École Nationale Supérieure des Télécommunications
Groupe des Ecoles des Télécommunications - membre de ParisTech
46, rue Barrault - 75634 Paris Cedex 13
Tél. + 33 (0) 45 81 77 77 - Fax + 33 (0) 1 45 89 79 06 - www.telecom-paris.fr

ENST 2007 S 002

Conference on the Economics of Information and Communication Technologies - 2007

Paris, June 21 - 22, 2007

TELECOM PARIS
école nationale supérieure des télécommunications

Proceedings of the 2007 Telecom Paris Conference on the Economics of Information and Communication Technologies

ENST Paris, France — June 21-22, 2007

Editor : Patrick Waelbroeck (ENST)
Organized by :
Department of Economics and
Social Sciences of Telecom Paris

ENST 2007 S 002

École Nationale Supérieure des Télécommunications

Program of the 2007 Telecom Paris conference on the economics of ICT

Paris, June 21-22, 2007

Schedule

	Thursday
9:00-9:30	Welcome speech
9:30-10:30	Michael D. Smith
11:00-12:30	Two-sided/Open Source
12:30-14:00	Lunch
14:00-15:00	Jacques Crémer
15:15-16:45	Pricing-Bundling/Empirical issues

Thursday evening: conference dinner at Altitude 95 on first floor of the Eiffel Tower

	Friday
9:00-10:00	Ivan Png
10:30-12:30	Telecom/Network
12:30:14:00	Lunch
14:00-16:00	Productivity/Piracy
16:00-16:30	Closing meeting

Invited Talks

Michael D. Smith (Carnegie-Mellon U.)
Discussant: Patrick Waelbroeck (ENST)

Jacques Crémer (U. Toulouse, IDEI)
Discussant: Marc Bourreau (ENST)

Ivan Png (National U. Singapore)
Discussant: Anne Duchêne (Drexel U.)

Sponsors

Program of the 2007 Telecom Paris conference on the economics of ICT
June 21, 2007

Two-sided markets

María Fernanda Viecens (U. Carlos III Madrid), "Pricing Strategies in Software Platforms: Video Consoles vs. Operating Systems"
Discussant: Thomas Trégouët (CREST-LEI) TBC

Andrei Hagiu (Harvard Business School), "Merchant or Two-Sided Platform"
Discussant: Marianne Verdier (ENST)

Open source

Yossi Spiegel (Tel Aviv U.), "The Incentive to Participate In Open Source Projects: A Signaling Approach"
Discussant: Emeric Henry (London Business School)

German Daniel Lambardi (GREMAQ), "Code Development: Open vs. Proprietary Source"
Discussant: Yann Ménière (Ecole des Mines, CERNA)

Pricing and bundling

Hanming Fan (Yale U.) and Peter Norman (U. North Carolina, Chapel Hill), "On Bundling by Natural Monopolies"
Discussant: Nicolas Sahuguet (HEC Montréal)

Eric Jahn (Goethe U. Frankfurt), "Competitive Brand Extension"
Discussant: Laurent Linnemer (CREST-LEI)

Liliane Karlinger (U. Vienna) and Massimo Motta (European U. Institute and U. Bologna), "Exclusionary Pricing and Rebates When Scale Matters"
Discussant: Philippe Choné (CREST-LEI)

Empirical issues in ICT industries

María Rosalía Vicente (U. of Oviedo) and Ana Jesús López (U. Oviedo), "An empirical analysis of broadband diffusion in the New Member States of the European Union"
Discussant: Lukasz Grzybowski (U. Alicante)

Gautam Gowrisankaran (Washington U. St. Louis) and Marc Rysman (Boston U.), "Dynamics of Consumer Demand for New Durable Goods"
Discussant: Sofronis Clerides (U. Cyprus)

Lapo Filistrucchi (Tilburg U. and U. Siena), "The Impact of Internet on the Market for Daily Newspapers in Italy"
Discussant: Ana Paula Martins (U. Catolica Portuguesa)

Program of the 2007 Telecom Paris conference on the economics of ICT
June 22, 2007

Telecom industries

João Vareda (U. Nova Lisboa), "Unbundling and Incumbent Investment in Quality Upgrades and Cost Reduction"
Discussant: Nicolas Schutz (Paris Jourdan Sciences Eco.)

Ángel Luis López (GREMAQ), "Dynamic Competition in Telecom Networks under the Receiver Pays Principle"
Discussant: Jean-Sébastien Bedo (France Telecom R&D)

Tseveen Gantumur (European U. Viadrina) and Andreas Stephan (European U. Viadrina and DIW), "Mergers & Acquisitions and Innovation Performance in the Telecommunications Equipment Industry"
Discussant: Denis Lescop (INT)

Network effects

Mikhail M. Klimenko (Georgia Institute of Technology), "International Trade Policies and Agreements on Compatibility Standards for Industries with Network Externalities"
Discussant: Marvin Sirbu (Carnegie-Mellon U.)

Gal Oestreicher-Singer (New York U., Stern) and Arun Sundararajan (New York U., Stern), "Network Structure and the Long Tail of Electronic Commerce"
Discussant: Kevin Mellet (France Telcom R&D)

Daniel CERQUERA (ZEW), "Durable Goods, Innovation and Network Externalities"
Discussant: Gal Oestreicher-Singer (New York U., Stern)

ICT and productivity

NKetcha Pierre Valere (U. Yaounde II), "Robust Estimation of ICT Effects on Cameroonian Firm's Productivity"
Discussant: Chiraz Karamti (ENST)

Jörg Ohnemus (ZEW), "Does IT-Outsourcing Increase Firm Success? A Firm-level Investigation for Germany"
Discussant: Thomas Houy (ENST)

Piracy

Michael R. Ward (U. Texas, Arlington), "Teaching Piracy"
Discussant: Rémi Douine (ENST)

Fabrice Le Guel (U. Nice Sophia Antopolis) and Fabrice Rochelandet (U. Paris 11, ADIS), "The willingness to pay for online music in the presence of copying: An empirical investigation"
Discussant: François Moreau (CNAM)

Contents

Pricing Strategies in Software Platforms: Video Consoles vs. Operating Systems

María Fernanda Viecens
Universidad Carlos III de Madrid*

March, 2007

Abstract

We study software platforms for which the total amount that users spend depends on the two-sided pricing strategy of the platform firm, and on the pricing strategy of application developers. When setting prices, developers may be constrained by one of two margins: the demand margin and the competition margin. By analyzing how these margins affect pricing strategies we find some conditions which explain features of the market of operating systems and its differences with the one corresponding to the video consoles. The problem that arises when the platform does not set prices (as an open platform) is considered. We show that policy makers should promote open source in operating systems platforms but not necessarily in video consoles. We also analyze the incentives for a platform to integrate with applications as a function of the extent of substitutability among them and provide a possible explanation for the observed fact of vertical disintegration in these industries.

Keywords: two-sided markets, technology platforms, complements, vertical disintegration, competition policy.

JEL Classification: L10, L12, D40

*Address for Correspondence: *María Fernanda Viecens:* Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Spain, E-mail: mviecens@eco.uc3m.es. Part of the research for this paper was conducted during a stay at the Economic Department at Mannheim University. I thank members of the department for their hospitality, in particular Konrad Stahl. I also thank M. Ángeles de Frutos for her advise, support and encouragement. This paper has benefited from participants in the Jamboree meeting 2007, ENTER Program. The usual disclaimers apply.

1 Introduction

Many modern industries work around software platforms. Typical examples are operating systems for computers, personal digital assistants, smart mobile phones or videogame consoles. The usual feature is that they connect or attend different types of customers that benefit from the interaction among them, characterizing what is known in the literature as two (multi)-sided platforms. On the one side, developers write the applications or software that improve the value of the platform for the users. On the other side, users derive utility from consuming the system (the platform and the applications). Because of this, users are concerned about the system price, i.e., the total amount spent in the platform and the software. The system price will hence depend on the two-sided pricing strategy of the platform firm which in turn affects the market of complementary applications, and on the pricing strategy in the developers' market. This paper offers a model of a monopolist two-sided platform that allows us to analyze the pricing strategies it will adopt, the level of entry it will induce in the applications' market and the welfare it will generate. Furthermore, by considering that it can become either an open platform or a proprietary one, we will study the implications of having one or the other. Finally, issues related to the vertical structure of the platform and to the role of outside options will also be analyzed.

Two well known and widely used sofware platforms are video consoles and computer operating systems. In both, users care for the total charge of the system (platform and applications). Nevertheless they have followed quite different pricing strategies. Operating system platforms charge high prices to the users and subsidize developers. However, video console firms charge low prices to users and make profits on the developers' side.[1] We provide here a possible explanation for the difference based on the margin at which developers compete. When setting prices, developers may be constrained by one of two margins, the demand margin and the competition margin. As long as the demand margin binds, prices of developers affect the overall demand of the system and they set the price that maximizes their profits, a price that is lower than their marginal contribution to the users utility. In contrast, if competition margin binds, developers can not affect overall demand of the system and they are forced to set a price equal to their contribution to the users surplus.[2] What margin is binding depends on the number of applications in the market and on the level of substitutability among them. In particular, the competition margin is more likely to bind as long as users prefer a system with many applications and these are near substitutes. In the market of video console gamers state that price is very important in deciding what game to buy. Some of them report having a huge number of games and, for instance, among the ten top rated PlayStation 2 games, 3 of them belong to the adventure genre and 3 to the role-playing

[1] This issue is largely analyzed by Hagiu (2005).

[2] Lerner and Tirole (2004) introduce the two margins to analyze pricing startegies in patent pools.

genre.[3] These facts allow us to presume that developers writing for the video console are constrained by the competition margin. However, users of operating systems need a lower number of applications that indeed are far substitutes, like a text processor, a spreadsheet or a browser, so that we suspect the developers in this market are constrained by the demand margin. By analizing how these margins affect the pricing strategies and the profits of the platform, we find some conditions that may help to explain features of the market of operating systems and its differences with that corresponding to the video consoles, and shed some light on the different pricing routes they have followed. We observe that the platform price for users is higher when demand margin binds than when competition margin binds, and this is consistent with the observed fact that operating systems charge high prices to users, whereas video console firms charge low prices to them.

When considering the problem that arises if the platform does not set prices (as an open platform), our model allows us to contribute to the current enthusiastic discussion on whether governments should promote (as some of them do) open source platforms. Nowadays, 50% of European public administrations declare that they use some open source software and the figure is 35% for the USA. In addition, some large companies are also using open source programs.[4] The literature is not conclusive about recommendations. Hagiu (2005) shows that there is a tradeoff between the extent to which proprietary platforms internalize indirect network effects through profit-maximizing pricing and the two-sided deadweight loss they create. He shows that a proprietary platform may generate a higher level of product variety and welfare than an open platform. In contrast, Economides and Katsamakas (2006a) find that the variety of applications and social welfare is always larger when the platform is open source. We here show that outcomes may depend on the margin that binds. We find some results that suggest that policy makers should promote open source platforms where demand margin binds (as operating systems) but not necessarily in platforms where competition margin binds (as video consoles). In particular, we prove that if demand margin binds, a proprietary platform and an open platform will provide the same level of applications, so that the latter will generate more welfare for users. However, if competition margin binds a proprietary platform may generate a larger number of applications and higher welfare to users than an open platform.

In a book about empirical business and economics aspects of software based platforms, Evans, et. al. (2006) document that almost all the successful firms in these industries started being one-sided, producing applications at home, and later they disintegrated becoming in firms producing only the platform

[3] See *Game Daily: June 22, 2004* at http://www.gamedaily.com, *Video Game Culture: Leisure and Play Preferences of B.C. Teens - Summary of Findings* at http://www.media-awareness.ca, and www.gamespot.com, September 2006.

[4] In a sample of 600 large companies in USA, 35% use one or more "free" softwares and 39% of 300 European large firms do so. Forrester Consulting, in El Mundo Digital 22/11/2006. In Spain, for instance, some "Comunidades Autónomas" are supporting open source. In 2007, the public administration of Extremadura will start to work with Linux. Andalucía and the Basque Country are also heading in the same direction (El País Digital, 16/11/2006).

and supported by independent developers.[5] We here try to provide a possible explanation for this observed fact based again on the margin that binds for developers. We analyze the incentives of a platform to integrate with applications (becoming one-sided) as a function of the extent of substitutability among them. We derive some conditions about the relationship between the welfare effects of a merger and the degree of substitution of the applications. We also offer an explanation for partial integration and we show that in the long run the platform will be partially integrated with the killer applications for which demand margin will bind and will allow free entry for developers of other applications.[6]

Finally, we study the effects on incumbent platform strategies for facing the threat of an outside option that offers a surplus for developers or users. Examples of outside options for users of the video game consoles are those games that can be played in the computer or online in the internet.[7] Writing these games is the outside option that developers have to the video console. Outside options for a proprietary operating system are the open platforms such as Linux. It is developing quickly in terms of number, variety and quality of applications and availability of support and other complementary services. In this sense, Linux is now an outside option to Windows and nowadays it is considered a serious threat to the latter.[8] Thus, we can interpret the analyses as an option that competes or threatens the incumbent platform. Questions we try to answer with this analyses are, for instance, given Windows being the incumbent firm, is it the grow importance of Linux in the users' benefit? What about developers of softwares?. If Linux becomes more important so that the value of writing applications for it increases, is this profitable for them? We find that it would not be in the interest of the users to promote the outside options (i.e., online games or computer games) to the video game console since, whenever competition margin binds, a higher outside option value for the users may lead to a decrease in their surplus. However, an increase in the value for developers of writing for an open platform such as Linux or Google has a positive impact in the users' surplus.This is the case because if demand margin binds, an increase in the outside option of the developers will always increase the users surplus.[9]

The structure of the paper is as follows. We present the model of a monopoly platform in section 2, and in section 3 we analyze the developers problem. In section 4 we solve the problem of a profit platform and compare its performance

[5] Several facts that we cite along the article are documented by Evans, et. al. (2006).

[6] For instance, Microsoft produces operating system Windows and Office package. Nintendo wrote Mario Brothers, its killer game.

[7] Gamers report an average of 6,65 of hours spent per week on online-games and the home PC use of time explains 25% of children's and adult's games. http://www.cybersurvey.com/reports

[8] See www.cnn.com, World Business, "Reclusive Linux founder opens up", 19/05/2006, and "Microsoft vs. Open Source: Who Will Win?- HBS Working Knowledge, June 2005.

[9] In November 2006 Microsoft and Novell have signed a deal so that Linux programs can operate with Windows. Rivals will collaborate on technical development and marketing programs (The New York Times, 3/11/2006). A priori it seems the deal would benefit users and developers, but it warrants further analyses.

to that of an open platform in section 5. In section 6 we analyze incentives for integration and partial integration of the platform with applications. In section 7 we introduce outside options to the monopoly platform for developers and users. Finally, section 8 concludes.

2 A monopoly platform model

We assume that there is a monopoly platform and preferences of users are defined over the platform, its applications and an outside good. There is a measure one of users and their tastes for the platform are uniformly distributed along the unit interval. The utility of a user located at distance t from the platform is

$$U = V(M) + x - kt,$$

where M is the number of software varieties or applications, x is the numeraire good and k measures the degree of platform differentiation. It is further assumed that $V(M)$ is concave and increasing in M.[10]

Every user who purchases the platform consumes at most one unit of each application and maximizes her utility by choosing applications and consumption of the outside good subject to the constraint

$$\Sigma_{j=1}^{M} p_j + x + P^U = y,$$

where p_j is the price of a unit of application variety j, P^U is the charge that platform sets to the users and y is their income. A user's decision can be decomposed into two decision problems. First, the user sets her optimal basket of applications among the total number in the market,

$$G\left(M, \Sigma_{j=1}^{M} p_j, P^U\right) = \max_{M \leq N} \{V(M) - \left(\Sigma_{j=1}^{M} p_j\right)\} - P^U, \tag{1}$$

where N is the number of applications in the market, then the user buys the platform if and only if

$$G\left(M, \Sigma_{j=1}^{M} p_j, P^U\right) - kt \geq 0.$$

The users demand for the system (size of the network) is hence determined by

$$t^d = \frac{G\left(M, \Sigma_{j=1}^{M} p_j, P^U\right)}{k} \epsilon \left[0, 1\right].$$

Note that demand depends on the price that platform sets for the users, but also on the number and prices of applications.

[10] Similar utility functions are used by Church and Gandal, (1992, 1993, 2000) and Church et.al. (2003).

On the other side there are $\bar{N}$ potential developers of applications, each of them providing a single different application. Profits of developer of application i are given by

$$\pi_i = p_i t^d - F - P^D,$$

where F is a fixed cost of production, and P^D is the price that platform charges developers to allow them to write platform compatible applications.

Costs of the platform are assumed zero, so that platform profits are given by,

$$\Pi = P^U t^d + P^D N.$$

In this set-up we study the pricing strategies of the platform and developers. To do so we consider a game whose timing is as follows: in the first stage, the platform sets the charge to developers and these decide upon entry. In the second stage, the platform sets the price to the buyers. In the third stage, developers compete and set the prices for the applications to the buyers, then finally buyers decide if they buy the platform and the number of applications.

3 Application prices, users payments and System effects

When a user considers buying the platform, her decision will depend upon the prices set by developers. No user will purchase a video console without buying some video games, nor an operating system without buying the application softwares. Because of this we first study how developers set prices which will be a key point in our analysis. We then solve the second stage of the game at which the platform sets the price for users, taking N as given. Before that let us define two elasticities that will be used throughout the paper.

Ignoring the integer problem we define the elasticity of $V(N)$, a measure of the degree of substitutability of applications for the users,[11] as follows,

$$e_v(N) = \frac{V'(N)\,N}{V(N)}.$$

Since $V(N)$ is increasing and concave, it lies in the interval $(0,1)$. For a given N, we consider that applications to be near substitutes if $e_v(N)$ is sufficiently low.[12]

Similarly, let us define the elasticity of $V'(N)$,

$$\varepsilon_v(N) = \frac{V''(N)\,N}{V'(N)}.$$

[11] It has also been interpreted as a measure of "degree of preference for variety" (see Kühn and Vives (1999) and Hagiu (2005)).

[12] Our interpretation here is similar to the one in Lerner and Tirole (2004): given N patents and two surplus functions $V_1(\cdot)$ and $V_2(\cdot)$, such that $V_1(N) = V_2(N)$, applications are more substitutable for surplus function $V_1(\cdot)$ than for $V_2(\cdot)$ if $V_1'(\cdot) < V_2'(\cdot)$.

Given that $V(N)$ is concave, it follows that $\varepsilon_v(N)$ is negative. The relationship between these two elasticities is the content of next lemma

Lemma 1 $e_v(N)$ is increasing in N as long as

$$e_v(N) < 1 + \varepsilon_v(N), \tag{2}$$

and is decreasing if the other inequality holds.[13]

3.1 Equilibrium application prices

The problem faced by developers is similar to the problem faced by a licensor in a patent pool. In the context of patents, the licensor problem has been studied by Lerner and Tirole (2004). In their model, the surplus derived from using N patents is also a function $V(N)$, strictly increasing in N. They show that, when setting a licensing fee, an individual licensor may be constrained by either of two margins that they call the *competition margin* and the *demand margin.* In our context, developers are constrained in a similar way. If the developer can not increase her price without, because of this, being excluded from the set of applications selected by the users, (in user's problem (1)) then the competition margin binds. In contrast, demand margin is said to bind for developer i, if she can individually raise her price without being excluded but leading to a reduction in the overall demand for the system (effect on t^d). In particular, if the demand margin binds, a developer chooses a price $p_i = \widehat{p}$ such that

$$\widehat{p} = \arg\max_{p_i} \; \{p_i \frac{V(N) - P^U - (N-1)\widehat{p} - p_i}{k}\}. \tag{3}$$

Consequently,

$$\widehat{p} = \frac{V(N) - P^U}{(N+1)}.$$

In contrast, if the competition margin binds, the price that a developer sets is its marginal contribution to the users utility, i.e.,

$$\widetilde{p} = V(N) - V(N-1).$$

Note that $\widetilde{p}$ depends on $V(N)$ but neither on the demand of the system t^d nor on P^U.[14] Besides, $\widetilde{p}$ is always positive, whereas $\widehat{p}$ is not necessarily so, as it will depend on the value of P^U.

Next lemma follows immediately from propositions 1 and 4 in Lerner and Tirole (2004).

Lemma 2 *There exists a unique and symmetric equilibrium such that, if $\widetilde{p} <$ $\widehat{p}$, developers are constrained by the competition margin and charge equilibrium*

[13] For instance, functions $V(N) = \log(1+N)$ and $V(N) = (1 - \exp(-N))$ have $e_v(N)$ decreasing for all $N > 0$ and it is easy to show that they satisfy the reverse of (2) in all the relevant range of N. Function $V(N) = N^{\beta}$, with $\beta < 1$, presents constant elasticities, $e_v(N) = \beta$ and $\varepsilon_v(N) = \beta - 1$, then $e_v(N) = 1 + \varepsilon_v(N)$.

[14] If we ignore the integer problem, $\widetilde{p} = V'(N)$. Then, $\varepsilon_v(N)$ also represents the applications price elasticity to N when competition margin binds.

price $\widetilde{p}$, whereas if $\widetilde{p} > \widehat{p}$, developers are constrained by the demand margin and charge equilibrium price $\widehat{p}$.

As long as demand margin binds, developers set the price that maximizes their profits and this price is lower than their marginal contribution to the users' utility. In contrast, if the competition margin binds, the price that maximizes profits, as defined in (3), is higher than the marginal contribution to users surplus and then developers are forced to set a price equal to this contribution.

The consideration of both scenarios allows us to include in the analysis situations where the developers set the price that maximizes their profits and consider the reduction in the overall demand for the system when contemplating an application price increase (i.e. when demand margin is binding). Other papers in the literature, such as Hagiu (2005) and Church et. al. (2003), implicitly restrict their analyses to an scenario where the competition margin is always binding. In particular, Hagiu (2005) assumes that developers set prices for applications once users have bought the platform. Similarly, Church et. al. (2003) derive the equilibrium prices set by developers under the proviso that platform sales are invariant to application pricing.[15] Our contribution here will not only be to study the case in which the demand margin binds, but also the comparisons that will follow. Clearly, some of our results when the competition margin is the one that binds are similar to those found in these previous papers.

3.2 What is the binding margin?

We now try to establish what the conditions are that determine the margin that will bind, by using lemma 3.1, and the equilibrium values of prices $\widehat{p}$ and $\widetilde{p}$.

Lemma 3 *Developers are constrained by the competition margin if the platform sets a price to the buyers such that*

$$P^U < V(N) - \widetilde{p}(N+1). \tag{4}$$

If the opposite inequality holds, developers are constrained by the demand margin.

A closer look at (4) allows us to determine the binding margin as a function of the primitives in the model.

Proposition 1 *If*

$$e_v(N) < \left[1 - \frac{1}{\sqrt{N+1}}\right], \tag{5}$$

the competition margin will bind. If the opposite inequality holds, the demand margin will bind.

[15] In Church et. al. (2003), $V(N) = N^{\beta}$. For this utility function they show that the Nash equilibrium in developers' prices is given by $p(N) = V'(N)$ when $N > 1$ and $\beta \leq \frac{1}{2}$, so that, in our terminology, the competition margin binds.

Proof: See Appendix A.

The proposition above shows that the degree of substitution among applications and the number of developers determine the margin that binds. As long as applications are near substitutes the competition margin is more likely to bind. The same occurs when N is large, as the following corollary shows.

Corollary 1 *If $e_v(N)$ is non-increasing there exists N^* such that if $N < N^*$ the demand margin binds and if $N > N^*$ the competition margin binds. However, if $e_v(N)$ is striclty increasing, N^* may fail to exist, so that the demand margin always binds.*[16]

Proof See Appendix A.

From proposition 1 we deduce that those developers that write applications which are not near substitutes or are indeed complements will tend to compete in the demand margin. Similarly, those systems composed by a very high number of applications are more likely to have developers competing in the competition margin.

Using the results above, if one looks at the observed facts in the video game industry discussed in the introduction,
1) 76% of gamers state that price is very/somewhat important in deciding what game to buy,
2) From a survey of over 1,000 game consumers it is known that around 19.10% of them purchase 1 or 2 games per month, 26.50% purchase 1 every two month and 6.90% 3 or more per month,[17]
3) Some players report having more than 50 games,
4) Among the ten top rated PlayStation 2 games, 3 of them belong to the adventure genre and 3 to the role-playing genre. Among the ten top rated Xbox 360 games, 2 of them belong to the Ice Hockey genre.
Facts 1 and 4 suggest that there exists a near substitution between the games. Facts 2 and 3 show that consumers usually own a system of console and video games composed of many applications.

If we compare these facts with those observed for systems of operating systems and applications (i.e. Windows) we find that it is not easy to find a consumer using a huge number of applications.[18] Moreover, applications are far substitutes (and sometimes complements). A user may need a text processor and a spreadsheet and also a browser. Then, we presume that developers writing for an operating system are constrained by the demand margin whereas those writing for the video console are constrained by the competition margin.

[16] This is the case for instance for $V(N) = N + \sqrt{N}$ for which demand margin always binds. Note that if N^* exists, it is defined by $N^* = \left(\frac{1}{1-e_v(N^*)}\right)^2 - 1$.

[17] Zelos Group Survey: What Do Gamers Want? Everything. Electronic Gaming Business, Nov 19, 2003. http://www.findarticles.com

[18] Evans et.al. (2006) point out that, as opposed to the case of video consoles, "there's probably not much correlation between the number of applications that someone uses on a computer and the value that person places on that computer".

3.3 Users prices and system effects

In the second stage of the game, the platform sets the price for users, taking N as given. When the demand margin binds, the platform will set a price to the users as[19]

$$\widehat{P}^U = \frac{V(N)}{2}. \tag{6}$$

It then follows that the price set by developers will be $\widehat{p} = \frac{V(N)}{2(N+1)}$.

Meanwhile, if the competition margin binds, the optimal price that platform chooses for the users is

$$\widetilde{P}^U = \frac{V(N) - \widetilde{p}N}{2}. \tag{7}$$

Equations (6) and (7) put in evidence the existence of "system effects" in the industry. These effects arise when the value of one component depends on complementary components in the system.[20] The presence of system effects is reflected in the price that the platform sets to users which increases with the number of applications. In addition, when competition margin binds $\widetilde{p}$ affects $\widetilde{P}^U$ because of the complementarity between the applications and the platform. In particular, for a given N, when the price of the applications increases, the benefit that the platform makes per user decreases.[21]

When the competition margin binds, the relative charges paid by users can be expressed as a function of $e_v(N)$,

$$\frac{\widetilde{P}^U}{\widetilde{P}^U + \widetilde{p}N} = \frac{1 - e_v(N)}{1 + e_v(N)}.$$

Lemma 4 *As long as applications are more substitutes, applications will be relatively less expensive, and the platform can charge users more.*

When substitution is strong on the developers' side, prices in this market are very low and the platform takes advantage of this situation setting a higher price for the platform. Lemma 4 implies that it is profitable for the firm selling the console to accept games that compete among them or are near substitutes, which is consistent with the observed practice in the video game industry as stated in fact 4.

The relative charge paid by users for the platform when demand margin binds is given by

$$\frac{\widehat{P}^U}{\widehat{P}^U + \widehat{p}N} = \frac{N+1}{2N+1}.$$

[19] See proof of proposition 1.

[20] See Evans and Schmalensee (2001). System effects are a clear feature of software platforms where the user buys a system (platform and applications) and cares for the total charge of the system.

[21] Note also that for the same N, $\widehat{P}^U > \widetilde{P}^U$, it is consistent with the observed fact that operating systems charge high prices to users whereas video console firms charge low prices to them.

Note that when demand margin binds, relative charges depend only on N whereas it depends on both, N and $V(N)$, when competition margin binds. The next proposition presents how relative charges vary with N.

Proposition 2 *If demand margin binds, the relative payment made by users to the platform is decreasing in N. However, if the competition margin binds, the relative payment is increasing in N whenever $e_v(N)$ is decreasing in N.*

As N increases, users tend to spend more on the bulk of applications when demand margin is binding. The same occurs, whenever the competition margin binds provided that $e_v(N)$ is increasing in N. However, $\frac{N+1}{2N+1} > \frac{1}{2}$, meaning that more than one half of the money that users spend in the system goes to the platform when demand margin binds. Meanwhile, it may occur that $\frac{1-e_v(N)}{1+e_v(N)} < \frac{1}{2}$ if $e_v(N) > \frac{1}{3}$.

When setting the price to users, the platform should optimally preserve this ratio, if not, a competitor with a better pricing strategy may easily overcome the incumbent's advantages.[22]

Let the users demand elasticity with respect to the price by the platform be $E_p = \frac{\partial t^D}{\partial P^U}\frac{P^U}{t^D} = -1$ and the elasticity of demand with respect to the number of applications be $E_s = \frac{\partial t^D}{\partial N}\frac{N}{t^D} = -\frac{\epsilon_v \widetilde{\varepsilon}_v}{1-\epsilon_v}$. The ratio $-\frac{E_s}{E_p}$ measures the effect of platform price equivalent to a 1% increase in N.[23] In the users' interest, a 1% increase in the number of applications is equivalent to a $\frac{\epsilon_v \widetilde{\varepsilon}_v}{1-\epsilon_v}$% price cut.[24] This ratio is increasing in ϵ_v and $\widetilde{\varepsilon}_v$. That is to say that an increase in N is more valued as long as it conveys a reduction in developers applications prices and applications are near complements.

4 Developers Entry and Welfare: Profit Platform vs Open Platform

In the first stage a proprietary platform sets a price to the developers that then decide upon entering the market. If the platform is open, this price is zero.[25] One could think that the platform, through the choice of prices for developers, determines the number of applications. However, this assertion may not always

[22] For instance, in the market for video players, VHS overcame Beta after six years of higher installed base by Beta. The strategy of the winner was a widespread licensing of VHS and a low- priced VHS player, compared with a high-priced Beta player and restricted licensing (See Economides 2006).

[23] Note that if $V(N) = N^{\beta}$ the ratio is $\frac{-E_s}{E_p} = 1$.

[24] Clements and Ohashi (2005) have computed this ratio for the USA video game industry. They find that a 1% increase in game titles is equivalent (in average) to a 2.3% price cut of the console price.

[25] An open platform will charge zero to both users and developers. Nevertheless, we will assume that developers set positive prices to users for their applications. Applications for open platforms like Linux are often free for consumers. However, there are also several applications that are not free that are offered for Linux operating system (Economides and Katsamakas (2005a)).

be true. In particular, if developers' gross profits (i.e., $p(N)\, t^d(N)$) are increasing in the number of applications, then the platform can not affect entry which will equal $\bar{N}$. This is the case when the positive indirect network effect more than compensates the direct negative effect of competition. An additional developer exerts a positive effect on other developer's profits, explained by the fact that more participation by one side (i.e., developers) induces more participation by the other side (i.e., users), which benefits customers and makes them more willing to participate.[26] Consequently, whenever developers' gross profits are increasing at $\bar{N}$, the platform would charge a price $P^D(\bar{N}) = p(\bar{N})\, t^d(\bar{N}) - F$ and its profits will be

$$\Pi = P^U(\bar{N})\, t^d(\bar{N}) + \left(p(\bar{N})\, t^d(\bar{N}) - F\right)\bar{N}.$$

In contrast, if developers' gross profits are decreasing at $\bar{N}$, because the positive effect on the demand is compensated by the negative effect on the price, then there is a one-to-one relation between N and P^D, so that the platform rather than maximizing profits over P^D can do so directly over N. The platform will hence optimally choose N to maximize its profits given by

$$\Pi = P^U(N)\, t^d(N) + \left(p(N)\, t^d(N) - F\right) N. \tag{8}$$

From the expression above it is clear that an increase in N affects the profits of the platform in two ways, through the profits made on users (first term in (8)) and through the profits made on the developers (second term in (8)). How these effects depend on the degree of substitution between the applications that developers offer is quite clear when looking at the profits made on the developers' side. If substitution is strong, their profits, gross of P^D, are lower, then the surplus that the platform may extract from them is also lower (or even negative if it is optimal for the platform to subsidize the developers, i.e., $P^D < 0$). Regarding the profits made on the users' side, recall that both $\tilde{P}^U$ and $\hat{P}^U$ are increasing in N. In addition, the positive effect of entry on $\tilde{P}^U$ and $\tilde{t}^d = \frac{\tilde{P}^U}{k}$ is higher when substitution between developers is higher (whenever $\frac{\partial \tilde{p}}{\partial N} = V''(N)$ is high). When N increases $\widetilde{p}$ decreases, and this additional effect is taken into account by the platform when allowing access to the developers, becoming an additional incentive to promote entry. The optimal level of entry will depend on the margin that binds.

If demand margin binds, the platform will optimally choose $\hat{N}$ such that it solves

$$\frac{V(N)\, V'(N)\,(2N+1) - (V(N))^2 \frac{N}{N+1}}{2k\,(N+1)^2} = F, \tag{9}$$

[26] Farrell and Klemperer (2004) state that an indirect network effect arises whenever the indirect benefit outweighs any direct loss from more participation by one's own side. Thus, following this definition, there is an indirect network effect among developers as long as profits are increasing in N.

whereas if competition margin binds, it will choose $\tilde{N}$ such that

$$\frac{V\left(\tilde{N}\right)V'\left(\tilde{N}\right)-\left(V'\left(\tilde{N}\right)\right)^{2}\tilde{N}\left[1+\varepsilon_{v}\right]}{2k}=F. \tag{10}$$

The discussion above is the content of next lemma.

Lemma 5 *Let $G(N)$ $(H(N))$ stand for the developers' gross profits when demand (competition) margin binds,* and *let N^{*} be such that if $N<N^{*}$ the demand margin binds and if $N>N^{*}$ the competition margin binds. Assume* $H(N^{*})>G(N)$ for all N.[27] *The patterns of equilibrium entry in a proprietary platform will depend on the binding margin and the size of $\bar{N}$. In particular:*

i) If $H(N^{*})>G(N)$ for all N and *$\bar{N}>N^{*}$, in any stable equilibrium of developers' entry the competition margin will always bind and the level of entry will be*

$$N=\begin{cases}\bar{N} & \text{if} \quad H'\left(\bar{N}\right)>0\\ \min\left(\tilde{N},\bar{N}\right) & \text{if} \quad H'\left(\bar{N}\right)<0,\end{cases}$$

where $\tilde{N}$ solves (10).

ii) If $\bar{N}<N^{*}$, the level of entry will be

$$N=\begin{cases}\bar{N} & \text{if} \quad G'\left(\bar{N}\right)>0\\ \min\left(\hat{N},\bar{N}\right) & \text{if} \quad G'\left(\bar{N}\right)<0,\end{cases}$$

where $\hat{N}$ solves (9).

Proof: See Appendix A.

When the platform is open there are no platform prices to affect agents decisions (recall that now $P^{U}=P^{D}=0$), so that developers will enter until their profits are zero, i.e.,

$$p(N)\,t^{d}(N)-F=0.$$

Lemma 6 *Let $G^{o}(N)$ $(H^{o}(N))$ stand for the developers' gross profits when demand (competition) margin binds in an open platform and let N^{o*} be the N that determines the binding margin. Then,*

i) $G'^{o}\left(\hat{N}^{*}\right)=G'\left(\hat{N}^{*}\right)=0$

ii) $\hat{N}^{*}=N^{o*}<N^{*}$

iii) $G^{o}\left(\hat{N}^{*}\right)=H^{o}\left(\hat{N}^{*}\right)$

Proof: See Appendix A.

Point *i)* implies that the maximum in gross profits when demand margin binds occurs at the same N in both types of platforms. Point *ii)* implies that

[27] This is not a restrictive assumption, all the surplus functions that we are considering here satisfy it.

if demand margin binds, gross developers profits are increasing. Whereas, if competition margin binds, profits may be increasing or not. Note that a comparison of outcomes under open and proprietary platforms is not direct for the range of $N \epsilon \left(\hat{N}^*, N^*\right)$ as competition margin will bind under an open platform whereas the demand margin binds under a proprietary platform. Finally, point *iii)* shows that developers' profits are continuous at the point where the change from a margin to the other occurs.

If gross profits are increasing at $\bar{N}$, then $\bar{N}$ developers will entry. If not, the number of developers is determined by

$$\frac{V\left(\tilde{N}^o\right)V'\left(\tilde{N}^o\right) - \left[V'\left(\tilde{N}^o\right)\right]^2 \tilde{N}^o}{k} = F. \tag{11}$$

The next proposition compares the levels of entry that occur in each case and the effect on users' welfare.

Proposition 3 *i) If demand margin binds, a proprietary platform and an open platform will provide the same level of N, so that the latter will generate more welfare for users.*

ii) If competition margin binds a proprietary platform may generate a larger number of applications and higher welfare to users than an open platform.

Proof See Appendix A.

For comparison purposes, consider now the problem solved by a benevolent social planner. She would choose the optimal number of applications, N^{FB}, to maximize social welfare given by

$$W^* = \int_0^t V(N)\,dz - \int_0^t kzdz - FN,$$

where $t^{FB} = \frac{V\left(N^{FB}\right)}{k}$.

The first order necessary condition yields the first best allocation,

$$\frac{V\left(N^{FB}\right)V'\left(N^{FB}\right)}{k} = F. \tag{12}$$

Condition (12) that determines the first best level of N equalizes the marginal benefit with the marginal cost of an additional application. The former is the marginal utility enjoyed by users ($V'\left(N^{FB}\right)$ times the size of the market t^{FB}), whereas the latter is the fixed cost of producing one more application. Then if $\bar{N} < N^{FB}$ social planner chooses $\bar{N}$ and chooses N^{FB} otherwise.

As long as $\bar{N} < \hat{N}$ entry is $\bar{N}$ and equals N^{FB}. The same occurs when competition margin binds and $\bar{N} < \tilde{N}$. Then, when the effect of N on platform profits is strong (and this is more likely when e_v is high) the platform will tend to generate the same level of entry as the social planner.

Proposition 4 *Assume* $\bar{N} > \max\left(N^{FB},\ \hat{N},\ \tilde{N}\right)$. *If demand margin binds, a proprietary platform chooses a level of N smaller than the first best. However, if competition margin binds the comparison is not conclusive.*

Proof: See Appendix A.

Proposition 3 and 4 yield some insights into policies regarding the emergence of open source platforms competing with platforms such as Windows (Linux is the classic one, but there are also some others like Google which offer programs for free). In contrast, we do not observe the emergence of open platforms in the market of video consoles. The propositions above suggest that policy makers should promote open source in platforms like operating systems but not necessarily in those like video consoles.

