accessed December 15, 2005).

close substitutes for paid copies, particularly in music, suggesting that unpaid consumption would simply displace paid consumption. While MP3 files contain less information than, say, CDs, many listeners find the compressed files to contain nearly the same information. And possession of an unpaid music file allows its user all the activities available to the owner of a purchased CD.

Movies are different in a few respects that suggest less displacement by unpaid use. First, while an unpaid copy of a movie, on CD or DVD, contains the same information as a legal DVD and allows a very similar experience to home video ownership, home video is an imperfect substitute for theatrical viewing. Second, as economic theorists have emphasized, information sharing can stimulate paid consumption if the sharing remains small-scale[3]. The logic is simple: two isolated persons valuing a movie DVD at $11 each will not purchase it for its $20 price. If they plan to share, then the sum of their valuations exceeds the price, and they will purchase it together. The large size of video files has kept peer-to-peer sharing at small scale, largely through traditional "sneakernet" file transmission.[4] The possibility for file-sharing to stimulate paid demand therefore seems more real in movies than in music.

To date there have been numerous academic studies of the effect of file-sharing on revenue in the music industry.[5] Most, but not all, studies find that file sharing depresses sales, although much less than one for one. That is, much of the music consumption occurring through file sharing would not have occurred in the absence of

---

[3] See Bakos, Brynjolfsson, and Lichtman (1999) and Varian (2000).
[4] A user dons his sneakers and carries a DVD from his friend's house to his DVD-burning computer. See http://en.wikipedia.org/wiki/Sneakernet, accessed December 15, 2005.
[5] See, for example Blackburn (2004), Oberholzer-Gee and Strumpf (2004), Zentner (2004), Rob and Waldfogel (2004), Bounie, Bourreau and Waelbrock (2005), Hong (2005), Hui and Png (2003), and Leibowitz (2003).

file sharing. As a result, file sharing tends to increase social welfare by more than the revenue loss to sellers (ignoring any effects on the supply of music). This study examines the displacement question in the context of the motion picture industry. Using survey data on movie consumption by about 500 University of Pennsylvania college students, we ask whether unpaid consumption (of burned or downloaded copies) stimulates or depresses demand for various forms of paid consumption, including theatrical, television, rental, and DVD ownership. The surveyed population is far from representative of the US population – respondents are richer and are more likely to have broadband Internet connections as well as access to DVD burning hardware. While results are therefore not representative of the entire US, the effects documented for this population are suggestive of effects that might operate more broadly as technologies continue to diffuse.

Using a variety of cross-sectional and longitudinal empirical approaches, we find large and statistically significant evidence of displacement. In what we view as the most appropriate empirical specifications, we find that unpaid first consumption reduces paid consumption by about 1 unit. Unpaid second consumption has a smaller effect, about 0.20 units. These estimates indicate that unpaid consumption, which makes up 5.2 percent of movie viewing in our sample, reduced paid consumption in our sample by 3.5 percent.

I. Theoretical Background

Prior to file sharing, each consumer has some willingness to pay for each movie. The distribution of these consumer valuations make up the demand curve for the movie.

The advent of file sharing has two kinds of effects. First, even holding constant the distribution of valuations, with file sharing, all kinds of consumers – with valuations above or below the market price – can now get access without paying. The sales displacement and welfare effects of this – worked out in section I.1 below – depend on whether the unpaid consumption would have manifested itself as paid consumption in the absence of file sharing.

1. Supply and Demand

Supply and demand provides a useful framework for thinking about the sales displacement and welfare consequences of unpaid consumption. For this discussion suppose that the marginal cost of production and distribution is zero.

First, consider the market prior to file sharing, in figure 1a. Sellers sell B_0 units at a price of P. The area under the demand curve is divided into three regions, consumer surplus (the consumer valuations in excess of the price they pay), revenue ($P^* B_0$), and deadweight loss (foregone mutually beneficial transactions for units with consumer valuation above the marginal cost but below the price, so no transaction occurs). When unpaid consumption is feasible, its effect depends on which units are obtained without payment, whether they are those with valuation above or below P.

Start by considering an extreme case in which only units with valuation below P are obtained without payment. This results in the situation in figure 1b. Sellers continue to sell B_0 units at a price of P. Another S units are obtained without payment. Hence, in this situation, unpaid consumption does not displace paid consumption, and $\frac{\partial B}{\partial S} = 0$.

Here, the deadweight loss prevailing prior to unpaid consumption becomes consumer surplus, while sellers remain as well off as before. File sharing makes society unambiguously better off.

Next, consider a case where all of the unpaid consumption is drawn from situation where consumer valuations exceed P. This is depicted in figure 1c. Now ($B_0 - S$) units are transacted at P, while S units are consumed without payment. The sum of producer and consumer surplus is the same as before, as is deadweight loss. But $S*P$ is transferred from producers to consumers. Here, paid consumption falls by the exact amount of unpaid consumption, or $\frac{\partial B}{\partial S} = -1$.

A general lesson from this analysis is that the lower the rate of displacement ($\frac{\partial B}{\partial S}$), the more welfare gain that file sharing brings to society as a whole. As the rate of displacement grows (in absolute value, toward -1), the social gain from file sharing falls to zero. The effect of unpaid consumption becomes a pure transfer from sellers to buyers. The main goal of our paper is to estimate $\frac{\partial B}{\partial S}$.

An important caveat bears mention: it is possible that unpaid consumption displaces sales and changes the nature or volume of movies that get made. To the extent that our estimates have welfare implications, they are implications for welfare, given which movies are being made.

2. File Sharing and Changing Demand

In addition to allowing unpaid access regardless of one's valuation, file sharing can also stimulate demand via two channels. First, small-scale information sharing can stimulate demand if it makes a group of consumers willing, as a group, to purchase a product that no group member would have purchased alone. Two features of music file sharing prevent it from remaining small in scale, in turn making the suggestion that file-sharing would stimulate music sales rather far-fetched. First, file compression makes it fast to download music, particularly over broadband connections. A typical song can be downloaded in a few seconds. Second, with peer-to-peer (P2P), if one person purchases an album, it becomes available to all Internet-connected persons willing to download. As a result, file-sharing is potentially unlimited in scale.

File sharing in movies remains fundamentally different. Even compressed, a typical Hollywood movie is about 4 gigabytess. Using the newest downloading technologies (e.g. BitTorrent), downloading a movie takes between 30 minutes and 8 hours, depending on how many persons have already made the file available online.[6] Similarly, DVD copying requires about 90 minutes using a typical computer setup and DVD burner. DVD copying is generally quicker and results in higher quality copies than downloading. As a result, the dominant form of file sharing in movies is DVD copying, rather than downloading. By its nature, DVD copying is smaller in scale than P2P copying, making it more plausible that file-sharing could stimulate sales. On the other hand, while consumers listen to some albums repeatedly, they may wish to watch movies only once. That is, movies and music may differ in the extent to which initial consumption is a substitute or a complement for subsequent use. Perhaps even more than in music, the effect of file sharing on movie industry revenue is theoretically ambiguous.

---

[6] "Cable and DSL modem users can expect an average of 25 megabytes per hour, sometimes slower if the swarm is small with less than 2 seeders. On a good day with a big swarm, however, you can download a 5MB song within 3 minutes, and a 900MB movie within 60 minutes." http://netforbeginners.about.com/od/peersharing/a/torrenthandbook_4.htm , accessed 2/24/2005

Second, file-sharing may increase consumers' willingness to purchase by allowing him to sample the product prior to purchase. Shapiro and Varian (1999) emphasize that free samples of information products can stimulate subsequent paid consumption. The samples they envision are incomplete versions of the product. Even if unpaid access to movies stimulates interest in movies, it is not clear that consumers would be willing to pay for something that they have already obtained without payment. First, unpaid copies are generally the entire movie rather than an appetite-enhancing free sample. And, second, since unpaid DVD copies are nearly identical, there is no need to pay for an upgrade to a higher-quality experience to see the movie again. Still, it is possible that consumers, once they know they like a movie, would be willing to pay for a legal copy for to continue to watch.

3. Industry Background

In 1980 box office revenue provided the majority of revenue to the US motion picture industry. The box office share has declined steadily since then. The US film industry generated $45 billion in revenue in 2004. The box office share in 2004 was 17 percent, while home video accounted for nearly half.[7]

Between 2000 and 2004, overall industry revenue grew 39 percent in real terms.[8] In contrast to the music industry whose concerns about piracy are motivated in part by sharp recent revenue declines, the movie industry continues to experience robust growth in revenue, driven largely by the growth in DVD sales. Concerns about piracy, while

[7] See http://www.edwardjayepstein.com/mpa2004.htm, accessed December 13, 2005.
[8] Ibid.

substantiated in many instances (see http://www.mpaa.org/anti-piracy/), are not additionally fueled by falling industry revenue.

II. Data

The data for this study are derived from a series of surveys administered to over 500 Penn undergraduates in Spring and late Fall 2005. In each survey, students see a list of the top 50 movies from each of the previous 3 years (2002-2004 in the Spring, 2003-2005 year to date for the Fall). For each movie they are asked whether they saw it, as well as the sequence by which they saw it. Respondents can have seen a movie by any (and all) of 4 paid methods: **theater**, **television**, **rental**, and **purchase**; as well as by either or both of two unpaid means: by viewing a **downloaded** copy, or by viewing a **burned** copy of a legally obtained copy (generally DVD). Thus, for each movie seen, a respondent enters a "1" under the mode they first saw it, a "2" under the mode by which they saw it next (if they saw it more than once), and so on.

In addition to their movie consumption, the survey also asks about family income, respondent race and age, speed of Internet access, and three variables related to their interest in movies: how often they go to movie theaters, how many movies they own, and their level of interest in movies (on a five-point scale). The sample includes nearly 21,000 instances of movie consumption. For analysis we include only respondents who made no mistakes in reporting their sequences.[9] This brings the usable sample down to 470 individuals.

[9] We discard data for individuals who indicate that they saw a movie by multiple means but do not indicate the sequence.

Table 1 describes the sample. The mean age is 18, and the age range is between 17 and 22. Nearly a quarter of the respondents are Asian, 6 percent are black, and 6 percent are Hispanic. Respondents are from families of above-average income: Nearly 30 percent of the sample reports family income in excess of $250,000. Another 37 percent reports family income of $100,000-$250,000. Just 10 percent of respondents report family income below $50,000.

Nearly half (46%) report a "typical" level of interest in movies. Nearly a third report above-average interest in movies. Only a fifth report below average interest in movies. Respondents own a mean (median) of 38 (20) movies, and they go to movie theaters 1.6 (1) time per month. Internet access is ubiquitous among the respondents: 98 percent of respondents report high-speed access in 2005, up steadily from 70 percent in 2002.

For the 150 movies in each survey, respondents report an average of 54.2 instances of paid consumption and 3.0 instances of unpaid consumption. Theatrical consumption is the most common form, accounting for nearly half, followed by rental (20 percent), television (16 percent) and purchase (15 percent). The two forms of unpaid consumption account for 5.2 percent of consumption, the majority (3.1 of 5.2) burned and the remainder (2.1 of 5.2 percent) downloaded.

A particular movie is frequently viewed through more than one medium by a respondent. The bottom part of Table 1 breaks consumption down by medium and by viewing order. While respondents engage in 57.2 episodes of consumption, they see on average only 44.5 separate movies. Virtually all theatrical consumption occurs on first viewing; and theatrical viewing is the way that nearly 60 percent of consumption sequences begin. The other major medium for first viewing is rental (8.3 of 44.5, or 19 percent). The other media account for relatively few first viewings; on the other hand, first viewings via those media account for most of the viewings via those media. For example only 4 percent (1.8 of 44.5) first viewings use unpaid copies. But these viewings represent most of the unpaid consumption. To put this another way, unpaid consumption is rare; but when it occurs, it usually occurs on first viewing.

The vast majority of second viewings are on purchased copies, television, or rental copies. As with first viewings, unpaid consumption accounts for a small fraction of second viewings. At the same time, over a quarter of unpaid consumption occurs on the second viewing.

The first column of Table 2 shows consumption paths that begin with theatrical and rental consumption. The length of the sequences varies with the medium of initial consumption. Of the sequences beginning with rental, less than fifteen percent are viewed through another medium. Of the sequences that start in the theater, far more – over a third – continue to another medium, roughly 10 percent each to purchase, rental, and television with another few percent subsequently consumed via unpaid means.

III. Empirical Strategy

Before beginning the detailed analysis, it is helpful to distinguish two questions of interest. First, to what extent does an instance of unpaid consumption displace paid consumption? Said another way, what is the *rate* of displacement? Second – if there is displacement – how much paid consumption is displaced by unpaid consumption? Or, what is the *amount* of displacement? This is calculated as the volume of unpaid consumption times the rate of displacement.

We have two basic approaches to measuring the effect of unpaid consumption on paid consumption. First, we can make straight use of the cross section, asking whether people who engage in more unpaid consumption engage in less paid consumption. Implementing this approach requires a way of predicting how many movies respondents would have paid to consume in the absence of any unpaid consumption. The main challenge to this approach is that persons who like movies may like them through various media. That is, people who like movies may both like to pay for them and to consume them without paying. If it is not possible to control for all determinants of individuals' interest in movies, then unobserved heterogeneity will tend to induce a positive relationship between unpaid and paid consumption, even if the causal relationship is negative.

The structure of our data allows a second, panel approach. The data include individuals' movie consumption, by medium, yearly 2002(3)-2004(5). We can include individual fixed effects – which control for the individuals' vintage-constant interest in movies – then ask whether persons with higher unpaid consumption in some vintages have higher or lower paid consumption of movies from those vintages.

Our sequencing information allows us to use ask slightly more pointed questions. It makes no sense for a downloaded second viewing to affect first viewing. Indeed – to the extent that a second viewing by unpaid means mechanically follows paid theatrical consumption – naïve cross sectional regressions could perversely show positive relationships when analyzed without regard for sequence. We can more sensibly instead ask whether unpaid first consumption increases or decreases paid first consumption. And whether it increases or decreases second consumption. These are the basic strategies we employ.

1. Cross Sectional Approaches

The cross sectional approach to the displacement question asks whether people with more unpaid viewings have fewer paid viewings. As a first pass at this approach we can simply plot first viewings by paid means against first viewings by unpaid means, in Figure 2. The line is a smoothed description of the relationship, and it is essentially flat. One interpretation of this flat relationship is that unpaid consumption does not displace paid consumption. Yet, persons who like movies more might simply engage in more paid *and* unpaid consumption, which would tend to mask any substitution that operates. Indeed, when we break our respondents down by self-reported interest in movies, the persons with high interest in movies engage in both more paid and more unpaid consumption. To meaningfully measure displacement using a cross sectional approach, we need to find variables that predict the counterfactual paid consumption absent any unpaid consumption. We have three main variables for this task, the respondent's interest in movies on a five-point scale, the size of his or her home video library, and how frequently he or she sees a movie in a theater.

To implement this approach in a regression, define:

B_i^j = the number of paid movies that person i watches for the j^{th} time (i.e. the number of n^{th} viewings for this person),

S_i^j = the number of unpaid movies that person i watches for the j^{th} time;

X_i = characteristics of person i such as the level of interest in movies (fan dummy, size of movie collection, frequency of movie theater attendance);

ε_i^j = unobserved determinants of tendency to make j^{th} viewings for person i.

We can then estimate models of the form:

$B_i^1 = X_i\beta^1 + \alpha^1 S_i^1 + \varepsilon_i^1$, or (1)

$B_i^2 = X_i\beta^2 + \alpha^{22} S_i^2 + \alpha^{21} S_i^1 + \varepsilon_i^2$. (2)

In effect a model like (1) asks whether people with more unpaid first viewings have fewer paid first viewings, after controlling for determinants of paid first consumption. The second model is similar but examines second paid consumption, with possible relationships between both unpaid first and second consumption on paid second consumption. That is, not only can decisions to engage in second unpaid viewings affect paid second viewings. Unpaid first viewings, if they reduce the tendency for a second viewing, can also affect the tendency for a paid second viewing.

Unless the controls are exhaustive, unobserved heterogeneity is a looming concern with models of this sort: persons who watch movies without paying may simply like movies a lot, even beyond the capacity of our control variables to distinguish among persons in the sample. If so, then our estimate of the extent of displacement (the α coefficients) will be biased upward, that is, against a finding of sales displacement.

Table 3 reports results, overall and disaggregated by each paid medium. The first five columns are regressions of the number of first paid viewings in this medium on the number of first unpaid viewings (either downloaded or burned). Controls include the fan dummy, the movie library, and the individual's monthly theater visits. People with larger libraries, more frequent theater patronage, and higher levels of movie fandom have more paid first viewings. The coefficient on unpaid first viewings is negative but insignificantly different from zero. The second column examines theatrical first viewing. Here the coefficient of interest is -0.12, below its standard error of 0.16 and therefore not statistically significant.

Columns (6)-(10) are regressions of second viewings on controls and the numbers of unpaid first and second viewings. Persons with one more unpaid first viewing have 0.34 fewer second paid viewings. The coefficients for television, rental, and purchase are all negative, and the rental coefficient is large and significant (-0.26).

Column (6) also indicates that persons with more unpaid second viewings have more paid second viewings as well (0.42 and significant). Most of this reflects additional rental: persons who watch more unpaid second viewings engage in more rented second viewings.[10]

If one takes these models literally – and causally – then the effect of each additional first unpaid viewing is the sum of 0.01 and -0.34, or a reduction of 0.33 in paid viewings. Each second-time unpaid viewing, on the other hand, raises paid viewing by 0.42. There are many reasons to doubt a causal interpretation, however. First, as already emphasized, there is unobserved heterogeneity. People who like movies will tend to engage in more paid and unpaid consumption. This can look like a stimulating effect of unpaid consumption on paid consumption even if it is not.

[10] We also estimated models that disaggregate unpaid consumption into distinct downloaded and burned components. The distinct categories have statistically indistinguishable effects, so here and below we report only models with overall unpaid consumption as an explanatory variable.

More convincing estimates require ways of circumventing the unobserved heterogeneity. A natural strategy is instrumental variables. Our survey includes information on broadband and DVD-burner access which seem like promising instruments for unpaid consumption (see Rob and Waldfogel, 2004; Zentner, 2004). In our data these technological access variables bear no significant relationships with unpaid consumption. These variables fail as instruments for three reasons. First, broadband access is virtually ubiquitous, so there is little variation. Second, even if there were more variation in broadband access, most unpaid consumption copies are obtained via copying rather than downloading, which reduces the usefulness of access speed in predicting unpaid consumption. And while access to DVD-burning hardware would seem a promising predictor, whether one has access to this technology may be less important than whether one's friends have such access. We are left without a promising IV approach.

The sequential nature of consumption allows a different strategy. We can restrict attention to second viewings for the subset of movies with a first paid consumption, then ask how much paid second consumption is displaced by an unpaid second consumption. On average a third of movies first viewed at the theater are subsequently viewed by another means (see Table 2). By restricting attention to the movies first viewed, say, at the theater, we grow closer to holding constant individuals' interest in the movies in question. We can then ask how the number of paid second viewings, per movie first seen in the theater, relates to the number of unpaid second viewings, again relative to the number first seen in the theater.

Table 2 offers a simple implementation of this approach. The table divides survey respondents into two groups, those who engage in no unpaid consumption during the sample, those who engage in some. Table 2 also reports these sequences separately for persons who report no unpaid consumption as well as those engaging in unpaid consumption. A comparison of these groups' consumption sequences provides some suggestive evidence of displacement following a first (paid) viewing.

For persons with no unpaid consumption, 34.9 percent of movies first viewing in a theater are viewed a second time by another paid means. For the persons who engage in some unpaid consumption, only 30.1 percent of movies first viewed in a theater have a second, paid viewing, while 4.5 percent have an unpaid second viewing. That is, both groups follow up a first theatrical viewing with the same tendency to view the movie through a second medium. But the paid viewing by those who sometimes steal is lower by almost exactly the amount of their unpaid consumption. This suggests one-for-one displacement.[11] A similar exercise to second consumption following first viewing of a rental copy yields similar implied displacement.[12] While suggestive, these calculations control for nothing.

Table 4 implements a regression analogue to the comparison in Table 3. Column 1 is a regression of the ratio of paid second viewings to theatrical first viewings on the ratio of unpaid second viewings to theatrical first viewings. The coefficient on the unpaid ratio is –0.27 and nearly statistically significant, indicating that an unpaid second viewing displaces 0.27 paid second viewings. The second column regresses the number

---

[11] $\dfrac{\partial paid}{\partial unpaid} = \left(\dfrac{.301-.349}{.045-0}\right) \approx -1.$

[12] $\dfrac{\partial paid}{\partial unpaid} = \left(\dfrac{-0.043}{0.125-0.097}\right) \approx -1.5.$

of paid second viewings following a theatrical viewing on the number of unpaid second viewings following a first theatrical viewing, controlling for the number of theatrical first viewings. The displacement coefficient is again –0.27 and significant. Columns (3) and (4) repeat the exercise for movies with a first paid rental consumption. Given that subsequent consumption follows initial rental in only 15 percent of consumption sequences, we do not have much scope for finding relationships among the movies first viewed through rental. And, indeed, second unpaid consumption of movies first seen through rental has no statistically discernible displacement effect on second paid consumption.

These estimates that come closer to controlling for unobservable tendencies to buy movies provide some evidence of displacement. Unfortunately, this approach does not provide an estimate of the displacement effect of first unpaid viewing.

2. Longitudinal Approach

The problem with the cross sectional estimates is that the error term, containing the individual's unobserved tendency to pay for movies, is correlated with his tendency to engage in unpaid consumption. The solution is to purge the error of this component. Longitudinal data allow this, provided that the unobservable determinant of individuals' paid consumption is constant across vintages. Because these estimates avoid this important form of unobserved heterogeneity, this is our preferred estimation approach, particularly for estimating the effect of first unpaid consumption on first paid consumption.

The movies in the survey represent the 50 most popular (highest box office) films over each of the three years prior to the survey. We can calculate the number of paid and unpaid consumption episodes among the films of each vintage. We can then use the data as a panel, including a fixed effect to purge the error term of individuals' vintage-invariant interest in buying movies. In effect, we ask whether persons who watch more unpaid 2005 movies, relative to 2003 movies, watch fewer paid 2005 movies, relative to 2003 movies.

Figure 3 shows the relationship between the differences in paid and unpaid consumption of movies between the 2003 and 2005 vintages (for the Fall sample). The smoothed line summarizes the relationship, and it is clearly negative. Persons who consume more unpaid movies from 2005 relative to 2003 consume fewer paid movies from 2005 relative to 2003. This looks like sales displacement.

To test this more systematically in a regression framework, define:

B_{it}^{j}=number of paid j^{th} (first or second) consumption episodes by person i for vintage t movies,

S_{it}^{j}=number of unpaid j^{th} (first or second) consumption episodes by person i for vintage t movies,

μ_i= unobserved determinants of an individual's tendency to buy movies,

ϕ_t= vintage effect, and

ε_{it}= indiosyncratic error.

We can then estimate:

$$B_{it}^{1} = \alpha^{1} + \beta^{1} S_{it}^{1} + \mu_i + \phi_t + \varepsilon_{it}^{1} \quad \text{and} \tag{1}$$

$$B_{it}^{2} = \alpha^{2} + \beta^{22} S_{it}^{2} + \beta^{21} S_{it}^{1} + \mu_i + \phi_t + \varepsilon_{it}^{2}. \tag{2}$$

Table 5 reports results.[13] As before, the first five columns refer to first consumption episodes. The coefficient on unpaid consumption in the first column, -0.76 (s.e.=0.10) indicates that persons with one additional unpaid first consumption engage in 0.76 fewer paid first consumption episodes. Each of the by-medium coefficients is also negative and significant, the largest being -0.27 (theater) and -0.29 (rental).

The last five columns of the table examine the relationship between unpaid first and second consumption and paid second consumption. Here, the coefficient on unpaid second consumption is -0.19 (s.e.=0.09), while the coefficient on unpaid first consumption is -0.24 (s.e.=0.06).

Interpreting the coefficients causally, a first unpaid consumption reduces paid consumption by 1 unit (-0.76 + -0.24). Recall that most first consumption episodes are theatrical and that a third of these are followed by a second paid consumption episode. Hence the maximum displacement exceeds 1.

The longitudinal displacement estimate of unpaid second consumption of –0.19 is similar to the Table 4 estimates of roughly –0.27 following theatrical first consumption.

What do these estimates mean? On average in our data, consumption sequences include 1.3 viewings, so the maximum amount of conceivable displacement exceeds one.[14] An unpaid first consumption reduces paid consumption by 1 unit. This is not 100 percent displacement, but it is nearly 80 percent of the maximum possible displacement.

---

[13] The models are estimated combining data from the Spring ("early") and Fall ("late") surveys. Because early respondents report on movies 2002-2004 while late respondents report on movies 2003-2005, each group gets its own vintage dummies in these models.

[14] To see this imagine a first consumption in the theater followed, sometimes, by rental and/or purchase. Alternatively, one might begin by watching an unpaid copy and then engage on no additional paid consumption. A decision to begin with an unpaid copy can therefore displace more than one paid consumption.

An unpaid second consumption reduces second paid consumption by only about 0.2, or 20 percent of the maximum possible displacement, by contrast.

In our sample individuals have an average of 1.8 unpaid first consumption episodes. According to our longitudinal estimates, these unpaid first consumption episodes displace 1.4 paid first consumption episodes (=1.8*0.76) per person and 0.4 second consumption episodes (=1.8*0.24). Our respondents have an average of 0.9 unpaid second consumption episodes, which displace 0.2 (=0.9*0.19) paid second consumption episodes. Overall, then, unpaid consumption displaces an average of 2.0 paid consumption episodes per person in the sample. To put this another way, unpaid consumption reduces paid consumption episodes in the sample from a counterfactual 56.2 to the observed 54.2, or by 3.5 percent.

Conclusion

Unpaid consumption of movies is a small share of consumption, even in a sample of highly Internet connected and technically sophisticated college students from high-income families. Moreover, with the rapid growth of DVDs over the past few years, film industry revenue has grown sharply, even as theatrical attendance has been flat. Unlike the music industry, which has actually experienced a pronounced downturn which is arguably based on file sharing, the movie industry faces smaller outward signs of a problem.

Yet, the evidence in this paper indicates that the file sharing that occurs has sharp displacing effects on paid consumption. Our best estimate indicates that an unpaid first consumption reduces paid consumption by one unit, and an unpaid second consumption reduces paid second consumption by about a fifth of a unit. The movie sales displacement rates that we estimate are much larger than displacement estimates that we estimate for music (Rob and Waldfogel, 2004). The large rate of displacement suggests that the main effect of unpaid consumption in movies is simply to transfer revenue from sellers to buyers. Unlike in music, where much unpaid consumption would not have occurred without file sharing, in movies unpaid consumption simply displaces paid consumption. This suggests that unpaid movie consumption does not reduce deadweight loss.

Why would displacement rates be so high in movies? File sharing in movies remains clumsy. Persons can choose between slow downloads and relatively easy to defeat DVD file protection. If the means available for copying movies become easier to use, file sharing may become a very serious threat to the film industry. The difficulty of obtaining unpaid copies may explain the sharp displacement. Perhaps only highly motivated consumers – those placing high valuations on the movies they are copying – engage in file sharing today. If file sharing were easier (and therefore more nearly costless), it is possible that even persons placing low valuations on movies – and who would therefore not otherwise consume the movie if paying – would share. Then the displacement would be smaller.

However, even if the cost of obtaining movies without paying falls, movies remain different from music in their cost of consumption. Watching a movie requires a few hours of undivided attention, which is costly to people. Music, on the other hand, can occupy divided attention. Hence, a person can get some use out of music he values very little (he can give it an experimental listen while doing something else). To the extent that movies are costly to consume even when free, the extent of displacement per unpaid movie viewing may remain quite high.

Because the amount of unpaid consumption in the sample is small (about 5.2 percent), the amount of displacement (3.5 percent) is small as well. While the amount of displacement in the sample is small, our large estimates of the rate of displacement may give the movie industry a cause for concern as the technologies allowing unpaid consumption diffuse widely. Sharp sales displacement effects raise difficult questions about the marketing of information goods that future research will need to address.

References

Bakos, Yannis, Erik Brynjolfsson, and Douglas Lichtman. "Shared Information Goods." Journal of Law & Economics XLII (April 1999): 117-155.

David Blackburn (2004), "On-line Piracy and Recorded Music Sales" available at http://www.economics.harvard.edu/~dblackbu/papers/blackburn_fs.pdf

David Bounie, Marc Bourreau, Patrick Waelbroeck "Pirates or Explorers? Analysis of Music Consumption in French Graduate Schools." (2005) http://www.ecare.ulb.ac.be/ecare/ws/honorvictor/papers/waelbroeck.pdf

Hong, Seung-Hyun. "The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster." August, 2005. https://netfiles.uiuc.edu/hyunhong/www/napster.pdf

Hui, Kai-Leung and Ivan Png. "Piracy and the Legitimate Demand for Recorded Music." *Contributions to Economic Analysis*. Vol. 2, no. 1, article 11 (2003).

Liebowitz, Stanley J. "Copying and Indirect Appropriability." *Journal of Political Economy*, vol. 93, No. 5 (Oct 1985): 945-957

Liebowitz, Stan "Will MP3 Downloads Annihilate the Record Industry? The Evidence So Far." Unpublished Manuscript. University of Texas at Dallas. June 2003.

Oberholzer-Gee, Felix and Koleman Strumpf. "The Effect of File Sharing on Record Sales: An Empirical Analysis." Mimeo. Harvard Business School. January 2004.

Shapiro, Carl and Hal Varian. Information Rules. Harvard Business School Press. Cambridge, MA. 1999.

Varian, Hal. "Buying, Sharing, and Renting Information Goods." Journal of Industrial Economics. Vol. XLVIII, no. 4, Dec 2000, pp. 473-488

Zentner, Alejandro. "Measuring the Effect of Online Piracy on Music Sales." Unpublished Manuscript. University of Chicago. December 2003.

Table 1: Sample Characteristics

movie interest	a lot less	less	same	more	a lot more
N=469	6.0%	15.4%	46.3%	23.9%	8.5%

family income (000)	under $25	$25-$50	$50-$100	$100-$250	over $250
N=441	3.4%	7.0%	22.7%	37.0%	29.9%

	mean	median	N
home video library	37.6	20	463
age	18.7	18	470
monthly theater visits	1.6	1	461

DVD burner	no access	access w/o ownership	own one
N=541	21.1%	24.5%	54.4%

race	white	Asian	black	Hispanic
	59.6%	24.5%	5.5%	5.5%

	Viewing Modes on Nth Viewing			
viewing mode	all	first	second	third
theater	25.5	25.4	0.1	0.0
television	8.9	4.9	3.3	0.6
rental	11.2	8.3	2.7	0.2
purchase	8.5	4.0	3.8	0.6
total paid	54.2	42.7	9.8	1.4
downloaded	1.2	0.8	0.3	0.1
burned	1.8	1.0	0.6	0.2
total unpaid	3.0	1.8	0.9	0.3
total	57.2	44.5	10.7	1.7

Table 2: Consumption Sequences, Overall and by whether Engage in Unpaid Use

	Overall	No Unpaid Use	Some Unpaid Use
total	20687	9315	11372
Sequences Starting With			
theater	56.9%	58.5%	55.6%
television	11.0%	11.5%	10.6%
rental	18.7%	19.2%	18.3%
purchase	9.1%	10.0%	8.3%
downloading	1.7%	0.0%	3.0%
burning	2.3%	0.0%	4.1%
Following First Viewing in Theater			
no subsequent	65.3%	65.1%	65.4%
television	9.8%	10.7%	9.0%
rental	9.7%	9.8%	9.7%
purchase	12.8%	14.4%	11.4%
downloading	0.8%	0.0%	1.5%
burning	1.6%	0.0%	3.0%
Total paid	32.3%	34.9%	30.1%
Total unpaid	2.4%	0.0%	4.5%
Following First Viewing through Rental			
no subsequent	86.7%	87.5%	85.9%
theater	0.1%	0.1%	0.1%
television	6.4%	5.9%	6.9%
purchase	4.5%	6.5%	2.8%
downloading	0.6%	0.0%	1.1%
burning	1.8%	0.0%	3.3%
Total paid	11.0%	12.5%	9.7%
Total unpaid	2.3%	0.0%	4.3%

Table 3: Cross Sectional Displacement Estimates for First and Second Viewings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Viewing:	first buy	first theater	first tv	first rental	first purchase	second buy	second theater	second tv	second rental	second purchase
Unpaid First Viewings	0.0062	-0.1200	0.0926	-0.0030	0.0367	-0.3351	0.0120	0.0222	-0.2636	-0.1057
	(0.1965)	(0.1622)	(0.0766)	(0.1150)	(0.0677)	(0.1100)**	(0.0064)	(0.0695)	(0.0499)**	(0.0624)
Unpaid Second Viewings						0.4186	-0.0070	0.0136	0.5347	-0.1227
						(0.1587)**	(0.0092)	(0.1002)	(0.0720)**	(0.0900)
fan==2	9.6677	4.1441	1.8201	4.0491	-0.3456	1.4561	-0.0532	0.8772	0.4692	0.1630
	(3.4942)**	(2.8841)	(1.3620)	(2.0455)*	(1.2045)	(1.8299)	(0.1062)	(1.1556)	(0.8300)	(1.0380)
fan==3	16.9425	9.8594	1.6728	5.1814	0.2289	3.5706	-0.0086	1.0674	1.4291	1.0827
	(3.1999)**	(2.6412)**	(1.2473)	(1.8732)**	(1.1030)	(1.6757)*	(0.0973)	(1.0582)	(0.7601)	(0.9506)
fan==4	24.2538	15.5082	3.8026	5.5389	-0.5959	6.9474	-0.0038	3.4337	1.2330	2.2845
	(3.4618)**	(2.8574)**	(1.3494)**	(2.0266)**	(1.1933)	(1.8137)**	(0.1053)	(1.1454)**	(0.8227)	(1.0288)*
fan==5	28.1352	17.7141	3.9585	7.8630	-1.4004	6.5046	-0.0696	3.6289	0.5197	2.4256
	(4.0252)**	(3.3224)**	(1.5690)*	(2.3564)**	(1.3875)	(2.1080)**	(0.1224)	(1.3312)**	(0.9562)	(1.1958)*
own	0.0288	0.0045	-0.0029	-0.0061	0.0333	0.0320	0.0002	0.0008	-0.0001	0.0311
	(0.0120)*	(0.0099)	(0.0047)	(0.0071)	(0.0042)**	(0.0064)**	(0.0004)	(0.0040)	(0.0029)	(0.0036)**
often	3.2080	3.2493	-0.5044	0.6606	-0.1975	1.1235	0.0114	0.1493	0.6326	0.3302
	(0.6959)**	(0.5744)**	(0.2713)	(0.4074)	(0.2399)	(0.3655)**	(0.0212)	(0.2308)	(0.1658)**	(0.2074)
Constant	19.2457	10.1857	3.3753	2.4537	3.2310	3.0394	0.0525	1.2220	0.6058	1.1591
	(3.0131)**	(2.4870)**	(1.1745)**	(1.7639)	(1.0386)**	(1.5780)	(0.0916)	(0.9965)	(0.7158)	(0.8951)
Observations	454	454	454	454	454	454	454	454	454	454
R-squared	0.28	0.25	0.04	0.04	0.13	0.22	0.01	0.06	0.18	0.21

Standard errors in parentheses. * significant at 5%; ** significant at 1%.

Table 4: Displacement on Second Viewing, Given First Paid Viewing

	(1)	(2)	(3)	(4)
	Paid 2nd / First at Theater	# Paid 2nd after Theater	Paid 2nd / First thru Rental	# Paid 2nd after Rental
Unpaid 2nd / Theater	-0.2743			
	(0.1454)			
# Unpaid 2nd after Theater		-0.2726		
		(0.1608)		
# First in Theater		0.3592		
		(0.0213)**		
Unpaid 2nd / Rental			-0.1641	
			(0.1305)	
# Unpaid 2nd after Rental				0.0182
				(0.0506)
# First thru Rental				0.1195
				(0.0111)**
fan==2	0.0172	-0.0757	-0.0848	-0.3767
	(0.0476)	(1.2739)	(0.0499)	(0.4733)
fan==3	0.0532	0.0661	-0.0644	-0.4225
	(0.0435)	(1.1817)	(0.0460)	(0.4383)
fan==4	0.1123	0.9524	-0.0511	-0.3488
	(0.0471)*	(1.3001)	(0.0496)	(0.4701)
fan==5	0.0575	-0.1557	-0.1185	-0.1298
	(0.0548)	(1.5111)	(0.0568)*	(0.5443)
own	0.0009	0.0260	0.0003	0.0021
	(0.0002)**	(0.0044)**	(0.0002)*	(0.0016)
often	0.0015	0.0606	-0.0012	0.0079
	(0.0095)	(0.2638)	(0.0098)	(0.0940)
Constant	0.2170	-2.1084	0.1606	0.1148
	(0.0410)**	(1.1135)	(0.0437)**	(0.4147)
Observations	452	452	397	397
R-squared	0.11	0.52	0.03	0.28

Standard errors in parentheses. * significant at 5%; ** significant at 1%

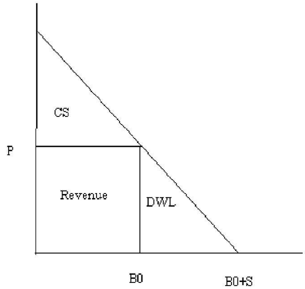
Table 5: Longitudinal Displacement Estimates for First and Second Viewings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Viewing:	first buy	first theater	first tv	first rental	first purchase	second buy	second theater	second tv	second rental	second purchase
Unpaid First Viewings	-0.7579	-0.2652	-0.1186	-0.2947	-0.0793	-0.2443	0.0013	-0.1188	-0.0351	-0.0917
	(0.0958)**	(0.0694)**	(0.0507)*	(0.0594)**	(0.0373)*	(0.0585)**	(0.0037)	(0.0415)**	(0.0283)	(0.0360)*
Unpaid Second Viewings						-0.1944	-0.0034	-0.0423	0.0176	-0.1663
						(0.0862)*	(0.0055)	(0.0611)	(0.0417)	(0.0530)**
early_d2003	0.9531	0.3567	0.2336	0.1297	0.2331	0.5010	-0.0136	0.1813	0.1098	0.2236
	(0.2818)**	(0.2041)	(0.1490)	(0.1748)	(0.1097)*	(0.1722)**	(0.0109)	(0.1220)	(0.0833)	(0.1058)*
early_d2004	-1.9830	-0.3221	-0.4335	-0.9954	-0.2320	-1.2785	-0.0138	-0.5000	-0.2274	-0.5373
	(0.2822)**	(0.2044)	(0.1493)**	(0.1750)**	(0.1099)*	(0.1718)**	(0.0109)	(0.1217)**	(0.0831)**	(0.1056)**
late_d2004	-0.2637	0.0741	-0.5439	0.1750	0.0311	-0.2853	0.0147	-0.3420	0.0669	-0.0248
	(0.2577)	(0.1867)	(0.1363)**	(0.1598)	(0.1003)	(0.1567)	(0.0099)	(0.1110)**	(0.0758)	(0.0963)
late_d2005	-6.3243	-0.9934	-1.7435	-2.3790	-1.2084	-3.2709	0.0149	-1.1645	-0.8565	-1.2648
	(0.2246)**	(0.1627)**	(0.1188)**	(0.1393)**	(0.0875)**	(0.1365)**	(0.0086)	(0.0967)**	(0.0660)**	(0.0839)**
Constant	13.9864	7.5657	1.9289	3.0669	1.4249	3.7833	0.0189	1.3420	0.9495	1.4729
	(0.1334)**	(0.0967)**	(0.0706)**	(0.0828)**	(0.0520)**	(0.0824)**	(0.0052)**	(0.0583)**	(0.0398)**	(0.0506)**
Observations	2164	2164	2164	2164	2164	2164	2164	2164	2164	2164
Number of individuals	541	541	541	541	541	541	541	541	541	541
R-squared	0.43	0.04	0.14	0.24	0.15	0.35	0.00	0.11	0.15	0.19

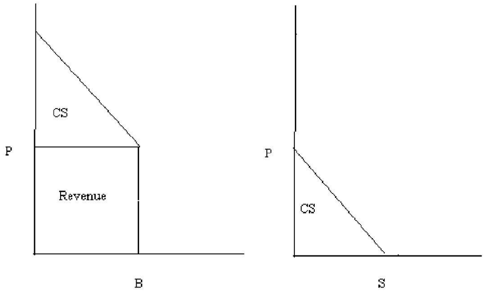
Standard errors in parentheses. * significant at 5%; ** significant at 1%

Figure 1

a: No Unpaid Consumption



1.b Paid and Unpaid: Low Valuation File Sharing



.