5 Integration and the margin

Assume now that the platform firm can also develop its applications at zero marginal cost and at a fixed cost F per application. Then, if the platform is integrated, meaning that one firm produces the platform and the N applications, its system price will be

$$P^I = \frac{V(N)}{2}$$

and profits will be

$$\Pi^I = \left(\frac{V(N)}{2}\right)^2 \frac{1}{k} - FN.$$

We have shown that when integration is absent and demand margin binds, the resulting system price is $\frac{V(N)}{2}\left(1 + \frac{N}{N+1}\right)$ which is larger than P^I. The rational behind the result is clear: under separation there is a double marginalization as neither the platform nor the developers take into account the reduction of sales of the others when raising the price so that an inefficiently large price arises.

However, if integration is absent and competition margin binds, the resulting system price is $\frac{V(N)+V\'(N)N}{2}$ which gets close to P^I as $V\'(N)$ gets close to zero, which is the case when applications are very substitutes.

Proposition 5 *Inefficiencies of disintegration tend to disappear as long as competition margin binds and applications are near substitutes.*

Consider total profits of the firm. If demand margin binds, these are

$$\Pi^{DMB} = \left(\frac{V(N)}{2}\right)^2 \frac{1}{k}\left(\frac{2N+1}{N^2+2N+1}\right) - FN < \Pi^I,$$

so that the platform will always prefer being integrated in order to get developers to aware of the impact of their pricing strategies on the other developers and on the platform profits. Note that under separation even if the platform can control N through P^D, it can not control the price developers set.

If competition margin binds, profits are

$$\Pi^{CMB} = \left(\frac{V(N) - \widetilde{p}N}{2}\right)^2 \frac{1}{k} + \widetilde{p}\left(\frac{V(N) - \widetilde{p}N}{2}\right) N\frac{1}{k} - FN.$$

Again, as long as $\widetilde{p} = V'(N)$ tends to zero (because the extent of substitutability among applications is great or N is very high), profits tend to Π^I.[28]

The results above are consistent with the observed phenomena that initially platforms are vertically integrated and later disintegrate. Recall from corollary 1 that there exists N^* which determines the margin that is going to be binding. When the industry is less developed (initial steps of the industry with N low) the platform strictly prefers being integrated. As the industry evolves and the number of developers available in the market increases, the competition margin is likely to bind, prices of applications will be $V'(N)$, decreasing in N, and at this stage of the industry, the platform will be more willing to disintegrate.[29] As the market of developers matures and becomes more competitive, the firm can concentrate on producing only the platform. Note that other alternative explanations are offered in the literature for the phenomena of vertical disintegration that not can be explained within this model. For instance, Stigler notes that firms need to arise vertically integrated since technology is not familiar in the market. When the industry grows, production process are well known and scale of the market allows specialization, such that disintegrating is profitable. [30] Another different explanation for no integration is given by Gawer and Henderson (2005), when discussing Intel's strategy. They suggest that managers were aware of how important the generation of complements was to the success of Intel's business; however, although it is in the interest of the platform to enter complementary markets, the platform knows that this could discourage entry by new firms.[31] A more trivial explanation comes from the fact that the platform does not always possess the requisite capabilities to produce some of the complementary goods.[32]

5.1 Partial Integration

A widely observed fact in software industries is that some computer software are clearly more useful or more commonly used than others, Office software and

[28] In particular, the necessary condition is that $V'(N)\,N$ be decreasing, i.e., $\varepsilon_v(N) > 1$.

[29] PDA's were born as "smart agendas" offering a limited number of applications. Then, they evolved to become "small computers". Something similar has ocurred in the mobile phone industry. In addition to the traditional communication service, today they allow for hundreds of applications. See "What is a Window Mobile" in www.microsoft.com.

[30] George Stigler, "The division of labor is limited by the extent of the market", Journal of Political Economy 59 (June 1951), quoted by Evans, et. al.(2006).

[31] Dave Johnson, a director of Intel, explained: "The market segment gets hurt if third parties think: "Intel, the big guys, are there, so I do not want to be there..."... it is not what we want, because we are trying to encourage people to do these complementary things". Gawer and Henderson (2005), pp. 18.

[32] Claude Leglise, director of the Developer Relation Group, responded: "Intel has no corporate competence in entertainment software. We do not know how to do video games, so forget it". Gawer and Henderson (2005), pp. 13.

Messenger are illustrative examples. At the same time, some video games are the most popular (killer games) in the market, so that applications' contributions to total surplus may be different. To incorporate this feature into our model, in what follows we allow applications to be heterogeneous.

Assume that each application i has a contribution $N_i \in [0, N]$, with the normalization

$$\sum_{i=1}^{N} N_i = N.$$

Note that $N_i = 1$ will bring back the homogeneity we have considered so far. Let us further assume that $\frac{\partial N_i}{\partial i} > 0$ and let us define $V(\cdot)$ by $V\left(\sum_{i=1}^{N} x_i N_i\right)$, where $x_i = 1$ if user buys application i and $x_i = 0$ otherwise. The next lemma is inspired in proposition 6 in Lerner and Tirole (2004).

Lemma 8 *Assume that gross surplus of users by applications is* $V\left(\sum_{i=1}^{N} x_i N_i\right)$, *where* $x_i = 1$ *if users buy application* i *and* $x_i = 0$ *otherwise, with* $\frac{\partial N_i}{\partial i} > 0$. *Then, there is a mass* $0 \leq n \leq N$ *of developers that are constrained by the competition margin and charge a price* $\widetilde{p}_i = V_i'$, *their marginal contribution to the total surplus. The rest of the developers are constrained by the demand margin and all of them set the same price* $\widehat{p} = \frac{V(N) - P^U - \int_0^n \widetilde{p}_i d_i}{N - n + 1}$. *Finally, the platform sets* $P^U = \frac{V(N) - \int_0^n \widetilde{p}_i d_i}{2}$.

When the platform decides P^U, it defines the value n, i.e., the mass of developers that will be constrained by the competition margin. For every $i \in [0, n]$ it must hold that $\widetilde{p}_i = V_i' < \widehat{p}$ and that V_i' is increasing in i. If $n = 0$, we have that every developer is constrained by the demand margin. Analogously, if $n = N$ every developer is constrained by the competition margin.

Proposition 6 *In the long run the platform will be partially integrated with the killer applications for which demand margin will bind, and will allow free entry for developers of other applications.*

This proposition may help us to explain why platforms are often partially integrated, most of them with the core application. Microsoft produces operating systems and some of the applications (i.e. Office package). Nintendo wrote Mario Brothers, the killer game of one of its consoles. In the US the proportion of games developed in house is about 10% for GambeCube and 8% for PlayStation and Xbox.

6 Platform Competition: the role of outside options

Up to now we have assumed that a monopolist platform (either proprietary or open) provides a good with no competition at all. Nevertheless in many industries, either open and proprietary platforms coexist, or there are several for-profit platforms competing to attract both users and developers. We now extend our basic framework by assuming that a proprietary and an open platform operate in the same industry.[33] Our aim here is to analyse how a firm that offers a proprietary solution will respond to changes in an outside option that provides a positive surplus or profit to their clients (i.e., to users and developers). We analyze how the monopoly reacts in terms of prices and we abstract from other strategies such as investment.[34]

A user who purchases the open platform gets a net surplus $v = V(Z) - h$, where $V(Z)$ measures the utility users derive given the applications written for the open platform and h is an exogenous cost (interpreted as a transportation cost or a cost of learning to use this outside good). Consequently, users will purchase the proprietary platform as long as

$$V(\mathcal{N}) - kt \geq v > 0.^{35}$$

The game is solved as in previous sections but considering v. In what follows we provide some comparative statics analyses to changes in v in order to study its impact on users welfare. We start assuming that the competition margin binds. Then, we move to an scenario where the demand binds. We restrict the analyses to values of N for which developers' profits are decreasing so that the proprietary platform can affect entry.

Consider the impact of a change in v on developers' profits and on the number of applications. The condition that arises when the platform at the first stage maximizes with respect to N is

$$\frac{V(N)V'(N) - (V'(N))^2 N \ [1 + \ \varepsilon_v]}{2k} - \frac{V'(N)v}{2k} = F. \tag{13}$$

and from the comparison with equation (10) it follows that the monopolist will reduce enty due to the term $\frac{V'(N)v}{2k}$. This term is decreasing in N and smaller as long as applications are very substitutes.

It means that developers of video consoles may not have incentives to increase the value of v (i.e., writing applications for computers or online games) because the monopolist may react reducing the level of entry and thus the incentives for them. However, this response will not be important whenever the games are near substitutes.

By taking into account its impact on entry, the next proposition provides results on the impact of outside options on users surplus.

Proposition 7 *Whenever the competition margin binds, a higher outside option value for the users may lead to a decrease in their surplus. In contrast,*

[33] Since the open platform is considered non-profit, we will assume that it behaves myopically and hence does not play a best response against the pricing strategies by the proprietary platform. In contrast, the proprietary platform will take into account the presence of the open platform when deciding upon its pricing strategies.

[34] Economides and Katsamakas (2006b) study investment incentives of platforms and developers in a proprietary system and in an open source one.

[35] Note that v is used to proxy for the extent of product market competition.

if the demand margin binds, the impact on users' surplus of a higher outside option will generally be positive.

Proof. See Appendix A.

Regarding the other side of the market, we now assume that developers can obtain a profit of w when writing applications for the open platform. Note that nothing changes if developers are allowed to write for both platforms (i.e., to multihome). In that case developers get a higher total profit but the strategies of the proprietary platform do not change. Results are different if we assume that developers are forced to choose one of the platforms (i.e., to singlehome) due, for instance, to contractual arragements. Thus, developers will enter the market of the proprietary platform as long as

$$\pi_i = p_i t^d - F - P^D \geq w.$$

The effect of an increase in w is analogous to an increase in the fixed cost, so it clearly leads to a reduction in the level of N.

Proposition 8 *If the competition margin binds, an increase in the outside option of the developers will always reduce the users' surplus. However, if the demand margin binds, an increase in the outside option of the developers will always increase the users' surplus.*

Proof. See Appendix A.

We have shown that reinforcing competition pressure for developers when competition margin binds leads to a reduction in the users welfare. Results are quite different when demand margin binds. Promoting the benefits that writing for Linux has for the developers (sometimes interpreted as a "reputation effect"[36]) would be in favour of the users.

Let us provide an illustrative example. Consider $V(N) = N^{\beta}$ where $\beta = 0.45$ and a fixed cost $F = 0.14$. A value $\beta = 0.45$ determines that competition margin is binding as long as $N > 2.3$ and we restrict the analyses to this range of N. For a value $v = 0.1$, the surplus of the users is 0.71 whereas for an increase $\Delta v = 0.05$, the new users surplus is 0.59. It represents in terms of elasticities that a 1% increase in the users outside option implies a 19% decrease in the users surplus.[37]

To compare the effects of w and v, consider now $\beta = 0.25$ (so that competition margin binds as long as $N > 0.8$) and a fixed cost $F = 0.075$. Given the initial values $w = v = 0.1$, we find that a change in v (i.e., $\Delta v = 0.05$) exerts a direct impact on P^U equal to $\frac{\partial P^U}{\partial v} = -\frac{1}{2}$, whereas there is no direct impact when w changes (i.e., $\Delta w = 0.05$). However, when we compute the total effect, considering the indirect one by the effect on N, we find that $\frac{\partial \widetilde{P}^U}{\partial v}\frac{v}{\widetilde{P}^U} = -0.01$ and $\frac{\partial \widetilde{P}^U}{\partial w}\frac{w}{\widetilde{P}^U} = -0.03$, meaning that, under these parameters, the monopolist decides to reduce the price more for users when there is an outside option for the developers than when there is one for the users themselves.

[36] Economides and Katsamakas (2005.b). Other motivations are explained in "Microsoft vs. Open Source: Who Will Win?- HBS Working Knowledge, June 2005.

[37] The exercise has been computed assuming $k = 1$.

7 Conclusions

We have solved a model that provides some results for a better understanding of the two-sided pricing strategies of a platform that sells a good whose value depends on the applications sold in a market of developers. We note that when setting prices the developers are constrained by two margins: the demand margin and the competition margin. What margin is binding depends on the number of applications in the market and on the level of substitutability among them.

We find that if the demand margin binds, policy makers should promote open source platforms. However this is not necessarily the case when competition margin binds.

We consider the case where applications are asymmetric in the users' surplus and we find that in the long run the platform will remain integrated with the applications for which demand margin binds and will leave for third-party developers the production of applications for which competition margin binds.

Finally, we find that it would not be in the interest of the users to promote the value of outside options for the platform when competition margin binds. However, an increase in the value of the outside option for developers would have a positive impact on the users surplus if demand margin binds.

References

[1] Armstrong, M. (2005), “Competition in two-sided markets”, forthcoming RAND Journal of Economics.

[2] Clements, M. and Ohashi, H. (2005), "Indirect Netwok Effects and the Product Cycle: Video Games in the US, 1994–2002", The Journal of Industrial Economics, 53, 4, pp. 515-542.

[3] Church, J. and Gandal, N. (1992), "Network Effects, Software Provision and Standardization", Journal of Industrial Ecomics, 40, 1, pp. 85-103.

[4] Church, J. and Gandal, N (1993), "Complementary Network Externalities and Technological Adoption", International Journal of Industrial Organization", 11, pp. 239-260.

[5] Church, J. and Gandal, N (2000), "Systems Competition, Vertical Merger and Foreclosure", Journal of Economics and Management Strategy, vol. 9, Issue 1, pp. 25-51.

[6] Church, J. Gandal, N. and Krause, D. (2003), "Indirect Network Effects and Adoption Externalities", CEPR WP 3738.

[7] Economides (2006), "Competition Policy in Network Industries: An Introduction", in Dennis Jansen (ed.), *The New Economy and Beyond: Past, Present and Future,* Edward Elgar, pp.

[8] Economides and Katsamakas (2006a), "Two-Sided Competition of Proprietary vs. Open Source Technology Platforms and the Implications for the Software Industry", Management Science vol. 52, No. 7, pp. 1057-1071.

[9] Economides and Katsamakas (2006b), "Linux vs. Windows: A Comparison of Application and Platform Innovations Incentives for Open Source and Proprietary Software Platforms", in Juergen Bitzer and Philipp J.H. Schroeder (eds.), *The Economics of Open Source Software Development*, Elsevier Publishers.

[10] Evans and Schmalensee (2001), "Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries", NBER WP 8268.

[11] Evans, D. Hagiu, A. and Schmalensee, R. (2004), "A Survey of the Economic Role of Software Platforms in Computer-Based Industries", CESIFO working paper.

[12] Evans, D. Hagiu, A. and Schmalensee, R. (2006), "Invisible Engines", MIT Press.

[13] Farrell, J. and Klemperer, P. (2005), "Coordination and Lock-In: Competition with Switching Costs and Network Effects", forthcoming in the Handbook of Industrial Organization.

[14] Gawer and Henderson, (2005), "Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel", NBER WP 11852.

[15] Kühn, K. and Vives, X. (1999). "Excess entry, vertical integration, and welfare", RAND Journal of Economics Vol. 30, No. 4, pp. 575-603.

[16] Hagiu, A. (2005). "Two-Sided Platforms: Pricing and Social Efficiency", Harvard Business School and Research Institute of Economy Trade and Industry working paper, Cambridge, Mass.

[17] Legros, P. and Stahl, K. (2002), "Global vs. Local Competition", CEPR WP 3333.

[18] Lerner, J. and Tirole, J. (2004). "Efficient Patent Pools", American Economic Review, 94, 3, pp. 691-711.

[19] Nocke, V., Peitz, M., Stahl, K. (2004), "Platform Ownership", CEPR, DP 465726.

[20] Rochet, J.C. and Tirole, J. (2006), "Defining Two-Sided Markets", forthcoming RAND Journal of Economics.

Appendix A

Proof of proposition 1

To show the result we compute the profits that each situation generates for the platform, then we compare them and deduce the optimal strategy for the platform. If the platform sets a price that satisfies $P^U < V(N) - \widetilde{p}(N+1)$, then the competition margin will bind for the developers and platform profits will be

$$\widetilde{\Pi}^{PU} = P^U \left[\frac{V(N) - \widetilde{p}N - P^U}{k} \right].$$

The price that maximizes profits, given the constraint, is

$$\begin{aligned} P^U &= \frac{V(N) - \widetilde{p}N}{2} \text{ if } \widetilde{p} < \frac{V(N)}{N+2}, \text{ and} \\ P^U &= V(N) - \widetilde{p}(N+1) \text{ if } \widetilde{p} > \frac{V(N)}{N+2}. \end{aligned}$$

If the platform sets a price such that $P^U > V(N) - \widetilde{p}(N+1)$, so that demand margin will bind for the developers, platform profits will be

$$\widehat{\Pi}^{PU} = P^U \left[\frac{V(N) - P^U}{k(N+1)} \right].$$

The price that maximizes profits, given the constraint, is

$$\begin{aligned} P^U &= \frac{V(N)}{2} \text{ if } \widetilde{p} > \frac{V(N)}{2(N+1)}, \text{ and} \\ P^U &= V(N) - \widetilde{p}(N+1) \text{ if } \widetilde{p} < \frac{V(N)}{2(N+1)}. \end{aligned}$$

Comparing above the profits we observe that if $\widetilde{p} < \frac{V(N)}{2(N+1)}$ the price that generates highest profits for the platform is $\widetilde{P}^U = \frac{V(N) - \widetilde{p}N}{2}$. If $\widetilde{p} > \frac{V(N)}{N+2}$, the platform will optimally choose $\widehat{P}^U = \frac{V(N)}{2}$. Finally, whenever the relevant interval is $\frac{V(N)}{2(N+1)} < \widetilde{p} < \frac{V(N)}{N+2}$, if $\widetilde{p} < \frac{V(N)}{N}\left[1 - \frac{1}{\sqrt{N+1}}\right]$ the platform will set $\widetilde{P}^U = \frac{V(N) - \widetilde{p}N}{2}$ and will set $\widehat{P}^U = \frac{V(N)}{2}$ otherwise. It follows that the competition margin will bind if $\widetilde{p} < \frac{V(N)}{N}\left[1 - \frac{1}{\sqrt{N+1}}\right]$, and this occurs whenever $e_v(N) < \left[1 - \frac{1}{\sqrt{N+1}}\right]$, as claimed.

Proof of corollary 1

Note that the function $\left[1 - \frac{1}{\sqrt{N+1}}\right]$ is increasing in N, equals zero at $N = 0$, and goes to one as N goes to infinity. Since $e_v(N) \epsilon (0,1)$, if $e_v(N)$ is a non increasing function, it will necessarily cross $\left[1 - \frac{1}{\sqrt{N+1}}\right]$. However, if $e_v(N)$ is an increasing function, a crossing point may not exist.

Proof of lemma 5

When the demand margin binds, developers will enter until profits are zero so that it is satisfied

$$\left(\frac{V(N)}{2(N+1)}\right)^2 \frac{1}{k} - F - P^D = 0.$$

If competition margin binds, the developers zero profit condition will be

$$V'(N)\left(\frac{V(N) - V'(N)N}{2k}\right) - F - P^D = 0.$$

Consequently, let $G(N) = \left(\frac{V(N)}{2(N+1)}\right)^2 \frac{1}{k}$ and $H(N) = V'(N)\left(\frac{V(N)-V'(N)N}{2k}\right)$.

i) In a stable equilibrium, profits are zero and decreasing. Consider now an equilibrium such that $G(N) = F + P^D$ (so that demand margin binds). Since $G(N) < H(N^*)$, when $\bar{N}$ is sufficiently large a coalition of developers will enter to obtain (at least) profits $H(N^*)$, and the result follows.

Then,

1) if $H'(\bar{N}) > 0$ gross developers profits are strictly increasing so that entry is $\bar{N}$.

2) if $H'(\bar{N}) < 0$ gross developers profits are strictly decreasing so that the platform will choose $N = \min\left(\tilde{N}, \bar{N}\right)$, and the result follows.

ii) We must distinguish two cases. 1) If $G'(\bar{N}) > 0$ gross developers profits are increasing and entry is $\bar{N}$.

2) If $G'(\bar{N}) < 0$ gross developers profits are decreasing so that the platform will choose $N = \min\left(\hat{N}, \bar{N}\right)$, and the result follows.

Proof of lemma 6

Note that $G^o(N) = \left(\frac{V(N)}{(N+1)}\right)^2 \frac{1}{k}$ and $H^o(N) = V'(N)\left(\frac{V(N)-V'(N)N}{k}\right)$. Result i) follows trivially. Note that the concavity of V ensures that $\hat{N}^*$ always exists. ii) Note that $\hat{N}^*$ solves $V'\left(\hat{N}^*\right) = \frac{V\left(\hat{N}^*\right)}{\hat{N}^*+1}$. The equality $\hat{N}^* = N^{o*}$ follows from the fact that in an open platform competition margin binds as long as $V'(N) < \frac{V(N)}{(N+1)}$. To prove that $\hat{N}^* < N^*$ recall from corollary 1 that N^* satisfies $V'(N^*) = \frac{V(N^*)}{N^*}\left(1 - \frac{1}{\sqrt{(N^*+1)}}\right)$. Since $V'(N)$ is decreasing and $\frac{V(N)}{(N+1)} > \frac{V(N)}{N}\left(1 - \frac{1}{\sqrt{(N+1)}}\right)$ for all N, it follows that $\hat{N}^* < N^*$. Part iii) follows from straigthforward computations.

Proof of Proposition 3

i) The first statement follows from point i) in lemma 7 (profits of developers are increasing under both regimes for the same range of N) then in both cases entry will equal $\overline{N}$. If demand margin binds, with a proprietary platform the system price is $P^U + pN = \frac{V(N)}{2} + \frac{V(N)}{2(N+1)}N$, that is higher than $\frac{V(N)}{N+1}N$, the system price with an open platform, so that the second statement follows.

ii) From the comparison between (10) and (11), it follows that as long as $e_v > \frac{1}{1-\varepsilon_v}$ (i.e.$V'(N) - V''(N)N > \frac{V(N)}{N}$), a profit platform yields a higher N than the open platform. The second statement is proven by the fact that when competition margin binds, the users' surplus (net of kt) is increasing in N. The condition $e_v > \frac{1}{1-\varepsilon_v}$ imposes that $\varepsilon_v < -1$ since $e_v < 1$. An example for which a proprietary platform yields a higher N than an open platform is given by

$$V(N) = \begin{cases} (1 - \exp(-0.05N)) & \text{if} \quad N \leq 7 \\ (0.8 - \exp(-0.1N)) & \text{if} \quad N > 7, \end{cases}$$

with $F = 0.0045$.[38] The proprietary platform chooses $N \simeq 25$ whereas the open chooses $N \simeq 24$. The competition margin binds for all $N > 13$.

Proof of Proposition 4

The first statement follows from the comparison between (9) and (12). The second statement follows from the comparison between (10) and (12) and the fact that as long as $-e_v(1+\varepsilon_v) > 1$ (i.e.$-V''(N)N > \frac{V(N)}{N} + V'(N)$) the proprietary platform may generate excess of entry. As in the previous proof, the condition $-e_v(1+\varepsilon_v) > 1$ requires $\varepsilon_v < -1$ and it is more stringent than the condition in the proof of proposition 3. It does not contradict the condition to be in the competition margin $e_v < 1 - \frac{1}{\sqrt{N+1}}$, nor the condition for a maximum in the social planner problem, $e_v <\mid \varepsilon_v \mid$, nor the fact that $V(N)$ is concave.

Proof proposition 7

Given a user t, if competition margin binds, her surplus gross of kt is equal to $V(N) - pN - P^U$. We observe that this surplus will be increasing (decreasing) in v as long as $\left[1 - V''(N)N\frac{\partial N}{\partial v}\right] \lesseqgtr 0$, and the first statement follows. To prove the second statement note that if demand margin binds, users' prices are: $\widehat{p} = \frac{V(N)-v}{2(N+1)}$ and $\widehat{P}^U = \frac{V(N)-v}{2}$. The platform optimally chooses the N that maximizes profits

$$\frac{1}{k(N+1)}\left(\frac{V(N)-v}{2}\right)^2 + \left(\left(\frac{V(N)-v}{2(N+1)}\right)^2\frac{1}{k} - F\right)N.$$

Note that expression (9) can also be written as

$$\frac{V(N)}{2k(N+1)^2}\left[V'(N)(2N+1) - V(N)\frac{N}{N+1}\right] = F,$$

and when the outside option appears it transforms in

$$\frac{V(N)}{2k(N+1)^2}\left[V'(N)(2N+1) - V(N)\frac{N}{N+1}\right] -$$
$$\frac{v}{2k(N+1)^2}\left[V'(N)(2N+1) - (2V(N)-v)\frac{N}{N+1}\right] = F.$$

[38] The equilibrium occurs at $N > 7$ so that $V(N) = (0.8 - \exp(-0.1N))$. Note that $V(N) = (1 - \exp(-0.05N))$ if $N \leq 7$ ensusres that $V(0) = 0$.

So, the effect on N of v will depend on the second term. If this is positive, the monopolist will reduce N whereas if this is negative the impact on N will be positive. Both situations may occur; however since platform profits are lower for each N, the most likely case is that the monopolist will reduce N.

Now, note that whenever demand margin binds and there is an outside option v, the users surplus, gross of the cost kt, equals

$$V(N) - \widehat{p}N - \widehat{P}^U = V(N) - \frac{[V(N) - v]\,N}{2(N+1)} - \frac{[V(N) - v]}{2}$$

The first derivative of this surplus with respect to v is going to be positive as long as $\frac{\partial N}{\partial v}\left[\frac{V'(N)}{N+1} - \frac{V(N)}{(N+1)^2} + \frac{v}{(N+1)^2}\right] + \frac{2N+1}{N+1} > 0$. The second term of the left hand side of the inequality is always positive. However, the term in brackets is negative as long as $\epsilon_V < \left[1 - \frac{v}{V(N)}\right]\frac{N}{N+1}$, and this is the case along the relevant range of N (when gross developers profits are decreasing). The first term will be positive if $\frac{\partial N}{\partial v} < 0$ (the most likely case) and negative otherwise, so that the result follows.

Proof proposition 8

If the competition margin binds the effect of an increase in w on users surplus is equal to $-\frac{V'''(N)N\frac{\partial N}{\partial W}}{2} < 0$ and the first statement follows. To prove the second statement, note that the surplus is decreasing in N if $e_v < \frac{N}{N+1}$ and this occurs for the relevant range of N. Given that $\frac{\partial N}{\partial w} < 0$, the second statement follows.

Merchant or Two-Sided Platform?

ANDREI HAGIU *

Strategy Unit, Harvard Business School

Abstract

This paper provides a first pass at comparing two polar strategies for market intermediation: “merchant” mode – buying from sellers and reselling to buyers - and “two-sided platform” mode – enabling affiliated sellers to sell directly to affiliated buyers. The merchant mode is more profitable when the chicken-and-egg problem for the two-sided platform is more severe and when the degree of complementarity among sellers’ products is higher. The platform mode is preferred when seller investment incentives are important or when there is asymmetric information regarding seller product quality. We discuss these tradeoffs in the context of several prominent digital intermediaries.

1 Introduction

With ever more sophisticated logistics and the rise of information technologies, intermediaries and market platforms have become increasingly ubiquitous and important agents in the digital economy. Sites such as Alibaba, Amazon, eBay, Google Video, i-mode[1], iTunes, Rakuten[2], YouTube, etc. help connect tens of thousands of sellers to tens of millions of buyers. While market intermediation is not a new phenomenon, the digital economy has revealed that there can be two polar types of intermediaries: “merchants”, who acquire goods[3] from sellers and resell them to buyers, and “two-sided platforms”, who allow “affiliated” sellers to sell directly to “affiliated” buyers.

This paper is a first pass at clarifying the differences between these two forms of market organization by intermediaries and the economic tradeoffs involved, which we show are more profound than the presence of absence of indirect network effects.

The main difference between the classic form of market intermediaries – which we will call merchants from now on – and two-sided platforms is that pure merchants, by taking possession of sellers’ goods, take full control over their sale to consumers. By contrast, pure two-sided platforms leave that control entirely to sellers and simply determine buyer

---

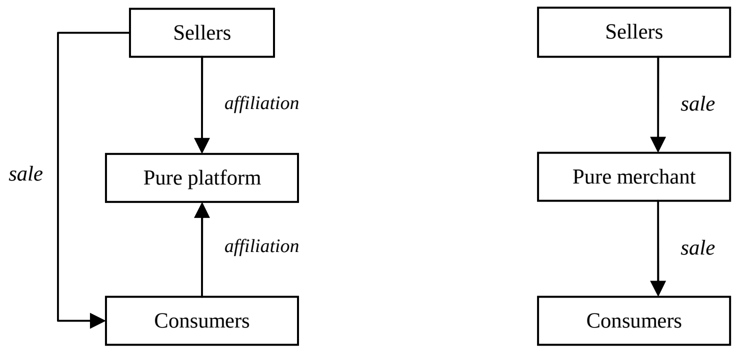
* Harvard Business School, Morgan Hall #212, Cambridge, MA 02163, USA. Email: ahagiu@hbs.edu I am grateful to two anonymous referees for improving this paper. All errors are mine.

[1] The world’s most successful mobile Internet service, launched in 1999 by NTT DoCoMo, Japan’s leading mobile operator. As of January 2007, i-mode connects more than 50 million users to 100,000 sites.

[2] Japan’s largest online shopping mall, offering access to more than 18,000 merchants, selling more than 18 million products.

[3] This may either mean physical ownership or simply the right to sell (exclusive or not) for digital goods.

and seller affiliation with a common marketplace. The following figure depicts this difference:



To give a few examples, retailers like Walmart and Amazon are (mostly) merchants. At the opposite end of the spectrum, eBay is a pure two-sided platform. More interesting are intermediaries, such as Apple's iTunes digital music store, which exhibit both platform and merchant features and therefore lie in-between these two extremes. Although Apple does not take physical or full legal "possession" of the songs it distributes (the rights remain with music publishers), it does obtain the right to repackage and price them as it sees fit on iTunes. The 99 cents per song policy is entirely Apple's unilateral decision (and is increasingly contested by music studios), designed to provide simplicity of usage to consumers and promote sales of Apple's associated digital music player, the iPod[4]. Thus, even though the combination iPod/iTunes exhibits two-sided indirect network effects (music publishers obtain higher profits by signing a distribution deal with Apple's iTunes when more consumers buy iPods, and vice versa), the extent of control over pricing and distribution that Apple maintains make iTunes more similar to a merchant such as Walmart, rather than a pure two-sided platform, such as eBay.

Using a simple framework, we formalize the economic tradeoffs between the pure (one-sided) merchant "mode" and the pure two-sided platform "mode". First, we show that, unlike the pure merchant mode, the pure two-sided platform mode exhibits indirect network externalities between sellers and consumers (buyers). As a consequence, sellers may be unwilling to affiliate with the platform because they anticipate other sellers will not do so, leading to low consumer demand for the platform, which ex-post justifies a non-affiliation decision. In such cases, the merchant model helps "break" these unfavorable seller expectations by eliminating indirect network effects and achieves higher total profits. On the other hand, however, the merchant model may entail higher costs per seller – inventory, risk, management, etc. – so that a platform model is more desirable when one wants to achieve higher product variety (and the risk that seller hold unfavorable expectations is not too high).

Second, the merchant mode also dominates the platform mode whenever there are significant complementarities and/or substitutability between sellers' products that sellers pricing independently (on a platform) do not internalize. By taking control over pricing (as well as advertising, distribution, bundling, etc.) decisions, the merchant can create more

[4] Indeed, Apple's profits from the iPod/iTunes combination come largely from sales of iPod, where the company enjoys margins higher than 20 percent. On the music side, it is estimated that Apple makes less than 10 cents for every song.

value and extract more profits from consumers. In addition, we also show that a platform faces an inherent hold-up problem when contracting with sellers before selling to consumers, since it does not take into account seller profits when it sets its access price to consumers, therefore tends to excessively limit consumer adoption. By definition, a merchant fully internalizes seller profits[5].

Third, whenever sellers can make quality enhancements to their products *after* contracting with the intermediary and before selling to consumers, the platform mode is more desirable since it preserves seller investment incentives by making them the residual claimants of those investments. Similarly, when quality is uncertain, the intermediary is better off devolving the corresponding risk at least partially back to the sellers, by using a platform mode.

1.1 Related literature

In distinguishing between the merchant and two-sided platform modes of market intermediation, this paper connects for the first time two previously separate strands of economics research. On the one hand, the literature on market microstructure and intermediation (Rubinstein and Wollinsky (1987), Stahl (1988), Biglaiser (1993), O'Hara (1995), Spulber (1996a) and (1996b), Rust and Hall (2001)) concerns itself with the merchant form of intermediation. The focus has been on understanding how intermediaries help homogeneous product markets clear by setting bid and ask prices to match supply with demand and by providing liquidity (O'Hara (1995)), as well as on deriving the welfare effects of the presence of intermediaries. Several papers have also studied the role of such intermediaries in quality certification and shown that they exhibit economies of scale when reducing adverse selection concerns (Biglaiser (1993)). Some papers also distinguish between marketmakers – which post publicly observable bid and ask prices – and middlemen – whose bid and ask prices can only be discovered through costly search (Rust and Hall (2001)). However, all intermediaries studied in this literature are assumed to buy products from sellers and resell to buyers at the posted prices, subject to the constraint that demand does not outstrip supply for any intermediary at any point in time[6]. In addition, there are no meaningful affiliation decisions of buyers or sellers with intermediaries since each can only conduct one-shot spot transactions with an intermediary: buy or sell a unit of the product at the announced ask (respectively bid) price. Hence, there are never any indirect network effects between buyers and sellers. Also, given the focus on homogeneous product markets, the only relevant transaction variable is price and it is *always* determined by the intermediary.

On the other hand, the recent literature on two-sided markets has focused exclusively on pure two-sided platform intermediaries, emphasizing the indirect network effects which arise between the two sides of the market when the latter have to affiliate with the platforms in order to be able to transact with one another (Armstrong (2006), Caillaud and

---

[5] This is true in the simple framework we present in this paper because we assume away the existence of seller transactions with consumers outside of the intermediary. If such transactions existed (for example, through alternative sales channels or simply brand name), then even the merchant mode would leave certain externalities on seller profits uninternalized.

[6] This is mainly because the literature was focusing on brick-and-mortar intermediaries such as retailers or wholesalers channeling physical goods, or stock exchanges, where short positions cannot – usually – outstrip long positions for long. By contrast, in many of the digital markets we have in mind, such considerations are largely irrelevant: inventory is not an issue with digital goods.

Jullien (2003), Evans (2003), Hagiu (2006a) and (2006b), Rochet and Tirole (2003) and (2006), Schmalensee (2002)). Virtually all of these papers focus on the pricing structure chosen by two-sided platforms in order to internalize (partially) network externalities. While this literature takes the existence of indirect network effects as given, we show that it crucially depends on the nature of contracts between the intermediary and sellers. With pure buy-out contracts (merchant mode), the externalities disappear entirely, since they are fully internalized by the intermediary. More importantly, we show that there is in fact a continuum of intermediary types between a pure merchant and a pure two-sided platform, depending on the extent of control over buyer-seller interactions left to sellers. “Control” can be thought of as encompassing three important dimensions: control over strategic variables (pricing, advertising, distribution, etc.); sharing of economic risk (is the risk borne by the sellers or by the intermediary?) and “ownership” of buyers (how salient are individual sellers’ “brands” relative to the intermediary’s “brand” when buyers make their affiliation decisions?). A pure two-sided platform leaves control to sellers, whereas a merchant takes over full control.

As a consequence, our framework suggests an implicit definition of “two-sidedness”, relying on the division of control between sellers and intermediaries, rather than on the effects of the pricing structure chosen by the intermediary, as is the case with the definition proposed by Rochet and Tirole (2006). In contexts with platforms intermediating transactions between buyers and sellers, the Rochet-Tirole definition is overly inclusive with respect to ours: indeed, even the pricing structure chosen by a pure merchant (that is, its bid and ask prices) affects the total volume of transactions conducted. The key difference is that the Rochet-Tirole framework presumes a platform intermediating transactions between buyers and sellers *without* taking full control over buyer-seller transactions. Thus, there is a sense in which our framework augments the Rochet-Tirole definition by identifying a space in which the boundary between what is two-sided and what is not depends on a richer set of factors than indirect network externalities and pricing structure effects. It is worth emphasizing again that our framework implies that two-sidedness is not a 0-1 notion: rather, there is a continuum of forms of intermediation. The position along this continuum can be thought of as a strategic decision for the intermediary, involving the tradeoffs that we explore in the rest of the paper.

The remainder of the paper is organized as follows. We present our basic model in section 2, then we proceed to use variations of this model in order to formalize the respective effects of indirect network externalities, product complementarities/substitutability and investment incentives on the tradeoff between merchants and two-sided platforms in sections 3,4 and 5. Section 6 uses the insights drawn from the formal analysis to discuss several real world examples. Section 7 concludes.

2 Basic modeling framework

There is one intermediary which makes it possible for n identical sellers to deliver their products to consumers (buyers). The intermediary can choose between two ways of functioning: a *merchant* (one-sided) mode and a *two-sided platform* mode. Under the merchant mode, the intermediary buys sellers’ products by offering a buyout bid B^S for each seller and resells the goods to consumers for an individual price $p^M(n)$ that it chooses. In the two-sided platform mode, the intermediary charges each seller an “access” or "affiliation" fee P^S, in exchange for which sellers can sell their goods directly to the consumers affiliated with the intermediary for an individual price $p^P(n)$. The price $p^P(n)$ is determined by competition between n sellers for the consumers affiliated with the two-sided platform. We assume each seller is of measure 0 (that is, n is a continuum), so that it does not take into account the effect of its price on overall user demand for the platform. We also implicitly assume unaffiliated sellers cannot sell their products to consumers (affiliated or unaffiliated). This last assumption rules out competition among intermediaries – this would introduce strategic effects, some of which are explored in Hagiu and Lee (2006). We also rule out direct sale by sellers to consumers. This possibility would introduce externalities exerted by the direct sales channel over the sales through the intermediary, which would complicate the analysis significantly. We leave this extension of our model for future research.

Regardless of the mode – platform or merchant – chosen by the intermediary, consumers have to obtain access to (affiliation with) the intermediary in order to be able to purchase the sellers' products: we denote by P^C the consumer access fee charged by the intermediary. This assumption is made in order to keep the two modes of intermediation as similar as possible to each other and focus the comparison on other factors. There is no a priori reason why a pure two-sided platform should be in a better position to charge consumer access fees than a pure merchant, or vice versa. In many cases, this fee is equal to 0: for instance, Amazon does not charge users for browsing its website.

Consumers buy either from the intermediary itself if the latter chooses to function in a merchant mode, or directly from the sellers affiliated with the intermediary, if the latter chooses the two-sided platform mode.