1c. High Valuation File Sharing

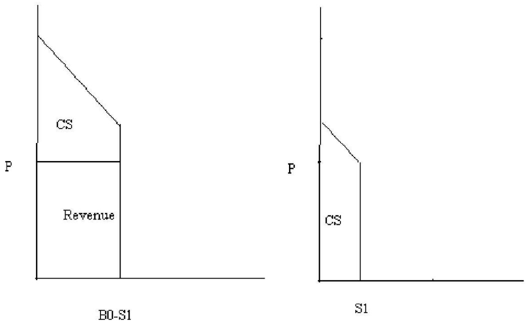


Figure 2: Paid and Unpaid First Viewings

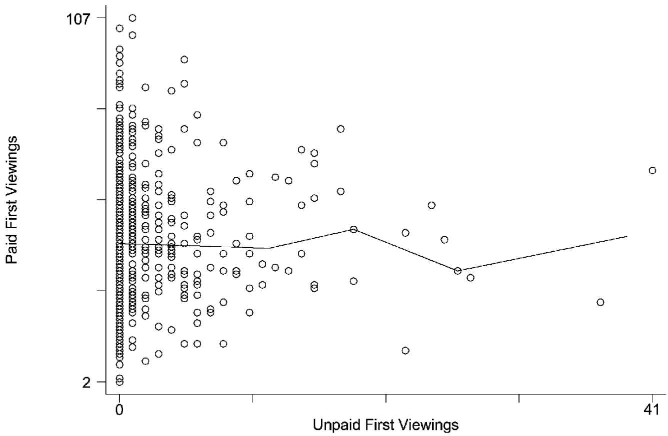
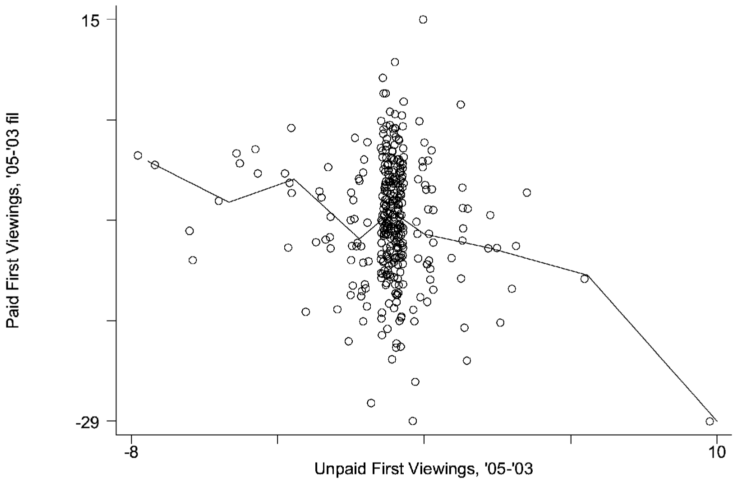


Figure 3: Changes in Paid and Unpaid First Viewings, 2003 to 2005 (Fall Sample)



Economic and Technical Drivers of Technology Choice: Browsers*

Timothy F. Bresnahan and Pai-Ling Yin

First Draft: Nov 05, 2003

This Draft: March 7, 2006

Abstract

The diffusion of new technologies is their adoption by different economic agents at different times. A classical concern in the diffusion of technologies (Griliches 1957) is the importance of raw technical progress versus economic forces. We examine this classical issue in a modern market, web browsers. Using a new data source, we study the diffusion of new browser versions. In a second analysis, we study the determination of browser brand shares. Both analyses let us examine the important of technical progress vs. economic forces. We find that the critical economic force was browser distribution with a complementary technology, personal computers (PCs). In both of our analyses, distribution had a larger effect on the rate and direction of technical change than technical browser improvements. This shows an important feedback. Widespread use of the Internet spurred rapid expansion of the PC market in the late 1990s. Our results show that the rapid expansion in PCs in turn served to increase the pace of diffusion of new browsers and thus move the economy toward new mass market commercial Internet use.

* The authors would like to thank 2 extremely helpful anonymous referees and seminar participants at the University of Toronto; Stanford-Berkeley IO Fest 2004; Japanese FTC; Federal Reserve Bank of Chicago Standards and Public Policy Conference; R&D, Education and Productivity Conference: An International Conference in Memory of Zvi Griliches (1930-1999); Kiel Munich Workshop on the Economics of Information and Network Industries, and the NBER Productivity Lunch.

Timothy F. Bresnahan
Landau Professor in Technology and the Economy
Stanford University
Department of Economics
579 Serra St.
Stanford, CA, 94305 USA
tbres@stanford.edu

Pai-Ling Yin
Harvard Business School
Morgan Hall 241
Soldiers Field
Boston MA 02163
pyin@hbs.edu

1) Introduction

A new invention creates a technological opportunity. The diffusion of the new technology to the economic agents who will use it determines the rate and direction of realized technical change in the economy.

The classical model of diffusion of a new technology (Griliches 1957) emphasizes adopters' incentives. Valuable technical progress is an incentive to adopt; the greater the advance of the latest technology over earlier alternatives, the more rapidly it will tend to diffuse. Yet adopters of many new technologies wait rather than adopt immediately.[1] They are held back by the fixed costs of learning about new technologies, of installing them, or by the costs of adapting technologies to particular uses. Actions by sellers of a new technology, such as distribution and marketing, can reduce the influence of these fixed costs and speed up diffusion. Ultimately, the pace of diffusion, and thus the pace of realized technical change in the economy, is determined not only by the pace of technical progress but also by economic forces like adoption costs and the success of market distribution (Griliches 1957).

The direction of technical progress is often determined by the same forces. The relative rate of diffusion of competing technologies depends on adopters' incentives. All else equal, a superior technology or a technology that is more effectively marketed and distributed will be adopted more quickly, while a harder-to-install technology will be adopted more slowly. The pace of adoption of competing technologies is particularly important in industries with *de facto* standards, like computing. The competing technology that diffuses more rapidly will often set the standard.

[1] Hall (2003) provides a modern survey of the literature documenting that many technologies diffuse slowly and seeking to explain the wide variation in the rate of diffusion.

In this paper, we study the diffusion of new and improved versions of World Wide Web (WWW) browsers from 1996 to 1999. We focus on commercial browsers from Microsoft and Netscape. The classic issues in the diffusion of technology appear in new forms.

The invention and commercialization of the browser triggered the widespread use of the Internet and created the opportunity for commercial applications such as Internet e-commerce, entertainment, and advertising-supported web pages. These commercial applications exploited improvements in each of several new browser versions released over the period we study. These applications could only reach a mass market if the improved browser versions were running on the personal computers (PCs) of many consumers. Thus, the pace of widespread diffusion of new browser versions determined the rate at which the Internet came to serve a mass market with valuable new commercial applications. For that reason, we study the diffusion of new and improved browser versions.

The interaction between Netscape and Microsoft led to another important market outcome. Widespread use of a browser of one brand would set a *de facto* standard for connection between the Internet and PCs. A Netscape standard seemed likely during the early stages, but the "browser war" ended with a *de facto* Microsoft standard in place.

Browser technical progress thus gives us two phenomena to study. We examine the pace of diffusion of new browser versions in order to study rate of technical progress in transformation of the Internet into a mass market commercial platform. We examine the reversal of the two main commercial browser brands' market positions in order to study the setting of a *de facto* standard.

The diffusion of commercial improvements to the browser, while very important, has not yet been studied by economists. In particular, no explanation has been proposed for a phenomenon noted by market participants at the time: later versions of both brands of browser diffused more slowly than earlier versions.

In contrast, the browser market brand switch from Microsoft to Netscape has been widely studied[2], but the explanation remains controversial because of the antitrust trial.

---

[2] See Cusumano & Yoffie (1998) , Shapiro & Varian (1999), and Windrum (2000) for analysis of the browser war events. See Manes & Andrews (1993) Wallace (1997), and Bank (2001) for Microsoft-

The controversy follows classical considerations in the diffusion of technology. One side argued that faster technical progress by Microsoft led to the reversal, while the other side emphasized the role played by distribution and marketing. Turned into a behavioral hypothesis about browser users, the two sides differed on whether users adopted the latest and greatest browser (technical progress explanation) or tended to use the browser that came with their computer (distribution convenience explanation).

In this paper, we create a new dataset and use it to study the determinants of both the pace of diffusion of new browser versions within brand and the shift in brand shares. We are able to measure the impact of technical progress and of distribution on both outcomes. This industry gives us the opportunity to study distribution in particular detail because of contracts which required or blocked the distribution of particular browsers. Distribution and technical progress vary separately because the contracts differentially affected the distribution of the same browser to users of different kinds of computers.

Our model of diffusion is entirely conventional, except for a particular feature of the PC industry in this era. The rapid expansion of demand for PCs meant that the browser was also diffusing into a rapidly expanding field of potential adopters instead of diffusing into a fixed field of potential adopters.

We reach the same conclusion, in separate analyses, about the pace of diffusion of new versions within brand and about brand choice. In each case, distribution played a larger role than did technical progress in determining the market outcomes. We quantify the effect of these forces in our analysis.

2) Browser Market Background

In this section, we review the market background with the goal of making clear what technical progress we study and why it is important.

---

centered histories. See Liebowitz and Margolis (1999), Fisher (1999) Schmalensee (1999a and 1999b), and Bresnahan (2002), as well as the interchange among Evans and Schmalensee (2000a, 2000b) and Fisher and Rubinfeld (2000a, 2000b) for analysis closely related to the antitrust case. Jenkins et al. (2004) studies the effect of browser brand tying and exclusion on brand choice, but by modeling dynamic investment decisions by Netscape and Microsoft rather than user decisions.

a) Browser Invention

A web browser lets an individual computer user easily find, observe, and retrieve information on the Internet. With a browser, users can access information like news, entertainment, maps, etc. The browser is also the users' gateway to new mass market online applications, including electronic commerce, email, and so on.

It had been evident to market participants for many years that a technology to connect users to online content would be valuable. Many computer and telecommunications industry firms attempted to create mass markets in online applications in the late 1980s and early 1990s, with little success. Demand for browsers grew quickly because the browser fulfilled this long-felt need.[3]

The Internet dates back to the 1970s, but a number of inventions in the early 1990s transformed it in a direction more suitable for mass market use. These included first the World Wide Web (WWW) and then the browser.[4] The WWW is largely defined by a set of standard "protocols" for connecting computers together.[5] One of the things a browser does is let a personal computer communicate through those protocols.

These inventions largely came out of academic science. Mosaic, the first browser, was invented at the University of Illinois. As a result, the key protocols for the Internet and the WWW were largely in the public domain and the technologies at the heart of the browser itself were not protected by any strong intellectual property rights.

In 1993 and 1994, an "Internet Mania" took off. Especially in universities, users demanded browsers in growing numbers, acting on their incentive get online. At this stage, these were simple freeware browsers like Mosaic. Users' incentive to get online grew as more online information and applications appeared. There were more and more web sites – simple ones, at this stage -- for them to visit. "Webmasters" (those who built websites) also benefited from greater browser usage. More browser usage meant a larger audience for online information and more customers for online commerce. The "mania"

---

[3] See Mowery & Simcoe (2002) for more on the history of the Internet's development, and Wiseman (2000) and Vogelsang & Compaine (2000) for essays on the economic and political impact of a commercialized Internet.

[4] Gottardi (2003) presents a diffusion model showing the impact of the browser on Internet use.

[5] There is an alphabet soup of protocols and standards in the Internet and in the WWW. We will avoid discussion of these specifics and expansion of these acronyms, to the extent possible. There are also a set of semi-public standard-setting bodies for these protocols, like the W3C, to which we pay little attention, since the important standard-setting activity in the era we study was *de facto* and commercial.

was a positive feedback loop between the growing number of websites and the growing number of users.

Very few people had used the Internet until the advent of the browser as a connection technology. With the browser, many more people could use the Internet easily. The Internet Mania spread beyond universities creating a large-scale demand for connecting to the Internet.[6] Internet service providers (ISPs) saw their market grow as a result. A retail ISP sector emerged and spread very rapidly (Downes & Greenstein 2002). Many existing online services, such as America OnLine (AOL), began to transform themselves, also offering ISP services (Swisher 1998). Others were entrepreneurial entrants.

Access to the Internet made the PC more valuable. Web browsing, email, and instant messaging were among the new and important applications that raised the value of PCs. The overall effect was an increased demand for PCs.[7]

These developments transformed two technologies. The PC had been used primarily for non-networked applications such as word processing and spreadsheets; it was transformed into a connected communications node. The Internet had largely been used by a small number of public or academic users; it was transformed into a ubiquitous commercial technology.

The role of interactions among complementary technologies as the center of these events is widely studied. Greenstein (2004), in a forthcoming book on Internet geography, describes the importance of complementary infrastructures in Internet diffusion. This useful literature has not yet addressed the end users' browser adoption decisions.

b) The Commercial Browser

Several of the inventors of the browser sought to commercialize it by founding Netscape Communications. In 1994, they released their first browser, which we abbreviate as NS1. Over the next several years, Netscape continued technical progress in the browser ranging from better display ("frames," and so on) to more secure interaction

[6] Forman et al. (2005) discuss the Internet diffusion process.
[7] Goolsbee and Klenow (2002) examine computer demand in this era, finding that externalities across households, email, and Internet use are important drivers of demand.

with servers (computers that hosted websites). This technical progress would make the browser and the WWW more commercially valuable by enabling more complex applications such as e-commerce and advertising-supported websites.

Netscape introduced five major versions of its browser between 1994 and the end of the "browser war" in 1999. Each included new technical progress of value to commercially oriented webmasters, and each provided new features valued by users. We list these major versions in Table 1.[8] For ease of comparison, we rename Netscape version 4.5 as version 5.

Table 1 Major Netscape browser versions in our analysis

NS1	NS2	NS3	NS4	NS4.5 ("NS5")
includes 1.x, 0.x	includes 2.x	includes 3.x	up to 4.4x	includes 4.6, 4.7

We study the diffusion of new browser versions to users. We do not study the parallel adoption of new and improved WWW technologies by webmasters. The distinction matters in thinking about technical progress in browsers. Some types of technical progress in browsers, such as "rendering" images more quickly, was directly useful to the user and thus provided an immediate incentive to adopt. However, many improvements in browsers were designed to permit webmasters to make more advanced web sites. This type of technical progress would only give users an incentive to adopt after websites took advantage of the improvements. Webmasters, in turn, could only get a wide audience for their more advanced websites if there was widespread usage of new and improved browsers. As a result, the rate of diffusion of the newest browser version depended on the time it took for webmasters to a) become convinced that the newest version would be adopted by a sufficient number of users and b) subsequently release advanced websites.

Webmasters cared about the rate of diffusion of newer browsers for a number of reasons, including the need to make websites work with both older and newer versions.

[8] There are many minor browser releases, updates and so on. We aggregate numerous small versions into major browser versions in our empirical analysis in order to have each version represent a substantial advance over the earlier one. The minor versions we include in each major version are listed in the last row of the Table. For example, Netscape Navigator versions numbered less than 2 (1.45, 0.98, etc.), are aggregated to major version NS1 in our analysis.

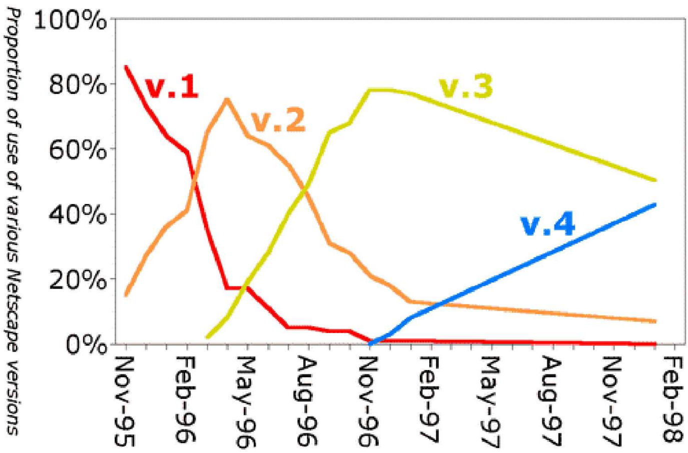
Most importantly, the advanced websites that they designed often had commercial applications, and widespread diffusion was necessary to reach a mass market. Statistical tools appeared that would let a webmaster look at a particular website's "server log" to see what versions of browsers its customers were using. A marketing research industry, with firms such as AdKnowledge and InterSe, emerged to measure the pace of diffusion of new browser versions and to report it to webmasters.

Webmasters noted a slowing pace of diffusion of new browser versions.[9] This presented them with a tradeoff between offering the best available web pages and gaining the highest volume of usage. In a website / magazine aimed at webmasters, Nielsen (1998) made a browser versions diffusion graph which we display as Figure 1. Nielsen used data from InterSe and AdKnowledge because those tracked the commercial websites of greatest importance to webmasters contemplating new features. In this figure, he shows the share of each major version of Netscape browsers out of total usage of all Netscape browsers. His conclusion was that Netscape users were adopting the newer versions more slowly than they adopted earlier versions.[10]

[9] See Neilsen (1998). Also see, for example, the discussion thread "Broswer usage stats" [sic] in the online discussion forum comp.infosystems.www.authoring.site-design of September 2nd to 4th, 1997.

[10] Neilsen also finds slowing adoption for IE.