Let $V(n)$ denote the gross utility each user derives from having access to n products through the intermediary. We assume $V(n)$ is increasing and either linear or strictly concave in n. The implicit assumption is that each user will consume all products available. One could write a richer model, in which each user consumes only a subset of the products available, but our main insights remain unchanged. Denote then by $F(.)$ the cumulative distribution function of a horizontal differentiation parameter across the population of users – which can be interpreted as the distance to the intermediary in taste space or the opportunity cost of visiting the intermediary. Then the number of users which choose to affiliate with the intermediary when the price of each good is p and the intermediary charges P^C for consumer affiliation, is given by:

$$N^C = F\left(V(n) - np - P^C\right)$$

Each seller incurs a fixed cost f to develop his product for the intermediary and a fixed distribution cost c when he makes his product available through a two-sided platform[7]. If, however, the intermediary is a merchant, the latter takes over the distribution costs. We assume that a merchant's fixed cost of distributing n products is $C(n)$. If the merchant benefits from economies of scale, then $C(n) < nc$, whereas if each individual

---

[7] f and c are the same for all sellers. The substance of our analysis would be unchanged if we allowed f or c to be distributed according to non-degenerate distributions.

seller is more efficient at distributing its own product than the merchant, then we have $C(n) > nc$.

Thus, when selling through a merchant offering bid B^S, each seller makes total profits:

$$\pi^M = B^S - f$$

When selling through a platform, those profits are:

$$\pi^P(n) = p^P(n)N^C - P^S - f - c = p^P(n)F\left(V(n) - np^P(n) - P^C\right) - P^S - f - c$$

Thus, the key difference in this very basic framework is that under the pure platform mode, sellers care about the number of consumers patronizing the intermediary, whereas under the pure merchant mode, they only care about the buyout bid.

The timing of the general pricing/adoption game we consider throughout the paper is:

1. The intermediary announces seller access fee P^S or buyout offer B^S (depending on the chosen mode) and P^C, the access price it will charge consumers in the third stage.[8]

2. Sellers decide whether or not to accept the intermediary's offer and those who do incur the cost f to make their products available through that intermediary.

3. Consumers decide whether or not to affiliate with the intermediary.

4. Under the merchant mode, the intermediary chooses price $p^M(n)$ and under the platform mode, affiliated sellers choose price $p^P(n)$; in both cases, affiliated consumers decide whether or not to buy seller products.

Note that we implicitly assume that all bargaining power lies with the intermediary when it makes pricing or buyout offers to sellers, that is, it will always charge a price that makes sellers just indifferent between accepting or not[9].

Finally, we take n as exogenously given throughout the paper. While nothing would change in our simple framework if n were endogenously chosen by the intermediary, it is important to note that in a richer model, with risk and uncertainty incorporated, the choice of n may turn out to have significant implications.

3 Indirect network externalities

In this section, we assume for simplicity that $V(.)$ is linear, that is, $V(n) = v_0 + n \times v$, and $p^P(n) = p$ exogenously given[10], with $v > p > 0$. In other words, sellers' products are identical but independent of one another.

---

[8] We assume that the intermediary can commit to P^C in period 1) in order to abstract – for now – from a hold-up problem which arises when commitment is not possible (see section 5).

[9] Introducing variable degrees of bargaining power would simply make the difference between merchants and two-sided platforms less stark in this framework without changing the main insights.

a) Two-sided platform mode

In this case, sellers' profits from affiliating with the platform are:

$$\pi(n) = pF\left(v_0 + n(v-p) - P^C\right) - P^S - f - c$$

Clearly, there are indirect network externalities among sellers, which means that multiple equilibria exist. The following analysis relies on Hagiu (2006a) for the notions of favorable and unfavorable seller expectations and corresponding solution concepts.

If seller expectations are "favorable" to the platform, then every seller expects all other sellers to affiliate with (adopt) the platform whenever:

$$pF\left(v_0 + n(v-p) - P^C\right) - P^S - f - c \geq 0$$

and in that case he will adopt the platform whenever this condition holds as well. It is easily seen that this is an equilibrium strategy for sellers given P^S and P^C, therefore the platform can charge:

$$P^S = pF\left(v_0 + n(v-p) - P^C\right) - f - c$$

This leads to total profits:

$$P^C N^C + nP^S = \left(P^C + np\right)F\left(v_0 + nv - np - P^C\right) - n(f+c)$$

which the platform maximizes over P^C to obtain:

$$\Pi_F^P = \max_{P^C} \left\{P^C F\left(v_0 + nv - np - P^C\right) + npF\left(v_0 + nv - np - P^C\right)\right\} - n(f+c)$$

If expectations are "unfavorable", then every seller expects no other seller will affiliate with the platform unless[11]:

$$pF\left(v_0 - P^C\right) - P^S - f - c \geq 0$$

It is easily seen that adopting, if and only if, this condition holds, is also an equilibrium strategy for sellers sustained by unfavorable beliefs. In this case, the platform can only charge:

$$P^S = pF\left(v_0 - P^C\right) - f - c$$

and attracts all sellers. Total platform profits are then:

$$\Pi_{NF}^P = \max_{P^C} \left\{P^C F\left(v_0 + vn - np - P^C\right) + npF\left(v_0 - P^C\right)\right\} - n(f+c)$$

---

[10] This could be the price of the products in an alternative, larger channel, or simply the monopoly price in a context with elastic and independent consumer demands for seller products.

[11] Recall that each individual seller is of measure 0.

and clearly:

$$\Pi^{P}_{NF} < \Pi^{P}_{F}$$

b) Merchant mode

The intermediary behaves as a merchant in the sense that instead of trying to "attract" sellers, it simply buys their products and resells them to consumers. Given our assumption that the intermediary has all the bargaining power in both modes, a seller will accept the bid if and only if:

$$B^{S} \geq f$$

Therefore, in this case, there are no indirect externalities to speak of. Each seller only cares about the bid she is being offered, not about the number of consumers that the merchant will be able to attract. Nor does any seller care about what the other sellers do.

The optimal strategy for the merchant is then to set $B^{S} = f$ and merchant profits are simply:

$$\begin{aligned}\Pi^{M} &= \max_{P^{C},p} \left\{ \left(P^{C} + np\right) F\left(v_0 + nv - np - P^{C}\right) \right\} - nf - C(n) \\ &= \Pi^{P}_{F} - \left[C(n) - nc\right]\end{aligned}$$

We have thus proven the following proposition:

Proposition 1 *The merchant mode is strictly preferred to the two-sided platform mode when the probability of unfavorable seller expectations is high enough or when the economies of scale in distribution are sufficiently large.*

The insight contained in this result is quite straightforward: it is easier (cheaper) to convince sellers to sell their products outright than to affiliate to a platform and sell the products to consumers themselves because the first option eliminates coordination issues.

On the other hand, however, a merchant generally incurs higher costs per seller, corresponding for instance to inventory and risk undertaken when taking possession of sellers' products, and to higher operational complexity (all products go through the merchant). Thus, the tradeoff between the merchant mode and the two-sided platform mode in this simple setting is between the higher operational costs of buying and selling products, and the higher costs of "convincing" sellers to affiliate. This suggests that intermediaries, especially for new goods, will generally start under a merchant format and, as a critical mass of sellers become affiliated, move towards a more "open", platform mode, which allows intermediaries to offer a broader variety of products. If, on the other hand, there are persistent economies of scale associated with centralized distribution, then the merchant mode will be more appropriate for increasing product variety.

In order to focus on other factors driving a wedge between the relative profitabilities of the two modes of intermediation, we will assume away from now on all distribution costs ($c = 0$; $C(n) = 0$), so that the only fixed costs are development costs.

4 Pricing distortions, product complementarity/substitutability

Another factor which typically makes the merchant mode more desirable for the intermediary is the existence of pricing distortions.

The first type of pricing distortion is the one introduced by independent seller pricing in a two-sided platform mode. Indeed, whenever sellers' products are complementary (substitutable), by pricing independently, sellers fail to internalize complementarity (substitutability) effects, leading to too high (too low) prices (that is, $p^P(n) \gtrless p^M(n)$).

To see this, assume P^C is exogenously fixed to $\overline{P}$ in both the merchant and the platform mode – this assumption helps us abstract from the hold-up issue analyzed below. In this case, under a two-sided platform mode, the intermediary sets:

$$P^S = p^P(n)F\left(V(n) - np^P(n) - \overline{P}\right) - f$$

leading to[12]:

$$\Pi^P = \left[\overline{P} + np^P(n)\right]F\left(V(n) - np^P(n) - \overline{P}\right) - nf$$

whereas under the merchant mode:

$$\Pi^M = \max_p \left\{\left[\overline{P} + np\right]F\left(V(n) - np - \overline{P}\right)\right\} - nf$$

Clearly, $\Pi^M > \Pi^P$ because the merchant has the flexibility to internalize the complementarity (subtitutability) between seller products.

Note that if $\overline{P}$ were not exogenously given and could be chosen freely by the intermediary, then the two modes lead to the same level of profits. The additional degree of freedom that the merchant possesses in choosing p is not necessary for reaching the first best level of profits in this simple model. Of course, in reality, that additional degree of freedom can make a big difference, as soon as the interval in which P^C can vary is limited (for example, Internet digital music stores usually cannot charge *access* fees to users) or consumers are vertically differentiated so that using a razor-and-blades pricing strategy is optimal. This is especially true when there are other variables that the merchant can choose (advertising levels, bundling strategies, store layout and design, etc.), which can create value above and beyond what the sellers can create by acting independently. Again, the tradeoff is that the merchant mode may incur higher operational costs than the platform mode.

The second type of price distortion is the one which arises when the platform cannot credibly commit to P^C in the first stage, upon contracting with sellers. In this case, the timing changes to:

(1) The intermediary announces seller access fee P^S or buyout offer B^S (depending on the chosen mode).

---

[12] From now on, we assume away the possibility of unfavorable seller expectations and focus on favorable expectations.

(2) Sellers decide whether or not to accept the intermediary's offer and those who do incur the cost f to make their products available through that intermediary.

(3) The intermediary sets the access price P^C for consumers and the latter decide whether or not to affiliate with the intermediary.

(4) Under the merchant mode, the intermediary chooses price $p^M(n)$ and under the platform mode, affiliated sellers choose price $p^P(n)$; in both cases, affiliated consumers decide whether or not to buy seller products.

As shown in Hagiu (2006b), a hold-up problem arises between the platform and sellers. Indeed, in the third stage, the platform sets:

$$P_2^{C*} = \arg\max_{P^C} P^C F\left(V(n) - np^P(n) - P^C\right)$$

which fails to take into account seller profits $np^P(n)F\left(V(n) - np^P(n) - P^C\right)$. As a result, P^C will be set too high relative to the price which maximizes joint profits, that is:

$$P_2^{C*} > P^{C*} = \arg\max_{P^C} \left(P^C + np^P(n)\right)F\left(V(n) - np^P(n) - P^C\right)$$

This is ex-ante reflected in P^S: the platform has to lower the affiliation fee for sellers in order to compensate them for the platform's subsequent failure to internalize their profits:

$$P^S + f = p^P(n)F\left(V(n) - np^P(n) - P_2^{C*}\right) < p^P(n)F\left(V(n) - np^P(n) - P^{C*}\right)$$

By contrast, the merchant mode allows the intermediary to fully internalize seller profits (it buys them out!) from the third stage on:

$$\Pi_2^M = \max_{P^C, p} \left(P^C + np\right)F\left(V(n) - np - P^C\right)$$

yielding total merchant profits from the perspective of stage 1 :

$$\Pi^M = \left(P^{C*} + np^P(n)\right)F\left(V(n) - np^P(n) - P^{C*}\right) - nf$$

For example, with linear consumer demand $F(u) = u$, we obtain:

$$\Pi^P = \frac{1}{4}\left[\left(V(n)\right)^2 - \left(np^P(n)\right)^2\right] - nf < \Pi^M = \frac{1}{4}\left(V(n)\right)^2 - nf$$

Thus, the merchant mode does strictly better than the platform mode, even in the absence of seller coordination problems.

One could argue that an easy way for the platform to get around this hold-up problem is to charge sellers variable fees (or royalties), which is often the case with real-world intermediaries. Indeed, assume now that under the two-sided platform mode, the intermediary can charge both the fixed fee f and a royalty ρ, proportional to the price $p^P(n)$ charged by sellers to consumers, $0 \le \rho \le 1$. In the second stage, the platform sets:

$$P^{C*}(\rho) = \arg\max_{P^C} \left[P^C + n\rho p^P(n)\right] F\left(V(n) - np^P(n) - P^C\right)$$

And in the first stage, the platform can therefore charge:

$$P^S = (1-\rho) p^P(n) F\left(V(n) - np^P(n) - P^{C*}(\rho)\right) - f$$

leading to total platform profits:

$$\Pi^P = \max_{\rho} \left[P^{C*}(\rho) + np^P(n)\right] F\left(V(n) - np^P(n) - P^{C*}(\rho)\right) - nf$$

It is then easily seen that setting $\rho = 1$ achieves the first-best level of profits, so that:

$$\Pi^P(\rho = 1) = \Pi^M$$

But $\rho = 1$ simply means that the platform is the residual claimant of all second-stage seller revenues (from selling to consumers), which is equivalent for all practical purposes to a merchant mode, that is, the intermediary taking possession of sellers' products).

It goes without saying that in reality, it is rarely feasible to charge $\rho = 1$, that is, to extract all revenues from sellers. There are several important reasons for this: the need to preserve seller investment incentives (when the latter need to invest in enhancing the quality of their products *after* having contracted with the platform) and asymmetric information about product quality, which requires the sellers to bear at least some of the risk associated with their products ($\rho = 1$ transfers all the risk to the platform-intermediary).

5 Hold-up vs. preserving innovation incentives

In the previous sections, we have seen that seller coordination issues notwithstanding, there are generally good reasons for the intermediary to function as a merchant in order to internalize product complementarities. If there were no countervailing forces to shift the balance in favor of the two-sided platform mode, one would observe many more merchant-type intermediaries.

Two such countervailing forces are asymmetric information and the need to preserve seller incentives to invest in product quality. On the first point, it is sufficient to point out that even though eBay may be able to extract higher payments from users by bundling and pricing together some of the products offered by its sellers, it would simply not be reasonable to buy all of these products, given the little information it has about them and their sellers.

The second point is related but less straightforward. It is useful to think of the example of videogames here: why don't manufacturers of consoles function as pure merchants for all the games sold on their platforms?[38] The problem is that the contractual arrangements with game developers occur long before the console and the games are finished. If the console maker was to agree to buy out a game developer at this stage, the latter would no longer have any incentives to continue working on improving his game. A better arrangement may then be for the console maker to allow the developer to be the residual claimant of the revenues derived from the game and perhaps extract some royalty payments in order to deal with its own incentives to expand the user market for consoles. We formalize this tradeoff in what follows.

Assume that the quality of seller products is variable. The variant of the model that we use here draws upon Hagiu (2006b). Under the platform mode, each seller can produce a product of quality q at fixed cost $f = \frac{cq^2}{2}$. If the intermediary opts for the merchant mode and takes possession of seller products, it can invest itself in improving product quality, at a cost $\frac{Cq^2}{2}$.

The timing of the game is now slightly different to reflect investments in product quality:

(1) The intermediary contracts with sellers (announces P^S or B^S).

(2) The sellers or the intermediary (depending on who owns the products at this stage) invest in product quality.

(3) The intermediary sets the access price P^C to consumers.

(4) Sellers or the intermediary sell products to consumers.

By symmetry, all sellers' products will be of the same quality, regardless of the mode chosen by the intermediary. We assume that when the common quality is q, consumer gross surplus from the n goods is $V(nq)$, where $V(.)$ is increasing and concave. This implies that $p^P(n,q) = qV'(nq)$. In what follows, we will use $V(x) = Ax^\beta$, $0 \leq \beta < \frac{1}{2}$. Finally, we also assume consumers' demand for affiliation with the platform/merchant is linear: $F(u) = u$.

The following proposition contains the key result of this section.

Proposition 2 *The total profits obtained under a platform mode, Π^P, are higher than the total profits obtained under a merchant mode, Π^M, if and only if:*

$$(1-\beta)(1+\beta)^{\frac{1+\beta}{1-\beta}}\left(\frac{c}{C}\right)^{\frac{\beta}{1-\beta}} < 1 \qquad (1)$$

Proof See appendix.

First, note that when $C \leq c$, that is, the merchant is at least as efficient in investing in product quality as the sellers themselves[13], then the merchant mode is always preferred since $(1-\beta)(1+\beta)^{\frac{1+\beta}{1-\beta}} > 1$ for all $\beta > 0$. This is simply because the merchant mode avoids the hold-up problem while still providing the optimal quality level.

If on the other hand C is sufficiently high relative to c, then the platform mode is preferred by the intermediary. Indeed, in that case, the gains from devolving control over and providing incentives for quality investments to sellers outweigh the drawback of incurring the hold-up problem.

It should be clear that all of the above results hold when n is not exogenously given, but chosen by the intermediary implicitly through its choice of P^S.

6 Discussion

The formal analysis in the previous three sections makes it clear that the tradeoff between the merchant mode of intermediation and the two-sided platform mode runs far deeper than the presence of indirect network externalities or absence thereof. In order to get a sense of how the distinction between the two modes works in real-world examples, we offer two representative mini-case studies in this section.

6.1 Amazon

Amazon, the world's leading online retailer, has undertaken a clear transition from a pure merchant to a two-sided platform-like intermediary over its 12-year history. When Jeff Bezos started the company in 1995, it was very close to a pure merchant mode, with the exception of certain contractual arrangements designed to transfer some inventory risk back to book wholesalers. After developing a sophisticated e-commerce and database infrastructure, Amazon started its "marketplace" initiative in 1999. Under this initiative, the company began allowing some of its suppliers[14] to operate their own storefronts on the Amazon.com website. Today, its contractual relationships with merchants fall roughly into 5 categories: zShops for small merchants, Merchants@Amazon.com, Merchants.com, syndicated stores and marketing deals, the last four targeting large business sellers. These new ways of contracting with suppliers represented a significant change in the way Amazon perceived its intermediation business. As explained by Jeff Bezos:

> "One of the things we had to learn through zShops and auctions was that we needed to think of ourselves as serving two distinct sets of customers. We pride ourselves in being "customer-centric", but for years "customers" meant "buyers". As we began to operate auctions and zShops we realized that these third-party sellers were equally important customers. And it took a little while for the organization to learn what their needs were and how we could best serve them." (Leschly *et al* (2002), p.7)

---

[13] Again, this may be because the merchant benefits from economies of scale in accumulating some know-how over and above what each individual seller can achieve by himself.

[14] Amazon refers to its suppliers as "merchants", but in order to maintain consistency and avoid confusion, we will call them "suppliers".

The five contractual relationships described above differ in the share of inventory risk and control Amazon is taking on. For instance, under the Merchants@Amazon.com arrangement, suppliers can sell products through Amazon's website while still performing many of the commerce functions themselves, such as maintaining ownership of and setting prices on inventory. By contrast, under the Merchants.com arrangement, Amazon operates third-party supplier websites (for example, www.Target.com) but takes inventory in its distribution centers and completes order fulfillment functions itself. In both cases, Amazon charges a fixed fee and a commission per item sold, ranging from 5% (when Amazon does not take significant inventory functions and risks) to 15% (when it does).

It is interesting to note a straightforward parallel between the third-party, digital storefronts on Amazon.com in the online world and the Wal-Mart's practice of renting out shelf space to some of its suppliers in the brick and mortar world. Indeed, instead of buying all products and reselling them, thus effectively taking ownership, Wal-Mart rents shelf-space to some suppliers such as Kellogg and Coca-Cola. These suppliers are responsible for stocking, displaying, pricing and advertising their merchandise, within the space allocated by Wal-Mart. The efficiency gains come from providing suppliers with incentives to use price, advertising, and display to maximize profits, part of which Wal-Mart can extract through higher rents. On the other hand, some efficiency is lost because suppliers do not fully internalize the effects of their in-store actions on other suppliers (see section 4). Kellogg, Coke and others are more than a little interested in the traffic Wal-Mart generates, since this determines how many consumers are likely to stop by their stands and eventually buy their products. It makes sense to think of them as "on board" the Wal-Mart two-sided platform.

6.2 Digital music and videos

We now turn to intermediaries of digital goods, for which inventory risk becomes irrelevant. In this context, the only thing that matters for the distinction between merchants and two-sided platforms is the allocation of control rights – between sellers and the intermediary – over strategic variables such as pricing, display, bundling, etc.

As pointed out in the introduction, although at first glance the iTunes digital music store looks like a two-sided platform, Apple's almost absolute control over music and movie pricing ($0.99 per song, $1.99 per TV series episode, $9.99 and $14.99 per movie downloaded), as well as over the consumer interface, makes it quite merchant-like. Apple's contention is that having a unified, easy-to-use interface, as well as a very simple pricing scheme, are critical to attracting consumers to the iTunes store, and that these attributes may be lost under a more decentralized control structure.

It is instructive to compare the respective modes of distribution of music content and of videogame content. Even at the very beginning of a new console generation, videogames are rarely sold under a merchant mode by the console manufacturers, with the exception of the few titles that every console buys out in order to ensure their exclusivity and solve the initial "chicken-and-egg" problem (consistent with our section 3 above). The reason has less to do with product variety (consumer demand for variety is arguably equally high in both markets) than with the significant asymmetry of information and need to preserve developer incentives to invest resources in enhancing game quality in the videogame case. By contrast, at the time music labels contract with digital distributors such as Apple's iTunes or Real Networks' Rhapsody, the "quality" of the songs (that is, their popularity) is usually known, therefore the intermediary mode chosen is determined by other considerations, such as the ability of a merchant to extract higher profits from consumers through creative pricing and bundling schemes.

An approach similar to Apple's in digital music has been adopted by many of the leading Internet video portals – currently the fastest growing online sector. Google Video/YouTube[15] is a case in point. Here, it is important to first distinguish between consumer generated media (CGM) and professional and semi-professional video content, produced by content publishers. The former constitute the majority of videos watched online today, however, the latter are growing faster and most industry experts agree that this is where the most interesting revenues potential lies. CGM is entirely free, therefore, largely irrelevant to our discussion. For professional content, Google/YouTube restricts providers to charging $1.99 or $3.99.

However, some competitors, such as Brightcove.com, have already opted to differentiate themselves from Google through the flexibility they offer content providers in choosing their revenue models – pay-per-view, pay-per-download or advertising-supported. In addition, Brightcove even allows its content providers to have store fronts on its website from which users can "jump" directly to the content providers' websites. Brightcove's vision is thus to "own" users by maintaining user accounts and placing itself at the nexus of the exploding market for Internet videos, without attempting to exert too much control over the transactions themselves.

It is interesting to ask what needs to happen in order for the pure two-sided platform strategy to become dominant for all of these sites. Presumably, over time, consumers will become more accustomed to such services, so that there will be less value created by uniform and centralized pricing.

The key insight which emerges from the preceding discussion and formal analysis is that the distinction between the two modes of intermediation – merchant vs. two-sided platform – is contractual. Where a given intermediary is located along the merchant-platform continuum depends on the allocation of control rights over the decision variables impacting the sale of products to consumers, on the sharing of economic risk and on the allocation of consumer ownership between suppliers and the intermediary.

7 Conclusion

In this paper, we have shown that the tradeoff between the merchant form of intermediary organization and the two-sided platform form is affected by several fundamental economic factors: indirect network effects between buyers and sellers; asymmetric information between sellers and the intermediary; investment incentives and product complementarities/substitutability. The following table summarizes all of the relevant factors and their effects on the desirability of each of the two modes of intermediary organization.

Economic issue	**Platform mode**	**Merchant mode**
Unfavorable expectations by sellers	-	+
Asymmetric information (sellers)	+	-

[15] Google famously acquired YouTube in October 2006 for $1.6 billion, but for now, the two portals continue to co-exist side by side.

Strong complementarity or substitutability between sellers' products	-	+
Need for ongoing investments by sellers	+	-
Need for ongoing investments by the intermediary	-	+
Uncertainty (consumer demand)	+	-
Consumer demand for product variety	+	-
Consumers unfamiliar with seller products	-	+
Suppliers have significant "contact" with consumers outside the intermediary (for example, brand name)	+	-

Table 1: Platform mode vs. merchant mode

Note that this table and the analysis above hold true for a monopoly intermediary. With competing intermediaries, more subtle strategic issues may arise which are beyond the scope of this paper. For example, Hagiu and Lee (2006) study the strategic effects of intermediaries devolving control over content pricing to content providers, in a context with intermediaries competing for content and consumers.

Clearly, this paper is only a preliminary treatment of the economic tradeoffs enumerated above. The main contributions have been to establish a connection between the recent literature on two-sided markets and the one on market intermediaries and to identify the factors driving the choice between the two polar forms of market organization for intermediaries. This should provide the initial starting point for future economics research on this topic.

8 References

Baumol, W. and J. Sidak (1994) "The Pricing of Inputs Sold To Competitors," *Yale Journal on Regulation*, 11: 171-202.

Armstrong, M. (2006) "Competition in Two-Sided Markets," *Rand Journal of Economics*, 37: 668-691.

Biglaiser, G. (1993) "Middlemen as Experts," *Rand Journal of Economics*, 24: 212-223.

Caillaud, B., and B. Jullien (2003) "Chicken and Egg: Competition Among Intermediation Service Providers," Rand *Journal of Economics*, 34: 309-328.

Evans, D. S. (2003) "The Antitrust Economics of Multi-Sided Platform Markets," *Yale Journal on Regulation*, 20: 325-82.

Evans, D. S., A. Hagiu and R. Schmalensee (2006) "Invisible Engines: How Software Platforms Drive Innovation and Transform Industries," *MIT Press*, October.

Hagiu, A. (2006a) "Pricing and Commitment by Two-Sided Platforms," *Rand Journal of Economics*, 37: 2006.

Hagiu, A. (2006b) “Two-Sided Platforms: Pricing, Product Variety and Social Efficiency,” mimeo, Harvard Business School.

Hagiu, A. and R. S. Lee (2006) “Exclusivity and Control,” mimeo, Harvard Business School.

Hagiu, A and D. B. Yoffie (2006) “Brightcove and the Future of Internet Television,” Harvard Business School Case N9-707-457.

Leschly, S., M. J. Roberts, W. A. Sahlman (2003) “Amazon – 2002,” Harvard Business School case 9-803-098

O’Hara, Maureen (1995) *Market Microstructure Theory*. Blackwell: Oxford.

Rochet, J.-C., and J. Tirole (2003) “Platform Competition in Two-Sided Markets,” *Journal of the European Economic Association*, 1: 990-1029.

Rochet, J.-C., and J. Tirole (2004) “Two-Sided Markets: Where We Stand,” *Rand Journal of Economics*, 37: 2006.

Rubinstein, A. and A. Wolinsky (1987) “Middlemen,” *Quarterly Journal of Economics*, 102: 581-593.

Rust, J. and G. Hall (2003), “Middlemen vs. Market Makers: A Theory of Competitive Exchange,” *Journal of Political Economy*, 111: 2003.

Schmalensee, R. (2002) “Payment Systems and Interchange Fees,” *Journal of Industrial Economics*, 50: 103-122.

Spulber, D. F. (1996) “Market Making by Price-Setting Firms,” *Review of Economic Studies*, 63: 559-580.

Spulber, D. F. (1996) “Market Microstructure and Intermediation,” *Journal of Economic Perspectives*, 10: 135-152.

Stahl, B. (1988) “Bertrand Competition for Inputs and Walrasian Outcomes,” *The American Economic Review*, 78: 189-201.

Yoffie, D. B. (2005) “Walmart, 2005,” Harvard Business School Case 9-705-460.

9 Appendix

Proof of proposition 2 Under the merchant mode, the intermediary can buy the goods at cost f in the first stage, invest in quality q and the sell them to consumers. In the second stage, it makes total profits:

$$\max_{P^C}\left\{\left(P^C+nqV'(nq)\right)\left(V(nq)-nqV'(nq)-P^C\right)\right\}=\frac{\left(V(nq)\right)^2}{4}$$

Therefore, total merchant profits from the perspective of stage 1 are:

$$\max_{q}\left\{\frac{\left(V(nq)\right)^{2}}{4}-n\frac{Cq^{2}}{2}-nf\right\}=\max_{q}\left\{\frac{A^{2}n^{2\beta}q^{2\beta}}{4}-n\frac{Cq^{2}}{2}-nf\right\}$$

leading to $q^{M}=\left[\frac{\beta A^{2}}{2C}\right]^{\frac{1}{2-2\beta}}n^{\frac{2\beta-1}{2-2\beta}}$ and:

$$\Pi^{M}=(1-\beta)\frac{A^{2}}{4}\left[\frac{\beta A^{2}}{2C}\right]^{\frac{\beta}{1-\beta}}n^{\frac{\beta}{1-\beta}}-nf \tag{2}$$

A two-sided platform sets P^{S} and ρ in the first stage, which determine the quality q chosen by sellers in the second stage, anticipating the price P^{C} that the platform will charge consumers in the third stage. Third stage platform profits are:

$$\max_{P^{C}}\left(P^{C}+\rho nqV'(nq)\right)\left(V(nq)-nqV'(nq)-P^{C}\right)=\frac{1}{4}\left[V(nq)-(1-\rho)nqV'(nq)\right]^{2}$$

and consumer adoption of the platform is:

$$N^{C}=\frac{V(nq)-(1-\rho)nqV'(nq)}{2}$$

In the second stage, all sellers choose the same quality q (by symmetry), given by[16]:

$$(1-\rho)V'(nq)N^{U}=cq$$

or:

$$(1-\rho)V'(nq)\frac{V(nq)-(1-\rho)nqV'(nq)}{2}=cq$$

Using $V(nq)=A(nq)^{\beta}$, we obtain $q^{P}(\rho)=\left[\frac{A^{2}\lambda(1-\lambda)}{2c}\right]^{\frac{1}{2-2\beta}}n^{\frac{2\beta-1}{2-2\beta}}$, where $\lambda=\beta(1-\rho)$. The platform can then charge:

$$P^{S}=(1-\rho)q^{P}V'(nq^{P})\frac{V(nq^{P})-(1-\rho)nq^{P}V'(nq^{P})}{2}-\frac{c(q^{P})^{2}}{2}-f$$

This leads to the following expression of platform profits:

[16] Recall our assumption that sellers are small enough so that they ignore the effects of their price and quality investments on total consumer demand for the platform.

$$
\begin{aligned}
\Pi^P &= nP^S + \frac{1}{4}\left[V\left(nq^P\right) - \left(1-\rho\right)nq^P V'\left(nq^P\right)\right]^2 \\
&= \frac{1}{4}\left[\left(V\left(nq^P\right)\right)^2 - \left(\left(1-\rho\right)nq^P V'\left(nq^P\right)\right)^2\right] - n\frac{c\left(q^P\right)^2}{2} - nf
\end{aligned}
$$

Plugging in the expression of $q^P(\rho)$ above, straightforward calculations yield:

$$
\Pi^P(\rho) = \lambda^{\frac{\beta}{1-\beta}}\left(1-\lambda\right)^{\frac{1}{1-\beta}}\frac{A^2}{4}\left(\frac{A^2}{2c}\right)^{\frac{\beta}{1-\beta}} n^{\frac{\beta}{1-\beta}} - nf
$$

Optimizing over λ (implicitly over ρ), we obtain the optimal royalty rate chosen by the two-sided platform:

$$
\rho^* = \frac{\beta}{1+\beta}
$$

Clearly, $0 < \rho^* < 1$. The optimal royalty rate for the platform trades off the need to provide sufficient investment incentives to sellers (which requires low ρ) against the need to overcome the hold-up problem induced by third-stage pricing (which requires high ρ).

Total platform profits are then:

$$
\Pi^P = \frac{\beta^{\frac{\beta}{1-\beta}}}{\left(1+\beta\right)^{\frac{1+\beta}{1-\beta}}}\frac{A^2}{4}\left[\frac{A^2}{2c}\right]^{\frac{\beta}{1-\beta}} n^{\frac{\beta}{1-\beta}} - nf \tag{3}
$$

Comparing the expressions of profits under the two different modes (2) and (3), we have $\Pi^P > \Pi^M$ if and only if condition (1) holds.

NET Institute*

www.NETinst.org

Working Paper #05-23

October 2005

The Incentive To Participate In Open Source Projects: A Signaling Approach

Yossi Spiegel
Tel Aviv University

* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

The incentive to participate in open source projects: a signalling approach*

Yossi Spiegel†

Current draft: October 30, 2005

Abstract

This paper examines the incentives of programmers to contribute to open source software projects on a voluntary basis. In particular, the paper looks at this incentive changes as (i) performance becomes more visible to the relevant audience, (ii) effort has a stronger impact on performance, and (iii) performance becomes more informative about talent. In all three cases, it is shown that whether we start from a stable interior equilibrium or an unstable interior equilibrium.

*The financial support of the NET insititute (http://www.NETinst.org) is gratefully acknowledged. For helpful discussions, I thank Jean Tirole.

†Recanati Graduate School of Business Adminstration, Tel Aviv University, email: spiegel@post.tau.ac.il, http://www.tau.ac.il/~spiegel

1 Introduction

Open source software (OSS) is a computer program whose source code - the instructions for the program, written in a human readable format - is distributed free of charge and can be modified, extended, adapted, and incorporated into other programs with relatively few restrictions. OSS is a rapidly expanding phenomenon: some OSS such as the Apache web server, dominate their product categories. In the personal computer market, the OSS such as the operating system Linux and the web browser Firefox gain rapid popularity. It is estimated that there are currently 29 million users of Linux worldwide and there were over 50 million downloads of Firefox.[1]

Apart from having millions of OSS users, there are also tens of thousands of participating programmers who contribute to various OSS projects, and there is also a growing number of firms who sell services, support, and documentation for OSS. The majority of the programmers who participate in OSS projects are unpaid volunteers. For example, Hars and Ou (2002) have surveyed 81 individuals involved in open source projects and found that only 16% received any direct monetary compensation for their contribution. This raises obvious questions about the incentives and motivations of the participating programmers who do not receive direct compensation for their efforts. There are three main, mostly complimentary, explanations for the willingness of programmers to contribute to OSS projects. The first two involve intrinsic motivations while the third involves extrinsic motivations.

The first explanation is that programmers simply like to be involved in open source projects, either because they simply enjoy being creative, or due to a sense of obligation or community related reasons. Indeed, a web-based survey conducted by Lakhani and Wolf (2003) reveals that the responding programmers were mainly driven by enjoyment-based intrinsic motivations.

The second explanation involves another type of intrinsic motivation. According to this explanation, system managers (e.g., users of Apache) who need improvements in software and are willing to make these improvements on their own. They then share these improvements with others in their community. A model along these lines is offered by Johnson (2001), who views

[1] See http://counter.li.org/estimate.php for the estimate on Linux and www.mozilla.org/products/firefox for the estimate on Firefox.

participation in OSS projects as a private provision of a public good (see Bessen, 2004, for a related model).

The third explanation, suggested by Lerner and Tirole (2002), is that programmers are willing to contribute to OSS projects in order to signal their ability to future employers, venture capitalists, or to peers and thereby boost their human capital or get ego gratification. Fershtman and Gandal (2004) examine a large data set on programmers' participation in OSS projects and argue that their findings are consistent with the hypothesis that programmers who contribute to OSS projects are driven by extrinsic motivations such as their desire to enhance their social status within the programmers' community or by their desire to signal their ability to potential employers. Hann et al (2004), examine a longitudinal data set of participant contributions made and accepted into three Apache open source projects for the period 1998 to 2002. They find that more contributions to the Apache open source projects do not result in wage increases for contributors. On the other hand, successful participation in the form of a higher status in the merit-based ranking within the Apache open source community is associated with a 13% - 27% increase in wages, depending on the rank attained. These findings are robust to various model specifications and remain true even after controlling for work and programming experience. Hann et al argue that their results are consistent with the notion that a high rank within the Apache Software Foundation is a credible signal of the productive capacity of a programmer.

Drawing on the "career concerns" literature (e.g., Holmström, 1999), Lerner and Tirole (2002) conjecture that the signalling incentive will become stronger as (i) performance becomes more visible to the relevant audience, (ii) effort has a stronger impact on performance, and (iii) performance becomes more informative about talent. The purpose of this paper is to examine these conjectures in the context of a formal model. The main finding in the paper is that the model always admit a no-effort equilibrium in which firms do not expect programmers to exert effort in order to contribute to OSS projects, and programmer in turn do not exert such effort. However, the model may also admit an even number of interior equilibria, half of which are stable and the other half is unstable. The analysis shows that the three conjectures are correct only if we start from a stable interior equilibrium but are incorrect if we start from an unstable interior equilibrium.

There are two closely related papers that also argue that programmers participate in

OSS projects in order signal their abilities to prospective employers. The two papers however differ from the current paper both in terms of their set up and in terms of their main focus. Lee, Moisa, and Weiss (2003) consider a model in which programmers need to choose between joining closed source software firms or OSS projects. If they join software firms, their wage reflects the expected productivity of all programmers who join software firms (talented ones and less talented ones). On the other hand, if they join OSS projects, they forgo current wages, but can signal their productivity to software firms and hence boost their future wages. The main focus of their analysis is on the relative sizes of the closed source system and the open-source system. In particular, their show that an open-source system will never exist alone in the market because mediocre programmers, who cannot benefit from signaling their talent, will always prefer to joint closed source software firms. On the other hand, a closed-source system can exist alone in the market, especially if the population of talented programmers is relatively small.

Leppämäki and Mustonen (2004) consider a model in which programmers signal their talent to software firms by choosing how many lines of code to contribute to an OSS project. As in the traditional Spence signalling model, talented programmers have a lower cost of writing lines of code. Consequently, in a separating equilibrium, only talented programmers contribute to the open source project and their contribution is chosen so as to deter untalented programmers from mimicking them. The model departs from the traditional Spence signalling model in that the freely available OSS project imposes either a positive or a negative externality on the commercial software offered by firms. The externality in turn affects the wages that software firms are willing to offer agents and hence the marginal benefit to signalling. Leppämäki and Mustonen focus on the effect of the externality on the incentive of talented agents to contribute to the OSS project. In particular, they show that if the OSS is a substitute (complement) for the commercial software then the contribution of talented programmers will end up being lower (higher) than in the case where OSS and the commercial software are independent of each.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 shows that the model can give rise to multiple equilibria and characterizes them. Section 4 study the comparative static properties of the model and in particular examines how the incentive to contribute to OSS projects is affected by the visibility of the contribution to prospective employers, by the sensitivity of performance to effort, and by how informative is the performance about talent. Finally, I examine the effect of intrinsic motivation to contribute to OSS projects in Section 5.

2 The model

Consider a competitive job market with a large number of agents, each of whom is either "talented" (i.e., has a high productivity) or "untalented" (i.e., has a low productivity). The marginal productivity of each talented agent if he is hired is w, while the marginal productivity of an untalented agent if he is hired is normalized to 0. Under full information, the wage of each agent is equal to his marginal product. Hence, the wage of talented agents is w while the wage of untalented agents is 0.