Figure 1 Nielsen's Netscape browser diffusion graph



To study the adoption of technologies which supported those mass market commercial applications, we focus on the diffusion of version 2 and later commercial browsers from Netscape and Microsoft. As our sample begins, some users are already using pre-commercial browsers (such as Mosaic) and version 1 browsers from Netscape and Microsoft. We measure technical progress relative to the base of version 1 commercial browsers.[11]

c) The "Browser War"

We now turn to a more controversial part of the history, the "browser war" between Netscape and Microsoft.[12]

Netscape introduced the browser as a modular component. That meant that the Netscape browser ran on all kinds of computers, including Macintoshes, UNIX machines, the latest version of Windows PCs, and earlier versions of Windows. (Introduction dates

[11] Of course, nothing is ever an entirely new application, and one might follow Nordhaus (1997) by modeling the technical level of a broad "online services" category which could include the predecessors to the widely used Internet. Before the widespread use of the Internet, however, online services reached approximately 2.5% of their current total users and had a fundamentally different approach to technology that lacked the universality and openness of the modern Internet.

[12] See footnote 2 for studies of this period.

for Netscape browsers on a variety of computers can be seen in Table 3.) A user could use a Netscape browser regardless of the web content accessed or kind of computer she used. The open modular component strategy also meant the browser's "protocols" for communicating with web sites were open and documented. Netscape encouraged the development of server software to communicate with the browser by other firms and also wrote such software itself.

Netscape's goal was ubiquitous distribution of its browser in order to set a standard for web communications. It made distribution agreements with firms selling new computers, ISPs, online services, and others. Netscape also made its browser available for free download on its website and for sale in retail stores.

The second commercial browser firm of any importance was Microsoft, producer of Internet Explorer (IE). Through the period of the Internet Mania, Microsoft did not supply browsers. Once Microsoft realized the browser was a competitive threat to the Windows operating system (OS), it began a rapid program of imitating Netscape's innovations. Released in August of 1995, the first version of Internet Explorer, IE1, was a poor imitation. After that, Microsoft worked to catch up with improved versions of its browser. The major versions of Microsoft browsers used in our analysis are listed in Table 2.

Table 2 Major Microsoft browser versions in our analysis

IE1	IE2	IE3	IE4	IE5
includes 1.x	includes 2.x	includes 3.x	includes 4.x	includes 5.x

Browser quality improved over time at both innovator Netscape and imitator Microsoft. It is clear that quality increase was more rapid at Microsoft, but whether Microsoft caught up in quality is debatable. Schmalensee (1999a and 1999b) and Liebowitz and Margolis (1999) use measures that show Microsoft catching up by version 3, while Fisher (1999) and Bresnahan (2002) argue that the catch-up occurred later, if at all.[13]

[13] Liebowitz and Margolis (1999) and Schmalensee (1999a and 1999b) used a relative browser quality index based on expert opinion. They proceeded by examining software reviews in personal computer magazines and counting the number of reviews which recommended Netscape, those which were

The impact of the differential rate of technical progress in the two brands of browsers is also controversial. Schmalensee (1999a and 1999b) and Liebowitz and Margolis (1999) argue that differential rates of technical progress explain browser brand market share changes over time. In contrast, Bresnahan (2002), Fisher (1999), and Fisher and Rubinfeld (2000a) argue that distribution played a substantial role in the browser brand shift.[14] Examining these very different hypotheses is one of the empirical goals of this paper.

The marketing and distribution strategies of the two browser firms had similarities and differences. Both Netscape and Microsoft attempted to distribute their browsers widely and charged zero prices at the margin.[15] Each hoped to have a ubiquitous browser in order to set standards for connecting personal computers to the Internet.[16]

There were important differences between the two firms' strategies as well. Microsoft began by introducing its browser only for the latest version of Windows. IE1 was introduced at the same time as a new version of Windows, Windows 95, and only worked with that OS. At the time, the vast majority of PC users were running earlier versions of Windows, notably Windows 3.1, with a minority running Apple Macintosh. These existing PC users could not use IE1. Starting with IE2 in late 1995, Microsoft introduced new browser versions both for the newest version of Windows and for other kinds of computers (sometimes with a lag).

Because we are interested in widespread diffusion, we focus on four mass market OSs for PCs. Three are from the Windows family (Windows 3.1[17], Windows 95, and

---

tied or mixed, and those which recommended Microsoft. Liebowitz and Margolis used similar data for many other applications categories. Fisher (1999) based his analysis in market research conducted by Microsoft and in the opinions of Microsoft marketing managers, both from documents in the antitrust case. Bresnahan (2002) quotes extensively from the Microsoft marketing managers.

[14] This and many other economic issues in the antitrust case were debated in a useful point/counterpoint format by Fisher and Rubinfeld (2000a and 2000b) and Schmalensee and Evans (2000a and 2000b).

[15] This applies to marginal prices in our sample period. Earlier, Microsoft's plan had been to distribute Internet Explorer in a Windows "plus pack" and charge separately for it. Microsoft abandoned that plan after recognizing a competitive threat from the widespread use of the Internet and never charged separately for Internet Explorer. Netscape, at the beginning of our sample, used a plan of price discrimination in favor of individual users, who got prices of zero, while corporate users were charged a price. Netscape later abandoned this in favor of zero prices to all users.

[16] For a review of the standards literature, see David & Greenstein (1990) and Stango (2004).

[17] In our data, Windows 3.1 includes older versions recorded as "16 bit Windows."

Windows 98), and the fourth is Apple Macintosh. These OSs are the ones most likely to be used by individual end users.

In Table 3, we report the introduction date for each of the major browser versions on each OS, gleaned from the suppliers' websites. There are some gaps in the table because of supplier choice (IE1 was available only for Windows 95) or because the browser was obsolete at the time of the OS's introduction (the 1994 version of Netscape's browser, NS1, was not relevant to the 1998 Windows 98). The table clearly shows Netscape's strategy of writing for all existing OSs and Microsoft's changing strategy.

Another aspect of Microsoft's distribution and marketing strategy was to compel widespread distribution of IE browsers and to prevent widespread distribution of Netscape browsers. Microsoft imposed contractual restrictions on sellers of computers (which required them to distribute IE with new computers) and used threats to stop computer sellers from distributing Netscape with new computers. Microsoft signed contracts with ISPs and online services (such as AOL) which required them to distribute IE to subscribers and strictly limited distribution of Netscape. These contractual agreements grew more restrictive and more widespread over the course of our sample period.[18] For example, the default browser distributed with Macintosh computers was Netscape until August of 1997, when a contract between Microsoft and Apple came into effect, making IE the default browser on Macs.

Table 3 Basic browser timing facts:
Introduction date for each browser on each operating system

[18] For a general overview of this topic with many links to original antitrust case documents, see Bresnahan (2002) or US Department of Justice et al. (1999) Section V. B.

	Windows 3.1	Windows 95	Windows 98	Macintosh
NS1	Dec. 1994	Aug. 1995		Dec. 1994
NS2	March 1996	March 1996		March 1996
NS3	Aug.1996	Aug. 1996		Aug. 1996
NS4	June 1997	June 1997	Aug. 1998	Sep. 1997
NS5		Oct. 1998	Oct. 1998	Oct. 1998
IE1		Aug.1995		
IE2	April 1996	Dec. 1995		April 1996
IE3	Dec. 1996	Aug.1996		Jan. 1997
IE4	Feb. 1998	Oct. 1997	Aug. 1998	Jan. 1998
IE5	March 1999	March 1999	March 1999	2003

The efficacy and purpose of the distribution restrictions was hotly disputed in the antitrust case. Davis & Murphy (2000) argued that the bundling of IE with Windows was itself user-friendly technical progress rather than a restriction on distribution, Microsoft's economic expert, Richard Schmalensee, argued that other means of distribution, such as downloading a browser over the Internet or buying it in a retail store, rendered the restrictions on computer manufacturers and on ISPs and online services irrelevant.

In contrast, the government emphasized the effect of distribution and restrictions on distribution.[19] Adopting the analysis of distribution advanced by Fisher (1999), it rejected Microsoft's defense contention that distribution was entirely unrelated to the success of IE.

Much of the government's argument was drawn from browser marketing documents inside Microsoft. The marketing managers selling browsers focused on two elements of the fixed costs as particularly important. One was simple distribution convenience. Users might conveniently get a browser with a new computer or when they

[19] Two forces led to the government winning this part of the antitrust case. One was the large volume of Microsoft documents and testimony concerning distribution and restrictions on it. A Microsoft browser marketing executive said in court that the purpose of restrictions was this: "we did specifically ask that ISPs distribute Internet Explorer by itself when they distributed Internet Explorer, so that we would not lose all of those side-by-side user choices." Cameron Myhrvold (1999), 2/10/99, at 62:7-25. The other reason was the Court's rejection of Schmalensee's statistical analysis.

signed up for Internet service. Less conveniently, they might download a new browser from the web and install it on their own computer. The second fixed cost identified by marketing managers was the complexity of installing a new piece of software on a computer.

Computer users are heterogeneous in the degree to which adopting new technologies involves large fixed costs. This heterogeneity in tastes is reflected in heterogeneous inertial behavior. Users vary in their valuation of convenience. They may have a faster or slower web connection.[20] Users can vary in their ability to install new software on their computer or in their confidence in their ability. For this reason, less-sophisticated users tend to avoid new software if it requires installation. Less sophisticated users also tend to be uninformed about new products and may have a fixed cost simply of learning of the existence of a new browser.

At one extreme, there are users who occasionally get new computers and use the software that came with their computer until they get the next one. At the other extreme are people who always download the latest version of software. In between are people who download the latest version or get it in a store when they come across a task they cannot complete without it. Our dataset does not have individual-user regressors (e.g., experience on the web, using a browser at work, etc.) which would allow us to examine this heterogeneity in detail. Instead, the testable implication of heterogeneous fixed costs is that some users might delay adoption of a new browser version.

The prevalence of these distinct behaviors depends not only on the size of the fixed costs but also on their distribution in the population. Distribution of fixed costs of adoption in the population is the determinant of whether inertia in diffusion is a quantitatively important force. It is also the determinant of the role of mass distribution and marketing. Since these costs are distributed in a population containing many new users (who were brought into the expanding field of diffusion by technical progress), introspection about the costs of downloading is an unpromising method. We rely instead on data.

[20] Over the period we study, modem speeds grew faster while browsers also grew larger, so the time costs of a download remained roughly constant for the average user.

We are able to examine the quantitative importance of mass distribution in two dimensions. Our largest contribution comes in examining the role of distribution vs. raw technical change in the diffusion of new browser versions.[21] We also revisit the distribution vs. technical change controversy in our empirical work on browser brand choice. Our contribution there consists of using better data and of examining measures of both technical progress and distribution.

3) Model

The goal of our empirical work is to examine the roles of distribution and technical progress in both the diffusion of new versions of each brand of browser and in the brand shift.

Several different technical and economic forces can accelerate the diffusion process. A new technology may be a substantial improvement over the one it replaces. Agents may all learn about the technology at about the same time. Network effects may lead to bunching of adoption times. Agents may experience similar costs and benefits of the new technology, so they reach the decision to switch to it at around the same time. Switching costs and inertia, if they are present, may also be broadly similar across agents. These forces favor a rapid adoption process once diffusion has begun.

The diffusion process can be slowed by a number of forces as well. The costs of adopting a technology may be high, or the costs sunk to an older technology may be high, for some users. Information about the new technology may be poor at first. Agents may wait for better versions of a technology and will have an incentive to wait if there are fixed costs of adopting a new version. For valuable new technologies, economic institutions are likely to arise to partially redress these retarding forces. Reliable information sources and effective distribution channels, for example, may overcome the inertia that these retarding forces build around older technologies.

Our model of diffusion of new versions is closely related to the technology adoption and diffusion model of Griliches (1957). Like him, we emphasize the economic return to a technology adopter. Like him, we use aggregate data to study the diffusion of

[21] Our closest predecessor is Goldfarb (2004), who looks at the role of universities on Internet adoption.

browsers into a similarly situated population. Like him, we study a diffusion process that takes place over time and use cross section differences in the environment of groups of adopters as exogenous predictors in the rate of diffusion. Indeed, much of our economic interpretation follows his, notably in seeing the adoption decision as limited by the frictions of distribution (possibly overcome by marketing and distribution) and advanced by the attractiveness of new technologies. Where we differ primarily is in context, in examining the direction as well as the rate of technical progress, and in our emphasis on the changing size of the population that might adopt the technology (the expanding field of adopters).

The browser diffused into a rapidly expanding field of potential adopters. The widespread use of the Internet caused by the browser meant that a large number of people were buying new PCs and opening new ISP subscriptions. We examine the possibility that this expansion of the diffusion field affected the rate and direction of diffusion of browsers.

a) Browser Version Analysis (within Brand): Pace of Diffusion

In our diffusion analysis, like any other, the dependent variable measures the tendency to choose the new technology. Conceptually, in our study, that means the tendency to choose the newest version of a brand of browser available to a particular customer. We condition on the kind of computer (OS) used and the brand of browser chosen; i.e. we focus on the choice of the newest browser version within the set of available versions of a particular brand of browser.

The start date of the diffusion process for each version of each brand of browser on each OS is the availability date given in Table 3. Call the newest version of each brand of browser at a given date, b*. The dependent variable is a function of the share of all users of a particular brand of browser on operating system o at date t who are using b*, $S_{b^*ot|brand}$. Our model is

$$\ln\left(\frac{S_{b^*ot|brand}}{1-S_{b^*ot|brand}}\right) = X_{b^*ot}\beta + \varepsilon, \qquad \text{Equation 1}$$

where X_{b^*ot} denote regressors associated with the newest browser of a particular brand on operating system o at time t, and ε is an iid error term.

The relevant concept of technical progress in this within-brand model is the improvement of each new version, from a user perspective, over the previous versions. We will measure technical progress by the rate at which the new version diffuses. Accordingly, the Xs always include $INTR_{b*ot}$, the number of months since the newest browser was introduced on the operating system. This allows the coefficient on $INTR_{b*ot}$ to capture one potential cause of the slowing rate of diffusion seen in Figure 1: decreases over time in the rate of improvement of new browser versions.

Of course, the coefficient on $INTR_{b*ot}$ measures all the forces that tend to make a technology diffuse that are not captured in other time-varying Xs. For example, we cannot control for website improvements which led users to want the latest browser. We nonetheless interpret a browser that diffuses faster (has a higher coefficient on $INTR_{b*ot}$) as one that has advanced technically relative to pre-existing browsers. This interpretation almost certainly overestimates the importance of technical progress as a cause of browser diffusion, particularly for later browser versions given that improvements in websites is an omitted variable.

The other Xs in our model are measures of distribution. The relevant concept of distribution is whether the browser came with the user's computer (or with their ISP account). A user who got a new version of a browser with their computer will not need to bear any costs of learning that it exists, downloading it, or installing it. In the data section below, we define regressors which measure the probability that a particular browser was distributed with a users' computer or by their ISP.

Our first goal is to estimate the distinct effects of these technical progress and distribution measures. Our second analytical goal stems from the systems nature of the PC and Internet industries and from the importance of some large-scale feedback loops. The upturn in PC demand resulting from the invention of the browser occurred rapidly enough potentially to impact the diffusion of improvements to the browser. So, too, did the rapid rise in ISP subscriptions. We contribute to the literature by quantifying the impact of rapidly increasing demand for a complementary product (PCs) on the diffusion of browsers.

b) Browser Brand Analysis: Direction of Diffusion

There is another aspect to users' browser choice other than version: brand. In our brand models, we estimate a linear probability model. The dependent variable, denoted $S_{IE,ot}$, is IE's share of commercial browsers in a particular month t on a particular OS o. The numerator is the number of users of all versions of IE. The denominator is users of all versions of IE plus all versions of Netscape. Our model is

$$S_{IE,ot} = X_{b^{*}ot}\beta + \varepsilon \,, \qquad \text{Equation 2}$$

where $X_{b^{*}ot}$ denote regressors associated with the newest browser of a particular brand on operating system o at time t, and ε is an iid error term.

Our interpretation of this linear probability model is entirely standard. In the brand analysis, the relevant concept of technical progress is relative, i.e., how much faster is IE improving than Netscape? Similarly, the relevant concept of distribution is relative, i.e., how much better distribution is IE gaining on a particular OS than is Netscape? Except that they are relative, our measures of technical progress and distribution in the brand analysis follow our measures in the version diffusion analysis.

The zero marginal prices for browsers mean that it is not possible to address browser quality by hedonic pricing methods (Griliches 1988). As a result, we will use Schmalensee's (1999a) relative browser quality measure as a regressor in part of our empirical work. He measures "leadership" as a dummy variable equal to one when the market share of a brand of browser (or other application) is largest. Leadership only changes once in browsers, so we replace his variable with share. Our variable, Rel_Qual_Jrnl, is the average share of journalists' recommendations for and against IE at time t. The concerns of the hedonic literature about valuation based on quality assessments by experts limit its usefulness, so we will also use a much less restricted set of quality measures.[22]

[22] In the case of PC industry magazine reviews, there is a particular interpretational problem related to the distinction between a reviewer's assessment of best technology and his recommendation. In markets with network effects, and especially in PC software markets, it is commonplace to recommend that users choose the product which is going to be most popular even if it is not the best. See, e.g., "Nine Timeless Tips for Tech Buyers," Kevin McKean, *PC World*, June, 2002, p. 19. ">>**Tip Lean toward what's popular**: What a shame to have to offer this advice. But the best technology doesn't always win in the market." [emphasis in original].

We do not attempt to control for network effects in browser brand choice, since our data are aggregate. We instead employ a reduced form model for the aggregate share of a brand. This allows us to estimate the equilibrium effects of changes in technical progress and distribution (some of which may flow through network effects) but not to separately estimate users' valuation of network effects vs. technical progress or distribution.

Conceptually, the two analyses described in the last two subsections attempt to measure the relative importance of the technical progress and distribution theories explaining the pace and direction of diffusion and the events of the marketplace in the late 1990s.

4) Data Set

We use aggregate data for browser usage based on browsing at a web site at the University of Illinois Urbana-Champaign (UIUC).[23] In this section, we describe the data source (since it is novel) and our sample, dependent variable, and regressors.

The UIUC computer center keeps monthly logs made by its World Wide Web servers. We chose UIUC because it has maintained those logs consistently since early in the history of widespread use of the Internet. Our sample begins with the oldest available data, April 1996, and ends in December 1999. We end at that date because it is the first year-end that is (a) clearly after the end of the browser war and (b) well into the diffusion of commercially capable browsers.

Users browsing the UIUC Web site are not randomly drawn from the population of all Web users. For the most part, they are people interested in browsing engineering students' web pages. The advantage is that we study technology use over a period of time by a growing body of similar users. Users browsing at UIUC early and late in our sample are reasonably like one another: they are technically proficient. The disadvantage is that the users we sample are more likely to prefer new technology than the general population. They are likely to have lower costs of getting software onto their computers and downloading it. Thus, we are likely to overestimate the impact of technical

[23] We are grateful to Ed Kubaitis for archiving the logs and for technical advice.

improvements on choice of browser and to underestimate the impact of distribution relative to the general population.

Our sample dates vary across the four OSs. For each OS, we restrict the sample to the time period in which it is economically important, which we define as the time from the official introduction of the OS to the introduction of its replacement's replacement.[24] For Macintosh, user agent codes do not permit us to distinguish OS versions, so the entire period is included. The sample dates are shown in Table 4.

Those dates exclude users who run an OS before its official release date (beta testers and the like.) They also exclude users who got their computer long before browsing. Neither group seems likely to be representative of mass market users' tradeoff between distribution convenience and technical excellence.