Under asymmetric information, it is common knowledge that the fraction of talented agents in the population is α, but firms cannot tell the agents' types apart before hiring them. To signal their types, agents can engage in some activity before they are hired by firms. Participation in an OSS project provides a good opportunity for talented agents to signal their ability due to the resulting exposure they get from peers. Specifically, I assume that when agents participate in an OSS project, they can either succeed (i.e., "solve a problem") or they can fail (i.e., "fail to come up with satisfactory results"). In particular, if an agent is talented and exerts effort e in the OSS project, his probability of success is $p(e, \gamma)$, where γ is a shift parameter. With probability $1 - p(e, \gamma)$ the agent fails. On the other hand, if the agent is untalented, his action succeeds with probability p_0 which is independent of his effort level. Since untalented agents cannot boost their probability of success, they do not exert any effort.

In and of itself, the activity does not benefit the firms nor the agents directly (for now I ignore the intrinsic motivation to participate in the OSS project). The only advantage of the activity from the firms and the agents' perspective is that it generates a signal on the agents' types. Firms cannot observe directly observe the efforts that the agents exert; rather they can only (imperfectly) observe whether the agent's activity has succeeded. In particular, firms observe a successful action with probability β. With probability $1-\beta$, as well as when the activity fails, firms observe nothing. Hence, β is a measure of the visibility of the agents' performance to potential employers. Whenever firms observe nothing, they cannot discern whether the agent has participated in the OSS project and did not succeed or whether he did not participate at all.

Using subscripts to denote partial derivatives, I make the following assumptions on the probability that a talented agent will succeed:

A1 $p_e(e, \gamma) > 0 > p_{ee}(e, \gamma)$

A2 $\lim_{e \to \infty} p_e(e, \gamma) = 0$

A3 $p(0, \gamma) = p_0 \geq 0, \quad \lim_{e \to \infty} p(e, \gamma) = 1$

A4 $p_\gamma(e, \gamma) > 0, \quad p_{e\gamma}(e, \gamma) > 0$

Assumption A1 says that effort raises the probability of success but does so at a decreasing rate. Assumption A2 implies that at the limit as e increases, the marginal effect of effort on the probability of success goes to 0. This assumption will ensure the existence of a solution to the maximization problem of agents. Assumption A3 says at one extreme, if talented agents do not exert effort, then their probability of success is equal to that of untalented agents, while on the other extreme, if their effort increases indefinitely, their probability of success approaches 1 in the limit. Assumption A4 implies that the shift parameter γ raises both the probability and the marginal probability that the activity will succeed. Hence, when γ increases, effort has a stronger impact on an agent's performance.

The payoff of each agents is increasing with his wage and decreasing with his effort level:

$$U = w - e.$$

3 Equilibrium

I now look for a Perfect Bayesian equilibrium in which talented agents exert effort, untalented agents do not exert effort, and the beliefs of firms are consistent with the agents' strategies. To characterize this equilibrium, suppose that firms believe that the effort of talented agents is $\widehat{e}$. Then, conditional on observing a successful action, firms believe that the agent is talented with probability

$$q(\widehat{e},\gamma \mid s)=\frac{\alpha p(\widehat{e},\gamma)}{\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0}. \tag{1}$$

On the other hand, if firms do not observe a success, they cannot tell whether (i) the agent is talented, exerted effort, and failed, or (ii) the agent is talented, exerted effort and succeeded, but his success was unobserved, (iii) the agent is untalented and failed, or (iv) the agent is untalented, succeeded nonetheless, but his success was not observed. Hence, conditional on not observing a successful action, firms believe that the agent is talented with probability

$$\begin{aligned} q(\widehat{e},\gamma \mid n) &= \frac{\alpha\left((1-p(\widehat{e},\gamma))+(1-\beta)p(\widehat{e},\gamma)\right)}{\alpha\left((1-p(\widehat{e},\gamma))+(1-\beta)p(\widehat{e},\gamma)\right)+(1-\alpha)\left((1-p_0)+(1-\beta)p_0\right)} \\ &= \frac{\alpha(1-\beta p(\widehat{e},\gamma))}{1-\beta\left(\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0\right)}. \end{aligned} \tag{2}$$

Note that given Assumption A3, $q(0,\gamma \mid s)=q(0,\gamma \mid n)=\alpha$: if firms expect talented agents to exert no effort, then success or failure is not an informative signal about the agent's talent. Moreover, note that $q(\widehat{e},\gamma \mid s)$ approaches 1 as p_0 approaches 0: if untalented agents cannot succeed then success is a sure sign that the agent is talented.

Next, we need to find the effort level that talented agents will exert. To this end, note that since the labor market is competitive, the wage of agents is $q(\widehat{e},\gamma \mid s)w$ following an observed success and $q(\widehat{e},\gamma \mid n)w$ otherwise. Hence, the expected payoff of talented agents given their effort level, e, and given the belief of firms, $\widehat{e}$, is

$$U(e)=\beta p(e,\gamma)q(\widehat{e},\gamma \mid s)w+\left((1-\beta)p(e,\gamma)+1-p(e,\gamma)\right)q(\widehat{e},\gamma \mid n)w-e. \tag{3}$$

The first term on the left-hand side reflects the idea that with probability $\beta p(e,\gamma)$, the talented agent's action succeeds and his success is observed by firms. The second term on the left-hand side reflects the idea that with probability $(1-\beta)p(e,\gamma)$, the successful action of a talented agent is not observed by firms and with probability $1-p(e,\gamma)$ it fails altogether. In both cases firms cannot tell whether the agent is talented or not and they pay him a wage $q(\widehat{e},\gamma \mid n)w$. The last term on the left-hand side of the equation is the agent's cost of effort.

Assuming that there is a large number of talented agents, each will ignore the effect of his own effort level on $\widehat{e}$. Since Assumption A1 ensures that $U''(e) < 0$, the effort level that each talented agent will choose given the firms' beliefs, $\widehat{e}$, is defined implicitly by the following first order condition:

$$U'(e) = \beta p_e(e,\gamma)\Delta(\widehat{e},\gamma)w - 1 \leq 0, \qquad eU'(e) = 0, \tag{4}$$

where

$$\begin{aligned}\Delta(\widehat{e},\gamma) &\equiv q(\widehat{e},\gamma \mid s) - q(\widehat{e},\gamma \mid n) \\ &= \frac{\alpha p(\widehat{e},\gamma)}{\alpha p(\widehat{e},\gamma) + (1-\alpha)p_0} - \frac{\alpha(1-\beta p(\widehat{e},\gamma))}{1-\beta\left(\alpha p(\widehat{e},\gamma) + (1-\alpha)p_0\right)} \\ &= \frac{\alpha\left(1-\alpha\right)\left(p(\widehat{e},\gamma) - p_0\right)}{\left(\alpha p(\widehat{e},\gamma) + (1-\alpha)p_0\right)\left(1-\beta\left(\alpha p(\widehat{e},\gamma) + (1-\alpha)p_0\right)\right)},\end{aligned} \tag{5}$$

is the increase in the probability that firms assign to an agent being talented following an observed success. The expression $\beta p_e(e,\gamma)\Delta(\widehat{e},\gamma)w$ represents the marginal benefit from effort which is equal to the marginal effect of effort on the probability that a successful action will be observed, $\beta p_e(e,\gamma)$, times the extra wage that an agents gets in this event, $\Delta(\widehat{e},\gamma)w$. At an interior optimum, this marginal benefit must be equal to the marginal cost of effort, which is 1. But, if $\beta p_e(e,\gamma)\Delta(\widehat{e},\gamma)w$ is smaller than 1 for all positive effort levels, then the talented agent will not exert any effort.

Before proceeding, it is worth noting that

$$\Delta_{\widehat{e}}(\widehat{e},\gamma) = \frac{\alpha\left(1-\alpha\right)p_e(\widehat{e},\gamma)\left[\beta\alpha^2\left(p(\widehat{e},\gamma) - p_0\right)^2 + p_0(1-\beta p_0)\right]}{\left(\alpha p(\widehat{e},\gamma) + (1-\alpha)p_0\right)^2\left(1-\beta\left(\alpha p(\widehat{e},\gamma) + (1-\alpha)p_0\right)\right)^2} > 0. \tag{6}$$

That is, if firms believe that talented agents exert more effort, then observed success leads to a larger increase in the probability that firms assign to an agent being talented. Recalling that $\Delta(\widehat{e},\gamma)w$ is the extra expected wage that an agent receives following an observed success, this implies that as firms believe that talented agents exert more effort, they are willing to pay higher wages to agents who were observed to be successful. Moreover, since by Assumption A3, $q(0,\gamma \mid s) = q(0,\gamma \mid n) = \alpha$, then $\Delta(0,\gamma) = 0$. Hence, if firms believe that talented agents do not exert effort, then observed success does not increase their assessment that the agent is talented.

Let $BR(\widehat{e})$ denote the solution of (4). This function is the best-response of each talented agent against the firms' beliefs about his effort level. In equilibrium, the firms' beliefs must be consistent with the true efforts of the talented agents. Hence, the equilibrium effort level, e^*, is defined implicitly by the equation

$$e^* = BR(e^*). \tag{7}$$

In other words, the equilibrium is defined by the intersection of the best response function, $BR(\widehat{e})$, with the 45^o line in the $(e, \widehat{e})$ space. Given its central role in what follows, I now study the properties of $BR(\widehat{e})$ in the next lemma. To establish this lemma, I first make the following assumption on the marginal productivity of a talented agent if he is hired by a firm:

A5 The marginal productivity of a talented agent is such that

$$w > \overline{w} \equiv \frac{(\alpha + (1-\alpha)p_0)\,(1 - \beta\,(\alpha + (1-\alpha)p_0))}{\beta\alpha\,(1-\alpha)\,(1-p_0)\,p_0}. \tag{8}$$

Lemma 1: *Suppose that Assumption A5 holds. Then, the best response of talented agents against the firms' beliefs about their effort levels, $BR(\widehat{e})$ has the following properties:*

(i) *Suppose that $BR(\widehat{e}) = 0$ for all $0 < e \leq \widehat{e}_1$ and $BR(\widehat{e}) > 0$ for all $e > \widehat{e}_1$, where $\widehat{e}_1$ is implicitly defined by the equation $\beta p_0 \Delta(\widehat{e}_1, \gamma) w = 1$.*

(ii) *$BR'(\widehat{e}) > 0$ for all $e > \widehat{e}_1$ and $\lim_{\widehat{e}\to\infty} BR'(\widehat{e}) = 0$.*

Proof: (i) First, note that since $\Delta(0, \gamma) = 0$, $U'(e) = -1$ when $\widehat{e} = 0$, so $BR(0) = 0$. Otherwise, if $\widehat{e} > 0$, then $\Delta(\widehat{e}, \gamma) > 0$. Since $p_{ee}(e, \gamma) < 0$, $U'(e)$ is a strictly decreasing function of e for all $\widehat{e} > 0$. Assumption A2 implies that as e goes to infinity, $U'(e)$ goes to -1. Hence, $U'(e) = 0$ attains a unique interior solution if and only if

$$U'(0) = \beta p_0 \Delta(\widehat{e}, \gamma) w - 1 > 0, \tag{9}$$

where the equality follows because by Assumption A3, $p(0,\gamma) = p_0$.

Since $\Delta(0,\gamma) = 0$, condition (9) clearly fails when $\widehat{e} = 0$, and by continuity, it also fails for sufficiently small values of $\widehat{e}$. On the other hand, since $\Delta_{\widehat{e}}(\widehat{e},\gamma) > 0$, an increase in $\widehat{e}$ raises $U'(0)$. Recalling from Assumption A3 that $\lim_{e\to\infty} p(e,\gamma) = 1$, it follows that in the limit, as $\widehat{e}$ increases,

$$\lim_{\widehat{e}\to\infty} \Delta(\widehat{e},\gamma) = \frac{\alpha\,(1-\alpha)\,(1-p_0)}{(\alpha + (1-\alpha)p_0)\,(1-\beta\,(\alpha+(1-\alpha)p_0))}.$$

This implies in turn that (9) can be satisfied for a large enough $\widehat{e}$ if and only if

$$\lim_{\widehat{e}\to\infty} U'(0) = \frac{\beta\alpha\,(1-\alpha)\,(1-p_0)\,p_0 w}{(\alpha + (1-\alpha)p_0)\,(1-\beta\,(\alpha+(1-\alpha)p_0))} - 1 > 0. \tag{10}$$

A sufficient condition for $\lim_{\widehat{e}\to\infty} U'(0) > 0$ is that $w > \overline{w}$, where $\overline{w}$ is defined by (8).

Therefore, whenever $w > \overline{w}$, there exists a unique value of $\widehat{e}$, denoted $\widehat{e}_1$, such that $U'(0) > 0$ for all $\widehat{e} > \widehat{e}_1$ and $U'(0) < 0$ otherwise, where $\widehat{e}_1$ is implicitly defined by the equation $U'(0) = \beta p_0 \Delta(\widehat{e},\gamma) w - 1 = 0$.

This implies in turn that for all $\widehat{e} \le \widehat{e}_1$, $U'(e) < 1$ for all e so $BR(\widehat{e}) = 0$. On the other hand, for all $\widehat{e} > \widehat{e}_1$, $U'(e) > 0$ for sufficiently small values of e. Since $U'(e)$ is a strictly decreasing function of e and since $U'(e)$ goes to -1 as e goes to infinity, it follows that whenever $\widehat{e} > \widehat{e}_1$, there exists a unique value of e that solves the equation $U'(e) = 0$. Hence, $BR(\widehat{e}) > 0$ for all $\widehat{e} > \widehat{e}_1$.

(ii) As part (i) shows, $BR(\widehat{e}) > 0$ for all $\widehat{e} > \widehat{e}_1$ and it is defined implicitly by the equation $U'(e) = 0$. That is, $U'(BR(\widehat{e})) = 0$. Fully differentiating this equation with respect to $\widehat{e}$ and rearranging terms, yields

$$BR'(\widehat{e}) = -\frac{p_e(e,\gamma)\Delta_{\widehat{e}}(\widehat{e},\gamma)}{p_{ee}(e,\gamma)\Delta(\widehat{e},\gamma)} > 0, \tag{11}$$

where the inequality follows because $p_{ee}(e,\gamma) < 0$ and because $\Delta_{\widehat{e}}(\widehat{e},\gamma) > 0$. To complete the proof, note that as $\widehat{e}$ increases so does e. However, Assumption A2 shows that $\lim_{e\to\infty} p_e(e,\gamma) = 0$. Hence, $BR'(\widehat{e})$ goes to 0 as $\widehat{e}$ goes to infinity. ■

Using Lemma 1, I can now characterize the equilibrium effort level of talented agents. To this end, recall from (7) that the equilibrium condition is given by $e^* = BR(e^*)$. Since $BR(\widehat{e})$ passes through the origin, $e^* = 0$ is a solution to the equilibrium condition. Hence, there always exists a no-effort equilibrium in which talented agents are not expected to exert effort and in fact do not exert effort. The question is whether there are additional solutions to the equilibrium condition $e^* = BR(e^*)$?

To address this question, I present $BR(\widehat{e})$ in Figure 1, using Lemma 1. The figure shows $BR(\widehat{e})$ in the $(e, \widehat{e})$ space. As the figure shows, $BR(\widehat{e})$ coincides with the vertical axis for sufficiently small values of $\widehat{e}$. As $\widehat{e}$ increases above $\widehat{e}_1$, $BR(\widehat{e})$ increases with $\widehat{e}$. Since $BR'(\widehat{e})$ goes to 0 as $\widehat{e}$ goes to infinity, $BR(\widehat{e})$ eventually becomes very steep.[2] Figure 1 also shows the 45^0 line. The equilibrium effort level of talented agents is determined by the intersection of $BR(\widehat{e})$ with the 45^0 line. As the figure shows, there are in general two possibilities depending on the shape of $BR(\widehat{e})$.

The first possibility, illustrated in Figure 1a, arises when $BR(\widehat{e})$ intersects the 45^0 line only at $e = 0$. In this case, the model does not admit interior equilibria in which $e^* > 0$. A sufficient (though not necessary) condition for case (i) is that $BR'(\widehat{e}) < 1$ for all $\widehat{e} > \widehat{e}_1$. The second possibility, illustrated in Figure 1b, arises when $BR(\widehat{e})$ intersects the 45^0 line at least once from above at some $\widehat{e} > \widehat{e}_1$. In this case, we do have interior equilibria in which $e^* > 0$. But, since $BR'(\widehat{e})$ goes to 0 as $\widehat{e}$ goes to infinity, $BR(\widehat{e})$ must intersect the 45^0 line at least one more time but from below. Hence, if there are interior equilibria in which $e^* > 0$, then their number must be even. A necessary condition for the model to admit only two interior equilibria (apart from the no-effort equilibrium) is that $BR''(\widehat{e}) < 0$. Using (11), it follows that this is the case whenever

$$BR''(\widehat{e}) = -\frac{p_e(e,\gamma)}{p_{ee}(e,\gamma)} \frac{d}{d\widehat{e}} \left[\frac{\Delta_{\widehat{e}}(\widehat{e},\gamma)}{\Delta(\widehat{e},\gamma)} \right].$$

Since $p_{ee}(e,\gamma) < 0$, it follows that $BR''(\widehat{e}) < 0$ if and only if $\frac{d}{d\widehat{e}} \left[\frac{\Delta_{\widehat{e}}(\widehat{e},\gamma)}{\Delta(\widehat{e},\gamma)} \right] < 0$.

I summarize this discussion in the following Proposition:

[2] Note that since Figure 1 shows $BR(\widehat{e})$ in the $(e, \widehat{e})$ space, a steep curve is associated with small value of $RB'(\widehat{e})$.

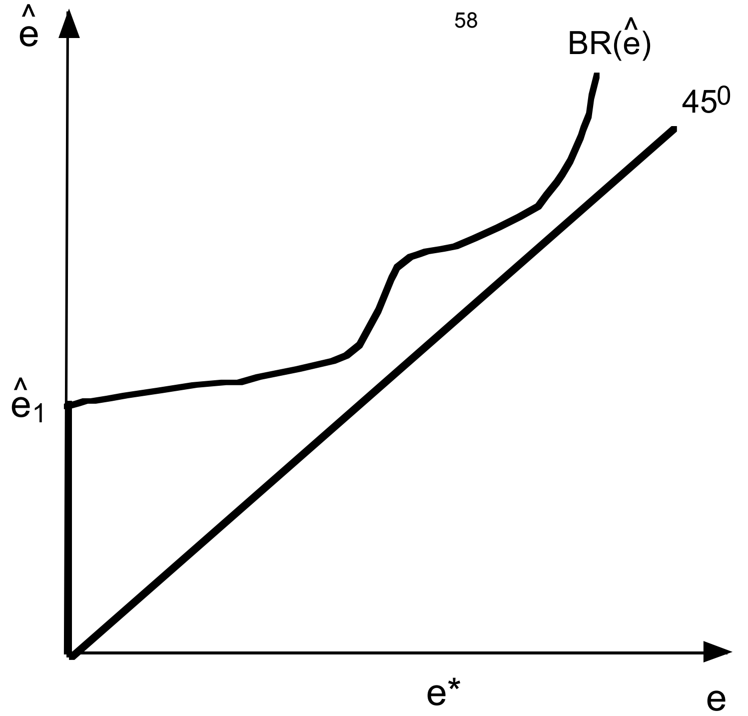


Figure 1a: No interior equilibria

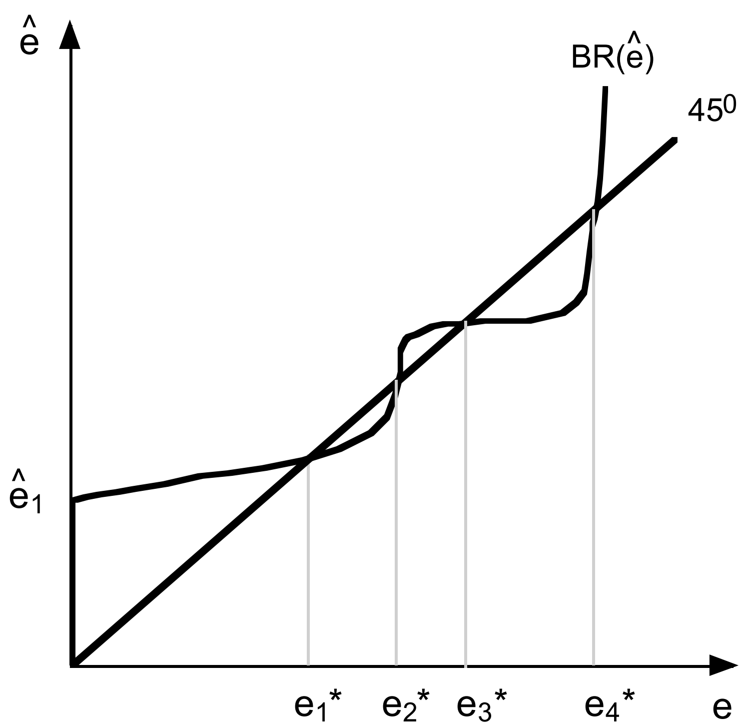


Figure 1b: Four interior equilibria

Proposition 1: *A sufficient condition for the no-effort equilibrium to be unique is that* $BR'(\widehat{e}) < 1$ *for all* $\widehat{e} > \widehat{e}_1$*. If however the model admits interior equilibria in which* $e^* > 0$*, then their number must be even.*

Next, suppose that there exist interior equilibria in which $e^* > 0$. Recalling that in equilibrium the firms' beliefs must be consistent with the true efforts of the talented agents, i.e., $\widehat{e} = e^*$, and substituting this equality into equation (4), the equilibrium effort level, e^*, is implicitly defined by

$$\beta G(e^*, \gamma)w = 1, \qquad G(e^*, \gamma) \equiv p_e(e^*, \gamma)\Delta(e^*, \gamma). \tag{12}$$

It should be noted that the left-hand side of equation (12) differs from the left-hand side of equation (4) because in the latter, the beliefs of firms about the efforts of talented agents are arbitrary, while in the former they are consistent with the true efforts of talented agents. Hence, $\beta G(e^*, \gamma)w$ can be interpreted as the marginal benefit of effort from an agents' point of view in equilibrium (i.e., given that firms hold correct beliefs about the agent's effort).

In the next section, I will study the comparative statics properties of e^*. Since the function $G(e^*, \gamma)$ plays a key role in that analysis, I now establish an important property of $G(e^*, \gamma)$.

Lemma 2: $G_e(e^*, \gamma) \geq (<) 0$ *as* $-\frac{p_{ee}(e^*,\gamma)}{p_e(e^*,\gamma)} \leq (>) \frac{\alpha}{w(1-\alpha)} \left(1 + \frac{p_0(1-\beta p_0)}{(p(e^*,\gamma)-p_0)^2}\right)$.

Proof: Straightforward differentiation reveals that

$$\begin{aligned} G_e(e^*, \gamma) &= p_{ee}(e^*, \gamma)\Delta(e^*, \gamma) + p_e(e^*, \gamma)\Delta_e(e^*, \gamma) \\ &= \left[\frac{p_{ee}(e^*, \gamma)}{p_e(e^*, \gamma)} + \frac{\Delta_e(e^*, \gamma)}{\Delta(e^*, \gamma)}\right] G(e^*, \gamma). \end{aligned} \tag{13}$$

But, using equations (5), (6), and (12), yields

$$\begin{aligned} \frac{\Delta_e(e^*, \gamma)}{\Delta(e^*, \gamma)} &= \frac{p_e(e^*, \gamma)\left[\beta\alpha^2\left(p(e^*, \gamma) - p_0\right)^2 + p_0(1 - \beta p_0)\right]}{\left(\alpha p(e^*, \gamma) + (1-\alpha)p_0\right)\left(1 - \beta\left(\alpha p(e^*, \gamma) + (1-\alpha)p_0\right)\right)\left(p(e^*, \gamma) - p_0\right)} \\ &= \frac{\alpha}{(1-\alpha)w}\left(1 + \frac{p_0\left(1 - \beta p_0\right)}{\left(p(e^*, \gamma) - p_0\right)^2}\right) \end{aligned} \tag{14}$$

Substituting from (14) into (13),

$$G_e(e^*,\gamma) = \left[\frac{p_{ee}(e^*,\gamma)}{p_e(e^*,\gamma)} + \frac{\alpha}{w\,(1-\alpha)}\left(1+\frac{p_0\,(1-\beta p_0)}{(p(e^*,\gamma)-p_0)^2}\right)\right] G(e^*,\gamma). \qquad (15)$$

The result follows by noting that the sign of $G_e(e^*,\gamma)$ depends on the sign of the square bracketed expression. This expression can be either negative or positive since $\frac{p_{ee}(e^*,\gamma)}{p_e(e^*,\gamma)}$ is negative by Assumption A1, while $\frac{\alpha}{w(1-\alpha)}\left(1+\frac{p_0(1-\beta p_0)}{(p(e^*,\gamma)-p_0)^2}\right)$ is positive. ■

Lemma 2 shows that $G(e^*,\gamma)$ may either increase or decrease with e. Intuitively, holding the belief of firms, $\widehat{e}$, constant, the marginal benefit of effort from an agent's point of view is decreasing with effort because effort raises the likelihood of success at a decreasing rate. Hence at first glance it would seem that $G(e^*,\gamma)$ should be decreasing with e. However, when talented agents exert more effort (and this is anticipated by firms), their extra wage following a success increases. This effect raises the marginal benefit of effort, which in turn implies that $G(e^*,\gamma)$ should be increasing with e. The first, negative, effect is more likely to dominate the second, positive, effect when α is small (there are few talented agents in the population) and when w is large (the productivity of talented agents is large); in both cases, the wage gap between success and failure is particularly large. Hence, $G_e(e^*,\gamma)$ is likely to be negative when α is small and when w is large and positive when α is large and w is small.

At a more technical level, note from (13) and (11) that

$$\begin{aligned} G_e(e^*,\gamma) &= \left[1+\frac{p_e(e^*,\gamma)\Delta_e(e^*,\gamma)}{p_{ee}(e^*,\gamma)\Delta(e^*,\gamma)}\right] p_{ee}(e^*,\gamma)\Delta(e^*,\gamma) \\ &= \left[1-BR'(e^*)\right] p_{ee}(e^*,\gamma)\Delta(e^*,\gamma). \end{aligned}$$

Since $p_{ee}(e^*,\gamma)\Delta(e^*,\gamma)$, it follows that $G_e(e^*,\gamma) \geq 0$ if $BR'(e^*) > 1$ and $G_e(e^*,\gamma) < 0$ if $BR'(e^*) < 1$. To interpret these conditions, note that $BR'(e^*)$ is just the slope of the best response function of talented agents against the beliefs of firms evaluated at the equilibrium effort level. When $BR'(e^*) > 1$, the best response at the equilibrium point, $BR(e^*)$, is steeper than the 45^0 line and hence cuts it from below. On the other hand, when $BR'(e^*) < 1$, $BR(e^*)$, is flatter than the 45^0 line and hence cuts it from above.

Notice that when $BR(e^*)$ cuts the 45^0 line from below (e.g., the equilibria e_2^* and e_4^* in

Figure 1b), the resulting equilibrium is stable in the sense that a Cournot tatônnement process will lead to a convergence to the equilibrium point starting from any close neighborhood of the equilibrium point. On the other hand, when $BR(e^*)$ cuts the 45^0 line from above (e.g., the equilibria e_1^* and e_3^* in Figure 1b), the resulting equilibrium is unstable. Hence,

Proposition 2: *Suppose that the model admits interior equilibria in which $e^* > 0$. Then, a given interior equilibrium is stable if $G_e(e^*, \gamma) < 0$ and unstable if $G_e(e^*, \gamma) \geq 0$.*

Since Lemma 2 indicates that $G_e(e^*, \gamma) < 0$ is more likely when α is small and w is large while $G_e(e^*, \gamma) \geq 0$ is more likely when α is large and w is small, one can conclude that stable equilibria are more likely when α is small and w is large while unstable equilibria are more likely when the reverse is true.

4 Comparative statics

Given Lemma 2, I now examine the conjectures of Lerner and Tirole (2002) that the signalling incentive of agents is stronger:

(i) the more visible the performance to the relevant audience,

(ii) the higher the impact of effort on performance, and

(iii) the more informative the performance about talent.

4.1 The effect of the visibility of performance on effort

To examine conjecture (i), recall that β is a measure of the visibility of the agents' performance to firms. Hence, I examine conjecture (i) by looking at the effect of an increase in β on e^*:

Proposition 3: *An increase in β which measures the visibility of the agents' performance to firms, increases the effort level that talented agents exert in stable interior equilibria but lowers it in unstable interior equilibria.*

Proof: Differentiating equation (12) with respect to e^* and β and rearranging terms,

$$\frac{\partial e^*}{\partial \beta} = -\frac{G(e^*,\gamma) + \beta \frac{\partial G(e^*,\gamma)}{\partial \beta}}{\beta G_e(e^*,\gamma)},$$

where

$$\begin{aligned}
\frac{\partial G(e^*,\gamma)}{\partial \beta} &= p_e(e^*,\gamma)\frac{\partial \Delta(e^*,\gamma)}{\partial \beta} \\
&= p_e(e^*,\gamma)\frac{\alpha(1-\alpha)(p(\widehat{e},\gamma)-p_0)}{(\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0)}\frac{\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0}{(1-\beta(\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0))^2} \\
&= p_e(e^*,\gamma)\Delta(e^*,\gamma)\frac{\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0}{(1-\beta(\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0))} > 0.
\end{aligned}$$

The sign of $\frac{\partial e^*}{\partial \beta}$ is equal to the sign of $-G_e(e^*,\gamma)$ which by Proposition 2 is positive in stable interior equilibria and negative in unstable interior equilibria. ■

Proposition 3 is illustrated in Figure 2. The equilibrium effort level, e^*, is attained at the point where $\beta G(e^*,\gamma)w$, which is the equilibrium marginal benefit of effort, cuts the horizontal line whose height is 1 and which represents the marginal cost of effort. An increase in β shifts the equilibrium marginal benefit of effort upward. Whether this leads to an increase or a decrease in e^* depends on whether $G(e^*,\gamma)$ is upward or downward sloping. When $G(e^*,\gamma)$ is downward sloping, which as Lemma 2 shows is likely to occur when α is small and w is large, an increase in β leads to an increase in e^*. On the other hand, when α is large and w is small, $G(e,\gamma)$ is likely to upward sloping so an increase in β lead to a decrease in e^*.

Proposition 3 shows that Lerner and Tirole's (2002) conjecture that the signalling incentive of agents will become stronger as their performance becomes more visible to the relevant audience is true only if the model admits interior equilibria and then only in interior equilibria that are stable. This a likely to be the case when there are few talented agents around (α is small) and when talented agents earn a high wage (w is high). Otherwise, this conjecture is incorrect: the signalling incentive of agents will become stronger as the agents' effort becomes less visible.

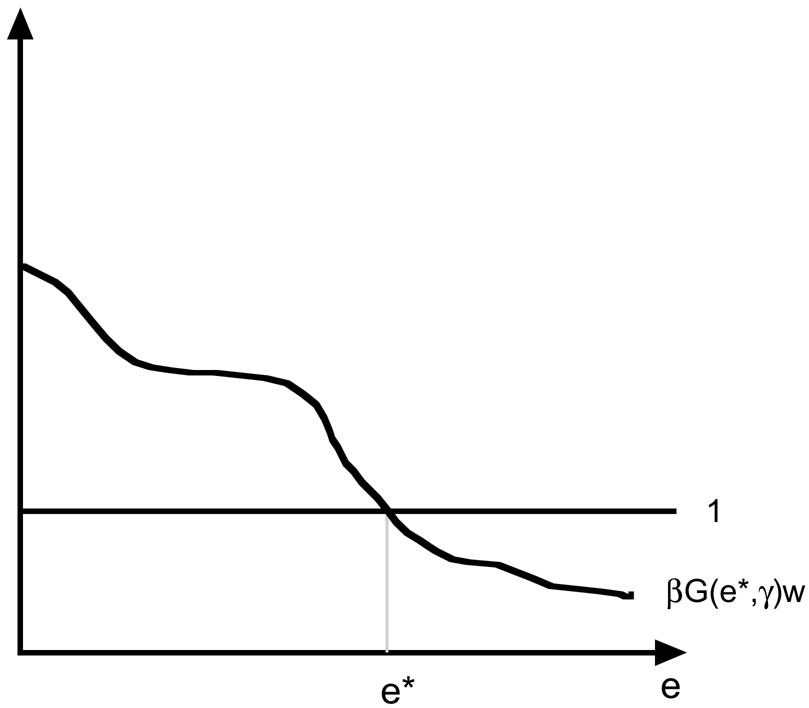


Figure 2a: α is small and w is large - G(e*,γ) is downward sloping

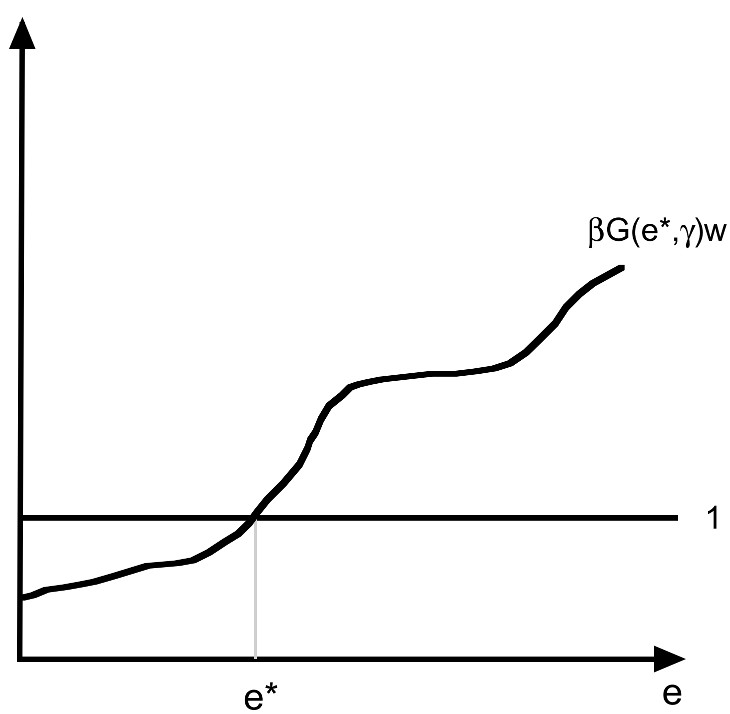


Figure 2b: α is large and w is small - G(e*,γ) is upward sloping

4.2 The effect of the sensitivity of performance to effort on effort

Next, I examine conjecture (ii) that the signalling incentive of agents will become stronger as effort has a stronger impact on performance. Recalling that effort has a larger impact on the probability of success when γ increases, it is obvious that in order to examine this conjecture I need to study the effect of an increase in γ on e^*:

Proposition 4 : *An increase in γ which ensures that effort has a larger impact on the probability that the action will succeed increases the effort level that talented agents exert in the activity in stable interior equilibria and decreases it in unstable interior equilibria.*

.

Proof: Differentiating equation (12) with respect to e^* and γ and rearranging terms,

$$\frac{\partial e^*}{\partial \gamma} = \frac{G_\gamma(e^*,\gamma)}{-G_e(e^*,\gamma)} = \frac{p_{e\gamma}(e^*,\gamma)\Delta(e^*,\gamma) + p_e(e^*,\gamma)\Delta_\gamma(e^*,\gamma)}{-G_e(e^*,\gamma)},$$

where $p_{e\gamma}(e^*,\gamma) > 0$ by Assumption A3, and

$$\Delta_\gamma(e^*,\gamma) = \frac{\alpha\,(1-\alpha)\,p_\gamma(\widehat{e},\gamma)\left[\beta\alpha^2\,(p(\widehat{e},\gamma)-p_0)^2 + p_0(1-\beta p_0)\right]}{(\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0)^2\,(1-\beta\,(\alpha p(\widehat{e},\gamma)+(1-\alpha)p_0))^2} > 0.$$

Hence, the sign of $\frac{\partial e^*}{\partial \gamma}$ is equal to the sign of $-G_e(e^*,\gamma)$ which is positive in stable interior equilibria and negative in unstable interior equilibria. ■

As in the case of an increase in β, an increase in γ shifts $G(e,\gamma)$ upward. Hence, an increase in γ will also lead to more effort by talented agents if $G(e,\gamma)$ is decreasing with e and to less effort if $G(e,\gamma)$ is increasing with e. Again, Lemma 2 shows that $G(e,\gamma)$ is decreasing with e if α is small and w is large, but increasing with e if α is large and w is small. Hence, Lerner and Tirole's (2002) conjecture that the signalling incentive will become stronger as effort has a stronger impact on performance is true only when initially, there are few talented agents around (α is small) and when talented agents earn a high wage (w is high). Otherwise, this conjecture is incorrect: the signalling incentive of agents will become stronger as the agents' effort has a weaker impact on their performance.

4.3 The effect of the informativeness of performance about talent on effort

Conjecture (iii) of Lerner and Tirole states that the signalling incentive of agents will become stronger as performance becomes more informative about talent. This conjecture can be examined by studying the effect of a change in the parameter p_0 on the equilibrium effort level of talented agents, e^*. This is because a decrease in p_0 implies that a successful agent is more likely to be talented; that is, when p_0 decreases towards 0, $q(\widehat{e}, \gamma \mid s)$, which is the probability that a successful agent is talented, increases towards 1.

Proposition 5 : *Suppose that $\Delta_{p_0}(e^*, \gamma) > 0$. Then a decrease in p_0 which ensures that performance is more informative about talent, increases the effort level that talented agents exert in the activity in stable interior equilibria but decreases it in unstable interior equilibria. If $\Delta_{p_0}(e^*, \gamma) < 0$ then the reverse is true.*

Proof: Differentiating equation (12) with respect to e^* and γ and rearranging terms,

$$\frac{\partial e^*}{\partial p_0} = \frac{G_{p_0}(e^*, \gamma)}{-G_e(e^*, \gamma)} = \frac{p_e(e^*, \gamma)\Delta_{p_0}(e^*, \gamma)}{-G_e(e^*, \gamma)},$$

where using the notation $x \equiv \alpha p(\widehat{e}, \gamma) + (1 - \alpha)p_0$,

$$\Delta_{p_0}(e^*, \gamma) = \frac{dq(\widehat{e}, \gamma \mid s)}{dp_0} - \frac{dq(\widehat{e}, \gamma \mid n)}{dp_0} < 0,$$

where the inequality follows because (1) implies that $\frac{dq(\widehat{e},\gamma|s)}{dp_0} < 0$ while (2) implies that $\frac{dq(\widehat{e},\gamma|n)}{dp_0} > 0$. Hence, the sign of $\frac{\partial e^*}{\partial p_0}$ is equal to the sign of $G_e(e^*, \gamma)$, which by Proposition 2 is negative in stable interior equilibria and positive in unstable interior equilibria. Hence, a decrease in p_0 raises e^* in stable interior equilibria and lowers e^* in unstable interior equilibria. ■

Like Propositions 3 and 4, a decrease in p_0 shifts $G(e, \gamma)$ upward. When $G(e, \gamma)$ is decreasing with e, which is the case in stable interior equilibria, this shifts induces more effort. On the other hand, in unstable interior equilibria, $G(e, \gamma)$ is increasing with e so the decrease in p_0 induces less effort. As before, whether the equilibrium is stable or not depends, among other things, on whether α is small and w is large or conversely. In any event, once again the conjecture is true only in stable interior equilibria but not true otherwise.

5 Intrinsic motivation for participation in OSS projects

Up to now I only considered extrinsic motivation for participation in OSS projects. Talented agents took place in these projects in order to try and generate positive signals about their talent and hence boost their prospects in the labor market. However, this view is obviously too narrow given that many participants contribute to OSS projects for other reasons like their sense of creativity, or the desire to solve problems that they face in performing daily tasks (like system managers). The question is how such intrinsic motivations are going to effect matters.

To address this issue, suppose that apart from their ability to boost their prospects in the labor market, agents also draw a positive utility v from successful contributions to OSS projects. Given v, the utility of talented agents becomes

$$U(e) = p(e,\gamma)\left[v + \beta q(\widehat{e},\gamma \mid s)w\right] + \left((1-\beta)p(e,\gamma) + 1 - p(e,\gamma)\right) q(\widehat{e},\gamma \mid n)w - e.$$

The effort level that each talented agent will choose given the firms' beliefs, $\widehat{e}$, is now defined implicitly by the following first order condition:

$$U'(e) = p_e(e,\gamma)\left[v + \beta\Delta(\widehat{e},\gamma)w\right] - 1 \leq 0, \qquad eU'(e) = 0.$$

As one can see, v raises the marginal benefit from effort and hence, other things being equal, it expands the set of parameters for which the model attains an interior equilibrium. Moreover, v shifts the best response function of talented agents outward in the sense that holding the belief of firms, $\widehat{e}$, constant, an increase in v leads to an increase in $BR(\widehat{e})$. Consequently, it is clear that an increase in v will lead to more effort in stable interior equilibria but less effort in unstable equilibria.

6 References

Bessen J. "Open Source Software: Free Provision of Complex Public Goods," Mimeo (available at http://ssrn.com).

Fershtman C. and N. Gandal (2004), "The Determinants of Output per Contributor in Open Source Projects: An Empirical Examination," Mimeo (available at http://ssrn.com)

Hars A. and S. Ou (2002), "Working for Free? Motivations for Participating in Open-Source Projects," *International Journal of Electronic Commerce*, 6(3), 25-39.

Holmström, B. (1999) "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169 182.

Hann I.-H., J. Roberts, S. Slaughter, and R. Fielding (2004), "An Empirical Analysis of Economic Returns to Open Source Participation," Mimeo, Carnegie-Mellon University.

Johnson J. (2002), "Open Source Software: Private Provision of a Public Good," *Journal of Economics and Management Strategy*, 11(4), 121-139.

Lakhani K. and R. Wolf (2002), "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects," MIT Solan Working Paper No. 4425-03.

Lee S., N. Moisa, and M. Weiss (2003), "Open Source as a Signalling Device - An Economic Analysis," WP Series Finance & Accounting, Johann Wolfgang Goethe-University, Frankfurt/Main.

Leppämäki M. and M. Mustonen (2004), "Signaling and Screening with Open Source Programming," Helsinki School of Economics, Working Papers W-377.

Lerner J. and J. Tirole (2002), "Some Simple Economics of Open Source," *The Journal of Industrial Economics*, 50(2), 197-234.

Code Development: Open vs. Proprietary Source*

(Preliminary and incomplete - Comments welcome)

German Daniel Lambardi†

March 15, 2007

Abstract

We develop a model in which a software is obtained from an initial amount of code after successfully overcoming a sequence of steps. The owner of the initial code must decide between carrying out these steps under an open or a proprietary source environment. Open source development will allow the initial code owner to save on developing cost aided by a community of "sophisticated" user-developer, however will imply lower future income on "unsophisticated" end-users. With this model we try to understand why some profit seeking firms may donate code and start open source projects. The dynamic structure of sequential code improvement will provide an alternative explanation on user - developers collaboration. Finally we introduce competition with an existing proprietary source alternative and we show that the incumbent might find optimal to lower prices to user-developers in order to deter entry or prevent the development of the alternative as Open Source.

*I would like to thank Jacques Crémer for his helpful comments and guidance. I am very grateful to Yossi Spiegel for his useful observations. I am solely responsible for any mistakes and errors.

†GREMAQ, Université de Toulouse glambardi@gmail.com

1 Introduction

The Open Source (hereafter denoted OS) Software phenomena has attracted considerable attention form the economist's community in the last few years. Several OS projects became successful alternatives to proprietary source (hereafter denoted PS) software and despite the heterogeneity of success it is now seen as a feasible way to develop code and produce software. Moreover, judging by some recent field studies (i.e.: Dahlander and Magnusson (2005), West and O'Mahony (2005), Bonaccorsi et. al (2004?)) OS development besides being feasible might also be profitable.

Numerous for-profit firms have decided to start OS projects, usually by releasing some valuable internally developed code and inviting an external community to join and collaborate with the code development[1]. The code is made available in a readable form for anyone to screen and make changes. At the same time an important amount of users spent their time developing the code and donating their progress to the community. The main motivation of this paper is to provide an explanation to this behavior by modeling the dynamic multistage process in which the code is developed.

The model assumes that a software is obtained from an initial amount of code, after successfully overcoming a fixed sequence of stages. The probability that the code advances from one stage to the next one is assumed to depend on the amount individuals devoted to the project at each stage. The owner of the initial code, a profit maximizing firm, must decide between carrying out these steps under an OS or a PS environment. Although we initially assume that this firm faces no competition from an alternative software, we later extend the model to capture this aspect.

Open source development will allow the firm to save on programming costs aided by the free programming time provided by a community of "sophisticated" user-developer (hereafter denoted UD). These savings come at a cost, since the royalty free distribution of OS will not allow the firm to make money by selling licences once the software is obtained. We will assume however that OS returns are not cero. OS firms are usually able to sell complementary

[1] West and O'Mahony (2005) call this kind of developments "spinout projects"

proprietary software as well as support and customization services to unsophisticated end-users (hereafter denoted EU). [2]

The PS development, on the contrary, faces higher development cost since the code owner must pay for each individual working on the progress of the software. On the other hand the firm will be able to sell licences once the software is obtained providing a higher future income than the OS alternative.

The set up considers that the UD are able to profit from the raw code at each stage. They derive utility after spending some programming time to adapt the code towards their needs. From the UD perspective each time the code advances to the following step, it becomes more valuable for them (i.e.: they derive more utility). The dynamic structure of the sequential code improvement will also help to understand the dynamic of cooperation: UD will collaborate with the OS firm because if their contributions are included in the next version of the code, then they will need to spend less time to adapt the code in the next stage.

Although the model is still in a basic stage, some results can be highlighted: a) If the firm decides to develop the code as OS, the overall probability to overcome all the stages to get the final software is smaller. This might hurt the welfare of EU and represents a threat that should be taken in consideration when doing welfare analysis of the OS phenomena b) UD might find optimal to collaborate in an OS project at early stages even if at that point they obtain negative utility in order to spend less time in the future adapting the code. c) When competition is introduced in the model and an existing alternative software is able to price discriminate, the incumbent might find profitable to lower the price charged to UD to deter entry or to avoid the development of the code as OS.

[2] see Dahlander and Magnusson (2005)

2 Basic Framework

A firm E owns a certain amount of code C_0 that can be used to produce a software S. To obtain S, the code has to increase and undergo certain transformations. This transformations are modeled as a finite number k of stages which the code must go through. At the final stage S is obtained. The overcoming of each stage requires individuals working on the development of the code, however this development is stochastic and the probability at any stage j that the code C_j advances one step towards C_{j+1} is given by.

$$\phi(N_j) = 1 - e^{-N_j}$$

where N_j is the amount of individuals working on the code development. This development can be done in an open or proprietary environment. If the firm choose to make the code open, some users will be able to read and understand the code and therefore cooperate with the development. By the same token the open environment reduces the ability of the firm to make profits on S due to the royalty free distribution.

The firm faces a population of potential code users. This population is composed by two categories of individuals: "sophisticated" User Developers (UD) and "unsophisticated" end users (EU). If the firm makes the code C_j open at stage j, the UDs are able to profit from it: they can use C_j to perform some private activity for which they derive utility after they spend some programming time to adapt the raw code. The EU group on the contrary can only profit from the code if S is obtained in the last stage.

To be more specific, the utility a UD derives from the finished software S is S. If the C_j is available at stage j the **gross** utility a UD derives from using it is $\alpha_j S$, with $0 < \alpha_j < 1$ and $\alpha_{j+1} > \alpha_j > \alpha_{j-1}$. Moreover each UD has to spend some programming time normalized to 1 to adapt the code towards their needs. With this we want to capture the idea that since C_j is not a "finished" product: it might need to be made compatible with other pieces of software the UD is already using, it might also contain an important amount of "bugs" that require the reprogramming of fractions of the code or the UD might need to develop some new features to use C_j. These modifications and developments represent potential improvements that can make C_j advance to C_{j+1} if they are shared with the rest of the community. For the moment we will assume that UDs are willing to cooperate by sharing their code modifications and they face no cost in doing so. A "reduced form" to model the UD collaboration at stage j is to say that from firm's E perspective it receives free help from an amount m_j of individuals working on the code development.

The desutility of a UD derived from the unit of time spent programming is l, therefore the **net** utility of using the code at stage j is

$$U_j = \alpha_j S - l$$

we further assume that UD are heterogeneous in the desutility l. The total amount of UDs is given by M, and the fraction of them that has l lower than a certain $\widetilde{l}$ is $G(\widetilde{l})$. For simplicity we will assume $G(l) \sim U(0, \bar{l})$ with $\bar{l} > S$.

At each stage j the decision of a UD to use C_j will depend on net utility $\alpha_j S - l$ compared to its outside option. This outside option will be 0 if no alternative software exists but will be positive with competition.

When S is obtained the firm gets some income. The amount of income will depend on the way the code was developed. Under Proprietary Source (PS) development the firm gets V, while under Open Source (OS) development gets δV with $0 < \delta < 1$. The parameter δ represents a shortcut to the idea that the firm E although it is not able to sell S, can still get some revenue from selling support services to EU or by selling some proprietary software complementary to S.

On the cost side, if the firm E develops the code as PS, she has to pay for the whole amount N_j of individuals working on the code development, while under OS development the firm gets free help from m_j individuals of the UD community. Therefore the probability of advancing to the next step is:

$$\phi(N_j) = \phi(n_j + m_j) = 1 - e^{-n_j - m_j} = 1 - e^{-n_j}\gamma_j$$

where

$$\gamma_j = e^{-m_j} < 1$$

and m_j is the amount UD devoted to the project at stage j while n_j is the amount of programmers paid by the firm. Of course under PS development $m_j = 0$ so $N_j = n_j$. Once decided between OS and PS development, the firm must decide at each stage j the amount n_j of programmers hired at wage w.

3 The simplest setting: Two stages and no competition

The utility derived by a UD at each stage (0 and 1) is given by:

$$U_0 \begin{cases} \alpha S - l & \text{if use OS} \\ 0 & \text{if not} \end{cases}$$

$$U_1 \begin{cases} S & \text{if code advanced} \\ \alpha S & \text{if code not advanced} \\ 0 & \text{otherwise} \end{cases}$$

The expression for U_0 follows easily from the basic framework description. The expression for U_1 might need some explanation. First we can see that if C_0 advances to C_1 and S is obtained, all UD's will use OS since the finished software does not require any programming so the associated desutility l is 0. Additionally if $l > \alpha S$ so that a UD does not use OS in stage 0, it will not use it in stage 1 if the code does not advance. Therefore a UD that has used the software at stage 0 does not need to spent programming time if the software does not advance at stage 1.

A UD will participate in OS development if at stage 0

$$\begin{aligned} 0 &< \alpha S - l => \\ l &< \alpha S \end{aligned}$$

The amount of UDs participating in OS development will be

$$m_0 = G(\alpha S)M = \frac{\alpha S}{\bar{l}}M$$

Firm E must choose between OS and PS development . The profit maximization problem under OS is:

$$\max_{n_0} \Pi_0 = (1 - e^{-n_0}\gamma_0)\delta V - wn_0$$

The first order conditions yields

$$\begin{aligned} 0 &= \gamma_0 \delta V e^{-n_0} - w \\ n_0^{os} &= \ln\left(\frac{\gamma_0 \delta V}{w}\right) = -m_0 + \ln\left(\frac{\delta V}{w}\right) \end{aligned}$$

The probability that the code advances will be given by

$$\begin{aligned} \phi(n_0^{os}) &= \left(1 - \gamma_0 e^{-\ln\left(\frac{\gamma_0 \delta V}{w}\right)}\right) \\ &= \left(1 - \frac{w}{\delta V}\right) \end{aligned}$$

so the expected profit under OS development will be

$$\begin{aligned} \Pi_0^{os} &= \left(1 - \frac{w}{\delta V}\right)\delta V - w \ln\left(\frac{\gamma_0 \delta V}{w}\right) \\ & \delta V - w\left(1 + \ln\left(\frac{\gamma_0 \delta V}{w}\right)\right) \\ & \delta V - w\left(1 + n_0^{os}\right) \end{aligned}$$

The profit maximization problem under PS is:

$$\max_{n_0} \Pi_0 = (1 - e^{-n_0})\delta V - wn_0$$

which yields the following results

$$\begin{aligned} n_0^{ps} &= \ln\left(\frac{V}{w}\right) \\ \phi(n_0^{ps}) &= \left(1 - \frac{w}{V}\right) \\ \Pi_0^{ps} &= V - w\left(1 + n_0^{ps}\right) \end{aligned}$$

At this point, some observations should be pointed out.

Proposition 1 *The firm E hires a lower amount of programmers n under OS*

Comparing n_0^{os} with n_0^{ps} we can see where does this reduction comes from. First because the revenue the firm gets at the final stage is now lower (δV_A instead of V_A). Second, there is a substitution effect: the firms perfectly offsets the help m_0 from the community of UD.

Proposition 2 *The overall probability of obtaining the software S is smaller under OS.*

The lower probability under OS is independent on the amount of labor provided by the UD. It comes exclusively from the lower revenue at the final stage (δ) that reduces the labor provision of the firm. EU might be hurt in terms of expected welfare if the firm chooses to develop as OS: Although EUs might face lower prices with OS, the probability that the software is obtained is lower. If the weight of EU is sufficiently high they might offset any welfare gain from the firm and the UDs derived from OS development

Proposition 3 *the firm E will develop the software as OS if the amount of individuals provided by the community exceeds the threshold value $\widetilde{m}$*

This is so because OS will outperform PS if

$$w\left(n_0^{ps} - n_0^{os}\right) > (1 - \delta)V$$

$$\begin{aligned} \ln\left(\frac{V}{w}\right)-\ln\left(\frac{\gamma_0\delta V}{w}\right) &> \frac{(1-\delta)V}{w} \\ \ln\left(\frac{V}{w}\right)-\ln\left(\frac{e^{-m}\delta V}{w}\right) &> \frac{(1-\delta)V}{w} \\ m &> \frac{(1-\delta)V}{w}+\ln\frac{1}{\delta}=\widetilde{m} \end{aligned}$$

4 Adding stages to the simplest setting

When there are only two stages, the development (one stage) could only be made under OS or PS. However if we add stages the firm might be willing to switch from one to the other. In general OS licences are more restrictive and in order to foster collaboration and adoption they usually forbid the code owner to make the code proprietary in the future. Therefore we are reasonably assuming that this is not a possibility in our model. However we can allow the firm to start the project as PS and then switch to OS development. A natural question is wether this is optimal for the firm or not

Proposition 4 *The firm will never find profitable to start as PS and then turn to the OS development*

This is a natural result from our modeling setup. A project that starts as OS versus one that starts as PS and then changes to OS will have the same final income but the later will have higher developing costs. A simple way to illustrate this is to analyze the three stage case (C_0 must advance to C_1 and then to C_2 where S is obtained). The two developing stages could be done as OS or alternatively the first one as PS and the second one as OS. For both cases the problem at the final stage is the same:

$$\max_{n_1}\Pi_1=(1-e^{-n_1}\gamma_1)\delta V-wn_1$$

The optimal n for both cases is given by $n_1^{os}=-m_1+\ln\left(\frac{\delta V}{w}\right)$. Where m_1 is defined by $m_1=\frac{\alpha_1 S}{l}M$. The expected profits is therefore. $\Pi_1^{os}=\delta V-w\left(1+n_1^{os}\right)$.

In the first stage the problem differs. If the firm develops OS in this step then the problem is:

$$\max_{n_0} \Pi_0 = (1 - e^{-n_0}\gamma_0)\Pi_1^{os} - wn_0$$

The optimal n is $n_0^{os} = -m_0 + \ln\left(\frac{\Pi_1^{os}}{w}\right)$. Where $m_0 = \frac{\alpha_0 S}{l} M$, with $\alpha_0 < \alpha_1$ so $m_0 < m_1$. So $\Pi_0^{os} = \Pi_1^{os} - w(1 + n_0^{os})$.

If on the firm develops the first step as PS then the problem is

$$\max_{n_0} \Pi_0 = (1 - e^{-n_0})\Pi_1^{os} - wn_0$$

The optimal n is $n_0^{ps} = \ln\left(\frac{\Pi_1^{os}}{w}\right)$ .So $\Pi_0^{ps} = \Pi_1^{os} - w(1 + n_0^{ps})$.

Since $n_0^{ps} > n_0^{os}$ then $\Pi_0^{ps} < \Pi_0^{os}$.

Since the OS development produces less final income, the whole point in choosing OS is to save on developing cost. Therefore the firm always values the help m provided by the OS community. A key element in this analysis is that this m is decided at each period and does not depend on the history or the future of the project since the UD's opportunity cost is 0.

Another thing to notice is that for PS or OS development it is always the case that $\Pi_j < \Pi_{j+1}$. Then if we compute at any $j-1$ the difference between PS and OS we have that

$$\Pi_{j-1}^{ps} - \Pi_{j-1}^{os} = \Pi_j^{ps} - \Pi_j^{os} + w(n_{j-1}^{os} - n_{j-1}^{ps})$$

so

$$\Pi_{j-1}^{ps} - \Pi_{j-1}^{os} = \Pi_j^{ps} - \Pi_j^{os} + w\left(\ln\left(\gamma_{j-1}\frac{\Pi_j^{os}}{\Pi_j^{ps}}\right)\right)$$

From this expression we can see that since $\gamma_{j-1} = e^{-m_{j-1}} < 1$, the difference between $\Pi_{j-1}^{ps} - \Pi_{j-1}^{os}$, if positive, is reduced as we move to stage 0. Therefore

Proposition 5 *Given two projects with the same final V, if one of them has more stages than the other it is more likely that OS development will outperform PS development*

Therefore, controlling for V, we would tend to see that projects requiring long development periods are carried on as OS.

5 Adding competition. The entry game.

We will consider that a Firm I is already selling a proprietary software in the market. Firm E must decide to enter or not the market and whether to develop a software as OS or PS. To make the problem interesting we assume that firm I can price discriminate between sophisticated users (potential UD of a OS alternative) and unsophisticated ones (EU). We also consider that the Firm E faces some entry costs F. To keep the problem simple we continue to assume a two stages developing process.

The timing of the problem is as follows

1. Incumbent PS firm I, must decide on the price p it charges to UD
2. Firm E decides between paying F to enter or not and whether to develop as OS or PS

   If firm E enters as OS, UD's chooses between paying for firm's I finished software or using firm E code. If firm E chooses PS, UD just buy firm's I software.
3. First period payoffs are realized. n and m are provided.
4. Second period payoffs are realized

The payoffs of the different agents are as follows:

a If firm E enters as PS, the final income V is shared in the following way, $\lambda^{ps}V$ goes to firm E and $(1-\lambda^{ps})V$ to firm I.

b If firm E enters as OS we will assume that the overall income of the industry is reduced in a proportion $(1-\delta)$ with $(\delta < 1)$ then $\lambda^{os}\delta V$ goes to firm E and $(1-\lambda^{os})\delta V$ to firm I. The variable δ captures the idea that the competition in the EU market is tougher under OS. Since OS development entails lower costs we will assume, to make the problem interesting, that $\lambda^{ps}V > \lambda^{os}\delta V$. Moreover, to simply expressions we will assume that $\lambda^{os} = \lambda^{ps} = \lambda$

c The utility derived at each period of a UD buying the PS software of firm I is

$$U_0^{ps} \begin{cases} S-p & \text{if buy} \\ 0 & \text{if not} \end{cases}$$

$$U_1^{ps} \begin{cases} S & \text{if buy in 0} \\ 0 & \text{if not} \end{cases}$$

d The utility of UD derived from using OS of firm E is

$$U_0^{os} \begin{cases} \alpha S - l & \text{if use OS} \\ 0 & \text{if not} \end{cases}$$

$$U_1^{os} \begin{cases} S & \text{if code advanced} \\ \alpha S & \text{if code not advanced} \\ 0 & \text{otherwise} \end{cases}$$

Again if a UD has chosen not to use OS in stage 0 he will not use it in stage 1 if the code has not advanced. Another thing to notice is that if the firm I faced no entry or the development is made PS, it could extract all the UD surplus by charging the price $p^m = 2S$. However if firm E's code is made OS then the maximum price for which there is demand is $p^{\max} = 2S - \phi S$. This is so since a UD can always wait until stage 1 to get the OS software without spending programming time and this happens with probability ϕ. Therefore the outside option for the UD is ϕS. From this reasoning we conclude that if firm E develops as OS then at stage 0 the whole population M of UDs will be divided between those that adopt OS those that buy I's software.

5.1 Determination of m

If the firm E has entered as OS, a UD will choose it if:

$$\alpha S - l + \phi(S - l) + (1 - \phi)(\alpha S - l) > 2S - p$$

$$l < \frac{p - (2-\phi)(1-\alpha)S}{2} = \widehat{l}$$

The UD with $l = \widehat{l}$ has desutility high enough such that he is indifferent between paying or spending time on OS

This implies the following amount of labor for the firm E (assuming an uniform distribution)

$$m(\phi, p) = G(p - (2-\phi)(1-\alpha)S)M. = \frac{p - (2-\phi)(1-\alpha)S}{\bar{l}} M$$

The expected signs $\frac{\partial m}{\partial p} > 0, \frac{\partial m}{\partial \phi} > 0, \frac{\partial m}{\partial \alpha} > 0$ are obtained.

Proposition 6 *The amount UD help the OS project receives is decreasing in the price of the proprietary alternative, increasing in the in the probability ϕ of advancing to the next step and increasing in the gross utility the software provides at the developing stage.*

5.2 Determination of n, ϕ and $\widetilde{p}$

for a given amount m the optimal choice of labor by the OS firm is

$$\begin{aligned} n_0^{os}(m) &= \ln\left(\frac{\gamma\delta\lambda V}{w}\right) \\ &= \ln\left(\frac{e^{-m}\delta\lambda V}{w}\right) \\ &= -m(\phi, p) + \ln\left(\frac{\delta\lambda V}{w}\right) \end{aligned}$$

for a given m and n the probability that the code advances next step is

$$\phi^* = \left(1 - \frac{w}{\delta\lambda V}\right)$$

Then the probability ϕ^* does not depend on n or m itself. Therefore for a given p we solve for m and n_0^{os}. The profit of firm E developing as OS is

$$\begin{aligned}
\Pi_0^{os}(p) &= \left(1-\frac{w}{\delta\lambda V}\right)\delta\lambda V - w\ln\left(\frac{\gamma\delta\lambda V}{w}\right) - F \\
&\quad \delta\lambda V - w\left(1+\ln\left(\frac{\gamma\delta\lambda V}{w}\right)\right) - F \\
&\quad \delta\lambda V - w\left(1+n_0^{os}(p)\right) - F
\end{aligned}$$

If the firm decides to develop as PS the optimal amount of labor is given by

$$n_0^{ps} = \ln\left(\frac{\lambda V}{w}\right)$$

Of course since the firm is no longer relying in the community of UD's, the amount of labor it hires does not depend on the price p the firm I charges to UD's. Therefore firm E PS profits will not depend also on p :

$$\begin{aligned}
\Pi_0^{ps} &= \left(1-\frac{w}{\lambda V}\right)\lambda V - w\ln\left(\frac{\lambda V}{w}\right) - F \\
&\quad \lambda V - w\left(1+\ln\left(\frac{\lambda V}{w}\right)\right) - F \\
&\quad \lambda V - w\left(1+n_0^{ps}\right) - F
\end{aligned}$$

Comparing Π_0^{ps} and $\Pi_0^{os}(p)$ firm E will prefer OS if

$$m(p) > \frac{(1-\delta)\lambda V}{w} + \ln\frac{1}{\delta}$$

therefore the p that makes firm E indifferent between OS and PS is

$$\begin{aligned}
m(p) &= \frac{(1-\delta)\lambda V}{w} + \ln\frac{1}{\delta} \\
\frac{p-(2-\phi)(1-\alpha)S}{\bar{l}}M &= \frac{(1-\delta)\lambda V}{w} + \ln\frac{1}{\delta} \\
\widetilde{p} &= \frac{\bar{l}}{M}\left(\frac{(1-\delta)\lambda V}{w} + \ln\frac{1}{\delta}\right) + (2-\phi)(1-\alpha)S
\end{aligned}$$

From the expression above we can see that $\widetilde{p}$ is positive. This is so because we are assuming that the raw code gives less utility to UD than firm I software at stage 0 (i.e. $\alpha < 1$) If this is not the case (i.e. having a $\alpha_0 > 1$ at the first stage and $\alpha_1 > \alpha_0$ at the final stage) then $\widetilde{p}$ could be negative. This would imply that firm I is "subsidizing" and paying some UD's to avoid OS development.

Proposition 7 *If the firm I, charges a prices $p \leq \widetilde{p}$ and the firm E decides to enter, the code will be developed as PS. If the firm I charges a prices $p > \widetilde{p}$ and the firm E decides to enter, the code will be developed as OS.*

5.3 Determination of p

The optimal problem of the incumbent firm is to decide p.

The profits for the firm I are given by

$$\Pi^I \begin{cases} \phi^{os}(1-\lambda)\delta V + (1-\phi^{os})V + p(1-G(p))M & \text{for } p > \widetilde{p} \text{ such that A chooses OS} \\ \phi^{ps}(1-\lambda)V + (1-\phi^{ps})V + pM & \text{for } p = \widetilde{p} \text{ such that A chooses PS} \\ V + pM & \text{for } p \geq \widetilde{p} \text{ such that A does not enter} \end{cases}$$

The minimum price firm I would charge is $\widetilde{p}$: under PS setting p further down has no effect on E's profit since it is not relying on UDs to develop the code. Therefore

Proposition 8 *entry deterrence using exclusively p can only be possible if at prices $p \geq \widetilde{p}$ (i.e.: at OS development) the firm E has negative profits (i.e.: fixed cost F are sufficiently high). If profits $\Pi_0^{ps}(\widetilde{p})$ of firm E are positive, the we can be sure that firm I will not be able to deter entry.*

If possible, deterrence will be optimal for firm I if

$$V + p^d M > \phi^{os}(1-\lambda)\delta V + (1-\phi^{os})V + p^{nd}\left(1 - G(p^{nd})\right) M$$

Where p^d denotes deterrence price and p^{nd} no deterrence price. Since $V > \phi^{os}(1-\lambda)\delta V + (1-\phi^{os})V$, it is required for deterrence to be optimal that $p^d > p^{nd}(1 - F(p^{nd}))$ and this is not always true. Notice first that, because ϕ^{os} does not depend on p, then the optimal p^{nd} just comes from

$$\max p\,(1 - G(p)\,M$$

so

$$p^{nd} = \frac{\bar{l} - (2-\phi)S(\alpha - 1)}{2}$$

On the other side p^d is just the highest price that makes $\Pi_0^{os}(p^d) \leq 0$. Therefore if $p^d > p^{nd}$ the condition for deterrence will hold trivially. On the other hand if $p^{nd} > p^d$ then the fraction of UD that will demand firm's I software should be such that:

$$(1 - F(p^{nd})) < \frac{p^d}{p^{nd}}$$

It might be more interesting instead, to think on what is needed to make entry deterrence **not** optimal:

$$V + p^d M < \phi^{os}(1-\lambda)\delta V + (1-\phi^{os})V + p^{nd}\left(1 - G(p^{nd})\right)M$$

Since $V > \phi^{os}(1-\lambda)\delta V + (1-\phi^{os})V$, $p^{nd}\left(1 - G(p^{nd})\right)M > p^d M$ must be big enough such that the inequality is reversed. Then, not only $p^{nd}\left(1 - G(p^{nd})\right) - p^d > 0$ is needed but also M should be big enough. This suggests that it might be more likely to observe entry accommodation of a OS alternative when the market of sophisticated users is rather big and on the contrary a more aggressive price behavior should be expected if this market is rather small.

Despite the fact that low prices to sophisticated users (compared to final users) by incumbent firms are probably natural due to price elasticity, the previous analysis also suggests that it might also hide some entry/OS development deterrence.

Finally if entry deterrence is not possible, then firm I must decide whether to set $p = \widetilde{p}$ to trigger PS development or set $p > \widetilde{p}$ and have OS development.

PS development will be preferred if

$$\phi^{ps}(1-\lambda)V + (1-\phi^{ps})V + \widetilde{p}M > \phi^{os}(1-\lambda)\delta V + (1-\phi^{os})V + p(1-G(p))M$$

or

$$(\phi^{ps} - \phi^{os}\delta)(1-\lambda)V + (\phi^{os} - \phi^{ps})V + (\widetilde{p} - p(1-G(p)))M > 0$$

Several trade off work at the same time and might offset each other. On the EU income side, although OS reduces industry income due to parameter δ, it also entails a lower probability of success ϕ^{os}, therefore the difference between PS and OS expected income remains undetermined. On the UD side, although $\widetilde{p} < p$, it is charged over all the UD.

6 Alternative formulations

UD instead of being heterogenous in l they could differ in the utility they derive from the code. The utility derived from a finished software S is θ. If the software is developed as OS, the **gross** utility derived from using the code at stage j C_j is $\alpha_j\theta$. Again they have to spend some programming time, normalized to 1, to adapt the code. Therefore the **net** utility of using the code at stage j is

$$\alpha_j\theta - 1$$

The total amount of UD is given by M, and the fraction that has θ higher than $\widetilde{\theta}$ is $F(\widetilde{\theta})$. For simplicity we will assume $F(\theta) \sim U(0,\overline{\theta})$ with $\overline{\theta} > 1$. The decision of a UD to participate in OS will depend on $\alpha_j\theta$ compared to 1 and to its outside option. At each stage only individuals such that $\theta > \alpha_j$ will participate.

A variation on this could be that the net utility to be $\alpha_j\theta t - \frac{1}{2}t^2$. And t is the amount of time the UD spends in the project. The optimal amount of time at each period would be given by $t^* = \alpha_j\theta$. Individuals with high θ are those who participate more in the project, and this participation is increasing in at each stage since $\alpha_{j+1}\theta > \alpha_j\theta$

6.1 Dynamics of cooperation

Here we develop the idea that UD will collaborate with the OS firm because if their contributions are included in the next version of the code, then they will need to spend less time to adapt the code. This is modeled assuming that in the next period it will be necessary to spend only ρ $(0 < \rho < 1)$ hour to use the code (ρ can be interpreted as a probability or the fraction of what was included in the code) Contributing, however, is not free, it implies a desutility c (this could be interpreted as for example as time spent submitting findings).

Assume 2 stages

$$U_0^{os} \begin{cases} \frac{1}{2}\theta - 1 - c & \text{if use OS and contribute} \\ \frac{1}{2}\theta - 1 & \text{if use OS and not contribute} \\ 0 & \text{otherwise} \end{cases}$$

$$U_1^{os} \begin{cases} \theta - \rho & \text{if code advanced and contribute} \\ \theta - 1 & \text{if code advanced and did not contribute} \\ \frac{1}{2}\theta & \text{if code not advanced} \\ 0 & \text{otherwise} \end{cases}$$

the condition for contribution is

$$\phi(1-\rho) > c$$

The UD with the minimum θ that will collaborate is

$$\widetilde{\theta} = \frac{c + \phi\rho + 1}{\frac{1}{2}\phi + 1}$$

assuming the worst situation, that is a c such that $c = \phi(1-\rho)$ then

$$\widetilde{\theta_w} = 2\frac{\phi + 1}{\phi + 2} > 1$$

since

$$\frac{\phi + 1}{\phi + 2} > \frac{1}{2}$$

This imply that the possibility of future labor cost reduction generates that some UD with negative first period utility engage in OS.

7 Related Literature

The paper that is more closely related to our way of modeling OS vs PS code development is Aghion, Dewatripont and Stein (2005). Their aim is to clarify the respective advantages and disadvantages of academic and private-sector research. As in our model an idea must overcome a fix number of stages to become a marketable product. With their model they determine the optimal path of academia/private sector stages the idea has to follow. While academia's (low focus) creative freedom implies a lower probability to advance stages, private research higher focus is more costly. Academia research is less expensive since lower wages paid to academics scientist reflect their willingness to forgo earnings in exchange for academic freedom. Their main finding is that private sector's "expensive" focus strategy only pays in later-stage research.

An important source of stylized facts on OS firms comes from Dahlander (2005). This paper provides a multiple case study on OS firms and the main goal is to address how OS firms generate returns and how that changes over time. From the six cases of small firms in Sweden and Finland, two of them where particularly useful to our model. One of the cases is related to a well established "second generation" OS firm that produces a database software (Mysql). The second firm exploits a webserver software solution (Roxen). Both OS projects were started by the firms and the software was developed using the support of UD. Firms now make profits on a combination of support services, software installation and customization (making the software "fit the customers") and from selling licences on complementary software. The paper stress the importance of being fist movers and building a significantly big "community" that will help on the product development. This is closely related to our threshold value $\widetilde{m}$ that makes OS development feasible.

The idea that firms might collaborate with and benefit from OS is not new in the OS literature. Schmidtke (2006) views OS phenomena as the private provision of a public good. The model suggests that although improvements in such a non-excludable public good cannot be appropriated, companies can benefit indirectly from the OS "quality" in a complementary proprietary segment. The model assumes that OS software already exist (no code development is modeled) and the author is interested in crowding in/out form public investment and pricing strategies on the complementary good under market entry.

The idea of competition between OS and a PS alternative has also been studied. Sen (2005) analyses the competition game between a freely available open source software (OSS), the commercial version of the same (OSS-SS) and a proprietary software (PS). Two dimensions characterize the software: its network benefits and the usability. Conditions under which PS dominates the market are analyzed and OSS-SS in not always found to hurt PS alternative. Our model differs from this one because we rather study the pricing strategy of the PS firm towards the UD to prevent the development of the OS alternative. Verani (2006) builds a model were firms compete in a differentiated duopoly. Each firm sells a good each made of two components, one of them a software. Demand depends in prices and quality of the "software" component. If goods are substitutes, the author finds that the investment in "software" quality is bigger under OS due to spillover effects across firms.

Finally, there has been a significant amount of literature to explain motivation of programmers to collaborate in OS development. Dewan et. al (2005) suggest a Principal (firm)-Agent (programmer) model with learning. The firm benefits from the programmer's participation in OS because he acquires skills that are useful for the firms own project. The firm cannot monitor the programmers effort division and too much OS attention might hurt firms project success. The programmer, on the other hand. wants to work in OS because it allows him to signal his talent to other firms and increase his wage. Spiegel (2005) also builds on a Principal - Agent model where programmers participate on OS to signal their talent to potential employers. In our model we point out another possible source of cooperation that differs from this literature: the dynamic of sequential code improvement and the possibility of saving programming time in future versions of the code.

Finally Athey and Ellison (2006) with a different modeling strategy also focus on the dynamics of OS. In their model, altruism and the anticipation of altruism are the key element that triggers OS development. The rate of decay of altruism jointly with the arrival of new programmers needs are the ingredients that govern the growth of the OS software. They also considers the effects of commercial competition on OS dynamics and they show commercial firms might reduce their prices to slow the growth of OS projects.

8 Conclusions

Although the model we presented here is still work in progress, we believe that gives some insights on why some for-profit firms decide to start OS projects and when this is more likely to happen. High development cost, long development periods, low profits on EU and a significant population of UD are important ingredients to trigger OS development. We also find that the overall success probability is lower under OS development. The introduction of an existing PS alternative in the model allows to stress the importance of the UD community on the OS development. The incumbent firm by reducing prices charged to sophisticated users, is able to reduce the size of the community of UD that collaborate on the OS development and in this way it can deter entry or induce the PS development. Finally the dynamic structure of sequential code improvement provides an alternative explanation on user - developers collaboration.

Besides the existing results we consider that this set up has a significant potential for future extensions and improvements:

- Competition between UD can be introduced to see to which extent the code development is affected and to study the rivalry conditions that allow OS success.
- The difference between UD and EU could be endogenized so the amount of EU could be made a choice variable for the firm (i.e.: a EU is just someone that verifies $\alpha S - l < 0$).

If the firm's income comes from selling support to EU, and the number of EU is reduced as the OS progresses, the firm might choose not to develop the code too much.

- Issues on Governance structure could be studied (i.e.: might the firm find profitable to forgo the control of the code development?)

9 References.

- Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein, (2005), "Academic Freedom, Private-Sector Focus, and the Process of Innovation", NBER Working Paper No. 11542.

- Athey, S and G. Ellison (2006), "Dynamics of Open Source Movements". Working Paper.

- Bonaccorsi A., Rossi C., and S. Giannangeli (2004), "Adaptive entry strategies under dominant standards.Hybrid business models in the Open Source software industry'

- Dahlander, L. and M. G. Magnusson (2005), "Relationships between open source software companies and communities: Observations from Nordic firms", Research Policy 34, pp. 481–493

- Dewan, Rajiv, Marshall Freimer, and Amit Mehra, (2005), "When a firm's employees work on open source projects"

- Lerner, Josh and Jean Tirole (2002),"Some simple economics of open source" Journal of Industrial Economics 50 no. 2, 197–234.

- Schmidtke, Richard, (2006), "Private Provision of a Complementary Public Good," Discussion Papers in Economics 964, University of Munich, Department of Economics.

- Sen, Ravi (2005). "A Strategic Analysis of Competition Between Open Source and Proprietary Software," Industrial Organization 0510004, EconWPA.