Table 4 Observed operating systems and periods of use

OS	Macintosh	Windows 3.1	Windows 95	Windows 98
Sample Dates	4/96—12/99	4/96—7/98	4/96—12/99	8/98—12/99
No. of Months	45	28	45	17

Our data processing procedures for using the web server logs are documented in Appendix A. Each time a user accesses a web page, the web server's log records information about the browser and OS the user is running in the "user agent" field. That, plus other information recorded by the web server, forms the basis for our dataset.

An observation in our dataset is a browser / OS / month triple. We aggregate the UIUC data to the level of users running the same OS and then calculate the usage shares of each major browser version of each brand. The sample sizes to estimate these browser usage shares are substantial.[25] We report the descriptive statistics on browser shares used to generate our dependent variable in tables and graphs.

The first graphs we look at are the browser version diffusion curves within each brand. We examine Windows 95, since all the versions of all the browsers were available on that OS. In Figure 2 we show S_{bot} (the within-brand shares of a browser version b on an operating system o at time t) on Windows 95 for IE browser versions. Our sample

[24] This is consistent with industry practice. Microsoft has a standing policy of only supporting the newest OS and its immediate predecessor, for example.

[25] The number of unique users rises from 35,757 in the first month of our data to 229,579 in the last month.

period begins just as IE3 is beginning to diffuse on Windows 95, and just before the share of IE2 peaks. The within-brand diffusion paths that are visible (those of IE3, 4, and 5) are flattening over time.

Figure 2 Market shares of major versions of Internet Explorer on Windows 95

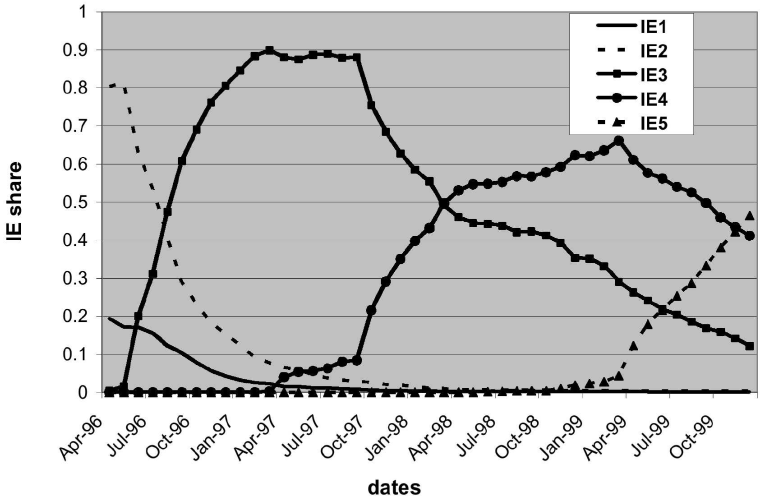
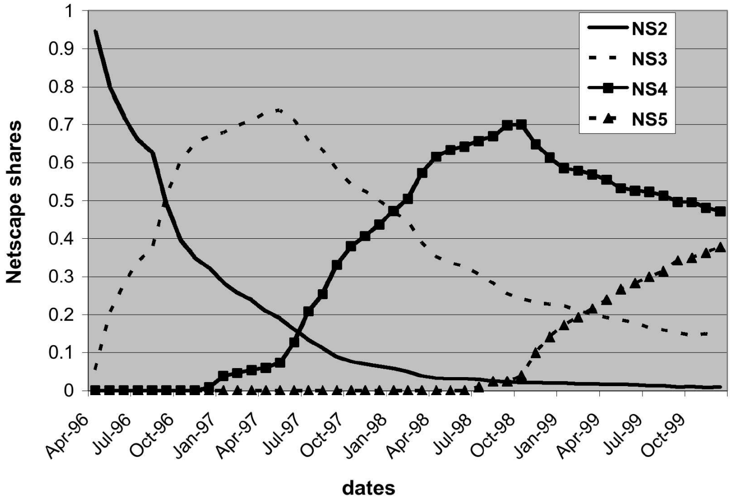


Figure 3 Market shares of major versions of Netscape on Windows 95



We graph similar data for Netscape browser within-brand shares on Windows 95 in Figure 3. Netscape 2 was the newest version when Windows 95 was released, so it has the high (nearly 1) share at the beginning of our sample period. Once again, the diffusion paths of the three later browsers (versions 3, 4, and 5) tend to flatten. The similarity between Figure 3, which is based on our UIUC data, and Nielsen's (1998) Figure 1, based on commercial websites, suggests we can compare our UIUC data set to the broader commercial market for browsers.

A more familiar fact about browsers in this era is that brand leadership shifted from Netscape to Microsoft. Market shares of IE versus Netscape across all OSs in our data show this shift (Figure 4). Our figure , based on the UIUC data, is similar to those made from more commercial web browsing sources, such as Henderson (2000). This suggests that our study of technically-oriented web users may be representative of the broader population.

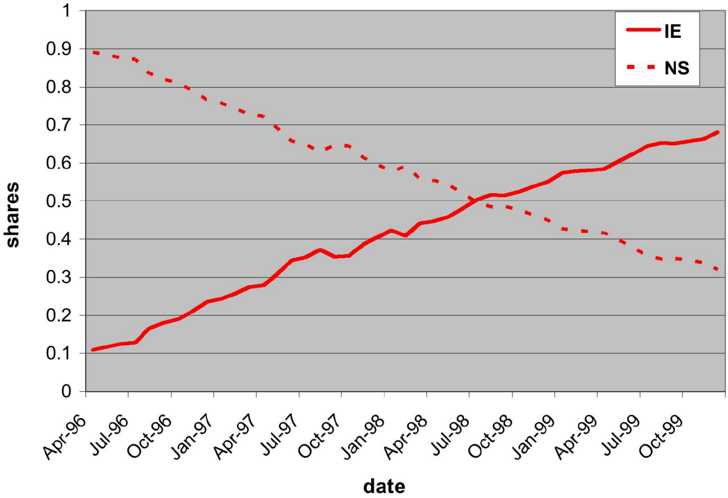
In his antitrust trial testimony, Richard Schmalensee (1999a) raised two objections to "hits" data like the data we use. He argued theoretically that positive feedback for a browser standard should be based on the number of users rather than the volume of usage. Since browser suppliers looked closely at hits data based on the observation that browser usage, not the number of users, is what matters to webmasters, the trial court rejected this theory, as do we. Schmalensee also raised the practical measurement problem that hits can be very hard to measure because of the practice of "caching" frequently viewed websites. It is unlikely that this is relevant to the UIUC websites, which do not host highly popular commercial websites but rather primarily the web pages of engineering students.

a) Regressors

Our distribution measures are based on the idea that convenience-oriented consumers will tend to use the software which came with their computer for a period of time. At a broad level, this amounts to looking at the shipments of new PCs as an explanation of browser utilization. By this means, we can distinguish the role of distribution convenience as distinct from adoption based only on technological progress in browsers. Consumers reacting to technical progress would respond to release of a browser version (with some time lag), whereas consumers reacting to distribution convenience will tend to respond to variation in the shipments browsers with PCs.

Conceptually, our distribution measures should be based on the browser which came with users' computers (or ISP accounts). We do not have the ideal data: direct measures of which browsers were distributed with the PCs of our users. We can, however, construct measures of the probability that any particular browser came with users' PCs by conditioning on the date at which we observe the browser being used and on the PCs' OS.

Figure 4 Market shares of Internet Explorer and Netscape on all OSs



We proceed in two steps. The first step defines measures under the (false) assumption that the newest versions of both brands of browser were shipped with all new PCs. The second step takes account of the history of restrictions on distribution.

In the first step, we measure the probability that a user of operating system o at time t got browser b with their computer as the probability that a computer running operating system o at time t was sold when browser b was the newest of its brand. This calculation does not use the UIUC data; it is based on the dates in Table 3 and on IDC and Microsoft data on the shipments of computers.[26]

Let N_{om} be new PC shipments in month m running operating system o. If the depreciation rate for computers is δ, the stock of computers running operating system o at time t is $\sum_{m=0}^{t} N_{om}(1-\delta)^{t-m}$. Denote the dates (in Table 3) when browser b was newest of

[26] IDC, a leading IT market research firm, does not separately report monthly shipments by OS. It reports monthly shipments of all PCs and annual totals by OS (IDC 2000a, 2000c, 2000d, 2000e, 2000h (additional data from 1996-1998 were used)). Fortunately, Microsoft internal documents detail the rate at which new versions of its OS replace old ones in the market place. For example, the Microsoft "OEM Sales FY '96 Midyear Review" gives the early history of Windows 95 vs. Windows 3.1 sales (Kempin 1998). This forms the basis for our allocation. We follow IDC by assuming 25% annual depreciation; lacking the retirements data they keep internally, we use a constant proportional depreciation assumption.

its brand on operating system o as the interval from m_1 to m_2. The probability that a computer running operating system o observed at time t was shipped when browser b was the newest of its brand is[27]

$$PCDISTR_{bot} \equiv \frac{\sum_{m=m_1}^{\max\{t,m_2\}} N_{om}(1-\delta)^{t-m}}{\sum_{m=0}^{t} N_{om}(1-\delta)^{t-m}}. \quad \text{Equation 3}$$

Here, the denominator is the stock of PCs running operating system o in use at time t. The numerator is the stock of PCs running operating system o in use at time t that were shipped when browser b was the newest of its brand.

For a given browser, $PCDISTR_{bot}$ varies both over time and across operating systems. A simple example will illustrate the definition and how it varies. The example, in Table 5, shows $PCDISTR_{bot}$ for IE version 1 and version 2 in late 1995 on Windows 95. As of November, IE1 was the only IE version which had been available for that OS, so it has $PCDISTR_{bot}$ of 1. In December, IE2 was introduced and, by the end of that month, 28% of Windows 95 computers ever distributed up to that time had been distributed in that month. Accordingly, $PCDISTR_{bot}$ for IE2 is 0.28 and $PCDISTR_{bot}$ falls for IE1.

Table 5 PCDISTR example

	$PCDISTR_{IE1\ Win95\ t}$	$PCDISTR_{IE2\ Win95\ t}$
November 1995	1	N.A.
December 1995	0.72	0.28

The time series variation in $PCDISTR_{bot}$ alone would not convincingly identify distribution convenience separately from "the shape of the S-curve."[28] Across operating

[27] for $t \geq m1$; $PCDISTR_{bot}$ = N.A. otherwise.

[28] Starting from the introduction date, $PCDISTR_{bot}$ first increases with t. After the browser version that succeeds "b" is introduced, $PCDISTR_{bot}$ begins to decrease. If the shipments of PCs were constant

systems, $PCDISTR_{bot}$ varies in a more promising way. A user running an older operating system is likely to have bought their PC earlier. Therefore, a newer browser version will have lower $PCDISTR_{bot}$ on an older operating system than on a newer one. Consider $PCDISTR_{IE4\ W95\ t}$ vs. $PCDISTR_{IE4\ W98\ t}$ in the months shortly after Windows 98 replaced Windows 95. At that point, IE4 had been the newest version of IE for several months. Among Windows 95 users, only those who recently bought a computer would have had IE4 come with it, but all users who had a Windows 98 computer would have had IE4 distributed with it. Thus, for months in autumn 1998, $PCDISTR_{IE4\ W98\ t} = 1$, but $PCDISTR_{IE4\ W95\ t}$ is near 0.43. In diffusion specifications which include INTR, the variation in $PCDISTR_{bot}$ conditional on the other regressors is largely cross-section variation.

In the second step, we deal with restrictions on distribution. Not all users who bought a new PC received both brands of browser. Instead, contracts between Microsoft and PC manufacturers sometimes required distribution and display of the latest version of IE. Also, there was pressure from Microsoft on PC manufacturers not to distribute or display Netscape browsers.[29]

A short history of the restrictions is in Appendix B – Restrictions on Distribution. Their dates are displayed in Table 11. That table shows the date at which "Must Carry" restrictions were in place requiring PC manufacturers to distribute and display IE. The variable MC_{om} is the fraction of PC manufacturers shipping operating system o in month m who were contractually required to carry IE. Table 11 also shows the dates of "Must Not Carry" provisions under which Microsoft pressured PC manufacturers not to distribute Netscape Navigator. The variable EX_{om} is the fraction of PC manufacturers shipping operating system o at time t who were not distributing Netscape's browser.

Combining the dates in Table 11 with information on the timing of PC shipments, we generate even sharper measures of the distribution of browsers with new computers.

---

over time (which they are not) this time series variation, together with time since introduction plus the logistic transform of the dependent variable, would determine the "shape of the S-curve."

[29] The contracts and the informal pressure were documented in the antitrust case, both in Microsoft internal documents and in its communications with outsiders such as OEMs. We base our regressors on these documents.

Descriptive statistics in Table 12 show the variation in these variables across OSs and over time.

The first variable measures the positive distribution advantage for Microsoft browsers from "Must Carry" provisions. We define $PCCARRY_{bot}$ as:

$$PCCARRY_{bot} \equiv \frac{\sum_{m=m_1}^{\max\{t,m_2\}} N_{om} MC_{om} (1-\delta)^{t-m}}{\sum_{m=0}^{t} N_{om} (1-\delta)^{t-m}} \quad \text{Equation 4}$$

$PCCARRY_{bot}$ is the probability that a computer running operating system o at time t (1) was shipped when browser b was the newest version of IE and (2) was shipped by a manufacturer subject to the "Must Carry" restrictions. It has the same definition as $PCDISTR_{bot}$ for IE except that we multiply the numerator by the "Must Carry" variable.

Since Microsoft always had "Must Carry" provisions in place for Windows 95 and Windows 98, $PCCARRY_{bot}$ is the same variable as $PCDISTR_{bot}$ for IE on those operating systems. These two variables differ on Windows 3.1, where $PCDISTR_{bot}$ is positive if small for some Internet Explorer browsers (which were new when late copies of Windows 3.1 were shipped) but $PCCARRY_{bot}$ is zero. They also differ on Macintosh, where restrictions were imposed only in the latter half of our sample.

In parallel, we define a variable, $PCEXCLU_{bot}$ for Netscape browsers based on EX_{om}. $PCEXCLU_{bot}$ is the probability that a computer running operating system o at time t was shipped (1) when browser b was the newest version of Netscape and (2) by a manufacturer that was pressured out of distributing Netscape browsers.

The cross-OS variation in $PCEXCLU_{bot}$ is broadly similar to that in $PCCARRY_{bot}$. Like $PCCARRY_{bot}$, $PCEXCLU_{bot}$ is distinct from $PCDISTR_{bot}$ for Netscape on the Macintosh OS, since the contract requiring Apple to make IE the default browser on the Macintosh took effect during our sample period. Like $PCCARRY_{bot}$, $PCEXCLU_{bot}$ is very different on Windows 3.1 than on other versions of Windows; it is always zero. The main difference between $PCARRY_{bot}$ and $PCEXLU_{bot}$ is that the "Must Not Carry" restrictions spread out over PC manufacturers over time, so that on Windows

95 and Windows 98, $PCEXCLU_{bot}$ is distinct from $PCDISTR_{bot}$. That said, most of the variation in $PCCARRY_{bot}$ and $PCEXCLU_{bot}$ in our data (conditional on other regressors) is the cross-OS variation.

There is controversy as to whether the distribution restrictions affected browser usage. The coefficients on $PCCARRY_{bot}$ and $PCEXCLU_{bot}$ permit us to test the hypothesis that the restrictions had no effect.

A second set of distribution variables are related to ISPs. ISPs, like PC manufacturers, saw a dramatic increase in their business in response to the widespread use of the Internet. ISPs distribute browsers and other network software to their customers. Just as some users may tend to use the software that came with their computers, they may also tend to use the network oriented software that came with their ISP subscription. The quantitative importance of this behavior will, of course, depend on the fraction of users who behave that way.

Accordingly, we define a distribution variable for ISPs, using ISP subscription data from IDC industry reports (IDC 2000b, 2000f, 2000g (additional data from 1996-1998 were used)). This variable, which we call $ISPDISTR_{bot}$ (distribution via ISPs) parallels $PCDISTR_{bot}$. It is defined by first calculating the stock of ISP subscribers and the flow of new subscribers. Then, for each browser on each OS at each time, we calculate the fraction of ISP users who were new subscribers when the browser was the newest of its brand available for that OS.

There is substantially less meaningful variation in $ISPDISTR_{bot}$ than in $PCDISTR_{bot}$, unfortunately. Both vary over time (when more computers are sold or more people get on the web). However, while $PCDISTR_{bot}$ also varies in cross section across operating systems, $ISPDISTR_{bot}$ varies only trivially across operating systems.

We also define a variable for Internet Explorer, $ISPTIED_{bot}$, based on the contracts between Microsoft and ISPs. These contracts required ISPs to distribute and display IE and not Netscape (for more details, see Appendix B – Restrictions on Distribution). The relationship between the $ISPTIED_{bot}$ variable and $ISPDISTR_{bot}$ is analogous to the relationship between $PCCARRY_{bot}$ and $PCDISTR_{bot}$: terms in the numerator are multiplied by the restriction variable.

A more detailed quantitative analysis of the ISP restrictions can be found in Fisher (1999). He compares ISPs bound by the restrictions to those not so bound and finds that the restrictions impact browser usage shares. Data limitations leave us unable to replicate his methods, because we cannot identify the (few) users who are using ISPs not bound by the restrictions. Another body of analyses of the ISP restrictions was carried out by Microsoft browser marketing personnel, who also concluded that the restrictions were effective.[30]

Another adjustment to the timing of diffusion is idiosyncratic to software. Often, software is "prereleased" in test versions before it is officially released to the market. We see these prerelease versions in our data. While the rate of prerelease usage of IE browsers is low, prerelease usage of Netscape browsers can be as high as 12 percent of all browser usage on an OS. If the user is running a prerelease version of an OS as well as a browser, the observation is not in our sample (as explained above). Otherwise, we add a regressor, $PREREL_{bot}$, to capture the effect of prerelease browsers, and change the definition of $INTR_{bot}$ to start from the prerelease date. The prerelease date is the date of the most significant "beta" test version of the browser. These changes leave our results largely the same when we include the $PREREL_{bot}$ dummy (which gets a substantial negative coefficient, as expected). As a result, we do not show these analyses in tables.

5) Estimates

We discuss our estimates in three steps. We first examine predictors of the rate of diffusion of new browser versions of each brand. We then examine the direction of technical change by looking at the determinants of browser brand share. In the subsequent section, we then use both those sets of estimates to analyze the quantitative importance of raw technological progress versus distribution and of the expanding field.

a) Within Brand Version Diffusion

We examine the diffusion of new versions of browsers within each brand. In this analysis, an observation is an OS and a month; e.g., browser users running Windows 95 in April, 1996.

[30] See U.S. Department of Justice et al. (1999) in the section entitled "Microsoft's internal analyses evidence the impact of its restrictions" for quotations from these analyses.

Many of the ordinary specification issues of discrete choice models apply here. For example, to avoid the situation where the share of the newest version is 1 by default, we start with the second browser version of the brand available on the OS.[31] Table 3 shows the dates for which each browser was newest on each OS.

All specifications include $PCDISTR_{b*ot}$, $INTR_{b*ot}$ and the interaction of $INTR_{b*ot}$ with the browser's version ($INTR\_V_{b*ot} = INTR_{b*ot} \times V_{b*}$, where V_{b*} takes the values 2, 3, 4, and 5 for each major browser version).[32]

Note that we include $INTR_{b*ot}$ and the interaction of $INTR_{b*ot}$ with V_{b*} but not V_{b*} itself. This is not an econometric oversight, but a specification decision. Including V_{b*} in Equation 1 would permit the long run penetration rate of newer versions of browsers to be lower or higher than older versions. While the long run penetration rate in a general diffusion study should vary (Griliches 1957) the diffusion of a new browser version should ultimately replace the previous version.