- Spiegel, Yossi (2005). "The Incentive To Participate In Open Source Projects: A Signaling Approach," Working Papers 05-23, NET Institute.
- West, J. and S. O'Mahony, (2005), "Contrasting Community Building in Sponsored and Community Founded Open Source Projects" Proceedings of the 38th Annual Hawai's International Conference on System Sciences, Waikoloa, Hawaii, January 3-6, 2005.
- Verani, Stephane, (2006). "Open Source Development in a Differentiated Duopoly," Economics Discussion / Working Papers 06-05, The University of Western Australia, Department of Economics.

On Bundling by Natural Monopolies

Hanming Fang* Peter Norman†

February 2007

Abstract

In the presence of fixed costs and informational constraints, commodity bundling may be necessary for the allocation to be constrained efficient. A profit maximizing monopolist that is allowed to use the bundling instrument may be better for the consumers than any incentive feasible provision mechanism that does not bundle. The parameter region where bundling dominates is enlarged relative to the benchmark case when there are no complementarities if goods are either complements or substitutes.

Keywords: Bundling, Fixed Costs, Exclusion, Efficiency, Complements, Substitutes.

JEL Classification Number: H41.

*Department of Economics, Yale University, P.O. Box 208264, New Haven, CT 06520-8264. Email: hanming.fang@yale.edu

†Department of Economics, University of North Carolina at Chapel Hill, 107 Gardner Hall, Chapel Hill, NC 27599-3305 normanp@email.unc.edu

1 Introduction

The purpose of this paper is to take a step towards the development of a normative framework for analyzing under which circumstances bundling by firms with market power can be justified from economic efficiency considerations.

Existing case law and many suggested legal tests all point to the need to investigate both *whether bundling is likely to be useful for anti-competitive purposes* and whether *there are any possible efficiency enhancing arguments for bundling* in the market in question. The first of these two questions is addressed in many studies of oligopoly models, for example Carlton and Waldman [7], Choi and Stefanadis [9], Matutes and Regibeau [22], Nalebuff [23], and Whinston [32]. In the law literature, bundling is commonly defended by arguing that the practice is common also in competitive markets, which may be explained by: i) cost reductions from bundling; ii) economics of scale or scope, or; iii) complementarities in preferences, or; some more elusive transaction cost arguments.[1] In contrast, economists have largely ignored the potential efficiency benefits, so, apart from obvious effects, we know little about how these are supposed to work.

We focus on the case with a monopolist with diminishing average costs: many goods that are commonly provided in bundles (for example cable TV, music and electronic journals) fit well in this natural monopoly setup. Moreover, arguments relating to economics of scale or scope are indeed often a significant point of debate in actual legal cases.[2]

We find that falling average costs can justify the use of commodity bundling in some cases. What is surprising about this is not that a case for bundling can be built on scale economics, but that there are cases when the desirability is completely unambiguous. It is possible that the *profit maximizing* allocation when commodity bundling is allowed is better for all agents than the *constrained social optimum* in a regime when commodities cannot be sold as a bundle. The striking implication of this is that an order to unbundle may make both consumers and the producer worse off, regardless of any other remedies that are added to the intervention. As our environment is one where we do not impose any ad hoc restrictions on the available policy instruments, this is a rather strong result.[3]

[1] A brief statement of many of the claimed efficiency benefits can be found in Thorne [29]. Also see Evans and Padilla [11], Evans and et al [30], and Kobayashi [18] for more extended discussions.

[2] Efficiency arguments based on cost or preference synergies that require bundled sales are also common. However, such arguments postulate either that the bundle is cheaper to produce than the components or that consumers view the bundle as a different product than the collection of all the components. Bundling is therefore desirable by *assumption* in these cases. Such synergies may sometimes be relevant (one of the most convincing cases is in Times-Picayune Publishing Co. v. United States [28] where the argument that bundling ads across editions saves on typesetting costs seems highly plausible), but it is unclear whether any further theoretical analysis is needed to understand the effects better.

[3] Examples where bundling has desirable effects for consumers due to the presence of fixed costs have been discussed before (see for example Brown and Alexander [6]). What is new in our analysis is that we take a mechanism design perspective, which allows us to characterize the best outcome that can be achieved without bundling with

We also investigate how our conclusions are changed if goods are either complements or substitutes. One motivation for this exercise is that it is commonly asserted that bundling is obviously desirable in case most consumers want to consume certain goods/components together. That is, the argument is that we observe that computers and software are sold in bundles because most users want pre-installed software.

It is not hard to see that this argument is nonsense, unless combined with some other rationalization for bundling. If consumers like pre-installed software, they may equally well purchase the software separately when customizing the computer, and the option of buying a particular computer without an existing operating system would benefit those who prefer other software.[4] Indeed, our analysis shows that the case for justifying bundling from economic efficiency when goods are complements is qualitatively the same as in the benchmark case without complementarities in preferences. However, we do find that it is "more likely" that bundling is efficient when goods are complements in the sense that the subset of parameter values for which bundling dominates separate provision is larger than in the benchmark case. Maybe surprisingly, this is true also in the case when goods are substitutes. Hence, one may argue that the important insight from considering complements and substitutes is that benefits are continuous in the degree of complementrity/substitutability, and that the special case when goods are neither complements or substitutes is useful to get a conservative estimate of the benefits.

1.1 Background

Bundling refers to the practice to sell two or more separate products as a package deal. Sometimes, purchasing the bundle is the only way to obtain the component goods, and this is usually referred to as *pure bundling*. In other cases, consumers have a choice between purchasing the components of the bundle as individual goods and buying the bundle, typically at a discount when compared with the prices of the individual goods. This practice is referred to a *mixed bundling* or as *bundled discounts*.

Pure and mixed bundling occur both in competitive and concentrated markets. In itself, bundling is legal under the current interpretation of US antitrust laws provided that no "forcing" is present: in a ruling on case about bundling of hospital care and anesthesiological services, the Supreme Court argued (See Jefferson Parish Hospital District No 2 v. Hyde [17]):

> Thus, the hospital's requirement that its patients obtain necessary anesthesiological services from Roux combined the purchase of two distinguishable services in a single

no restrictions whatsoever on the available contracts. Our results will therefore not go away if, say, additional instruments for price discrimination are introduced to the model.

[4]Sometimes, transaction costs, are added to this preference based explanation. In the special case of purchasing a computer online this seems completely implausible as the computer is customized by the consumer, who is already facing many options to add on the software package (but not to select "no operating system").

> transaction. Nevertheless, the fact that this case involves a required purchase of two services that would otherwise be purchased separately does not make the Roux contract illegal. As noted above, there is nothing inherently anticompetitive about packaged sales. Only if patients are forced to purchase Roux's services as a result of the hospital's market power would the arrangement have anticompetitive consequences. If no forcing is present, patients are free to enter a competing hospital and to use another anesthesiologist instead of Roux. The fact that petitioners' patients are required to purchase two separate items is only the beginning of the appropriate inquiry.

In contrast, bundling *is* a violation of US antitrust law if it can be established that bundling is used for an *unlawful exercise of monopoly power*. Exactly what this means operationally is somewhat blurry, but it seems clear is that a substantial market share is considered to be a necessary condition[5]

It is also evident that the courts do take potential efficiencies from bundling into account, and weigh these factors against anticompetitive effects. For example, in Jefferson Parish Hospital District No 2 v. Hyde [17] a point of debate was whether the contract was necessary for 24-hour anesthesiology coverage, in Times-Picayune Publishing Co. v. United States [28] it was recognized that requiring advertisers to place ads in both morning and evening editions saved costs of resetting the paper, in BSkyB decision dated 17 December 2002 [27] the main justification for allowing bundled discounts is that it is necessary for the seller to break even, implying that the programming would not be available without the bundling instrument.

Given this somewhat fuzzy legal landscape, it is hardly surprising that the pros and cons of bundling have been debated extensively in the recent literature on law and economics. Some studies (for example Ahlborn et al [2], Kuhn et al [19] and Nalebuff [24]) are arguing for increased use of economic theory as a guidance. Others, such as Evans et al [12] and Kobayashi [18] are more sceptical, and argue that economic theory is of limited value, largely because of unrealistic and restrictive assumptions that rule out any benefits from bundling.[6] This complaint echoes testimony in several anti-trust cases that argue that the existing simplified models abstract from crucial aspects of the problem. It is our hope that our slightly more flexible setup, which includes fixed costs and allows for complementarities, is a step in the right direction to make the economics literature more practically relevant.

[5] This is not unique to the US. The European commission has made several controversial decisions relating to both merger control and abuse of dominance where concerns about anticompetitive effects of commodity bundling have been important (publicized examples are the proposed GE/Honeywell and Tetra Laval/Sidel mergers and the order to unbundle the Microsoft media player from Windows). Again, the legal picture is rather unclear.

[6] It is also held that anticompetitive effects usually occur in restricted parameter ranges and that, as a practical matter, it is impossible to identify whether we are in a case where parameters are such that bundling is harmful or not.

2 The Model

Our model is related to Fang and Norman [14], but the focus here is policy oriented, and the model has been amended accordingly.[7] For tractability, we only consider a minimal type space where, for each good, consumers can only have a high or a low valuation. In addition:

1. In this paper, we assume that the set of agents is a continuum. This is *not* merely a technicality, as the continuum assumption makes it impossible to make aggregate outcomes depend on the behavior of finite sets of agents.[8] In our setup this simplifies the analysis considerably, because the provision decision becomes an "ex ante decision" instead of a function of the realized private information.[9]

2. The framework is extended relative to Fang and Norman [14] in that we allow the goods to be either substitutes or complements. A more minor extension is that the cost function is somewhat richer in that it allows for positive marginal costs.

2.1 Costs, Preferences, and Types

A continuum of ex ante identical agents have preferences over two indivisible goods, labeled good 1 and 2 respectively. Both goods have falling average costs. In order to produce any quantity of good j a fixed (per capita) cost $K^j > 0$ must be incurred. There is also a unit variable cost $c^j \geq 0$ for good $j = 1, 2$.

Consumers are privately informed about their preferences over the two goods, and, for simplicity, it is assumed that the willingness to pay for a good can take on only two values $\theta^j \in \left\{l^j, h^j\right\}$. Hence, the type space consists of four types and is given by

$$\Theta = \left\{\left(l^1, l^2\right), \left(l^1, h^2\right), \left(h^1, l^2\right), \left(h^1, h^2\right)\right\}. \tag{1}$$

The standard formulation in the literature assumes that the willingness to pay for the bundle equals the willingness to pay for the components. Here, we allow complementarities across goods, while still retaining the simple type space in (1), by assuming that the value of consuming both goods is scaled up or down relative to the sum of the single good valuations by a commonly known parameter. That is, we let $\mathrm{I}^j \in \{0, 1\}$ be a dummy that takes on value 1 when good j is consumed and assume that the utility of a type θ agent is given by

$$u\left(\theta^1, \theta^2, \mathrm{I}^1, \mathrm{I}^2, t\right) = \sum_{j=1,2}^{j} \mathrm{I}^j \theta^j + \mathrm{I}^1 \mathrm{I}^2 (\gamma - 1)\left(\theta^1 + \theta^2\right) - t, \tag{2}$$

[7] Other recent related papers are Geng et al [13] and Hellwig [16].

[8] This has conceptual implications for mechanism design applications. For example, with an atomless distribution there is no obvious way to write down well known mechanisms such as the Groves [15] mechanism.

[9] Fang and Norman [14] justifies the continuum analysis as an approximation of a large finite set of agents.

where $\gamma > 0$. If $\gamma < 1$ we will say that the goods are substitutes and if $\gamma > 1$ we will say that the goods are complements.

We denote the probability that an agent is of type $\left(j^1, k^2\right)$ by α^{jk}: the order of the superscripts indicates what valuation refers to which good. By assumption, α^{jk} is also taken to be the exact realized proportion of agents with type $\left(j^1, k^2\right)$ and we write

$$\alpha = \left(\alpha^{ll}, \alpha^{lh}, \alpha^{hl}, \alpha^{hh}\right) \in \Delta^4 \tag{3}$$

as shorthand for the distribution that is doing double duty as the probability distribution over types for a representative agent and the deterministic frequency distribution in the economy.

2.2 The Constrained Efficiency Problem

There are no benefits from treating different agents of the same type asymmetrically (see Proposition 1 in Fang and Norman [14]). Together with the assumption that the realized proportion of each type is deterministic this implies that we without loss of generality can focus on a simple class of mechanism consisting of:

1. a provision rule $\rho = \left(\rho^1, \rho^2, \rho^{12}\right)$, denoting the probability of providing goods 1 only, good 2 only, and both goods respectively, and;

2. a transfer rule $t : \Theta \to R$. We adopt the convention that $t(\theta)$ is the transfer from an agent with type $\theta \in \Theta$ to the mechanism designer, and;

3. a conditional allocation rule $x : \Theta \to [0,1]^3$, where $x(\theta) = \left(x^1(\theta), x^2(\theta), x^{12}(\theta)\right)$ are the probabilities that an agent with type θ consumes good 1 only, good 2 only and both goods respectively, conditional on the goods being provided. [10]

Utility is transferable (quasi-linearity), implying that any surplus that is created by a mechanism can be divided arbitrarily between the agents by use of ex ante lump sum transfers. We will therefore focus on the problem where the planner seeks to maximize the ex ante expected payoff of the representative consumer, which fully characterizes the set of *ex ante* Pareto optimal allocations.[11]. Substituting $(\rho, x(\theta), t(\theta))$ into (2) and taking expectations we can express this objective as

$$\mathrm{E}_\theta\left[U\left(\theta, \rho, x(\theta), t(\theta)\right)\right] = \mathrm{E}_\theta\left[\sum_{j=1,2}^{j} \underbrace{\rho^j x^j(\theta)}_{\substack{\text{Prob that } \theta \\ \text{consumes } j \text{ only}}} \theta^j + (\gamma - 1)\underbrace{\rho^{12} x^{12}(\theta)}_{\substack{\text{Pr that } \theta \\ \text{consumes both goods}}}\left(\theta^1 + \theta^2\right) - t(\theta)\right] \tag{4}$$

[10] It is possible to combine the provision and inclusion decisions in x and thereby "simplify" the setup by dropping ρ from the problem. However, the separation of provision and inclusion decisions makes the setup more transparent and intuitive.

[11] A larger set of Pareto efficient outcomes is obtained if *interim Pareto efficiency* is considered. See Ledyard and Palfrey [20].

As we assume that information about preferences is private, we maximize (4) under the constraint that truth-telling is optimal for each type,

$$\theta \in \arg\max_{y \in \Theta} U\left(\theta, \rho, x\left(y\right), t\left(y\right)\right) \text{ for all } \theta \in \Theta. \tag{5}$$

In addition, we assume that agents must be induced to participate in the mechanism voluntarily and that the only option to provide the goods is by financing the provision from the sales revenue. That is

$$\begin{aligned} U\left(\theta, \rho, x\left(\theta\right), t\left(\theta\right)\right) &\geq 0 && (6) \\ \text{for all } \theta &\in \Theta. \\ \mathrm{E}_{\theta}\left[t\left(\theta\right) - \rho^{1}\left(K^{1} + c^{1}\left[x^{1}\left(\theta\right) + x^{12}\left(\theta\right)\right]\right) + \rho^{2}\left(K^{2} + c^{2}\left[x^{2}\left(\theta\right) + x^{12}\left(\theta\right)\right]\right)\right] &\geq 0. && (7) \end{aligned}$$

To have any bite, (6) and (7) must be imposed simultaneously. Intuitively, this is because (7) says that the mechanism designer doesn't have access to outside subsidies. This is obviously not a binding constraint if lump sum transfers can be used, and the role of the consumer sovereignty assumption (6) is to rule out a solution where losses from marginal cost pricing are covered by lump sum taxes.

In terms of justifying the assumptions from first principles, (6) is the most problematic. Literally, the constraint says that no consumer can be made worse off by providing the goods. No modern society gives all citizens veto power on taxation, so the relevance of such a constraint may be debated. However, Hellwig [16] demonstrates that inequality aversion leads to qualitatively similar results as do the participation constraint (6), so we may think loosely about the assumption as a shortcut instead of societal preferences with curvature.

However, there is also a more practical justification for ruling out lump sum subsidies. In the end, we want to use the analysis to shed light on simple questions such as "when should the firm be allowed to bundle?". While the fundamental reasons may not be understood, it is nevertheless the case that policies involving subsidies from tax payers are simply not under consideration for many of these markets. In these cases, we can at least appeal to realism when defending the introduction of (6) and (7).

2.2.1 Why isn't Profits in Our Welfare Measure?

In the literature on regulation and procrurement it is common to consider a planner that seeks to maximize a weighted average of consumer surplus and profits. Here, we are only considering consumer surplus. The reason we are making such an assumption is that adding profits to the objective function will leave the solution unchanged *unless the weight on profits is higher than the weight on consumer welfare*. The basic insight is that any positive profit can be refunded to consumers, which will increase welfare unless there is slack in the constraints or the welfare weight on profits is significantly higher than the weight on consumer surplus. See Section 2.3 in Norman [26] for more detail.[12]

3 Linear Preferences

We begin the analysis with the case where $\gamma = 1$, meaning that the goods are neither complements or substitutes. This specification corresponds with with the setup in Adams and Yellen [1], McAffee et al [25] and an overwhelming majority of the studies on bundling and tying in the literature.[13]

3.1 Informationally Unconstrained Pareto Efficiency

To begin with, it is convenient to define β^j as the probability that an agent has a low valuation for good j,

$$\begin{aligned} \beta^1 &\equiv \alpha^{ll} + \alpha^{lh} \\ \beta^2 &\equiv \alpha^{ll} + \alpha^{hl}. \end{aligned} \tag{8}$$

First best efficiency is almost trivial in this case: conditional on good j being produced it is optimal to allow both types to consume good j for sure if $c^j < l^j$, whereas only agents with $\theta^j = h^j$ should consume the good if $l^j < c^j < h^j$. The informationally unconstrained planning solution is then simply to provide good j if and only if good j generates a positive net surplus:

Conclusion 1 *If $\gamma = 1$, then the informationally unconstrained social optimum is:*

1. *provide good j if and only if* $\max\left\{\beta^j l^j + \left(1-\beta^j\right) h^j - c^j, \left(1-\beta^j\right)\left(h^j - c^j\right)\right\} \geq K^j$
2. *if good j is provided and if $c^j < l^j$, then all consumers should be given access to the good.*
3. *if good j is provided and if$l^j < c^j < h^j$, then only the consumers with $\theta^j = h^j$ should be given access.*

[12] The key difference relative the regulation literature following Baron and Myerson [4] is that there are no unknown parameters in the firm profit function in our setup.

[13] There are exceptions. A few interesting examples can be found in Lewbel [21]. Chen and Nalebuff [8] consider a particular form of superadditivity, where the "add-on" is of value only if the consumer owns the "base product" (for example, anti-lock brakes without a car is presumably without value for most consumers). Venkatesh and Kamakura [31] is more similar to our model. They also introduce a single (commonly known) parameter that determines whether goods are complements or substitutes. The main difference with our work is that they only consider profit maximizing selling strategies, whereas our analysis is focused on economic efficiency.

3.2 Constrained Optima in the Absence of the Bundling Instrument

Our next task is to characterize the constrained optimum in a benchmark model where each good must be provided as a stand alone good. This gives us a best case scenario in case a firm is ordered to unbundle their products, which is of interest, in particular as we will show that this is sometimes worse than allowing a profit maximizing firm to operate without any constraints.

Step 1: Profit Maximization

As one of the constraints to the welfare optimization problem is a break even constraint it is useful to begin by noting that the profit maximizing production plan must be to sell at a price $p^j \in \{l^j, h^j\}$, or not to incur the the fixed cost at all. That is:

Conclusion 2 *If $\gamma = 1$ and if goods must be sold separately, then a profit maximizing plan is to;*

1. *provide good j and sell at price $p^j = l^j$ if and only if $l^j - c^j - K \geq \max\{0, (1-\beta^j)(h^j - c^j) - K^j\}$*
2. *provide good j and sell at price $p^j = h^j$ if and only if $(1-\beta^j)(h^j - c^j) - K^j \geq \max\{0, l^j - c^j - K\}$*
3. *not provide good j if $0 \geq \max\{l^j - c^j - K, (1-\beta^j)(h^j - c^j) - K^j\}$.*

In the first case, profit maximization is first best socially optimal. In the second case there is a distortion due to (over-)exclusion of the low valuation consumers if $c^j < l^j$, and in the third case there is a distortion due to under provision if the good is socially valuable.

Step 2: Welfare Maximization

If $l^j - c^j - K^j \geq 0$ it is immediate that setting $p^j \in [c^j + K^j, l^j]$ is socially optimal. We therefore assume that $l^j - c^j - K^j < 0$. In terms of the constrained optimal mechanism there are then three possibilities:

1. $(1-\beta^j)(h^j - c^j) - K^j < 0$, in which case good j is not provided in the constrained optimum,
2. $(1-\beta^j)(h^j - c^j) - K^j = 0$, in which case the constrained optimum is to provide good j and sell at price $p^j = h^j$.
3. $(1-\beta^j)(h^j - c^j) - K^j > 0$, in which case providing good j and selling at price $p^j = h^j$ generates a strict budget surplus. If $l^j \leq c^j$ it is first best optimal to only sell to consumers with high valuations, so selling at $p^j = h^j$ is also a constrained optimum.[14] However, if

[14] Strictly speaking, one can obviously argue that the budget surplus needs to be refunded to the consumers. That is, since we don't want to count firm profits as utility we could charge $p^j = (1-\beta^j)c^j + K^j$. As it is obvious that any positive profit can always be refunded without changing the consumption pattern we will ignore this issue in the rest of the paper.

$l^j > c^j$ it is socially desirable that low willingness to pay consumers get access to the good, and in this case it is possible to improve upon the profit maximizing mechanism by using a randomized mechanism as will be detailed below.

To understand how a randomized mechanism can be efficient, let x^j be the probability that a type l^j consumer consumes the good. As only the consumption patterns (and not the particular cost sharing arrangements) affect efficiency we may without loss assume that the low willingness to pay consumer pays $t^j\left(l^j\right) = x^j l^j$.[15] Provided that $l^j > c^j$ it is obvious that the higher the probability that type l^j gets access, the higher is social surplus. However, the problem is that the higher is the access probability for the low type, the less of the surplus can be extracted from the high type. That is, in order for it to be incentive compatible for type h^j to pay transfer $t^j\left(h^j\right)$ and get access with probability 1, a necessary and sufficient condition is that

$$h^j - t^j\left(h^j\right) \geq x^j\left(h^j - l^j\right). \tag{9}$$

As we are considering a case where social surplus is increasing in x^j, it follows that the incentive constraint (9) binds, implying that

$$t^j\left(h^j\right) = \left(1 - x^j\right)h^j + x^j l^j. \tag{10}$$

Substituting into the budget balance constraint (7), we can now express the budget surplus/deficit as a function of the probability that the low type gets access, x^j, and primitive parameters,

$$\begin{aligned} & \beta^j\left[t^j\left(l^j\right) - x^j c^j\right] + \left(1-\beta^j\right)\left[t^j\left(h^j\right) - c^j\right] - K^j \qquad (11) \\ = & \ \beta^j x^j\left(l^j - c^j\right) + \left(1-\beta^j\right)\left[\left(1-x^j\right)h^j + x^j l^j - c^j\right] - K^j \\ = & \ x^j\left(l^j - c^j\right) + \left(1-x^j\right)\left(1-\beta^j\right)\left(h^j - c^j\right) - K^j. \end{aligned}$$

In general, (11) may be increasing or decreasing in x^j. However, if it is increasing in x^j then the maximal revenue is $l^j - c^j - K^j$. This contradicts the assumption that we are looking at the case where $l^j - c^j - K^j < 0$ and $\left(1-\beta^j\right)\left(h^j - c^j\right) - K^j > 0$. Hence, it must be decreasing in x^j and there must exists some x^{*j} such that

$$x^{j*}\left(l^j - c^j\right) + \left(1 - x^{j*}\right)\left(1-\beta^j\right)\left(h^j - c^j\right) - K^j = 0 \tag{12}$$

Solving, we have that

$$x^{j*} = \frac{\left(1-\beta^j\right)\left(h^j - c^j\right) - K^j}{\left(1-\beta^j\right)\left(h^j - c^j\right) - \left(l^j - c^j\right)} \in (0,1), \tag{13}$$

where the conclusion that the probability is strictly between zero and one is using the parameter restrictions 1 $\left(1-\beta^j\right)\left(h^j - c^j\right) - K^j > 0$ and $l^j - c^j - K^j < 0$.

Summing up we can conclude:

[15] Only the expected payment matters, so an alternative is to only charge the low willingness to pay consumer if access to the good is granted, in which case the charge is l^j conditional on access and 0 otherwise.

Conclusion 3 *The possibilities for the unbundled constrained Pareto optimum are as follows:*

1. *If* $l^j \geq c^j + K^j$ *good* j *will be provided for sure and sold at a price* $p^j \in \left[c^j + K^j, l^j\right]$.

2. *If* $l^j - c^j - K^j < 0$ *and* $\left(1-\beta^j\right)\left(h^j - c^j\right) - K^j \geq 0$ *good* j *will be provided for sure.If, in addition:*

   A) $l^j > c^j$. *In this case type* l^j *will consume the good at price* l^j *with probability* x^{j*} *defined in (13) and pay a transfer of* $t^j\left(l^j\right) = x^{j*}l^j$ *(in expectation), and type* h^j *will consume the good for sure and pay* $t^j\left(h^j\right) = \left(1-x^j\right)h^j + x^j l^j$ *(in expectation).*

   B) $l^j \leq c^j$. *In this case selling to type* h^j *for sure at price* $t^j\left(h^j\right) \in \left[c^j + \frac{1}{1-\beta}K^j, h^j\right]$ *and not serving the other type is constrained optimal.*

3. *If* $l^j - c^j - K^j < 0$ *and* $\left(1-\beta^j\right)\left(h^j - c^j\right) - K^j < 0$ *good* j *cannot be provided at all without making a loss.*

That constrained optimum coincides with the unconstrained optimum in case 1 and 2A. In contrast, there is an inefficiency in case 2B, because customers that value the product above the unit cost are excluded from usage. In case 3 the constrained optimum is inefficient if and only if $\beta^j\left(l^j - c^j\right) + \left(1-\beta^j\right)\left(h^j - c^j\right) - K^j > 0$.

3.3 Example: Pareto Improving Bundling

A stark example of when bundling improves economic efficiency is when it is impossible for the monopolist to break even in the absence of the bundling instrument, but where profits can be made if the monopolist is allowed to bundle. To see that this is a possibility assume for simplicity that the goods are symmetric in all respects. That is, the goods enter symmetrically in preferences, meaning that $l^1 = l^2 = l$ and $h^1 = h^2 = h$. Moreover, the distribution over the typespace is symmetric, implying that $\alpha^{lh} = \alpha^{hl} = \alpha^m$. Finally, the costs are symmetric across goods, so that $c^1 = c^2 = c$ and $K^1 = K^2 = K$.

Given these symmetry assumptions, the interesting case is when,

1. $l - c - K < 0$

2. $(1-\beta)(h-c) - K < 0$

3. $\beta(l-c) + (1-\beta)(h-c) - K > 0$,

where we recall that $\beta = \alpha^{ll} + \alpha^m$. The first two inequalities mean that we are in the third case described in Conclusion 3, so neither good can be provided in the constrained optimum without bundling. The third inequality says that it is socially desirable to provide the goods. We will now demonstrate that introducing the bundling instrument sometimes alleviates this inefficiency, which ultimately is to be traced to the presence of fixed costs of production.

Consider the (non-random) fixed price pure bundling mechanism where consumers can get both goods at price $p = h + l$, but where neither good is available on its own. Types $\left(l^1, h^2\right), \left(h^1, l^2\right)$ and $\left(h^1, h^2\right)$ will then all have an incentive to purchase the bundle, so the revenue to the monopolist is $\left(\alpha^{lh} + \alpha^{hl} + \alpha^{hh}\right)(h + l) = \left(\alpha^{hh} + 2\alpha^m\right)(h + l)$. Hence, the monopolist makes a profit if

$$\left(\alpha^{hh} + 2\alpha^m\right)(h + l - 2c) - 2K \geq 0, \tag{14}$$

is satisfied.

Claim 1 *Fix c and K arbitrarily and assume that $\alpha^m > \frac{\alpha^{hh}\alpha^{ll}}{1-\alpha^{ll}}$.*[16] *Then, here exist pairs (l, h) such that (14) is satisfied, but where at the same time and at the same time*

$$\begin{aligned} l - c - K &< 0 \qquad (15) \\ \left(\alpha^{hh} + \alpha^m\right)(h - c) - K &= (1 - \beta)(h - c) - K < 0. \end{aligned}$$

Hence, there is a large set of parametrizations where allowing bundling is desirable for the consumers.

The parameter region where the pure bundling mechanism outperforms the constrained optimum when bundling is not allowed is depicted as the shaded region in Figure 1. To understand the figure, note that the first inequality in (15) is represented as the region to the left of the vertical line where $l - c = K$ and the second inequality as everything below the horizontal line where $(1 - \beta)(h - c) = K$. Since (14) is satisfied whenever (h, l) is above the line with slope negative 1 in the graph. It follows immediately that the region in the graph is the region where provision is zero in the absence of bundling, but where bundling leads to positive profits (and positive consumer surplus). This region is non empty whenever point C, the intersection between the lines $\left(\alpha^{hh} + 2\alpha^m\right)(h + l - 2c) - 2K = 0$ and $l - c - K = 0$, is located where (l, h) is such that $(1 - \beta)(h - c) < K$. This is intuitive, as this intersection is a point where either pure bundling or selling to all customers exactly breaks even. If $(1 - \beta)(h - c) \geq K$ at point C we would therefore be in a situation where bundling can break even only if either selling to all customers or selling to the high valuation customers would break even.

Hence, the question is when point C is located below the line where selling to high valuation customers only breaks even. To answer this, we simply calculate what the profit of selling to high valuation customers only is when $l - c = K$ and

$$h + l - c = \frac{2K}{\alpha^{hh} + 2\alpha^m} \Leftrightarrow h - c = \frac{2K}{\alpha^{hh} + 2\alpha^m} - K$$

[16] If valuations are independent, $\alpha_{hh} = \alpha^2, \alpha_{ll} = (1 - \alpha)^2$, and $\alpha_m = \alpha(1 - \alpha)$, and the inequality in Claim 1 is satisfied.

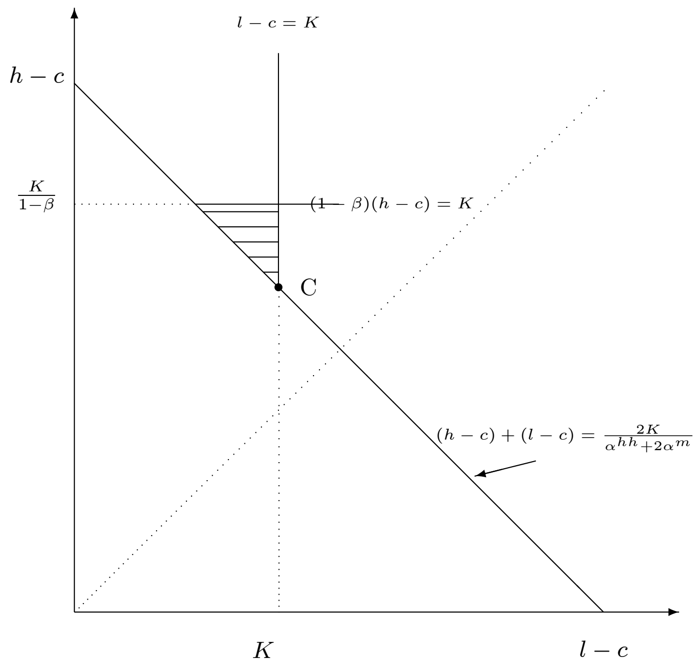


Figure 1: Bundling Allows Goods to be Sold without a Loss in Shaded Region

Hence, evaluated at point C we have that

$$\begin{aligned}
(1-\beta)(h-c)-K &= \left(\alpha^{hh}+\alpha^{m}\right)(h-c)-K \\
&= \left(\alpha^{hh}+\alpha^{m}\right)\left(\frac{2K}{\alpha^{hh}+2\alpha^{m}}-K\right)-K \\
&= K\left[\frac{2\left(\alpha^{hh}+\alpha^{m}\right)}{\alpha^{hh}+2\alpha^{m}}-\left(\alpha^{hh}+\alpha^{m}\right)-1\right] \\
&= K\left[\frac{\alpha^{hh}\alpha^{ll}}{1-\alpha^{ll}}-\alpha^{m}\right]
\end{aligned}$$

We conclude that the shaded parameter region in Figure 1 is non empty under the assumption that $\alpha_m > \alpha_{hh}\alpha_{ll}/\left(1-\alpha_{ll}\right)$.

The inequality requires that the valuations for the two goods are not too positively correlated. Intuitively, bundling increases revenue because there are more agents with at least one high valuation than there are agents with a high valuation for good j. The obvious downside is that the (per good) revenue for each paying customer decreases from h to $(h+l)/2$. Which effect is the strongest depends on details of the example, but if valuations are too strongly positively correlated there are too few agents with the mixed type for the increased sales to make up for the reduction in price even if $l-c$ is just barely below K.

3.4 Summary

We considered an example without any complementarities in either costs or preferences. Nevertheless, we found that it is possible that bundling may be necessary in order to break even. Clearly, the pure bundling mechanism used in the example is a feasible pricing strategy also for a profit maximizer, so we conclude that a profit maximizer is willing to operate in the market provided that bundling is allowed. Indeed, a profit maximizer can charge *at most* $h + l$ for the bundle, since a higher price would necessarily deter all types except hh from buying the bundle. One can show that selling the bundle at a price above $h + l$ and pricing the individual goods in a way so that hh has an incentive to buy the bundle makes a lower profit than the pure bundling strategy considered. The conclusion is that consumers must be weakly better off in the *profit maximizing* outcome with bundling than in the *constrained social optimum* without bundling. Hence, we have found circumstances where unbundling will harm *consumers* regardless of whether the unbundling is combined with other remedies or not (except, of course, production can be run by a regulated firm that is financed in part by lump sum transfers).

The underlying driving force is that bundling makes it easier for the monopolist to price discriminate. Usually, an improved ability to price discriminate is bad for the consumers as it implies that the monopolist can extract a larger share of the surplus. In our case, however, the presence of falling average costs makes a difference. Then, improved instruments for extracting consumer surplus can sometimes translate into provision of goods and services that would otherwise not be provided.

In a sense, the argument is simply a variation of the standard justification for patents. Monopoly profits create incentives for investments in "new ideas". Any change in the set of pricing instruments available to the monopolist that makes it possible to make a profit is therefore in the interest of the consumer if an active firm must make a loss before the change. Allowing the firm to bundle is one such possibility.[17] Clearly, one has to be cautious with this typ of argument. While an improved ability to price discriminate is desirable when pricing is done by an agent with benevolent preferences, this need obviously not be the case when we consider a for profit monopoly provider[18].

4 Bundling Complements

4.1 Preliminary Considerations

Informal discussions on bundling often argue that an obvious reason for bundling is that consumers value the bundle higher than the components. These arguments are usually based on the

[17] We focus on the possibility that allowing bundling may be good for the consumers. Obviously, this is often not the case, such as when the fixed costs are small enough so that goods are provided also when goods are sold separately.

[18] For an illustrative example, see Bergstrom and Bergstrom [5].

view that a final consumer good is to be viewed as an array of attributes. However, these preference based explanations are at best incomplete, as consumers could buy whatever combinations of parts they like if offered on an a la carte basis.

An obvious observation is that, in our setup, bundling can *never* be justified (as constrained social optima) unless average costs are falling. With constant (or increasing) average costs it is always socially optimal to use marginal cost pricing. This will at least break even, and doesn't involve any bundling.[19]

An immediate question is whether there are any intrinsic reasons to value a pair of features less if they are purchased as separate products, or if there are any savings on costs if the various features are assembled into a bundle prior to selling the product(s)? In the analysis below we assume that this is **not** the case. We make this assumption because it seems that in most cases the consumer should not care if features are bundled ex ante or ex post. That is, while the anti-lock breaks are useless without a car (a positive complementarity), the bundle consisting a given car model with anti-lock breaks should have the same value regardless of whether the anti-lock breaks are standard or optional, which is the assumption that we are making in the analysis below.

The reader should also remember that we are **not** comparing pricing by an integrated monopoly with that of independent "monopolies" selling complementary goods. This analysis, which dates back to Cournot [10], leads to the stark conclusion that the integrated monopoly is better for the consumers, because the complementarity in demand implies that there is a negative externality on the other market from an increase in the price.

Revenue Maximization

Consider the case with $\gamma > 1$, so that the two goods are complementary. Assuming that each good is sold to some customers, all the candidate revenue maximizing plans can then be summarized as in Table 1 below. In order to conserve space, some notation that is not completely self-explanatory is used. In particular, the allocations list the goods consumed by each type in the order $\left(h^1, h^2\right), \left(h^1, l^2\right), \left(l^1, h^2\right), \left(l^1, l^2\right)$, so $(12, \varnothing, 2, \varnothing)$ means that type $\left(h^1, h^2\right)$ consumes both goods, type $\left(l^1, h^2\right)$ consumes good 2, and the two other types consume nothing. Also, "sum" in the column $p^1 + p^2$ means that the sum of the prices is determined from the unique component prices. However, in some cases the component prices are not uniquely determined, and in the entry in the column $p^1 + p^2$ is the uniquely determined sum of the two prices, whereas the single price

[19] A monopolist will bundle if and only if bundling leads to a higher profit, and this could in some circumstances increase the consumer surplus relative to the non-bundling profit maximum for the monopolist. Such gains would obviously rely on details of the demand, and it seems contrived to build a case in favor of bundling based on an increase in consumer surplus as an unintended side effect of profit maxiization. We will not explore this possibility in this paper.

Allocation	p^1	p^2	$p^1 + p^2$
$(12, \varnothing, \varnothing, \varnothing)$	$\geq h^1$	$\geq h^2$	$\gamma\left[h^1 + h^2\right]$
$(12, 1, 2, \varnothing)$	h^1	h^2	sum
$(12, 1, \varnothing, \varnothing)$	h^1	$(\gamma - 1)\, h^1 + \gamma h^2$	sum
$(12, \varnothing, 2, \varnothing)$	$\gamma h^1 + (\gamma - 1)\, h^2$	h^2	sum
$(12, 12, 2, \varnothing)$	$\gamma\left[h^1 + l^2\right] - h^2$	h^2	sum
$(12, 1, 12, \varnothing)$	h^1	$\gamma\left[l^1 + h^2\right] - h^1$	sum
$(12, 12, 12, \varnothing)$ CASE A	$\geq h^1$	$\geq h^2$	$\gamma \min\left\{h^1 + l^2, l_1 + h^2\right\}$
$(12, 12, 12, \varnothing)$ CASE B	$\gamma l^1 + (\gamma - 1)\, h^2$	$(\gamma - 1)\, h^1 + \gamma l^2$	sum
$(12, 1, 2, 2)$	h^1	l^2	sum
$(12, 1, 2, 1)$	l^1	h^2	sum
$(12, 12, 12, 12)$	$\geq l^1$	$\geq l^2$	$\gamma\left[l^1 + l^2\right]$

Table 1: All possible profit maximizing selling strategies in the case that bundling is not allowed.

columns list the appropriate constraints on the individual prices.[20]

While the number of different cases is somewhat overwhelming, we note next that, in order to make a fair and interesting comparison with the problem in the absence of complementarities, only a few cases are of interest.