Observations for an OS are included only for the time periods shown in Table 4. We estimate by ordinary least squares after stacking the equations for each included browser on each OS. We have more observations ("N" in Table 7) for Netscape, since Netscape entered earlier and, especially early on, supplied more browser versions for more OSs. In addition to the coefficients of the estimating equation, Table 7 reports, in a line called Prob_Der_Mult, the multiplier that converts coefficients into probability derivatives at the mean values of the regressors (since the means of S_{b*ot}. are similar for the two brands of browser, the two values of Prob_Der_Mult are close as well). Since the units of the distribution variables are defined as shares, they have a quantitative interpretation in our regression. Consider a probability derivative like $\partial Pr(S_{b*ot|brand})/\partial PCDISTR_{b*ot.}$ That measures the tendency of users running a given brand of browser on

---

[31] On the Macintosh OS in the Netscape columns the dependent variable is never the share of NS1; it is first the share of NS2 in the months when it is the newest browser, then of NS3, and so on. But for IE on the Macintosh, we drop IE2; IE1 was never offered for that OS. Similarly, we drop NS1 in our Windows 3.1 observations in the Netscape analysis, and we drop IE4 in our Windows 98 observations in the Internet Explorer analysis, since that was the first available IE browser. Table 3 shows the dates for which each browser was newest on each OS.

[32] The use of cardinal values to represent different versions is a restrictive model of technical change. We also employ interaction terms between $INTR_{b*ot}$ and dummy variables for the browser versions. The estimated coefficients corroborate the magnitudes from employing $INTR\_V_{b*ot}$, so we employ $INTR\_V_{b*ot}$ in order to continue precise estimation of coefficients as we add more regressors.

operating system o to use the browser that was newest of its brand on their OS, thereby measuring the impact of PC manufacturer distribution.

Descriptive statistics of the dependent variables and regressors for the stacked samples may be found in Table 6. Estimates may be found in Table 7.

Table 6 Descriptive statistics for within-brand diffusion variables

		Mean	Std. Dev.	Min	Max
Netscape Browser Diffusion within Brand, n=131					
	S_{b*ot}	0.349	0.204	0.014	0.946
	$Ln[S_{b*ot}/(1- S_{b*ot})]$	-0.825	1.201	-4.225	2.866
	$PCDISTR_{b*ot}$	0.247	0.252	0.012	0.937
	$PCEXCLU_{b*ot}$	0.181	0.238	0	0.875
	$INTR_{b*ot}$	7.122	4.071	1	16
	$INTR\_V_{b*ot}$	29.069	19.711	3	75
	$ISPDISTR_{b*ot}$	0.158	0.121	0	0.423
Internet Explorer Browser Diffusion within Brand, n=109					
	S_{b*ot}	0.329	0.240	0.005	0.857
	$Ln[S_{b*ot}/(1- S_{b*ot})]$	-1.033	1.505	-5.323	1.794
	$PCDISTR_{b*ot}$	0.260	0.223	0.002	0.752
	$INTR_{b*ot}$	8.220	5.299	1	24
	$INTR\_V_{b*ot}$	30.780	21.342	3	96
	$ISPDISTR_{b*ot}$	0.217	0.140	0	0.571
	$ISPTIED_{b*ot}$	0.162	0.170	0	0.571

Table 7 Browser diffusion within brand

Dependent Variable: Logit of Share of Newest Version of Brand of Browser					
Regressor	Netscape			IE	
	(1)	(2)	(3)	(4)	(5)
C	-1.884 (0.174)	-1.932 (0.174)	-1.941 (0.197)	-3.125 (0.143)	-2.726 (0.128)
$PCDISTR_{b^{*}ot}$	2.216 (0.360)	4.648 (1.239)	2.195 (0.363)	4.104 (0.350)	3.570 (0.304)
$PCEXCLU_{b^{*}ot}$		-3.014 (1.472)			
$INTR_{b^{*}ot}$	0.488 (0.063)	0.370 (0.084)	0.550 (0.119)	0.209 (0.051)	0.529 (0.059)
$INTR\_V_{b^{*}ot}$	-0.102 (0.013)	-0.073 (0.019)	-0.109 (0.017)	-0.023 (0.013)	-0.074 (0.014)
$ISPDISTR_{b^{*}ot}$			-1.153 (1.885)		-9.850 (1.320)
$ISPTIED_{b^{*}ot}$					5.192 (0.902)
Prob_Der_Mult	0.21	0.21	0.21	0.19	0.19
N	131	131	131	109	109
R^2	0.42	0.44	0.42	0.75	0.84
Standard errors in parentheses					

The base model is reported in columns (1) and (4) of Table 7 for Netscape and IE respectively. The pattern of the estimates is largely the same. In each, we have a large, positive, precisely estimated coefficient on $PCDISTR_{b^{*}ot}$. This means that a higher rate of shipments of new computers increases the rate of diffusion of new browser versions (here, "new computers" refers to the ones that were shipped since the latest browser version was released).

The coefficients on $PCDISTR_{b^{*}ot}$ are economically large. At the mean of all the regressors, consider what the model predicts if we increase by 10% the fraction of users who got their PC since the newest version of their brand of browser was released. For IE, the model predicts an 8% increase in use of the newest version.[33] For Netscape, the figure is 5%. These results show that there is a strong tendency of users to run the version of their brand of browser which came with their PC, particularly for IE users. That also means that the pace of sales of new PCs strongly influences the diffusion of the latest browser version.

Table 7 columns (1) and (4) also have implications for technical progress in browsers. Each specification includes $INTR_{b*ot}$ and $INTR\_V_{b*ot}$. The interpretation in these two columns of $INTR_{b*ot}$ is that it measures all causes of the pace of diffusion other than distribution with new PCs. The positive coefficient of $INTR_{b*ot}$ and the negative coefficient of $INTR\_V_{b*ot}$ show that the pace of diffusion of new browser versions, holding fixed the stock of PCs, is declining. For both brands, the decline is rapid.

We interpret the decline in the pace of diffusion as measuring a deceleration in technical progress (since distribution has been held constant). It is a broad measure of technical progress, including improvements in websites as well as improvements in browsers.

In percentage terms, the decline is somewhat faster for Netscape in column (1) than for IE in column (4), a pattern that persists across the other specifications in Table 7. This table, which is entirely about the pace of diffusion of browser versions within brand, thus confirms once again that during our sample period, imitator Microsoft was catching up to innovator Netscape. IE had more rapid technical progress than Netscape.

In addition to our base specification, we report two other specifications for Netscape and one more for IE in Table 7. Column (2) adds the PC manufacturer exclusion restrictions measured by $PCEXCLU_{b*ot}$. This permits us to sharpen our test of the distribution hypothesis by examining what happened when Microsoft imposed "must not carry" Netscape restrictions on PC manufacturers. These effects are estimated in a specification which permits a declining rate of technical progress in Netscape browsers, so they do not confuse distribution restrictions with technical progress.

---

[33] We use the probability derivative multipliers and make comparisons at the means of the data in column (1) and column (4). An increase of 10% in $PCDISTR_{b*ot}$ for the newest version of Internet Explorer leads to an increase in that newest version's usage share in all Internet Explorer browsers of 7.8.% (the probability derivative is 0.19×4.104). For Netscape, an increase of 10% in $PCDISTR_{b*ot}$ leads to a smaller increase in the newest version's share, 4.7% (0.21×2.216).

Column (2) differs from (1) in two regards. First, a large, negative, significant coefficient on $PCEXCLU_{b*ot}$ shows that the "Must Not Carry" restrictions substantially reduced distribution of Netscape browsers. Second, the coefficient on $PCDISTR_{b*ot}$ is much larger in the Netscape estimates in column (2) than in (1). In column (1), this coefficient is the average effect of sales of new computers, averaged over times with and without Microsoft-imposed restrictions on distribution. In column (2), this coefficient measures the effect of sales of new computers without the contractual restrictions. Economically, the difference means that PC distribution was an important channel for Netscape users and that the rate of diffusion of the newest Netscape browsers would have been much greater if not for the exclusion restrictions.

There is no column like (2) reported for Internet Explorer browsers in Table 7; if there were, it would have $PCDISTR_{b*ot}$ and $PCCARRY_{b*ot}$. As we noted earlier, there is not enough independent variation in $PCCARRY_{b*ot}$ conditional on $PCDISTR_{b*ot}$, $INTR_{b*ot}$, and $INTR\_V_{b*ot}$ to estimate such a specification.

In columns (3) and (5) we add measures of ISP distribution. The ISP measures vary only across time, so they have much less meaningful variation in them. When we include a single $ISPDISTR_{b*ot}$ variable for Netscape in column (3), we get a negative coefficient, but not one that we can estimate precisely. The absence of $ISPTIED_{b*ot}$ as a regressor may lead the coefficient on $ISPDISTR_{b*ot}$ to reflect the exclusion of Netscape distribution through ISPs. However, adding $ISPTIED_{b*ot}$ puts too many regressors that vary together into that model to estimate any ISP effect precisely.

In column (5) we add both $ISPDISTR_{b*ot}$ and $ISPTIED_{b*ot}$ to an IE specification. The large positive coefficient on $ISPTIED_{b*ot}$ implies that Microsoft's distribution restrictions on ISPs increased the diffusion rate for new versions of Internet Explorer. The IE marketing manager who obtained the distribution restrictions in order to avoid side by side product comparisons with Netscape browsers (Myhrvold 1999) appears to have been correct in his assessment. Absent the restrictions, an increase in ISP subscriptions leads to a decrease in the rate of IE version diffusion (-9.85+5.192) though the sum is not all that precisely estimated. This suggests either (a) that Internet-oriented PC users dislike Microsoft technologies, or (b) that the variation in the ISP variables may be picking up other effects which are moving around over time. There is considerable reason to believe both of these stories and not enough information in the data to distinguish them.

We can precisely estimate the effects of distribution versus technical progress on diffusion of new versions of browsers when we have variation in both time series and across OSs, such as the PC manufacturer distribution variables. In other parts of our analysis, we have lower-quality variation in the regressors (as in the ISP variables which vary only over time), and our ability to measure their effects in these regressions is less. Nonetheless, the well-estimated coefficients of PC distribution and technical progress tell a clear story in which distribution is an important force.

b) Brand Shares: Direction of Diffusion

In Table 9, we examine the other demand dimension, browser brand shares. In all the columns of Table 9, the dependent variable is the share of IE browser usage in each month, and the denominator is Netscape plus IE browser usage. Descripive statistics can be found in Table 8.

In columns (1) - (3), labeled "aggregate," an observation is a month and the dependent variable is the aggregate share of IE browser usage by users running any of our four mass-market OSs, $S_{IE,t}$. In columns (4) - (7), labeled "each OS," an observation is an OS / month, and the dependent variable is $S_{IE,ot}$, the share of IE on that OS. The number of observations increases by this disaggregation, but not by 4 times many observations because we do not observe all the OSs in all the months (see Table 4). Our sample also varies by whether we include or exclude the time period in which version 5 of IE is newest (long sample).[34]

A literature (cited above) has already taken up the analysis of browser brand shares. We start with an analysis like the one presented by Schmalensee (1999a and 1999b) and Liebowitz and Margolis (1999). In column (1) we follow that earlier work as closely as possible, using the same time period and and copying their use of a single, predictor of Internet Explorer's share, $Rel\_Qual\_Jrnl_t$.

[34] In columns (4) - (7), estimation is by stacked OLS, whereas in columns (1)—(3), it is by OLS. In the short sample, we use a regressor based on Schmalensee (1999b). In the long sample, we extend his definition to version 5 browsers.

Table 8 Descriptive statistics for brand share variables

Variable				
Sample				
	Mean	Std. Dev.	Min	Max
Aggregate (n=45)				
$S_{IE,t}$	0.41	0.17	0.11	0.68
$Rel\_Qual\_Jrnl_t$	0.41	0.59	-1	1
D_3 Dummy for version 3 newest(1)	0.31	0.47	0	1
D_4 Dummy for version 4 newest	0.38	0.49	0	1
D_5 Dummy for version 5 newest	0.22	0.42	0	1
Each OS (n=135)				
$S_{IE,ot}$	0.306	0.264	0.002	0.941
$PCCARRY_{b*ot}$	0.251	0.228	0	1
(1) Our samples begin when version 2 browsers are newest, so that is the omitted category in the version dummies.				

Columns (2) and (3) continue to restrict the predictors of brand shares to relative quality measures, but use a larger sample size that includes version 5 browsers. Here we use two measures of relative quality. In column (2) we extend $Rel\_Qual\_Jrnl_t$ to the era of IE5. In column (3), we include browser version dummies, a less restrictive treatment of relative quality.

In columns (4) and (6), we continue to follow the specification of the earlier literature, using only relative browser quality as a predictor of brand shares. Here, however, we use observations for each OS.

These specifications replicate the findings emphasized by Schmalensee (1999a and 1999b) and Liebowitz and Margolis (1999). First, increases in relative browser quality, measured by $Rel\_Qual\_Jrnl_t$, predict increases in browser brand market shares. In these specifications, the coefficient is large and precisely estimated. Second, the predictive power of the model, measured by R^2, is high, at least in the aggregate model.

Table 9 Browser brand shares

Dependent Variable: Share of Internet Explorer in IE + Netscape Browser Usage Models Differ by Definition of an Observation							
	IE Share			IE Share			
	Aggregate			Each OS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.299	0.306	0.119	0.226	0.088	0.238	0.111
	(0.013)	(.014)	(0.030)	(0.020)	(0.011)	(0.020)	(0.016)
$Rel\_Qual\_Jrnl_t$	0.224	0.266		0.188	0.108	0.239	0.171
	(0.021)	(0.019)		(0.031)	(0.014)	(0.028)	(0.018)
Version 3 Newest Dummy			0.149				
			(0.034)				
Version 4 Newest Dummy			0.357				
			(0.033)				
Version 5 Newest Dummy			0.514				
			(0.035)				
$PCCARRY_{b*ot}$					0.623		0.584
					(0.029)		(0.042)
N	35	45	45	105	105	135	135
R^2	0.78	0.82	0.89	0.26	0.87	0.36	0.74

Liebowitz and Margolis (1999) interpret their findings, which closely parallel those in columns (1)-(4) and (6) of Table 9, as evidence that "Good products win."[35] They reject theories in which there are causes of brand leadership in software other than product quality, characterizing them as speculation. Their conclusion, they assert, comes from using superior methods to earlier scholars, particularly the "first systematic examination of real-world data from software applications markets. … [their] most important finding is the close relationship between market share change and product quality."[36] Schmalensee (1999a) agrees, writing "Netscape obtained fewer new users of Web-browsing software than Microsoft because its product did not keep pace with improvements in Microsoft's IE, and because it made numerous business and technical mistakes."[37] Schmalensee also contrasts his use of market outcome statistics to the government's reliance on internal Microsoft correspondence and documents in the

[35] Liebowitz & Margolis (1999), p. 135. They make similar analyses of several applications software markets.
[36] Liebowitz & Margolis (1999), p. 227.
[37] Schmalensee (1999b), slide C.

antitrust case. He rejects the government's theory that distribution matters in favor of the quality theory on this basis.

The earlier literature just cited and the specifications just examined all have in common that relative quality is the only cause of brand market shares explicitly considered in the data analysis. However, since both quality and distribution are potential causes of brand shares, the appropriate empirical approach is to include regressors associated with both causes.

We can undertake this approach in our each-OS sample. In columns (5) and (7) of Table 9, we continue to use the Schmalensee measure of relative quality and also include $PCCARRY_{b*ot}$, the Internet Explorer distribution advantage variable. These columns are otherwise identical to columns (4) and (6), respectively.

For identification of both a quality and a distribution effect, it is essential that we have cross section variation across the OSs as well as time series variation. $PCCARRY_{b*ot}$ varies across versions of Windows, and it varies over time on the Macintosh. As a result, in the each-OS sample, there is substantial independent variation in the distribution and quality measures. In column (7), for example, the correlation between $PCCARRY_{b*ot}$ and $Rel\_Qual\_Jrnl_t$ is 0.27. If we did not have cross-OS variation, however, we would not be able to pursue the analysis.

The result in both columns (5) and (7) is that the distribution advantage variable has a positive and precisely estimated coefficient. The estimates are economically significant. Increasing the percentage of OS users who obtained a computer on which Microsoft compelled the distribution and display of IE implies that the percentage use of IE rises by about 60% as much (0.62 or 0.58, depending on specification).

While quality continues to matter, it quantitatively matters less once we account for distribution. That is, specifications which look only at quality exaggerate its importance. Both quality and the distribution advantage for IE are trended upward, so the specification with only one regressor has omitted variable bias. Schmalensee (1999a and 1999b) and Liebowitz and Margolis (1999) examine only quality as a cause of market shares.[38] When we include a measure of distribution as well as their measure of quality, their restricted specification and its conclusion are rejected.

Our conclusion from the brand share analysis is the same as from the diffusion of new versions of the same brand. We find an important role for both technical progress and distribution. The two analyses are based on different phenomena. One comes from examining the diffusion of new versions of both brands of browsers, and the other comes from examining the brand shifts that occurred over time on different OSs. In each case, there is independent variation in the distribution variables. Our results show a common explanation. Users of computers respond to economic forces such as distribution convenience as well as to technical progress.

6) Quantitative Implications of the Estimates

Both the diffusion-within-brand results and the brand share results have estimates of the impact of distribution and of technical progress. We now examine the quantitative implications of the estimates for the relative importance of those two causes.

The first sense of the "relative importance" of technical progress is

$$\frac{\partial Y / \partial T}{\partial Y / \partial \text{dist}} \frac{\Delta T}{\Delta \text{dist}} \qquad \text{Equation 5}$$

where Y is a demand behavior – either a brand share or a rate of adoption of new browsers, dist is a distribution measure, T is a technical progress measure, and Δdist and ΔT are comparable changes in distribution and in technical progress. For any given scenario, defined by Δdist and ΔT, this sense of "relative importance" asks whether a hypothetical change in distribution of size Δdist would have more or less impact than a hypothetical change in technical progress of size ΔT.

The second sense of "relative importance" is more historically grounded. Suppose that demand behavior is different at different times, as in the slowing rate of browser diffusion within brand over time or the rising brand share for IE over time. We can ask how distribution and technical progress changed between those times, and what contribution to the brand share and the pace of new version diffusion each makes.

[38] Liebowitz and Margolis (1999) also examine evidence they say relates to tipping, but not in the context of a model that also includes quality measures.

a) Brand Share

We begin with the brand market share estimates in Table 9. Here our definition of Y in Equation 5 is the share of Internet Explorer, S_{IE}.

For Δdist, we consider the counterfactual scenario in which IE had no distribution advantage. For ΔT, we consider the counterfactual scenario in which IE did not catch up to Netscape in quality at all. To estimate ΔT we note that historically, the value of $Rel\_Qual\_Jrnl_t$ grew from -1 (all journalists recommended Netscape) to 1 (all Microsoft). So we use ΔT = 2, a change in $Rel\_Qual\_Jrnl_t$ of 2. To estimate Δdist, note that value of $PCCARRY_{b*ot}$ varied from 0 to 1, so we use Δdist = 1. These values for Δdist and ΔT represent large changes in the regressors (though these changes are still within our sample), so one should interpret the prediction with care.

Using estimates of the slopes from column (7) of Table 9, we get an estimate of the relative importance of technical progress of just under 0.6 (= 0.171/0.584 × 2/1). Technical progress is quantitatively less important as an explanation of brand shares than is distribution.[39]

We now turn to the second sense of relative importance. How much did brand shares shift over time as a result of distribution and of technical progress? We have already picked appropriate values of Δdist and ΔT to address this directly. However, quantitative interpretation of the numerator or the denominator of Equation 5 in this instance must be made carefully. To interpret them as the predicted market share change from a change in technological progress or in distribution is not obvious. Network effects mean that browser markets tip, so the predicted market share in a counterfactual experiment should be be near either 0 or 1, given the size of Δdist and ΔT.

The ratio of numerator to denominator has an interpretation whether or not there are network effects. It can be interpreted as showing the relative size of the impact of the historical change in IE product quality (technical progress) vs. the historical change in distribution advantages for IE. Our results imply a larger impact from the historical

[39] If we used column (5) of Table 9 instead, we would get a smaller estimate of the relative importance of technical progress. Both specifications lead to a much larger role for distribution than for technical progress.

distribution advantage than from the historical change in product quality – technical progress had about 0.6 as large an impact.

b) Within Brand Version Diffusion

We now turn to the relative importance of technology and distribution for the pace of diffusion within brand.

For this analysis, the relevant concepts of Y, ΔT, and Δdist are all time derivatives, so we label them with a dot above them. Our definition of $\dot{Y}$ is the rate of growth of the share of the newest brand, $\dot{S}_{b*}$. Our definition of $\Delta\dot{T}$ is the difference in rates of progress between adjacent versions of the browser. That corresponds to a change in V of -1 (negative because we are comparing the higher rate of technical progress earlier to the lower rate of technical progress later.) For the numerator of Equation 5 we note that (β_{INTR} +$\beta_{INTR\_V} \times$ V) is the model's prediction of $\dot{S}_{b*}$. Accordingly, the numerator is Prob_Der_Mult × $\beta_{INTR\_V} \times$ (-1).

Our conceptual definition of $\Delta \dot{dist}$ is the monthly rate of growth of the stock of PCs in use, and we use the mean monthly rate of growth of $PCDISTR_{b*ot}$ over all OSs and both brands in our sample; implicitly, we are comparing historical distribution to a counterfactual world of no distribution of browsers with new computers. In our sample, $PCDISTR_{b*ot}$ is growing at 3.2% per month. The denominator of Equation 5 is Prob_Der_Mult × $\beta_{PCDISTR} \times$ 0.032.

We are now ready to calculate Equation 5 for Internet Explorer based on the estimates in column (4) of Table 7. Noting that the probability derivatives cancel, this is $\beta_{INTR\_V} / \beta_{PCDISTR} \times -1/0.032 = -0.023/4.104 \times -1/0.032 = 0.18$. Quantitatively, the amount of technical progress represented by this measure of $\Delta\dot{T}$ is less important than distribution measured by $\Delta \dot{dist}$.

Making a parallel calculation for Netscape browsers based on $\beta_{PCDISTR}$ in column (1) of Table 7 does not make economic sense. $PCDISTR_{b*ot}$ in column (1) measures both the effect of more rapid distribution and the effect of limitations on distribution. A better set of estimates come from column (2) of Table 7, in which distribution is measured by $PCDISTR_{b*ot}$ and restrictions on distribution are measured by $PXECLU_{b*ot}$.

The first Netscape calculation is based on the coefficients of $PCDISTR_{b*ot}$ and $INTR\_V_{b*ot}$ in column (2). This corresponds to the question, how important is distribution vs. technical progress for Netscape browsers when there are no restrictions on distribution of Netscape? This is a within-sample calculation, since there were times and OSs (such as on the early days of the Macintosh) in which there were no restrictions on Netscape browser distribution.[40] Again using the same values for $\Delta \dot{dist}$ and $\Delta \dot{T}$ and again letting the probability derivative terms cancel, this is -0.073/4.648 × -1/0.032 = 0.49. Technical progress represented by this measure of $\Delta \dot{T}$ is once again quantitatively the less important force.

There is a clear difference between the IE and Netscape results; the impact of technical progress is closer to that of distribution for Netscape (0.49 vs. 0.18). Given the underlying economic and technological situation, this result is not surprising. First, our measure of $\partial \dot{Y} / \partial \dot{T}$ should be larger for Netscape. While both brands of browser show slowing technical progress across versions, in this era IE is catching up technically to Netscape, so the rate of decline of technical change for Netscape ($-\partial \dot{Y} / \partial \dot{T}$) should be larger. Second, distribution should be more important for IE, since contractual restrictions requiring bundling of IE make distribution comparatively important for that brand. Both real-world differences go in favor of a larger impact of technical change relative to distribution for Netscape, which is what we find.

A second calculation for Netscape contrasts the effects of restrictions on distribution to technical progress, based on the coefficients of $PCEXCLU_{b*ot}$ and $INTR\_V_{b*ot}$. This answers the economic question, which was a quantitatively more important predictor of the pace of adoption of Netscape browsers, technical progress or the restrictions on distribution imposed by Microsoft? Looking across the four eras in which the four major versions of Netscape browser were newest, we see monthly rates of growth of $PCEXCLU_{b*ot}$ (averaged across OSs) of 0.0%, 0.0%, 3.6%, and 3.3%. The obvious $\Delta \dot{dist}$ to use contrasts the early and late periods; we use $\Delta \dot{dist}$ = -0.035 (negative to have it in units distribution rather than blockage of distribution).

[40] Though there are requirements to distribute IE which may slow the distribution of Netscape. This will tend to bias downward this estimate relative to the concept we are attempting to quantify.

Accordingly, the value of Equation 5 is $-0.073/-3.014 \times -1/-0.035 = 0.69$. In this sense as well, the impact of technical progress is less than the impact of distribution (restrictions).

We now turn to the second sense of "relative importance," and ask which forces are historically important in explaining the slowing pace of diffusion over time. For Internet Explorer, the slowing pace of diffusion is shown in Figure 2. We use estimates from Table 7, column (4). Here the explanation of the change over historical time is simple. The pace of technical progress is slowing for IE, $\beta_{INTR\_V} < 0$. That slowing is not offset by the positive impact of more rapid distribution over time.[41]

The slowing pace of diffusion of new Netscape browsers (shown in Figure 1) can also be explained using estimates from Table 7, particularly column (2). Here, too, one cause is slowing technical progress, $\beta_{INTR\_V} < 0$. For Netscape browsers, two distribution coefficients matter. The coefficient of $PCDISTR_{b*ot}$ is 4.6; that of $PCEXCLU_{b*ot}$ is -3.0. The monthly rates of growth of $PCDISTR_{b*ot}$ during the era when our four major versions were the newest of their brand (averaged across OSs) were 4.4%, 3.7%, 3.0%, and 2.8%, respectively.[42] The monthly rates of growth of $PCEXCLU_{b*ot}$ during those three eras were 0.0%, 0.0%, 3.6%, and 3.3%. The trend in $PCDISTR_{b*ot}$ over time is slightly down, and the trend in $PCEXCLU_{b*ot}$ is upward. Accordingly, neither distribution coefficient works to offset the slowing pace of technical progress. Instead, the contribution of distribution forces is to further slow of the diffusion of Netscape browsers. Given the sizes of the coefficients and the trends in the Xs, more of this comes from the trend toward exclusion of Netscape browsers from distribution with new PCs than from slowing $PCDISTR_{b*ot}$ growth.

The last two subsections have examined the role of technical progress and of distribution from the perspective of our two models, and have found broadly similar conclusions: both forces matter, and the distributional forces quantitatively matter somewhat more.

[41] Historically, during the eras when IE2, IE3, IE4, and IE5 were the newest of their brand, the rates of change of PCDISTR (averaged across OSs) were 0.8%, 3.2%, 2.1% and 4.4%, respectively. That is an average rate of growth between versions of 0.8%, or 0.008. That implies a change in the rate of growth of the newest version of 0.00064 / month, so the upward trend in PCDISTR is too small to offset slowing technical progress.

[42] The trends in PCDISTR are different for the two brands because of weighting. IE in the later period has a bigger weight on Windows 98, which is rapidly growing.

c) Effect of PC Demand Growth on Browser Diffusion

In this subsection, we view these same quantitative findings from a different perspective. Since distribution with new PCs is quantitatively important, what role did the rapidly expanding field of diffusion for PCs play in the rate and direction of browser adoption? This analysis suggests aggregate implications of our estimates for economic growth and transformation.

The late 1990s saw a rapid expansion in the demand for information technology, brought about by the conversion of the Internet into a mass market technology for commercial applications. The social returns to that technology depended on the widespread diffusion of browsers suitable for mass market commercial applications. The private returns to that technology depended, in part, on which brand of browser would be in widespread use. Both the pace of diffusion of browsers and the direction in terms of brands were affected by the rapidly expanding field of diffusion, i.e., the growing installed base of PCs.

PC industry growth is typically rapid; it was extraordinary in the period we study. From January, 1995, to December, 1999, the installed base of personal computers doubled from 106 million to 213 million computers. How did that rapid growth affect the pace and direction of browser diffusion?

We compare that rapid growth to a replacement-demand scenario. We assume that one fifth of PCs go out of use each year, so that the gross flows of new PCs exceeds the net flow by enough to replace that.[43] Further, we assume that $PCDISTR_{b*ot}$, $PCCARRY_{b*ot}$, and $PCEXCLU_{b*ot}$ are all proportional to the rate of market growth. Finally, we assume that there is a new version of each brand of browser each year. We use derivatives from our browser diffusion model (Columns (2) and (4) of Table 7) at the mean of the data and fix technical progress at the level of a version 3 browser.

In the scenario of rapid growth found in our data during the sample period, the annual net rate of growth of PCs is 19% so the gross rate of new PCs is 39% (19% + 20% replacement demand). In the slower-growth scenario, the annual gross rate is 20%. These correspond to monthly growth rates of 2.8% and 1.5%, respectively.

---

[43] This is a steady-state assumption. In fact, during the historical era, retirements of PCs were less than a fifth of the stock because the stock consisted disproportionately of newer PCs.

First, as we can see by comparing the first two columns of Table 10, the impact of the more rapid growth of PC demand on Netscape version diffusion is small, raising the monthly diffusion rate from just under 4% (0.03 technology effect + 0.005 total distribution effect) to just over 4% (0.03 technology effect + 0.01 total distribution effect).

Table 10 The expanding field and browser diffusion

Effect on rate of diffusion of newest version	**Netscape**			**IE**	
	(1)	(2)	(3)	(4)	(5)
PC demand growth*	**Fast**	**Slow**	**Fast, no Exclusion**	**Fast**	**Slow**
Technology (INTR+INTR_V×3)	0.15	0.15	0.15	0.14	0.14
(xProb_Der_Mult)	**0.032**	**0.032**	**0.032**	**0.027**	**0.027**
Expanding Field (PCDISTR×PC growth)	0.13	0.07	0.13	0.11	0.06
Exclusion (PCEXCLU×PC growth)	-0.084	-0.045	0	0	0
Total Distribution ((Expanding Field + Exclusion) × Prob_Der_Mult)	**0.010**	**0.005**	**0.027**	**0.021**	**0.011**

* "Fast" = historical growth, "Slow"= replacement. Prob_der_mult = 0.21 for columns (1)-(3) and 0.19 for columns (4)-(5).

In contrast, the same change (shown in the columns (4) and (5)) increases the Internet Explorer monthly diffusion rate by a percentage point, from 4% (0.03 technology effect + 0.01 total distribution effect) to 5% (0.03 technology effect + 0.02 total distribution effect). It is easy to understand the large difference between these results. Given the coefficient of $PCDISTR_{b*ot}$, an increase in the pace of new PC demand leads to an increase in the pace of distribution of the newest version of IE that is 83% larger. For Netscape, an increase in $PCDISTR_{b*ot}$ together with a proportionate increase in $PCEXCLU_{b*ot}$ leads to a much smaller effect.

A related result can be seen in the contrast between columns (1) and (3). Removing the impact of the exclusion restrictions substantially increases the pace of diffusion of Netscape; the distribution effect rises from .01 to .03. Again, the intuition is simple. Exclusion substantially slowed the pace of diffusion of Netscape browsers in the historical world, despite the rapid pace of sales of new PCs.

The implicit calculation of the impact of the rapid growth of PC demand on brand shares compares the first (Netscape) to fourth (IE) columns. Given the high rates of new PC demand and the distribution advantages for IE, the stock of new browsers would soon become overwhelmingly IE. With slower growth (i.e., compare column 2 to the last column) the growth rates of the two brands would be much more similar.[44]

Even for a rapidly progressing technology like the browser at the height of the competitive "browser war," the role of the expanding diffusion field in overcoming the fixed costs of adoption was very important. Even with the restrictions on the distribution of one brand of browser – restrictions which had a quantitatively significant effect in the decline of Netscape– the expansion of the field was able to contribute to the widespread distribution of a new technology. Without the restrictions on OEM behavior imposed by Microsoft, the expanding field would have been even more powerful.

d) Implications

One quantitative conclusion is common to the brand share and newest-version diffusion estimates: both technical progress (in the broad sense) and distribution were important drivers of diffusion. The comparability of the economic impacts as measured in the brand share and diffusion models is not an artifact of our empirical specification. Those estimates use very different information in the data, suggesting that the comparatively large role for distribution is a robust finding.

The substantial role for distribution confirms what browser marketing executives in both Netscape and Microsoft learned from other kinds of quantitative evidence, such as surveys, and from their business experience. They viewed distribution as very important. (See Fisher (1999) and Bresnahan (2002) for quotations from internal documents.)

7) Our Economic Interpretation (and Its Limits)

The costs of adoption can significantly impact the diffusion of new technologies. In that case, users' pace of adoption will be open to influence by distribution and

---

[44] This calculation is not the most complex one that could be made. Inside Microsoft, Haas (1998) made a somewhat more complex brand growth calculation leading the same conclusion, that distribution with new PCs would push IE into the lead.

marketing as well as by the attractiveness of new technologies. In the particular technology we study (browsers), the distribution forces had a large impact on the rate and direction of technical progress.

That conclusion is unlikely to be a statistical artifact. If anything, our estimates systematically overstate the importance of technical progress, counting the attractiveness of the entire Internet as part of the technical attractiveness of browsers.

Our conclusion is particularly relevant for mass market technologies and platform technologies where the benefits of newer versions of the technology do not always accrue directly to the individual user. In these industries, individual user adoption costs play a large role in the pace of adoption.

As time passes and the speed of connection to networks improve, one source of individual user's fixed costs of new software adoption falls: the time costs of downloading a new browser.[45] There are other sources of fixed costs of adoption, such as the costs of learning about new versions, their features, their problems (e.g., security) if any, and so on. Our estimates do not permit us to separate the sources of fixed costs. It is not clear from our estimates, therefore, whether we would get the same relative importance of distribution and technological progress later on in this same technology. Further work may shed light on this issue.

The pace of diffusion of narrowly "technical" technologies that do not have a substantial user cost of learning, adapting, adjusting to, or obtaining technologies from the frontier, will be less influenced by distribution. Such "technical" technologies, however, are not the ones for which information technology typically generates new applications with mass market commercial value.

8) Conclusion

Both technical and economic forces affect the diffusion of a new technology. We study these forces in the diffusion of new and improved versions of commercial web browsers in the late 1990s. By exploiting data on browser usage and documented changes in technical progress (browser versions) and distribution methods, we quantify

[45] Over our sample period, browsers grew larger while online connections grew faster apace, so that the time cost for an average user was not strongly trended.

the significance of these two types of forces. We find that browser distribution via new personal computers (PCs) had a larger effect on the rate and direction of technical change than technical browser improvements.

That is a controversial finding in one of our analyses, of brand choice, and we have been careful to explain why we get a different answer from some earlier data analyses. Our main change is that we include measures of both technical change and distribution in the model. Since we include both causes, our conclusion that one cause mattered more than the other is an advance over the earlier analyses.

We also find the larger role for distribution in the diffusion of new and improved (suitable for complex commercial applications) browser versions. Finding the same forces at work in diffusion and in brand choice reinforces our view of the underlying forces. Enough browser users value distribution convenience that distribution could impact the rate and direction of technical progress.

The role of distribution in overcoming the transactions costs of adoption was magnified, in the case of browsers, by the increasing demand for PCs. While that sounds very narrow and specific, it is in fact an important general economic conclusion.

In complex systems like the Internet, the invention of a new general purpose technology will typically spur growth in complements. Rapid growth in complements – in this case rapid growth in the demand for PCs – in turn feeds back to growth in the new invention – in this case the browser. One mechanism for this feedback is a rapidly expanding diffusion field. Inertial forces like the costs of switching to and adopting the latest technology are less important when the diffusion field is rapidly expanding. This mechanism is particularly important in systems industries with general purpose technologies, like the PC and the Internet.

The economic boom of the late 1990s involved both real exploitation of positive feedback effects and a speculative bubble. The debate over the boom has focused on efforts to measure the productivity impact of investments in new information technology capital and on the role of a stock-market bubble in encouraging over-investment. Our results, though limited in scope to particular markets, point to another set of economic forces at work. One component of the investment boom in IT capital in the late 1990s was that the IT-using side of the economy was moving closer to an expanding technological frontier. This was due in part to positive feedback dynamics as analyzed in this paper, between new technologies like the browser and existing technologies like the PC. Demand for new technologies increased not only because they became more attractive but also because their adoption costs fell. Examination of the details of choices of particular technologies provides a microfoundation for understanding the overall economics of technical advance.

9) Appendix A – UIUC Data Details

When a computer accesses a page on a web server, it communicates to the server what browser and OS it is running. The server administrator can set the server to record that information in a log (see http://www.ews.uiuc.edu/bstats/months/). An example of the log results from September, 1996, is as follows:

Browser Versions - Top 40		
Specific Browser Version	Hosts	%
Mozilla/3.0 (Win95; I)	21598	8.4
Mozilla/2.0 (compatible; MSIE 3.0; Windows 95)	19515	7.6
Mozilla/2.02 (Win16; I)	10580	4.1
Mozilla/1.22 (compatible; MSIE 2.0; Windows 95)	9816	3.8

Web server logs record the Internet Protocol (IP) address of the computer accessing them. As a result, we use a unique IP address (or "host" in the odd lingo of server logs) as our definition of a single user. This will count multiple people who sit at the same computer as one user; the most important example of this is UNIX machines in university computer centers. This will count a user who gets a new IP address for each session as a single user and those who access the UIUC servers more than once in a month as multiple users; the most important example of this is dial-in users using ISPs. Users in offices, students in university dorms, and ISP subscribers who browse UIUC once a month will be counted correctly. In our first month of data, there were 35,757 unique hosts. This rose to 229,579 in the last month of our data.

When a computer accesses a web server, it passes a field called the "user agent" to the server. This field permits us to identify the browser and OS used by the accessing computer. We use the portion of the web log archive called "Browser Versions - All" which reports the number of "hosts" for each distinct user agent field. For each unique IP address (host) the archives record the last browser and OS used in that month.

There are thousands of distinct user agent fields, partly because there are many versions of browsers (especially early on, before conventions for the user agent field were set.) We aggregate all the distinct user agent codes linked to Microsoft or Netscape browsers to the major browser versions listed in Table 1 and Table 2.

Our algorithm for converting user agent codes into browsers and OSs begins with a Perl script used by Ed Kubaitis to make monthly statistical reports on the data we are using and with a number of browser detection software programs used on web servers. We make several changes to the script to catch oddities and exceptional cases.[46]

Many user agent fields are fairly simple to parse. For example "Mozilla/2.0 (compatible; MSIE 3.02; Windows 95)" refers to Microsoft Internet explorer version 3.02 running on Windows 95 and is coded as IE3 on Windows 95 in our classification.