Bundling with a Small Complementarity

We are interested in a case which is 1) easily comparable to the linear case, and; 2) the outcome in the absence of bundling is Pareto inefficient. To do this we assume that γ is close enough to unity, and that the two goods are fully symmetric. In this case the monopoly pricing problem when goods are sold separately is almost identical to the linear case if the complementarity is sufficiently small:

Claim 2 *Suppose that $l^1 = l^2 = l$, $h^1 = h^2 = h, \alpha^{lh} = \alpha^{hl} = \alpha^m > 0$, $c^1 = c^2 = c$ and $K^1 = K^2 = K > 0$, where we assume that $h > c$. Then, there exists $\delta > 0$ such that if $\gamma \in (1, 1 + \delta)$ a profit maximizing price for good $j = 1, 2$ in the absence of bundling is*

$$p^{j*} = \begin{cases} \gamma l & \text{if } \delta l - c \geq (1 - \beta)(h - c) \\ h & \text{if } \delta l - c < (1 - \beta)(h - c) \end{cases} \tag{16}$$

[20] Whether CASE A or CASE B applies depends on whether

$$h^1 + h^2 - \gamma \min\left\{h^1 + l^2, l^1 + h^2\right\}$$

is positive of negative. If the expression is negative and $p^1 + p^2 = \gamma \min\left\{h^1 + l^2, l^1 + h^2\right\}$, then either $h^1 + l^2$ has an incentive to purchase good 1 only or type $l^1 + h^2$ has an incentive to purchase to purchase good 2 only.

Claim 2 is easy to understand, but a proof is in the appendix for completeness. Obviously, if the goods are priced so that the low valuation customers will demand the good, then charging γl per good must be optimal as all consumers are willing to buy at this price. It only remains to argue that pricing at h is better than pricing at γh per good if γ is sufficiently close to unity, which is rather intuitive as the loss of consumers (a fraction α^m per good) is substantial no matter how close to unity γ is, but the price gain per unit sold converges to zero as γ converges to unity.

We conclude that it is impossible to break even if the two conditions

$$\begin{aligned} \delta l - c - K &< 0 \\ (1-\beta)(h-c) - K &< 0, \end{aligned} \tag{17}$$

are satisfied.

In spirit of the example in Section 3.3, contrast the non-bundling regime with a pure bundling mechanism that sells both goods to all types except ll. Because of the complementarity, it is now possible to charge $p = \gamma(h+l)$, implying that the monopolist can make a profit if

$$\left(\alpha^{hh} + 2\alpha^{hh}\right)[\gamma(h+l) - 2c] \geq 2K. \tag{18}$$

Comparing with the linear case, we first note that there are parameter configurations in terms of (l, c, K) such that

$$l - c - K < 0 \leq \delta l - c - K < 0 \tag{19}$$

for $\gamma > 1$, implying that it in some cases get easier to break even without bundling when there is a positive complementarity. On the other hand there are also parameter configurations in terms of (h, l, c, K, α) such that

$$\left(\alpha^{hh} + 2\alpha^{hh}\right)[(h+l) - 2c] - 2K < 0 \leq \left(\alpha^{hh} + 2\alpha^{hh}\right)[\gamma(h+l) - 2c] - 2K, \tag{20}$$

which means that it in some cases gets easier to break even with bundling.

At first blush, it therefore seems that our comparison is completely inconclusive. Sometimes the complementarity eliminates the need to bundle to break even, and sometimes the complementarity makes it easier for bundling to solve the problem of under provision. Also, all conditions are continuous in γ, so a small complementarity makes a small difference for the desirability of bundling. While this is true, there is nevertheless a sense in which a positive complementarity makes it "more likely" for bundling to help economic efficiency in that there is a larger subset of parameter vectors (h, l, α, c, K) for which the possibility to bundle is the only way to break even:

Proposition 1 *Fix c and K arbitrarily, assume that $\alpha^m \geq \frac{\alpha^{hh}\alpha^{ll}}{1-\alpha^{ll}}$, and assume that $\gamma > 0$ is small enough for (16) to apply. Let L be the set of (h, l) such that bundling is needed to break even in the linear case, let C be the set of (h, l) such that bundling is needed to break even with the complementarity, and let $\mu(\cdot)$ denote the Lebesgue measure (assuming that the set of admissible types is bounded by $\left[0, \overline{h}\right]$). Then, $\mu(L) < \mu(C)$.*

We will prove this by picture. Consider the graph in Figure 1, where the shaded area corresponds to parametrizations where bundling is needed for provision. We will argue that adding a small complementarity will increase the size of the area. To understand that, note that the area in Figure 1 are solutions to

$$\begin{aligned} l-c &< K && (21)\\ (h-c) &< \frac{K}{(\alpha^{hh}+\alpha^{m})} && (22)\\ (h-c)+(l-c) &\geq \frac{2K}{\alpha^{hh}+2\alpha^{m}} && (23) \end{aligned}$$

which is to be compared with solutions to (rearranging (17) and (18) to express the conditions in $(l-c)-(h-c)$ −space).

$$\begin{aligned} l-c &< \frac{1}{\gamma}\left[c\left(1-\gamma\right)+K\right] && (24)\\ (h-c) &< \frac{K}{(\alpha^{hh}+\alpha^{m})} && (25)\\ (h-c)+(l-c) &\geq \frac{1}{\gamma}\left[2c\left(1-\gamma\right)+\frac{2K}{\alpha^{hh}+2\alpha^{m}}\right] && (26) \end{aligned}$$

That is:

- the break even line for selling both goods separately to all types (the vertical line in Figure 1) shift inward by
$$K-\frac{1}{\gamma}\left[c\left(1-\gamma\right)+K\right]=\left(K+c\right)\frac{\gamma-1}{\gamma}$$

- the break even line for selling separately to the high valuation customers (the horizontal line in Figure 1) is unchanged.

- the break even line for pure bundling (the line with slope −1 in Figure 1) shifts inward by
$$\begin{aligned} &\frac{2K}{\alpha^{hh}+2\alpha^{m}}-\frac{1}{\gamma}\left[2c\left(1-\gamma\right)+\frac{2K}{\alpha^{hh}+2\alpha^{m}}\right] && (27)\\ =\; &\left(\frac{K}{\alpha^{hh}+2\alpha^{m}}+c\right)2\frac{\gamma-1}{\gamma}>2\left(K+c\right)\frac{\gamma-1}{\gamma}>\left(K+c\right)\frac{\gamma-1}{\gamma}. \end{aligned}$$

- Hence, we can illustrate the situation as in Figure 2, where the inwards shift in the line where the monopolist breaks even under bundling is larger than the inwards shift in the line where the monopolists breaks even if selling the goods separately to all types. Because (given that γ is not too large) nothing happens to the break even condition when selling to high types it follows that the triangular area where bundling is necessary for the firm to break even is enlarged. This is because the value for $h-c$ associated with point C decreases as a consequence of the inequality in (27): the angle at C is unchanged as the slope of the break even line with bundling is kept at −1, so the area of the triangle is enlarged if and only if moves downwards-which it does.

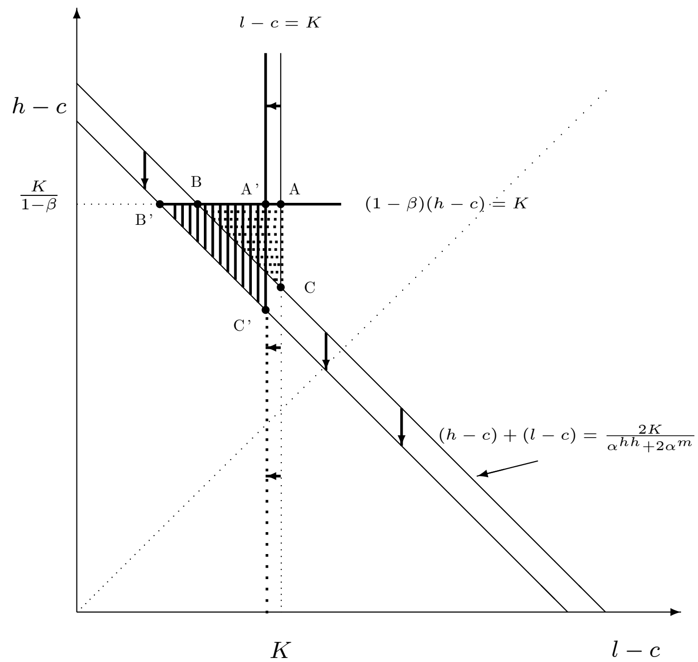


Figure 2: Comparing the desirability of Bundling for Complement Goods with the Linear Case

4.2 Summary

The most important insight from the analysis in this section is probably that whether goods are complements or not doesn't change anything qualitatively. The model is continuous in the degree of complementarity, and the crucial logic remains the same as in the linear case. However, Proposition 1 does establish that a small complementarity implies that bundling is necessary for a larger subset of the other parameters in the model.

5 Bundling Substitutes

Finally, we will briefly consider the case when $\gamma < 1$, so that the goods are substitutes and the marginal valuation falls when a second good is added to the consumption basket. Since the analysis of the non-bundling benchmark is similar to the case with complements we will leave some of the details to the reader. One verifies readily that:

Claim 3 *Suppose that $l^1 = l^2 = l$, $h^1 = h^2 = h, \alpha^{lh} = \alpha^{hl} = \alpha^m > 0$, $\alpha^{hh} > 0$, $c^1 = c^2 = c$ and $K^1 = K^2 = K > 0$, where we assume that $h > c$. Then, there exists $\delta > 0$ such that if $\gamma \in (1-\delta, 1)$ a profit maximizing price for good $j = 1, 2$ in the absence of bundling is*

$$p^{j*} = \begin{cases} \gamma l & \text{if } \delta l - c \geq (1-\beta)(\gamma h - c) \\ \gamma h & \text{if } \delta l - c < (1-\beta)(\gamma h - c) \end{cases} \tag{28}$$

The idea is similar to Claim 2, with the difference that selling at price h is dominated by selling at γh if γ is close enough to unity, as type hh would only buy one of the two goods if the price is h. Hence, the sells $\frac{1}{2}\alpha^{hh}$ fewer goods, which cannot be optimal when γ is near one.

We immediately observe that the conditions for bundling to dominate separate provisions are almost identical with the linear case,

$$\begin{aligned} \gamma l - c &< K && (29) \\ (1-\beta)(\gamma h - c) &< K && (30) \\ (\gamma h - c) + (\gamma l - c) &\geq \frac{2K}{\alpha^{hh} + 2\alpha^{m}}. && (31) \end{aligned}$$

That is, if (l^*, h^*) is such that bundling is necessary in the linear case, then bundling is necessary when goods are substitutes and $(l,h) = \left(\frac{l^*}{\gamma}, \frac{h^*}{\gamma}\right)$. As $\gamma < 1$ it is immediate that the parameter region where bundling is necessary is "stretched out", and again we have that it is "more likely" that bundling is necessary for efficient provision of the goods.

5.1 Bundling Due to Cost Synergies

Suppose that there are n consumers, but that each consumer is just like in the model above. Suppose the goods may either be bundled, which eliminates duplication in production, or sold separately. That is, the cost per unit is:

1. c per unit of the bundle if the goods are bundled.

2. c per unit of good 1 and c per unit of good 2 if the goods are sold separately.

Conditional on either bundling or separate salesAssume that $l < c < h$.

the per unit cost of selling good 1 is c

There is also a unit variable cost $c^j \geq 0$ for good $j = 1, 2$.

6 Discussion

This paper shows that fixed costs and an inability to perfectly price discriminate can create a plausible justification for commodity bundling. In anti-trust cases and policy discussions it is often argued that bundling may be efficient, but the efficiency enhancing effects mentioned are often synergies, benefits from reduced complexity and savings on transactions costs. Such effects tend to be rather elusive, and it is sometimes explicitly pointed out that benefits from bundling have the unfortunate feature of being very hard to measure. In contrast, documenting substantial fixed costs for the development of particular products seems feasible in many cases. One would also want to judge whether bundling is needed in order to break even, which, in principle, this boils down to estimating the parameters of a demand system. We therefore conclude that the *quantitative* relevance of the argument of this paper can be estimated using standard empirical methods provided that reliable cost data exists.

References

[1] Adams, W.J. and Janet L. Yellen. "Commodity Bundling and the Burden of Monopoly." *Quarterly Journal of Economics,* 90, 1976, 475 –98.

[2] Ahlborn, Christian, Evans David S. , and A. Jorge Padilla, "The Antitrust Economics of Tying: A Farewell to Per Se Illegality," *Antitrust Bullentin*, March 2004.

[3] Bakos, Yannis and Brynjolfsson, Erik. "Bundling Information Goods: Pricing, Profits and Efficiency," *Management Science*, 45, 1999, 1613-1630.

[4] Baron, David R., and Roger B. Myerson, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50(4), July 1982, 911-930.

[5] Bergstrom, Carl T, and Theodore C. Bergstrom. "The Costs and Benefits of Library Site Licenses to Academic Journals," *Proceedings of the National Academy of Sciences in the United States of America*, **101**, 2004, 897-902.

[6] Brown, Keith, and Peter J. Alexander, "Bundling in Cable Television: A Pedagogical Note With a Policy Option," *The International Journal on Media Management,* 6(3-4), 2004, 162-167.

[7] Carlton, Dennis W., and Michael Waldman, "The Strategic Use of Tying to Preserve and Create Market Power in Evolving Industries," *RAND Journal of Economics*, 33(2), Summer 2002, 194-220.

[8] Chen, Keith and Barry Nalebuff, "One-Way Essential Complements," *manuscript*, Yale School of Management, August 2006.

[9] Choi, Jay Pil, and Christodoulos Stefanadis, "Tying, Investment, and the Dynamic Leverage Theory," *RAND Journal of Economics*, 32(1), Spring 2001, 52-71.

[10] Cournot, A., *Reserches sur les Principles Mathematiques de la Theorie des Richesses* , Paris, Hachette, 1838 , translated as *Research into the Mathematical Principles of the Theory of Wealth*, Kelley, 1960.

[11] Evans, David, and A. Jorge Padilla, "Designing Antitrust Rules for Assessing Unilateral Practices: A Neo-Chicago Approach," *University of Chicago Law Review* 72-73, Winter 2005.

[12] Evans, David, A. Jorge Padilla, and Michael A. Salinger, "A Pragmatic Approach to Identifying and Analyzing Legitimate Tying Cases," in Ehlerman and I Attanasiu Eds. *European Competition Law Annual 2003: What is an Abuse of Dominant position*, 2004.

[13] Geng, Xianjun, Maxwell B. Stinchcombe, and Andrew B. Whinston, "Bundling Information Goods of Decreasing Value," *Management Science,* 51(4), April 2005, 662 - 667.

[14] Fang, Hanming and Peter Norman, "Optimal Provision of Multiple Excludable Public Goods," *manuscript*, Yale and UBC, April 2006.

[15] Groves, Theodore, "Incentives in Teams," *Econometrica*, 41, 1973, 617-631.

[16] Hellwig, Martin, F. "A utilitarian approach to the provision and pricing of excludable public goods", *Journal of Public Economics* ,89, December 2005, Pages 1981-2003.

[17] *Jefferson Parish Hospital District No. 2 v Hyde*, 466 U.S. 2, March 1984.

[18] Kobayashi, Bruce, "Does Economics Provide a Reliable Guide to Regulating Commodity Bundling by Firms? A Survey of the Economic Literature," *Journal of Competition Law & Economics*, Vol. 1 (2005), pp. 707-46. Available at: http://jcle.oxfordjournals.org/cgi/reprint/1/4/707?ijkey=7PAxdPovbfmj4Uo&keytype=ref.

[19] Kuhn, Kai-Uwe, Stillman, Robert, and Cristina Cafarra, "Economic Theories of Bundling and their Policy Implications in Abuse Cases: An Assessment in Light of the Microsoft Case," *University of Michigan John M. Olin Center for Law & Economics Working Paper #04-19,* September 2004. http://law.bepress.com/umichlwps/olin/art33 *Journal of Industrial Economics*

[20] Ledyard, John. and Thomas Palfrey, "A General Characterization of Interim Efficient Mechanisms for Independent Linear Environments," *Journal of Economic Theory*, forthcoming 2006.

[21] Lewbel, Arthur, "Bundling of Substitutes and Complements," *International Journal of Industrial Organization*, 3, 1985, 101-107.

[22] Matutes, Carmen, and Pierre Regibeau, "Compatibility and Bundling of Complementary Goods in a Duopoly," *Journal of Industrial Economics,* 40, 1992, 37-54.

[23] Nalebuff, Barry, "Bunding as a Barrier to Entry," *Quarterly Journal of Economics*, 119, 2004.

[24] Nalebuff, Barry, "Bundling and the GE-Honeywell Merger," in (John Kwoka and Lawrence White, eds.*The Antitrust Revolution*, 4th edition, Oxford University Press, 2003

[25] McAfee, R. Preston, John McMillan and Michael D. Whinston. "Multiproduct Monopoly,Commodity Bundling, and Correlation of Values." *Quarterly Journal of Economics,* 104, 1989, 371–84.

[26] Norman, Peter. "Efficient Mechanisms for Public Goods with Use Exclusion." *Review of Economic Studies*, **71**, 2004, 1163-1188.

[27] Office of Fair Trading, *BSkyB decision dated 17 December 2002: rejection of applications under section 47 of the Competition Act 1998 by ITV Digital in Liquidation and NTL*, July 2003, London, UK. Available at:
`http://www.oft.gov.uk/business/competition+act/decisions/bskyb.htm`

[28] *Times-Picayune Publishing Co. v. United States*, 345 U.S. 594, 1953.

[29] Thorne, John, "Discounted Bundling by Dominant Firms," George Mason Law Review, 13(2), 2005, 339-347.

[30] Evans, David, S., and Michael Salinger, "Why Do Firms Bundle and Tie? Evidence from Competitive Markets and Implications for Tying Law," *Yale Journal on Regulation*, 22(1) Winter 2005.

[31] Venkatesh R. and Wagner Kamakura, "Optimal Bundling and Pricing under a Monopoly: Contrasting Complements and Substitutes from Independently Valued Products," *Journal of Business*, 76(2), April 2003, 211-232.

[32] Whinston, Michael D., "Tying, Foreclosure, and Exclusion," *American Economic Review,* 80, 1990, 837-859.

A Omitted Proofs

A.1 Proof of Claim 2

Proof. We will take as granted that, because of the symmetry $p^{1*} = p^{2*}$. While this symmetry is not completely obvious, the most straightforward way to do without the symmetry is to add arguments for the asymmetric possibilities. This adds a few cases, but eventually boils down to the same logic as in (35) below, so we leave this to the reader. Obviously, the marginal price must be in the set $\{l, \delta l, h, \delta h\}$. It is immediate that that charging $p^{1*} = p^{2*} = l$ is dominated by charging $p^{1*} = p^{2*} = \gamma l$ as the good is sold to all consumers in both cases. The profit per good is then

$$\delta l - c - K. \tag{32}$$

The two remaining possibilities are to set $p^{1*} = p^{2*} = h$, which generates a per good profit of

$$(1-\beta)(h-c) - K = \left(\alpha^m + \alpha^{hh}\right)(h-c) - K \tag{33}$$

and p^{1*}, p^{2*} such that $p^{1*} + p^{2*} = \gamma 2h$, which generates a per good profit

$$\alpha^{hh}(\gamma h - c) - K \tag{34}$$

Subtracting (34) from (33) we see that the difference is

$$\alpha^m (h-c) - (\gamma - 1)\alpha^{hh} h \to \alpha^m (h-c) > 0 \tag{35}$$

as $\gamma \to 0$, implying that, there exists $\delta > 0$ such that (34) is larger than (33) if $\gamma \in (1, 1+\delta)$. The maximized profit is thus the larger of (32) and (33), which completes the proof. ∎

Exclusionary Pricing and Rebates When Scale Matters*

Liliane Karlinger† Massimo Motta‡

March 15, 2007

Abstract

We consider an incumbent firm and a more efficient entrant, both offering a network good to several asymmetric buyers. The incumbent disposes of an installed base, while the entrant has a network of size zero at the outset, and needs to attract a critical mass of buyers to operate. We analyze different price schemes (uniform pricing, implicit price discrimination - or rebates, explicit price discrimination) and show that the schemes which - for given market structure - induce a higher level of welfare are also those under which the incumbent is more likely to exclude the rival.

JEL classification: L11, L14, L42.

*We are indebted to Daniel Cerquera, Claude Crampes, Chiara Fumagalli, Jeanine Thal, and especially Patrick Rey, for helpful comments and suggestions. We are also grateful to seminar participants at Hitotsubashi University, Tokyo; CORE, Université Catholique de Louvain; University of Vienna; University of Linz; CEPR Conference on Applied Industrial Organisation, Madeira; 2nd CSEF Symposium on Economics and Institutions, Capri; 1st CRESSE Conference on 'Abuse of Dominance', Corfu; CCP Workshop 2005, University of East Anglia (Norwich, UK); Kiel-Munich Workshop on Economics of Internet and Network Industries; EMNet Conference 2005, Budapest; 2nd Workshop on Empirical IO, University of Innsbruck; European Economic Association Meeting 2006, Vienna; IDEI Conference on the Economics of the Software and Internet Industries, Toulouse; EUI, Florence. The usual disclaimer applies.

†University of Vienna, Economics Department at BWZ, Brünner Strasse 72, 1210 Vienna, Austria, email: liliane.karlinger@univie.ac.at

‡European University Institute and Università di Bologna. Address: European University Institute, Economics Department, Villa San Paolo, Via della Piazzuola 43, 50133 Florence, Italy, email: massimo.motta@eui.eu

1 Introduction

This paper deals with exclusionary pricing practices, that is anti-competitive pricing behavior by a firm endowed with a "dominant position" (as called in the EU), or with "monopoly power" (as called in the US). One such practice which has recently received renewed attention is *rebates*, i.e. discounts applicable where a customer exceeds a specified target for sales in a defined period.

There are different types of rebates, or discounts. They can be made contingent on the buyer making most or all of its purchases from the same supplier ("fidelity" or "loyalty" rebates), on increasing its purchases relative to previous years, or on purchasing certain quantity thresholds specified in absolute terms. It is on this last category of rebates that we focus here.

In the US, rebates have received a very favorable treatment by the courts for many years. Under US case law (see e.g. the Virgin v. British Airways (2001) case), loyalty rebates were said to promote competition on the merits as a rule, and it was for the plaintiff to demonstrate their anticompetitive effect.[1] However, the recent *LePage* (2003) decision - in which the Appeal Court reversed an earlier judgment and found 3M guilty of attempted monopolization for having used (bundled) rebates - may signal the willingness of the judges to use lower standards of proof for the finding of anticompetitive rebates.

In the EU, rebates have long been looked at with suspicion by the European Commission (which is the EU Competition Authority) and the Community Courts, which have systematically imposed large fines on dominant firms applying different forms of rebates.[2] But until the recent *Michelin II* judgment, dominant firms were at least allowed to grant pure quantity discounts, that is standardized rebates given to any buyer whose purchases exceed a predetermined number of units; *Michelin II*, instead, has established that even pure quantity discounts are anticompetitive if used by a dominant firm.[3,4]

One of the objectives of this paper is to take seriously the Community Court's assessment, and study whether rebates, in the form of pure quantity discounts, can have anticompetitive effects. A key feature of the environment we consider are scale effects, that we choose to model as scale economies on the demand side (but the main insights of the paper would also hold good with production scale economies, as we explain in Section 6).

More precisely, we study an industry exhibiting network effects, and we find that if rebates are allowed, an incumbent firm having a critical customer base

[1] See Kobayashi (2005) for a review of the US case law.

[2] For a review of the EU case law on rebates, see e.g. Gyselen (2003).

[3] Unless they are 'objectively' justified, that is unless the dominant firm can prove that the discount matches savings from transaction costs.

[4] The (almost) per se illegal status of exclusive contracts, rebates and discriminatory prices by dominant firms in the EU, as well as the difference relative to their treatment in the US (at least until recently), has led to a hot debate on the EU policy towards abuse of dominance. See Rey et al. (2005) for a contribution to the debate.

is more likely to exclude a more efficient entrant that can use the same rebate schemes but does not have a customer base yet. Rebates are a form of implicit discrimination, and the incumbent can use them to make more attractive offers to some crucial group of consumers, thereby depriving the entrant of the critical mass of consumers it needs (in our model, network externalities imply that consumers will want to consume a network product only if demand has reached a critical threshold).

Now, discrimination (implicit and even more so explicit discrimination) will allow the incumbent to play off the different groups of consumers against each other. This strategic use of price discrimination will exacerbate the coordination problems that buyers face, which in turn makes entry even more difficult for the new rival. Only very efficient entrants will be able to overcome the entry barriers that incumbents can raise in this manner.

To give an example of the type of industry that we have in mind, let us briefly review the *Microsoft Licensing Case* of 1994-95 (Civil Action No. 94-1564). Microsoft markets its PC operating systems (Windows and MS-DOS) primarily through original equipment manufacturers ("OEMs"), which manufacture PCs, and has agreements with virtually all of the major microcomputer OEMs. When discussing the substantial barriers to entry for potential rivals of Microsoft, the Complaint explicitly mentions "the difficulty in convincing OEMs to offer and promote a non-Microsoft PC operating system, particularly one with a small installed base". Moreover, "it would be virtually impossible for a new entrant to achieve commercial success solely through license agreements with small OEMs that are not covered by Microsoft's (...) agreements."

The US Department of Justice alleges that Microsoft designed its pricing policy "to deter OEMs from entering into licensing agreements with competing operating system providers", thereby reinforcing the entry barriers raised by the network effects that are inherent in this industry. In particular, the use of two-part tariffs, with high fixed fees and zero per-copy price, is considered strongly anti-competitive. Interestingly, though, the Final Judgment explicitly allows Microsoft to continue granting "volume discounts" (i.e. rebates), as long as Microsoft would use linear prices rather than two-part tariffs. Our paper, however, suggests that rebates, being a form of (implicit) price discrimination as well, can also be exclusionary.

Although rebates may have exclusionary effects, it is far from clear that they should be presumed to be welfare-detrimental, even if used by a dominant firm. As John Vickers, then Chairman of the UK Office of Fair Trading, put it:

> "These cases about discounts and rebates, on both sides of the Atlantic, illustrate sharply a fundamental dilemma for the competition law treatment of abuse of market power. A firm with market power that offers discount or rebate schemes to dealers is likely to sell more, and its rivals less, than in the absence of the incentives. But that is equally true of low pricing generally." (Vickers, 2005: F252)

Discriminatory pricing has similar contrasting effects. Consider for instance an oligopolistic industry. On the procompetitive side, it allows firms to decrease prices to particular customers, thereby intensifying competition: each firm can be more aggressive in the rival's customer segments while maintaining higher prices with the own customer base, but since each firm will do the same, discriminatory pricing will result in fiercer competition than uniform pricing, and consumers will benefit from it.[5] On the anticompetitive side, though, in asymmetric situations discriminatory pricing may allow a dominant firm to achieve cheaper exclusion of a weaker rival: prices do not need to be decreased for all customers but only for the marginal customers.[6]

This fundamental dilemma between, on the one hand, the efficiency effects (consumers would buy more and pay less) created by rebates and discriminatory pricing and, on the other hand, their potential exclusionary effects (rival firms would be hurt by such practices, and may be driven out of the market), is possibly the main theme of the paper. Indeed, we shall study here different pricing schemes that both an incumbent and a rival firm can adopt, and show that the schemes which - for given market structure - induce a higher level of welfare are also those under which the incumbent is more likely to exclude the rival. More specifically, we show that explicit price discrimination is the pricing scheme with the highest exclusionary potential (and hence the worst welfare outcomes if exclusion does occur), followed by implicit price discrimination (i.e., rebates, or pure quantity discounts) and then uniform pricing. However, for given market structure (i.e., when we look at equilibria where entry does occur), the welfare ranking is exactly reversed: the more aggressive the pricing scheme the lower the prices (and thus the higher the surplus) at equilibrium. This trade-off between maximizing the entrant's chances to enter and minimizing welfare losses for given market structure, illustrates the difficulties that antitrust agencies and courts find in practice: a tough stance against discounts and other aggressive pricing strategies may well increase the likelihood that monopolies or dominant positions are successfully contested, but may also deprive consumers of the possibility to enjoy lower prices, if entry did occur.

Although it deals with pricing schemes rather than contracts, our paper is closely related to the literature on anticompetitive exclusive dealing. However, in our model exclusion will arise although the incumbent and the rival firm *simultaneously* set prices and all the buyers can purchase at the same time.[7]

Since Segal and Whinston (2000) is probably the closest work to ours, let us be more specific on the differences with their work. Building on Rasmusen et al. (1991), they show the exclusionary potential of exclusive contracts when the

[5] See Thisse and Vives (1998). For a recent survey on discriminatory pricing, see e.g. Stole (2005).

[6] See e.g. Armstrong and Vickers (1993).

[7] Bernheim and Whinston (1998) analyze the possible exclusionary effects of exclusive dealing when firms make simultaneous offers, but in *non-coincident markets*: first, exclusivity is offered to a buyer in a first market; afterwards, offers are made to a buyer in a second market. In their terminology, our paper is looking at *coincident market* effects, which makes our analysis closer to Aghion and Bolton (1985), Rasmusen et al. (1991), Segal and Whinston (2000) and Fumagalli and Motta (2006). All these papers, however, study only exclusive dealing arrangements and assume that the entrant can enter the market (if at all) only *after* the incumbent and the buyers have negotiated an exclusive contract.

incumbent can discriminate on the compensatory offers it makes to buyers. Our study differs from theirs in several respects: (i) in their game the incumbent has a (first-mover) strategic advantage in that it is allowed to contract with buyers before entry occurs; (ii) if buyers accept the exclusivity offer of the incumbent, they commit to it and cannot renegotiate it even if entry occurs; (iii) buyers are symmetric and only linear pricing is considered. In our paper, instead, (i) the incumbent and the entrant choose price schedules simultaneously, (ii) buyers simply observe prices and decide which firm to buy from (therefore avoiding any problems related to assumptions on commitment and renegotiation); (iii) we explore the role of rebates and quantity discounts in a world where buyers differ in size. Yet, the mechanisms which lead to exclusion in the two papers are very similar: both papers present issues of buyers' miscoordination, and scale economies which are created by fixed costs in their model are created instead by network effects in ours (but in the concluding section, we explain that we obtain the same results by dropping network externalities and assuming that the entrant has still to incur fixed sunk costs).

Our paper is also related to Innes and Sexton (1993, 1994), who also analyze the anticompetitive potential of discriminatory pricing. In their papers, however, they consider a very different contracting environment, strategic variables, and timing of the game. In particular, after the incumbent made its offers, they allow the buyers to contract with the entrant (or to enter themselves), so as to create countervailing power to the incumbent's. Despite all these differences, Innes and Sexton's insight that discrimination helps the incumbent to 'divide and conquer' consumers reappears in our paper, even if we also allow for the entrant to use the same discriminatory tools available to the incumbent, and even if contrary to Innes and Sexton's (1994) finding, in our case a ban on discrimination cannot prevent inefficient outcomes: in our setting, exclusion can arise also under uniform linear pricing.

Finally, our paper is related to the literature on incompatible entry in network industries. The very nature of network effects provides a strong incumbency advantage, shielding dominant firms against competitors even in the absence of any anticompetitive conduct (Farrell and Klemperer (2006)). Crémer et al. (2000) show that compatibility is a key variable in determining whether or not an entrant can successfully challenge an incumbent. Under incompatibility, entry equilibria may not even exist, and when they exist, the incumbent is likely to maintain a higher market share than under compatibility. Thus, if compatibility is a choice variable, the incumbent can use it strategically to deter entry. Where incompatibility could be overcome through multi-homing, Shapiro (1999) argues that incumbents can use exclusive dealing contracts to block multi-homing, thus excluding a technologically superior firm. Our paper adds to this literature in showing that even simple price discrimination can be sufficient for an incumbent to deter a more efficient in such network industries.

The paper continues in the following way. Section 2 describes the model, Section 3 solves the model under the assumption that prices have to be non-negative. Three cases are analyzed: uniform pricing, explicit (or 3rd degree) price discrimination and implicit (or 2nd degree, or rebates) price discrimination. Section 4 studies the effects of the different pricing schemes on consumer surplus. Section 5 discusses some extensions of the model. First, we consider the possibility that firms subsidize customers' usage, i.e., can charge negative linear prices; then, we turn to the case of elastic (linear) demands, allowing for both linear and two-part tariffs; finally, we discuss the case of full (or buyer-specific) discrimination (in the base model we do not allow firms to discriminate across identical buyers). Section 6 concludes the paper.

2 The setup

Consider an industry composed of two firms, the incumbent I, and an entrant E. The incumbent supplies a network good, and has an installed consumer base of size $\beta_I > 0$. (The network good is durable: "old" buyers will continue to consume it but no longer need to buy it.) I incurs constant marginal cost $c_I \in (0,1)$ for each unit it produces of the network good.

The entrant can supply a competing network good at marginal cost $c_E < c_I$, i.e. it is more efficient than the incumbent. E has not been active in the market so far, that is it has installed base $\beta_E = 0$, but it can start supplying the good any time; in particular, when the game starts it does not have to sink any fixed costs of entry.

The good can be sold to $m+1$ different "new" buyers, indexed by $j = 1, \ldots, m+1$. There are $m \geq 1$ identical small buyers, and 1 large buyer.[8] Goods acquired by one buyer cannot be resold to another buyer, but they can be disposed of at no cost by the buyer who bought them (in case the latter cannot consume them). Side payments of any kind between buyers are ruled out. Define firm i's network size s_i (where $i = I, E$) as

$$s_i = \beta_i + q_i^1 + \ldots + q_i^{m+1} \tag{1}$$

i.e. the firm's installed base plus its total sales to all "new" buyers.

To simplify the analysis, we assume that demands are inelastic. (Section 5.2 presents the results for linear demand functions.) A buyer will either buy from the incumbent, or from the entrant (but not from both). The large buyer can consume at most $Q^l = 1-k$ units, while any small buyer can consume at most $Q^s = \frac{k}{m}$ units. Buyers exert positive consumption externalities on each other: If firm i's network size s_i is below the threshold level $\bar{s}$, consumption of i's good gives zero surplus to its buyer.[9] The goods produced by the two firms are incompatible, so that buyers of firm i do not exert network externalities on

[8] We assume $m \geq 1$ so as to allow for the large buyer to be smaller than the set of all small buyers (which in turn allows for the large buyer to receive better price offers) and to show that prices under rebates depend on the degree of fragmentation of small buyers (and converge to prices under explicit discrimination as $m \to \infty$).

[9] The assumption that a buyer's utility from consuming is positive only if the network in question reaches the threshold size $\bar{s}$ is designed to capture in an admittedly simple way the presence of network effects. Rather than assuming that the utility of a consumer increases continuously with network size, we assume a discontinuous formulation; this has the advantage

buyers of firm j. For a network good of sufficient size, large and small buyers have the same maximum willingness to pay of $\bar{p} = 1$.

We assume that

$$\beta_I \geq \bar{s} \tag{2}$$

i.e. the incumbent has already reached the minimum size, while the entrant's installed base is $\beta_E = 0$. In order to operate successfully, the entrant will have to attract enough buyers to reach $\bar{s}$.[10]

Let the unit prices offered by the two firms to a buyer of type $j = l, s$ be $p_I^j \leq 1$ and $p_E^j \leq 1$. Then, buyer j's demand functions for the incumbent's good, q_I^j, and for the entrant's good, q_E^j, are given by:[11]

$$q_I^j\left(p_I^j, p_E^j, s_I, s_E\right) = \begin{cases} Q^j & \begin{array}{c} \text{if } s_E \geq \bar{s} \text{ and } p_I^j \leq p_E^j \\ \text{or } s_E < \bar{s} \text{ and } -p_E^j < 1 - p_I^j \end{array} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$q_E^j\left(p_I^j, p_E^j, s_I, s_E\right) = \begin{cases} Q^j & \begin{array}{c} \text{if } s_E \geq \bar{s} \text{ and } p_E^j \leq p_I^j \\ \text{or } s_E < \bar{s} \text{ and } -p_E^j \geq 1 - p_I^j \end{array} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where the large buyer's demand is $Q^l = 1 - k$, while the typical small buyer's demand is $Q^s = \frac{k}{m}$. If $s_E \geq \bar{s}$ and there is a tie in prices, $p_E^j = p_I^j$, the buyer may either buy from I or from E (we allow for both possibilities).

The parameter $k \in (0, 1)$ is an indicator of the relative weight of the small buyers in total market size: $1 - k$ measures the large buyer's market share, while k measures the market share of the group of small buyers. Assume that $1 - k > k/m$, so that the large buyer's demand is always larger than a small buyer's demand (provided they both demand strictly positive quantities). Note that the assumption $1 - k > k/m$ implies an upper bound on k, namely

$$k < \frac{m}{m+1} \in \left[\frac{1}{2}, 1\right). \tag{5}$$

Total market size is normalized to 1: $m(k/m) + (1 - k) = 1$.

Define a buyer's net consumer surplus as gross consumer surplus minus total

that the old generation of buyers can be safely ignored when studying welfare effects: since we shall assume that they have already attained the highest level of utility, new buyers' decisions will never affect old buyers' utility.

[10] Note that if the entrant manages to reach the minimum size $\bar{s}$, then consumers will consider I's and E's networks as being of homogenous quality, even if $s_I \neq s_E$.

[11] These demand functions apply for general (positive or negative) prices. In the base model we restrict prices to be non-negative. Section 5 considers the case where prices can be negative.

expenditure:

$$CS^l\left(p_i^l, q_i^l, s_i\right) = \begin{cases} q_i^l(1-p_i^l) & \text{if } s_i \geq \bar{s} \text{ and } l \text{ buys } q_i^l \leq 1-k \\ (1-k) - q_i^l p_i^l & \text{if } s_i \geq \bar{s} \text{ and } l \text{ buys } q_i^l > 1-k \\ -q_i^l p_i^l & \text{if } s_i < \bar{s} \text{ and } l \text{ buys } q_i^l \leq 1-k \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$CS^s\left(p_i^s, q_i^s, s_i\right) = \begin{cases} q_i^s(1-p_i^s) & \text{if } s_i \geq \bar{s} \text{ and } s \text{ buys } q_i^s \leq \frac{k}{m} \\ \frac{k}{m} - q_i^s p_i^s & \text{if } s_i \geq \bar{s} \text{ and } s \text{ buys } q_i^s > \frac{k}{m} \\ -q_i^s p_i^s & \text{if } s_i < \bar{s} \text{ and } s \text{ buys } q_i^s \leq \frac{k}{m} \\ 0 & \text{otherwise} \end{cases}$$

The demand functions defined above can be derived from these expressions of net consumer surplus.