Others are more complicated. Some user-agent fields list multiple browsers and multiple OSs; often these appear as user-agent fields within user agent fields, demarcated by parentheses. For example, "Mozilla/3.0 (Windows 3.10;US) Opera 3.60b3 [en]" is a browser, Opera 3.6, pretending to be another browser, Netscape Navigator 3.0. This "spoofing" is common, since it lets a web server give pages to one browser as if it were another. All Microsoft Internet Explorer browsers, for example, begin their user agent field with the version of Netscape they imitate followed by "compatible." In our algorithm below, we refer to this problem as "multiple browser names."

The first step in our browser classification algorithm is to search for a list of names, like "Opera," which are neither IE nor Netscape and to classify them as "other." Regardless of where the name appears in the user-agent field, we classify these browsers as "other".

Our second step is to classify all of the remaining user agent fields as IE or Netscape. If the browser name contains "MSIE," we classify it as IE. This captures not only Microsoft-branded browsers, but co-branded browsers from AOL, etc., based on MSIE technology. Among the remaining browser names, they are classified as Netscape if they contain either "Netscape" or "Mozilla" but not "compatible." If there are multiple browser names in the user agent field, we use the outermost (not in parentheses) and leftmost (first) one. Remaining browsers are then classified as "other."

We also search the user-agent field for the OS that the computer is running. Just as multiple browser names are present, multiple OS names are sometimes present as well.

---

[46] Many of the improvements we make to the classification algorithm were not necessary to the purposes of contemporary web statistics. For example, there are versions of IE1 with "IE 4.40" in their user agent, years before the existence of IE4.

In addition to spoofing, sometimes a browser will identify itself as running on a list of OSs. We once again use the outermost and leftmost OS name.[47]

All our econometric results are based on the Microsoft and Netscape browsers, but some users use other browsers. Also, there are some non-human "users". Not every computer that accesses a web server is being used by a person; some are running automatic indexing programs or "spiders." Also, a few users change the user agent code, typically to express their engineering individuality. We exclude the spiders and anomalous user-agent fields to the extent possible, keeping only the genuinely distinct browsers (Opera, Mosaic, etc.).

10) Appendix B – Restrictions on Distribution

We model some but not all of the distribution restrictions documented in the antitrust case.

MC_{om} measures the fraction of PC manufacturers selling operating system o who were contractually bound to distribute and display Internet Explorer in month m. PC manufacturers selling Windows 95 and Windows 98 were required to distribute the newest version of IE with new PCs. Thus, Table 11 shows MC_{om} ("must carry") for the newest version of IE on those operating systems as 1 throughout the sample. PC manufacturers never had to distribute and display IE with Windows 3.1, but always had the MC_{om} restriction for later versions of Windows. For the Macintosh, the MC_{om} restriction activates with a contract that took effect in August, 1997. MC_{om} varies across the OSs in our data and, to a small degree, over time.[48]

We omit any measure of the increasing stringency of the must carry restrictions. Throughout 1995, Microsoft compelled distribution of IE with Windows 95; beginning in early 1996, Microsoft enforced restrictions which compelled display of IE as well, including (for example) putting an IE icon on the Windows desktop, under the "Windows Experience" marketing label. This led to strife with PC manufacturers and monitoring of

[47] Our sample means depart from Ed Kubaitis' statistics on this point. He assigns OSs based on a precedence system, classifying a user-agent field as coming from Windows 95 if that OS is named anywhere in the field, for example.

[48] According to Microsoft's legal theory, there is no such thing as a browser so that the MC contract provisions simply required OEMs to take all of Windows. While this claim is incorrect, that does not affect our empirical analysis of regressors based on MC_{om}.

compliance by Microsoft. Microsoft bans on valuable technology, such as the special screens which appear when a PC is first used, also led to strife; they were imposed because the screens sometimes mentioned Netscape.

We also omit direct restrictions on end users. Starting with IE3, Microsoft went beyond limitations on PC manufacturers and made it harder for the end user to remove IE from their computer. With IE4, this was even more difficult for users. Similarly, there were increasingly tight bundling strategies for end users who bought Windows without a computer, for example, to upgrade.

There are a number of nuances in the MC_{om} provisions of PC manufacturer contracts not captured here. The most important of these are (a) that the MC_{om} restrictions on PC manufacturers grew more and more complex and onerous over time and (b) that technical restrictions also made it more difficult for the user to remove IE from her computer over time.

To some degree, the requirement to distribute and display IE led PC manufacturers to distribute only IE. Others distributed two browsers.

Furthermore, Microsoft used threats to block PC manufacturers from distribution and display of Netscape Navigator. This leads to a second measure reported in Table 11. The table shows EX_{om}, a variable for restrictions which blocked PC manufacturers from distributing Netscape browsers.[49] This variable measures the fraction of PC manufacturers who agreed not to distribute or not to display Netscape browsers with new PCs. The Windows 3.1 and Macintosh values are the same as MC_{om} – always zero for Windows 3.1, and changing from zero to one on the Macintosh. For Windows 95 and Windows 98, PC manufacturers agreed to these restrictions only over time. We use findings of fact from the antitrust case reporting on the status of PC manufacturers who had agreed to those restrictions, interpolating linearly between reported dates.[50]

The effect of all these restrictions was to lower the number of PC manufacturers carrying two browsers. Netscape had distribution agreements with OEMs in 1995. By

---

[49] For this variable, the view of Microsoft's legal defense would matter, since Microsoft denied that there were any such restrictions. Again, this claim is incorrect and does not affect our analysis. For our purposes, EX_{om} is simply the basis for a regressor.

[50] See *Plaintiffs' Joint Proposed Findings of Fact*, US vs. Microsoft & New York vs. Microsoft (1999), Section VII, 364.4.1, i and 364.4.2, ii.

January, 1998, Microsoft had succeeded in blocking PC manufacturer distribution of Netscape browsers on almost all new computers.[51]

Table 11 Distribution restrictions history

PC manufacturers OS	$\mathbf{MC_{om}}$		$\mathbf{EX_{om}}$	
Windows 3.1	0		0	
Windows 95	1		0	m < Jan. 1996
			11/17	May 1996
			rising linearly to	
			56/60	Jan. 1998
			constant thereafter	
Windows 98	1		56/60	
Macintosh	0	m < Aug. 1997	0	m < Aug. 1997
	1	m ≥ Aug. 1997	1	m ≥ Aug. 1997
ISPs	0	m < Jan. 1997	0	m < Jan. 1997
	1	m > Dec. 1996	1	m > Dec. 1996

Microsoft also imposed distribution restrictions on ISPs and online services (OLSs) like AOL. Starting in 1996, Microsoft sought contracts with these firms by which they agreed to distribute IE and to stringently restrict distribution of Netscape browsers.[52] AOL, for example, agreed in March 1996 to such a distribution deal that took effect with the launch of IE3 that summer. By early 1997, fourteen of the fifteen largest ISPs had signed similar contracts (the holdout was Erols).[53] Warren-Boulton (1998) dates the ISP restrictions program as being in place by late 1996.

[51] See GX 421 (Kempin 1998).

[52] ISPs and online services could not offer or display any way for their customers to get Netscape browsers; if a customer demanded Netscape nonetheless, they could provide it, but not to more than 15% of total customers.

[53] See GX 93 (Hovstadius 1996) for remarks about IE distribution agreements circa Sept. 1996, GX 1833 (Sikka 1996) for a summary of all ISP browser distribution arrangements circa Dec. 1996, GX 440 (Chase 1997) for April 1997 remarks that ISPs object to the limitations on distribution, and GX 835 (Norberg 1997) for a late 1997 ISP distribution arrangements snapshot.

The restrictions on ISPs are simply represented by a date after which ISPs were required to distribute and display IE browsers and banned from displaying Netscape browsers or distributing them.

11) Complete Descriptive Statistics Table

Descriptive statistics are taken over months when an OS-browser combination exists according to Table 3 and the OS is within sample period as defined in Table 4.

Table 12 Descriptive Statistics for $PCDISTR_{bot}$, $PCCARRY_{bot}$, and $PCEXCLU_{bot}$

			$PCDISTR_{bot}$				$PCCARRY_{bot}$	$PCEXCLU_{bot}$
Browser	**OS**	**N=**	Mean	Std. Dev.	Min	Max	Mean	Mean
IE2	mac	45	0.108	0.030	0.017	0.150	0	0
	w31	27	0.075	0.016	0.024	0.088	0	0
	w95	45	0.257	0.199	0.089	0.742	0.257	0
NS2	mac	45	0.047	0.011	0.017	0.066	0	0
	w31	28	0.040	0.007	0.012	0.046	0	0
	w95	45	0.161	0.119	0.058	0.483	0	0.025
IE3	mac	36	0.108	0.035	0.012	0.151	0.041	0
	w31	19	0.064	0.025	0.017	0.087	0	0
	w95	41	0.416	0.143	0.104	0.694	0.416	0
NS3	mac	41	0.133	0.039	0.018	0.184	0	0
	w31	23	0.071	0.017	0.022	0.083	0	0
	w95	41	0.317	0.134	0.104	0.602	0	0.138
IE4	mac	24	0.198	0.114	0.015	0.374	0.198	0
	w31	5	0.005	0.002	0.002	0.008	0	0
	w95	27	0.412	0.126	0.072	0.536	0.412	0
	w98	17	0.694	0.303	0.248	1.000	0.694	0
NS4	mac	28	0.129	0.049	0.014	0.186	0	0.129
	w31	13	0.041	0.014	0.013	0.055	0	0
	w95	31	0.438	0.133	0.063	0.589	0	0.372
	w98	17	0.313	0.305	0.063	1.000	0	0.292
IE5	w95	10	0.096	0.046	0.022	0.158	0.096	0
	w98	10	0.520	0.199	0.150	0.752	0.520	0
NS5	mac	15	0.154	0.080	0.022	0.272	0	0.154
	w95	15	0.172	0.074	0.031	0.268	0	0.160
	w98	15	0.779	0.173	0.341	0.937	0	0.727

12) Bibliography

Bank, David. (2001) *Breaking Windows: How Bill Gates Fumbled the Future of Microsoft.* New York: Free Press, 2001.

Bresnahan, T. (2002) "The Economics of the Microsoft Case," working paper, Stanford University.

Cusumano, Michael A. and David B. Yoffie. (1998) *Competing on Internet Time: Lessons from Netscape and Its Battle with Microsoft.* New York: Free Press, 1998.

Chase, Brad. (1997) Email Re: Referral Server (GX 490) in *United States v. Microsoft Corporation,* Civil Action No. 98-1232.

David, Paul A. and Shane Greenstein. (1990) "The Economics of Compatibility Standards: An Introduction to Recent Research." *Economics of Innovation & New Technology.* 1(1-2). pp. 3-41.

Davis, Steven J. and Kevin M. Murphy. (2000) "A Competitive Perspective on Internet Explorer." *The American Economic Review* 90 (2): 184-187, May 2000.

Downes, Tom and Shane Greenstein. (2002) "Universal Access and Local Internet Markets in the US." *Research Policy* 31: 1035-1052, 2002.

Evans, Daniel S. and Richard L. Schmalensee. (2000a) "Be Nice to your Rivals: How the Government is Selling an Antitrust Case without Consumer Harm in *United States v. Microsoft*," in *Did Microsoft Harm Consumers?: Two Opposing Views.* Washington, D.C.: AEI Press, 2000.

Evans, Daniel S. and Richard L. Schmalensee. (2000b) "Consumers Lose if Leading Firms Are Smashed for Competing," in *Did Microsoft Harm Consumers?: Two Opposing Views.* Washington, D.C.: AEI Press, 2000.

Fisher, Franklin, (1999) Direct Testimony in *United States v. Microsoft Corporation,* Civil Action No. 98-1232 (TPJ), January 1999. http://www.usdoj.gov/atr/cases/f2000/2057.htm

Fisher, Franklin M. and Daniel L. Rubinfeld. (2000a) "*United States v. Microsoft:* An Economic Analysis," in *Did Microsoft Harm Consumers?: Two Opposing Views.* Washington, D.C.: AEI Press, 2000.

Fisher, Franklin M. and Daniel L. Rubinfeld. (2000b) "Misconceptions, Misdirection, and Mistakes," in *Did Microsoft Harm Consumers?: Two Opposing Views.* Washington, D.C.: AEI Press, 2000.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. (2005) Geographic Location and the Diffusion of Internet Technology. *Electronic Commerce Research and Applications* 4 (1): 1-10, Spring 2005.

Goldfarb, Avi. (2004) "The (Teaching) Role of Universities in the Diffusion of the Internet." University of Toronto, Rotman School of Management, Working Paper, September, 2004.

Goolsbee, Austan and Peter J. Klenow. (2002) "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." *Journal of Law and Economics* 45: 317-343, October 2002.

Gottardi, Giorgio. (2004) "Lo sviluppo di Internet. Un'analisi tramite modelli di diffusione (Growth of the Internet : A Diffusion Model Analysis. With English summary)." *L'Industria, Nuova Serie* 24 (April-June 2003): 223-248.

Greenstein, Shane M. (2004) "The Economic Geography of Internet Infrastructure in the United States." Northwestern University, Center for the Study of Industrial Organization, Working Paper #0046, August 2004.

Griliches, Z. (1957), "Hybrid Corn: An Exploration in the Economics of Technical Change" Econometrics, 25(4) pp.501-522.

Griliches, Z. (1988), "Postscript on Diffusion," Chapter 3 in Technology, Education, and Productivity, Basil Blackwell, New York.

Hall, Bronwyn H. (2003) "Innovation and Diffusion" Working Paper, University of California.

Haas, Randy (1998) "Browser Model for 3YO," Email to Brad Chase and others, February 6th. Government Exhibit 310.

Henderson, Rebecca E. (2000) Direct Testimony in *United States v. Microsoft Corporation,* Civil Action No. 98-1232, April 28, 2000.

Hovstadius, Bjorn. (1996) Trial Exhibit: Email Re: ISP Marketing Update (GX93) in *United States v. Microsoft Corporation,* Civil Action No. 98-1232. September 9, 1996.

IDC. (2000a) Client Operating Environments Market Forecast and Analysis, 2000-2004, IDC #22346, June 2000.

IDC. (2000b) Consumer Internet Service Provider Market Share Update, 1999, IDC #22065, April 2000.

IDC. (2000c) IDC's Quarterly PC Update: 1Q00 Review/2Q00 Outlook, IDC #22811, August 2000.

IDC. (2000d) IDC's Quarterly PC Update: 2Q00 Review/3Q00 Outlook, IDC #23224, October 2000.

IDC. (2000e) IDC's Quarterly PC Update: 4Q99 Review/1Q00 Outlook, IDC #22067, April 2000.

IDC. (2000f) Internet Service Provider Market Review and Forecast, 1999-2004, IDC #21203, December 1999.

IDC. (2000g) Network Insider: May 2000, IDC #22379, May 2000.

IDC. (2000h) Operating Environments Market Forecast and Analysis, 2000-2004, IDC #22597, July 2000.

Jenkins, Mark, Paul Liu, Rosa Matzkin, and Daniel L. McFadden. (2004) "The Browser War - Econometric Analysis of Markov Perfect Equilibrium in Markets with Network Effects." Working Paper. April 22, 2004, revised December 2004

Kempin, Joachim. (1998) Trial Exhibit: Microsoft OEM Sales FY '98 Mid-Year Review (GX 421) in *United States v. Microsoft Corporation,* Civil Action No. 98-1232.

Liebowitz, S. J.and Margolis, S.E. (1999) *Winners, Losers & Microsoft: Competition and Antitrust in High Technology*, The Independent Institute, Oakland, CA, 1998.

Manes, Stephen and Paul Andrews. (1993) *Gates: How Microsoft's Mogul Reinvented an Industry—and Made Himself the Richest Man in America.* New York: Doubleday, 1993.

McKean, Kevin. (2002) "Nine Timeless Tips for Tech Buyers." *PC World* (June 2002): p. 19.

Mowery, David C. and Timothy Simcoe. (2002) "The Internet," in *Technological Innovation and Economic Performance*, eds. Benn Steil, David G. Victor, and Richard R. Nelson. Princeton, New Jersey: Princeton University Press, 2002.

Myhrvold, Cameron. (1999) Direct Testimony in *United States v. Microsoft Corporation,* Civil Action No. 98-1232, February 10, 1999.

Nielsen, Jakob (1998) "Alertbox: The Increasing Conservatism of Web Users: March 22, 1998 http://www.useit.com/alertbox/980322.html

Norberg, Susan (preparer). (1997) Trial Exhibit: Netscape Competitive Analysis- ISP/OLS Channel Revenue Segment (GX 835) in *United States v. Microsoft Corporation,* Civil Action No. 98-1232.

Nordhaus, W. (1997), "Do Real Output and Real Wage Measures Capture Reality? The History of Lighting Suggests No," in Timothy Bresnahan and Robert Gordon, (eds.) The Economics of New Goods, University of Chicago Press, Chicago, pp. 29-70.

Schmalensee, R. (1999a) Direct Testimony in *United States v. Microsoft Corporation,* Civil Action No. 98-1232 (TPJ), January 1999. http://www.courttv.com/archive/trials/microsoft/legaldocs/ms_wit.html

Schmalensee, R. (1999b) Supplementary Exhibits to Direct Testimony (DX 2098) in *United States v. Microsoft Corporation,* Civil Action No. 98-1232 (TPJ), January 1999.

Shapiro, Carl and Hal R. Varian. (1999) "The Art of Standards War." *California Management Review* 41 (2): 8-32, Winter 1999.

Sikka, Ajay. (1996) Trial Exhibit: Email Re: North American ISP XLS for Jekk Raikes (GX 1833) in *United States v. Microsoft Corporation,* Civil Action No. 98-1232. December 19, 1996.

Stango, Victor. (2004) "The Economics of Standards Wars." *Review of Network Economics* 3 (1): 1-19, March 2004.

Swisher, Kara. (1998) *Aol.com: How Steve Case Beat Bill Gates, Nailed the Netheads, and Made Millions in the War for the Web.* New York: Random House, 1998.

U.S. Department of Commerce. (2002) Economics and Statistics Administration. National Telecommunications and Information Administration. *A Nation Online: How Americans are Expanding Their Use of the Internet,* February 2002.

U.S. Department of Justice et al. (1999) "Plaintiffs Joint Proposed Findings of Fact – Revised" in Civil Action No. 98-1232, US v. Microsoft and State of New York (and others) v. Microsoft. September 10, 1999.

Vogelsang, Ingo and Benjamin M. Compaine, eds. (2000) The *Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy.* Cambridge, Massachusetts: MIT Press, 2000.

Wallace, James. (1997) *Overdrive: Bill Gates and the Race to Control Cyberspace.* New York: John Wiley & Sons, 1997.

Warren-Boulton, Frederick. (1998) Direct Testimony in *United States v. Microsoft Corporation,* Civil Action No. 98-1232 (TPJ), November 1998.

Windrum, Paul (2000) "Back from the brink: Microsoft and the strategic use of standards in the Browser Wars" MERIT Working Paper, Universiteit Maastricht.

Wiseman, Alan E. (2000) *The Internet Economy: Access, Taxes, and Market Structure.* Washington, D.C.: Brookings Institution Press, 2000.

Dépôt légal : 2006 – 4ème trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 1242-5125 ENST S (Paris) (France 1993-9999)