Since both types of buyers have the same prohibitive price $\bar{p} = 1$, a monopolist who could charge discriminatory linear prices would set a uniform unit price $p_i^m = 1$. Thus, discriminatory pricing can arise only as a result of the strategic interaction between the incumbent and the entrant.

We assume that neither demand of the large buyer alone, nor demand of all small buyers taken together, is sufficient for the entrant to reach the minimum size:

$$\bar{s} > \max\ \{1-k, k\}\,. \tag{7}$$

In other words, in order to reach the minimum size, the entrant has to serve the large buyer plus at least one (and possibly more than one) small buyer.[12,13]

Note that only units which are actually consumed by a buyer count towards firm i's network size.

We also assume that the threshold level $\bar{s}$ is such that if the entrant sells to all $m+1$ new buyers, then it will reach the minimum size: $\bar{s} \leq 1$.

This, together with the assumption $c_E < c_I$, implies that the social planner would want the entrant (and not the incumbent) to serve all buyers.

The game. Play occurs in the following sequence: At time $t = 0$, the incumbent and the entrant simultaneously announce their prices, which will be binding in $t = 1$. At time $t = 1$, each of the $m+1$ buyers decides whether to patronize the incumbent or the entrant. We also assume that offers are observable to everyone, e.g. because they have to be posted publicly. Then, when the buyers have to decide which firm to buy from, the firms' offers will be common knowledge.

As for the prices that firms can offer in $t = 0$, in the base model (Section 3) we will restrict attention to linear pricing schemes, but we consider three different possibilities: (1) uniform prices (Section 3.1); (2) explicit (or third-degree) price

[12] If either $\bar{s} < 1-k$, or $\bar{s} < k$, then the miscoordination issues, which are at the heart of this paper, would not arise.

[13] We could assume in addition that $\bar{s} \leq 1-k+k/m$, so that the large buyer plus *exactly one* small buyer is sufficient for E to reach the minimum size. This assumption only changes the analysis of miscoordination equilibria when firms can charge negative prices.

discrimination (Section 3.2); and (3) the case of central interest, that is implicit (or second-degree) price discrimination, i.e. the case of standardized quantity discounts or "rebates" (Section 3.3).

Uniform pricing means that a firm must charge the same price to all buyers. Under explicit price discrimination, each firm can set one price for the large buyer, and a different price for the small buyers (all buyers of the same type will be charged the same price). Under implicit price discrimination, all buyers are offered the same price menu, where different prices apply depending on whether the buyer reaches a certain quantity threshold or not: if this menu is designed appropriately, buyers will self-select into different tariffs: small buyers will buy below the threshold, while the large buyer will buy above the threshold, and so the large buyer will end up paying a different price than the small buyers.

Explicit discrimination may not always be feasible, for instance because of informational constraints (firms cannot observe buyer types), or because of policy constraints (explicit discrimination is outlawed, as in the European Union). However, when buyers are asymmetric, pure standardized quantity discounts can induce de facto discrimination, and so allow firms to (imperfectly) replicate outcomes under explicit discrimination.

Section 5 will show that the main results are robust to changes in the assumptions we make in the base model on prices. There, we shall analyze the cases where prices can be negative, where demand is elastic, and where full price discrimination is allowed, that is firms can make buyer-specific offers (in the base model, we do not allow firms to discriminate among buyers of the same type).

3 Equilibrium solutions, under different price regimes

In this Section, we assume that firms set linear (and non-negative) prices, and we consider three different price regimes: uniform prices; explicit (3rd degree) price discrimination; implicit (2nd degree) price discrimination (i.e., rebates).

3.1 Uniform pricing

Assume that firms can only use uniform linear prices, p_i with $i = I, E$. Recall that any buyer's demand for E's good, $q_E^j(\ldots, s_E)$, depends on the size of E's network, s_E, which in turn depends on E's sales to the buyers, $\{q_E^1, \ldots, q_E^{m+1}\}$. Thus, in line with Segal and Whinston (2000), we find that our game has two types of pure-strategy Nash equilibria: one where all buyers (or sufficiently many) buy from the entrant, and one where all buyers buy from the incumbent.

The following proposition shows the highest prices that can be sustained in each of these two types of equilibria.

Proposition 1 *(equilibria under uniform linear prices) If firms can only use uniform flat prices, the following two pure-strategy Nash equilibria exist under the continuation equilibria as specified:*

(i) ***Entry equilibrium:*** *E sets $p_E = c_I$, I sets $p_I = c_I$, and in all continuation equilibria where $p_E \leq p_I$, all buyers buy from E.*

(ii) ***Miscoordination equilibrium:*** *I sets $p_I = p_I^m = 1$, E sets $p_E = p_E^m = 1$, and in all continuation equilibria where $p_E \leq p_I$, all buyers buy from I.*

The prices in (i) and (ii) are the highest that can be sustained in each type of equilibrium.

Proof: see Appendix A

Which type of equilibrium will eventually be played depends on the underlying continuation equilibria, i.e. on how buyers coordinate their purchasing decisions after observing the firms' offers:[14] If a buyer can rely on all other buyers patronizing E whenever E's offer is at least as good as I's, then it is perfectly rational for this buyer to buy from E as well. This, in turn, corresponds exactly to what all other buyers expect him to do, and so confirms the rationality of their own supplier choice. Under such a continuation equilibrium, the entry equilibrium of Proposition 1 (i) will arise.

If instead each buyer expects all other buyers to patronize I even when I's price is strictly higher than E's, then no buyer will want to buy from E: Recall that no individual buyer's demand is ever sufficient for E's network to reach the minimum size $\bar{s}$. Then, being the only buyer to buy from E means ending up with a good that has zero value to that buyer (no matter how cheap it is). Since buying from I still gives non-negative surplus, each buyer will want to buy from I, which then confirms all other buyers in their decision to buy from I as well.

Note also that under the miscoordination equilibrium, the incumbent can charge the monopoly price without losing the buyers to the entrant.[15] These equilibria are particularly troublesome, because they show that the highest possible price can persist even in the presence of a more efficient competitor.

The equilibria characterized in Proposition 1 represent extreme cases, in the sense that the underlying continuation equilibria are the most favorable ones for the firm that serves the buyers in equilibrium. *These equilibria are by no means the only equilibria that can arise in our game.*

For instance, there are other equilibria where all buyers do miscoordinate on the incumbent, but the latter can at most charge some price $\tilde{p}_I < p_I^m = 1$. Such an equilibrium can be sustained by continuation equilibria where buyers buy from I as long as $p_E \leq p_I \leq \tilde{p}_I$, but would switch to E if p_I exceeded $\tilde{p}_I$. Likewise, there are entry equilibria where the entrant must charge a strictly

[14] "Coordination" refers to the collective behavior under *individual* decision making; we do not allow buyers to meet in $t = 1$ and make a joint decision on which firm to patronize.

[15] In this situation, the entrant is indifferent among all prices $p_E \geq 0$ it could charge, and might as well offer the monopoly price, which weakly dominates all other possible equilibrium prices.

lower price than c_I to induce buyers to coordinate on E. *For the rest of the paper, we will focus on those continuation equilibria which are the most profitable ones for the firm that eventually serves the buyers.* The motivation for this choice is two-fold: First, these equilibria are the Pareto-dominant ones from the point of view of the firms. Second, from a policy point of view, the equilibria with the highest profits are those which cause most concern.

Finally, there can also be equilibria where both I and E offer the same price, and a critical number of buyers patronize E (so that E reaches the minimum size), while the remaining buyers buy from I. These equilibria can only be sustained by very specific continuation equilibria, and we will not consider them in the following sections of this paper.

3.2 Explicit (3rd degree) discrimination

In this section, we first analyze miscoordination equilibria and then entry equilibria.

3.2.1 Miscoordination equilibria

Proposition 1 gives us the equilibrium for the case of uniform linear pricing. Assume now that the two firms can do 3rd degree (or explicit) discrimination, i.e. each firm chooses a pair of prices $\left(p_I^s, p_I^l\right)$, one price for the large buyer, and another for the small buyers (this is partial discrimination: firms cannot offer different prices to buyers of the same size).

With respect to the uniform pricing case, *nothing changes in the miscoordination equilibria*, the most profitable of which is for the Incumbent still the one where $p_I^s = p_I^l = p_I^m = 1$,[16] while the entrant sets $p_E^m = 1$ and all buyers buy from I. Clearly, the incumbent would have no incentive to deviate from this solution. No buyer would deviate either: if any of them decided to accept a lower price offered by the entrant given that all others buy from the incumbent, he would have zero surplus and would reduce his utility.

Proposition 2 *(miscoordination equilibria under explicit discrimination) Let each firm choose a pair of prices (p_I^s, p_I^l), one for each type of buyer. Under the appropriate continuation equilibria, the miscoordination equilibrium where all buyers buy from I exists for all parameter values. The highest sustainable prices are $p_I^s = p_I^l = p_E^s = p_E^l = p_I^m = 1$.*

Proof: Consider the following continuation equilibria: Following offers where either $p_E^s \leq p_I^s$ or $p_E^l \leq p_I^l$, or both, all buyers buy from I. Then, even if the entrant can charge different prices to both groups (where both prices may be strictly lower than I's prices), no single buyer will have an incentive to switch to the entrant as long as he expects all other buyers to buy from I: E's network cannot reach the minimum size with only one buyer, so its good gives

[16] Note that in our model the monopoly price charged by a firm under explicit discrimination will be the same for all buyers. This is clearly a special feature of the model, which simplifies the analysis without losing much insight.

zero utility, and as long as E charges a non-negative price for it, I's offer will (weakly) dominate E's offer. The rest of the proof is analogous to the proof of Proposition 1.□

Thus, the possibility to price discriminate does not allow the entrant to solve the miscoordination problem. Hence, miscoordination equilibria will continue to exist even if we allow for explicit price discrimination.

3.2.2 Entry equilibria

For entry equilibria, things change relative to the uniform pricing case. To fix ideas, start with the candidate entry equilibrium where both firms charge c_I and all buyers buy from the entrant (we have seen that this is an entry equilibrium in the uniform linear pricing case). This equilibrium can be disrupted by the incumbent setting a price $c_I - \epsilon$ to one category of buyers and the monopoly price to the other category: the loss made on the former would be outweighed by the profits made on the latter. Indeed, under this deviation the former category strictly prefers to buy from I, thus preventing the entrant from reaching the minimum size, and the latter category would then prefer to buy from I rather than from the entrant, since they would derive zero utility from buying from E.

Therefore, an entry equilibrium can exist only if it is immune to the deviations outlined above, i.e. if the entrant's prices to both large and small buyers are so low that the incumbent cannot profitably undercut either of the two prices while charging the monopoly price to the other group. This implies that the highest prices that the entrant can charge in any entry equilibrium will be strictly below c_I. Thus, for an entry equilibrium to exist, the efficiency gap between entrant and incumbent must be large enough.

Proposition 3 *(entry equilibria under explicit discrimination) Under explicit price discrimination, entry equilibria only exist if*

$$c_I \geq \min\left\{\frac{1+c_E}{2}, k + c_E, 1 - k + c_E\right\}.$$

The highest prices that the entrant can charge in any such entry equilibrium are

$$p_E^s = \begin{cases} \frac{c_I-(1-k)}{k} & \text{if } c_I \geq 1-k \\ 0 & \text{if } c_I < 1-k \end{cases} \quad \textit{and} \quad p_E^l = \begin{cases} \frac{c_I-k}{1-k} & \text{if } c_I \geq k \\ 0 & \text{if } c_I < k \end{cases}.$$

Proof: see Appendix A

Figure 1 illustrates the results of Proposition 3 (recall that miscoordination equilibria exist for all parameter values). For given k, the figure shows that the larger c_I with respect to c_E, the more likely for entry to be an equilibrium of the game. The effect of k on equilibrium outcomes is slightly more complex. In particular, entry is more likely at very low levels and very high levels of

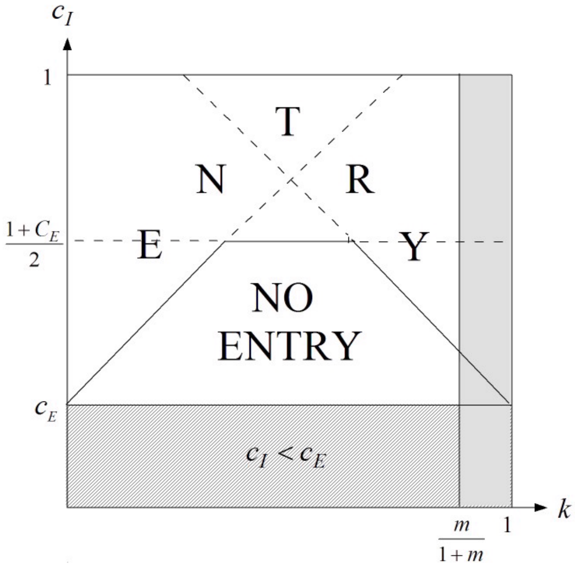


Figure 1: Regions where entry equilibria exist and do not exist under explicit price discrimination (the grey areas are outside of the parameter space)

k. This is because the entry equilibrium may be disrupted by the incumbent's offers to the buyers; such offers are made by discriminating among the buyers, for instance by extracting rents from small buyers and offering surplus to the large (or vice versa). If for instance k is very small, the incumbent is not able to extract much surplus from the small buyers and accordingly the best offer to the large buyer will not be very attractive. Instead, it could extract a lot of surplus from the large buyer and could in principle make a princely offer to the small buyers. However, small buyers account for a small proportion of demand (k very small), and the large rent earned from the large buyer can only be passed on to small buyers through price cuts on each unit they buy. But since prices are restricted to be non-negative here, the incumbent will soon hit the $p_I^s = 0$ constraint when k is small. Thus, the incumbent can only transfer a small part of the rent from large to small buyers, and so the entrant will find it easier to match the incumbent's best offers, hence entry equilibria will exist. The same argument can be used symmetrically to explain results for the case where k is close to 1. Of course, we shall see below that when prices are not restricted to be non-negative, these effects will disappear, and k will affect results monotonically.

To sum up:

- Exclusionary equilibria always exist, and the highest sustainable prices are exactly the same as under uniform linear pricing.[17]

- Entry equilibria only exist if c_I is high enough relative to c_E. When they exist, note that the highest sustainable equilibrium prices are always strictly below c_I (which is the highest sustainable equilibrium price under uniform linear pricing).

- With respect to uniform pricing, thus, price discrimination (i.e. a more aggressive pricing strategy): (a) on the one hand, makes exclusion more likely; (b) on the other hand, for given market structure, results in (weakly) lower prices.[18]

3.3 Implicit (2nd degree) discrimination (or rebates)

Let us now consider the case where firms cannot condition their offers directly on the type of buyer (large or small), but have to make uniform offers to both types which may only depend on the quantity bought by buyer $j = 1, \ldots, m+1$:

$$T_i(q_i^j) = \begin{cases} p_{i,1} q_i^j & \text{if } q_i^j \leq \bar{q}_i \\ p_{i,2} q_i^j & \text{if } q_i^j \geq \bar{q}_i \end{cases} \tag{8}$$

(If the buyer buys exactly the threshold quantity, $q_i^j = \bar{q}_i$, the firm may either charge $p_{i,1}$ or $p_{i,2}$.) Each buyer can now choose his tariff from this price menu by buying either below the sales target $\bar{q}_i$ or above it.

It is well-known that such quantity discounts or rebates, when applied to buyers who differ in size, will be a tool of (de facto) discrimination, even if the schemes as such are uniform. But to achieve discrimination, the tariffs have to be set in a way that induces buyers to self-select into the right bracket, with small buyers voluntarily buying below target, and the large buyer choosing to buy above it.

Consider the case where $\bar{q}_i < 1-k$. Then, the large buyer can either buy $1-k$ units at price $p_{i,2}$, which yields total surplus $CS^l\left(p_{i,2}, 1-k\right) = (1-k)(1-p_{i,2})$, or he can buy the threshold quantity $\bar{q}_i$ at price $p_{i,1}$ (i.e. a quantity which falls short of his actual demand at this price), in which case his net consumer surplus is $CS^l\left(p_{i,1}, \bar{q}_i\right) = (1-p_{i,1})\bar{q}_i$. If $p_{i,1}$ is sufficiently lower than $p_{i,2}$, it may be worthwhile for the large buyer to buy fewer units than he wants in return for a lower per unit price.[19]

[17] This is an artifact of the model. In more general models, monopoly prices will be different at the explicit discrimination equilibrium. However, the result that miscoordination equilibria will always exist and that at one of those equilibria the monopolist is able to charge monopoly prices would still be valid.

[18] In the entry equilibria, prices are strictly lower; in the exclusionary equilibrium, prices to both groups of buyers are the same as under uniform pricing.

[19] Assume that each buyer is only allowed one transaction. This rules out the possibility that a large buyer makes "multiple small purchases" so as to buy a large amount of units at the lower price. Presumably, important transaction costs may be invoked to justify this assumption, which in a way is nothing else than the counterpart of the assumption that a small buyer cannot buy a large quantity and then resell it to others. In both cases, it is arbitrage which is prevented.

Next, consider the case where $\bar{q}_i > k/m$. A typical small buyer $j = s$ will then have to choose between buying k/m units at price $p_{i,1}$, which yields total surplus $CS^s(p_{i,1}, k/m) = (1 - p_{i,1})(k/m)$, or buying the sales target $\bar{q}_i$ at price $p_{i,2}$ (i.e. a quantity which exceeds his actual demand at this price), giving total surplus of $CS^s(p_{i,2}, \bar{q}_i) = \frac{k}{m} - p_{i,2}\bar{q}_i$. In this case, if $p_{i,2}$ is sufficiently lower than $p_{i,1}$, the small buyer will want to purchase more units than he can actually consume in order to qualify for a lower unit price.[20]

We say that firm i's offer satisfies the *"self-selection conditions"* if the large buyer prefers to buy above the threshold, and the small buyers prefer to buy below the threshold, i.e. if

$$\begin{aligned} CS^l(p_{i,2}, 1-k) &\geq CS^l(p_{i,1}, \bar{q}_i) \\ \text{and } CS^s(p_{i,1}, k/m) &\geq CS^s(p_{i,2}, \bar{q}_i) \end{aligned} \tag{9}$$

For any offer that satisfies the self-selection condition, denote $(p_{i,1})$ by (p_i^s), and $(p_{i,2})$ by (p_i^l), for $i = I, E$.

We now look for the equilibria that arise in this game when both firms can use quantity discounts.

3.3.1 Miscoordination equilibria

Proposition 4 *(miscoordination equilibria under rebates) Let firms use rebates as defined in (8). Under the appropriate continuation equilibria, the miscoordination equilibrium exists for all parameter values, and the highest (monopoly) prices can be sustained at equilibrium.*

Proof: analogous proof as for the above cases.

Miscoordination arises under rebates for the same reason as under uniform pricing and explicit discrimination, which were discussed at length above. In the most profitable equilibrium, the incumbent does not actually offer a discount to either of the two groups, but charges the same (monopoly) price $p_I^s = p_I^l = 1$ to both large and small buyers. This offer trivially satisfies the self-selection conditions defined in (9).

3.3.2 Entry equilibria

The implicitly discriminatory effect of rebates gives rise to an exclusionary mechanism similar the one under explicit discrimination. Since buyers are asymmetric, they can be induced to self-select either into the high-quantity or the low-quantity bracket of the price menu, thus allowing the incumbent to de facto price-discriminate between them. This in turn enables the incumbent to offer a below-cost price to one group, thus winning their orders, while making up for

[20] Recall that we exclude reselling of units between buyers (while allowing for free disposal), so the only thing a small buyer can do with units he cannot consume is to throw them away.

the resulting losses by charging a high price (possibly the monopoly price) to the other group.

The major difference between explicit and implicit discrimination lies in the self-sorting conditions, which reduce the range of prices that the incumbent can offer. Consider for instance the case where, under explicit discrimination, the incumbent charges the monopoly price $p_I^s = 1$ to the small buyers, and $p_I^l = 0$ to the large buyer. Clearly, this offer does not satisfy the small buyers' self-sorting condition: At a zero price, the small buyers would always prefer to "buy" above the quantity threshold (i.e. receive a large quantity for free, and dispose of the units they cannot consume) rather than paying $p_I^s = 1$ (or any other positive price) for a small quantity.

Likewise, an offer where $p_I^s < c_I$ and $p_I^l = 1$ cannot be replicated through a rebate tariff: in this case, it is the large buyer who would prefer to buy below the threshold and enjoy a positive surplus on the (few) units he consumes, rather than buying above the threshold and being left with zero surplus.[21]

Thus, while rebates still have exclusionary potential, the incumbent's deviation offers will be less aggressive under rebates than under explicit discrimination, allowing for entry equilibria to be sustained where they do not exist if firms can explicitly price discriminate.

Proposition 5 *(entry equilibria under rebates) Under rebates as defined in (8), entry equilibria only exist if*

(i) $c_E < \frac{1}{2(m+1)}$ *and* $c_I \geq \min\left\{c_E(1+m), k + c_E, \frac{m}{1+m} + c_E - k\right\}$

(ii) or if $c_E \geq \frac{1}{2(m+1)}$ *and* $c_I \geq \min\left\{\frac{m+(1+m)c_E}{1+2m}, k + c_E, \frac{m}{1+m} + c_E - k\right\}$

The highest prices that the entrant can charge in any such entry equilibrium are

$$p_E^s = \begin{cases} 1 - \frac{m(1-c_I)}{k(m+1)} & \text{if } c_I \geq 1-k-k/m \\ 0 & \text{if } c_I < 1-k-k/m \end{cases}$$

$$p_E^l = \begin{cases} \frac{c_I - k}{1-k} & \text{if } c_I \geq \frac{k(1+m)}{m} \\ \frac{c_I}{(1-k)(m+1)} & \text{if } c_I < \frac{k(1+m)}{m} \end{cases}$$

Proof: see Appendix A

Corollary 6 *The parameter space for which entry equilibria exist under explicit discrimination is a proper subset of the parameter space for which entry equilibria exist under rebates.*

Proof: see Appendix A

[21] Such a rebate scheme may appear as somewhat unorthodox, since buyers are "rewarded" for buying little and "penalized" for buying a lot. However, this is a deviation offer which will never be made in equilibrium.

Figure 2 illustrates the results of the analysis of entry equilibria under rebates and non-negative prices for the case where $c_E \geq \frac{1}{2(m+1)}$ (recall that miscoordination equilibria exist for all parameter values). We see that the region where

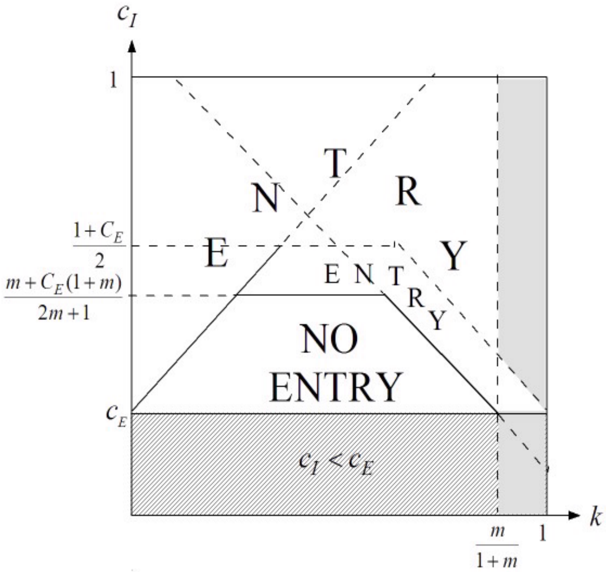


Figure 2: Regions where entry equilibria exist and do not exist under rebates (i.e. implicit price discrimination), compared to explicit discrimination

entry equilibria do not exist is smaller under rebates than under explicit discrimination. While nothing changes for low values of k (rebates exactly replicate the outcome under explicit discrimination), exclusion becomes more difficult for intermediate and high values of k. Intuitively, given m, the large buyer becomes smaller and smaller the higher k is, and so he becomes more and more similar to the small buyers, making it difficult to discriminate between them through rebates without violating any of the self-sorting conditions.

Note that as m grows, so that a single small buyer becomes smaller and smaller, both the efficiency thresholds and prices under rebates converge to the values under explicit discrimination. In the limit case where $m \to \infty$, the self-selection constraints play no role: the large buyer will never want to behave like a small buyer whose demand is infinitely small, and vice versa for the small buyer, and so the implicit and explicit discrimination cases coincide.

4 Consumer Welfare

Recall from Section 2 that, in our model, entry is always socially efficient, because the entrant produces at a lower marginal cost than the incumbent. Thus, all miscoordination equilibria are inefficient. The higher production costs associated with having a less efficient firm serve the buyers are the only source of inefficiencies in our model: buyers have inelastic demand functions, so they will always consume the efficient quantities, no matter how high the prices are.

Yet, prices do matter, as they determine consumer surplus, which is often considered the objective function of antitrust agencies. Now, comparing equilibrium prices across different price regimes is not straightforward because each price regime gives rise to multiple equilibria, both entry and miscoordination equilibria, and each of these can be sustained by a broad range of prices. The approach we take here is to compare the "worst case scenarios" *given* market structure, i.e. the highest sustainable prices under each price regime given that either the incumbent or the entrant serves the buyers.

Proposition 7 *(consumer surplus)*

(i) Miscoordination equilibria: Under all three price regimes (uniform pricing, explicit discrimination, and rebates), the highest equilibrium price is the monopoly price, and so consumer surplus is the same:

$$CS^j_{\text{explicit}} = CS^j_{\text{implicit}} = CS^j_{\text{uniform}} = 0 \text{ for } j = s, l.$$

(ii) Entry equilibria: At the highest sustainable prices under each regime, consumer surplus is maximal under explicit discrimination, intermediate under rebates, and minimal under uniform pricing:

$$\begin{aligned} CS^l_{\text{explicit}} &\geq CS^l_{\text{implicit}} > CS^l_{\text{uniform}} > 0 \text{ with strict inequality if } c_I < \frac{k(1+m)}{m} \\ CS^s_{\text{explicit}} &\geq CS^s_{\text{implicit}} > CS^s_{\text{uniform}} > 0 \text{ with strict inequality if } c_I \geq 1 - k - k/m \end{aligned}$$

Proof: Under all three price regimes, buyers consume the same quantities. Thus, their consumer surplus is solely determined by the price they pay: the higher the price, the lower is consumer surplus.

(i) follows immediately from Propositions 1,2, and 4.

(ii) The following table shows the prices buyers pay under each of the three price regimes. The inequalities follow from simple algebra.

Table 1: Highest Sustainable Prices in Entry Equilibria

	Uniform		Implicit		Explicit
Large Buyer:					
$c_I < k$	$p^l_E = c_I$	$>$	$p^l_E = \frac{c_I}{(1-k)(m+1)}$	$>$	$p^l_E = 0$
$c_I \in \left[k, \frac{k(1+m)}{m}\right)$	$p^l_E = c_I$	$>$	$p^l_E = \frac{c_I}{(1-k)(m+1)}$	$>$	$p^l_E = \frac{c_I - k}{1-k}$
$c_I \geq \frac{k(1+m)}{m}$	$p^l_E = c_I$	$>$	$p^l_E = \frac{c_I - k}{1-k}$	$=$	$p^l_E = \frac{c_I - k}{1-k}$
Small Buyers:					
$c_I < 1 - k - k/m$	$p^s_E = c_I$	$>$	$p^s_E = 0$	$=$	$p^s_E = 0$
$c_I \in [1 - k - k/m, 1 - k)$	$p^s_E = c_I$	$>$	$p^s_E = 1 - \frac{m(1-c_I)}{k(m+1)}$	$>$	$p^s_E = 0$
$c_I \geq 1 - k$	$p^s_E = c_I$	$>$	$p^s_E = 1 - \frac{m(1-c_I)}{k(m+1)}$	$>$	$p^s_E = \frac{c_I - (1-k)}{k}$

The ranking of consumer surplus is the reverse of the ranking of prices.□

The main results can be summarized as follows:

- Under uniform pricing, entry equilibria exist for all parameter values, but the price charged by the entrant is higher than any of the prices under rebates or explicit discrimination.
- Under rebates, entry equilibria exist for a larger region of the parameter space than under explicit discrimination. Indeed, if an entry equilibrium exists under explicit discrimination, it will also exist under rebates; but explicit discrimination may allow the incumbent to break some entry equilibria that would exist under rebates.
- When entry equilibria exist, the prices charged by the entrant to both groups of buyers are (weakly) higher under rebates than under explicit discrimination.

5 Extensions

In this Section, we shall deal with a number of extensions to the basic model. First, we shall analyze in Section 5.1 how results change when we consider the possibility that firms subsidize consumption, i.e. can charge negative prices. This makes the pricing behavior of both firms more aggressive. Not surprisingly, the Incumbent will be able to exclude entry for a wider region of parameter values, but the basic trade-off between exclusion and lower prices acquires now an important dimension. Indeed, the possibility of setting negative prices, i.e. of subsidizing buyers for *using* the product, gives an important tool to the entrant to disrupt miscoordination equilibria. Contrary to the base model (where prices were constrained to be non-negative), if negative-price *discriminatory* offers can be made, miscoordination equilibria do not always exist. In particular, unless the gap between incumbent's and entrant's costs is sufficiently small, miscoordination equilibria do not exist, and if they exist they can be sustained only by lower than monopoly prices.

Next, Section 5.2 will deal with the case of elastic demands, allowing for both linear prices and two-part tariffs.[22] So far, we have assumed that demands are inelastic for simplicity. One possible problem with these demands is that unless a productive inefficiency occurs, total welfare is the same at high or low prices. It is true that lower equilibrium prices will lead to a better social outcome unless consumer surplus and producer surplus have exactly the same weight in the objective function (and most antitrust authorities tend to maximize consumer welfare, not total welfare), but it is still important to look at how our results extend to a setting where demands are elastic.

[22]Because of space limitations, we only summarize our findings, without presenting the full treatment. The analytics for the elastic demand case and the case of full discrimination are available from the authors upon request.

Finally, Section 5.3 discusses the case where firms are allowed to discriminate even among buyers of the same type, i.e. to set different prices for different small buyers. If uniform pricing is at one end of the extreme, this fully discriminatory price regime is at the other end.

5.1 Allowing for usage subsidies

In this section, we keep inelastic demands, but relax the assumption that prices must be non-negative.

5.1.1 Uniform prices

Under uniform price offers, the results are the same as in the base model. The *miscoordination equilibrium* cannot be disrupted by negative price offers, because the entrant cannot profitably offer negative prices to all buyers. For the same reason, the *entry equilibrium* will also exist for all parameter values. Therefore, *Proposition 1 still holds good.*

5.1.2 Explicit price discrimination

We consider first miscoordination equilibria and then entry equilibria.

Miscoordination equilibria The possibility to offer negative prices changes dramatically the analysis of miscoordination equilibria. Consider for instance a natural candidate equilibrium, that is the miscoordination equilibrium prevailing under uniform (non-negative) prices: $(p_I^s = 1, p_I^l = 1)$ and all buyers buy from the incumbent. Under positive prices, this miscoordination equilibrium is sustained by any continuation equilibrium where firm I sets $p_I^s = p_I^l = p_I^m = 1$, firm E sets, for instance, $p_E^l = p_E^m = 1$, $p_E^s = 0$, and all buyers buy from I. This is an equilibrium because if a small buyer, who is offered a zero price by the entrant, decided to switch to the entrant given that all others buy from the incumbent, he would get zero surplus, because the entrant does not reach critical mass and hence the utility derived from consuming the product would be zero. Therefore, the entrant would have no incentive to deviate either.

But this reasoning does not hold any longer when negative prices are admitted. Suppose that firm I sets $p_I^s = p_I^l = 1$. If firm E sets $p_E^l = p_I^l - \varepsilon = 1 - \varepsilon$ and $p_E^s < 0$, then all buyers will buy from the entrant. Indeed, by buying from the entrant each small buyer would receive a strictly positive surplus $(k/m)(-p_E^s) > 0$ even if nobody else consumed the product. Therefore, they will want to consume in order to receive the payment. But since it is a dominant strategy for the small buyers to consume the product, the large buyer will now prefer to buy from the entrant as well, since the critical network size will be met, and since $CS^l(p_E^l) = (1-k)(1-p_E^l) > CS^l(p_I^l) = 0$.

More generally, a miscoordination equilibrium with prices (p_I^s, p_I^l) will not exist if the entrant can offer a negative price $p_E^s < 0$ to the small buyers such that

$CS^s(p_E^s, s_E < \bar{s}) > CS^s(p_I^s, s_I \geq \bar{s})$ while slightly undercutting the incumbent's offer to the large buyer, $p_E^l = p_I^l - \varepsilon$.[23]

Proposition 7 *(miscoordination equilibria under negative prices) Let $\bar{s} > (1-k) + \frac{k}{m}$. Let buyers buy from the incumbent whenever it is not a dominant strategy to buy from the entrant. Then, if both firms charge negative prices, a miscoordination equilibrium will only exist if $c_I \leq k + c_E$.*

(i) If $c_E \leq 1-k$, the equilibrium is characterized by

$$\begin{aligned} p_I^l &= 1, p_I^s = 1 - \frac{1}{k}\left[1-k-c_E\right] \\ p_E^l &\in [0,1], p_E^s = -\frac{1-k-c_E}{k} \end{aligned}$$

(ii) If instead $c_E > 1-k$, the equilibrium is characterized by $p_I^l = p_I^s = 1$, and $p_E^l = p_E^s = 1$.

Proof: see Appendix

Figure 3 illustrates in the space (k, c_I) the region where the miscoordination equilibrium arises, for the case $c_E < 1/2$. It shows that this equilibrium exists only if c_I is sufficiently close to c_E.

The main conclusions from the analysis are that:

(1) when negative prices are possible, then allowing for explicit discrimination disrupts miscoordination equilibria when c_I is sufficiently high.

(2) When a miscoordination equilibrium exists under explicit discrimination (with linear prices which can be negative), the incumbent will not be able to enjoy the monopoly outcome ($p_I^s = 1, p_I^l = 1$), unless $c_E > 1-k$; the incumbent needs to lower its prices to prevent the entrant from stealing its buyers.

Compared to uniform pricing regimes, where a miscoordination equilibrium which reproduces the monopolistic outcome is always possible, allowing for negative prices has the effect of both rendering miscoordination equilibria less likely, and, where such equilibria survive, of reducing the equilibrium prices at those equilibria. Note that in this case, p_I^s may even be below-cost, i.e. $p_I^s < c_I$!

Entry equilibria The analysis of entry equilibria when we allow for negative prices requires just a small modification of the problem already analyzed in Section 3.2 above, i.e. allowing for p_I^s and p_I^l to take negative values.

[23] In the case where $\bar{s} \leq (1-k) + \frac{k}{m}$, the entrant might as well charge a negative price to the *large* buyer, while matching I's offer to the small buyers. In this case, as soon as E attracted the large buyer, E needs just one more buyer to reach the minimum size. Thus, any small buyer will find it optimal to buy from E as well, and the miscoordination equilibrium is broken. This is not the case if $\bar{s} > (1-k) + \frac{k}{m}$, where the entrant needs more than one small buyer to reach the minimum size, so that attracting the large buyer is not sufficient to solve the coordination problem among the small buyers. For simplicity, we will focus on this "asymmetric" case here.

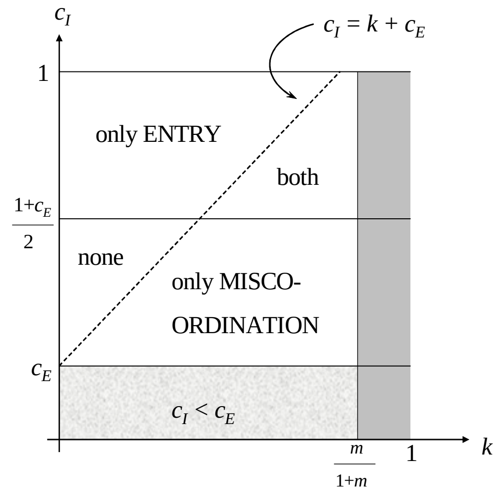


Figure 3: Regions where miscoordination equilibria and/or entry equilibria (or none) exist under negative prices, for $c_E < 1/2$

Proposition 8 *(entry equilibria under negative prices) If both firms can use explicit price discrimination and charge negative prices, entry equilibria only exist if*

$$c_I \geq \frac{1 + c_E}{2}.$$

The highest prices that the entrant can charge in any such entry equilibrium are

$$p_E^s = \frac{c_I - (1-k)}{k} \text{ and } p_E^l = \frac{c_I - k}{1-k}.$$

Proof: see Appendix A

Figure 3 illustrates entry equilibria. Note that under negative pricing, the incumbent can prevent entry for a larger region of parameter values than under non-negative prices: in the latter case, entry can also occur for values $c_I < \frac{1+c_E}{2}$, whereas under negative prices, the efficiency threshold shifts to $c_I = \frac{1+c_E}{2}$ everywhere.

The figure also shows that under explicit discrimination, there might be a situation where, for given c_E and k, for c_I sufficiently close to c_E a miscoordination equilibrium exists, for intermediate values of c_I no equilibrium in pure strategies exists, and for high values of c_I only the entry equilibrium will exist. (To be precise, such a situation exists if $c_E < 1/3$). For high values of k, there exists an area of parameter values where both miscoordination and entry equilibria will coexist.

To interpret these results, recall that under uniform pricing both entry and miscoordination equilibria exist under all parameter values. This multiplicity of equilibria in the base case makes it difficult to identify precise policy implications. However incomplete (depending on the values of c_E, there may also exist other regions where no equilibria exist under explicit discrimination, or where multiple equilibria exist also under explicit discrimination), the following Table allows to fix ideas. It shows that for relatively high efficiency gaps between incumbent and entrant, if explicit discrimination schemes are allowed consumer welfare will always be (weakly) higher than under uniform pricing (miscoordination equilibria never exist, and entry equilibria are characterized by (weakly) lower prices). For relatively low efficiency gaps between incumbent and entrant, though, the impact on consumer welfare is not unambiguous: at equilibrium, the incumbent will always serve, and the desirability of explicit discrimination schemes depends on which equilibrium would prevail under uniform pricing: if under uniform pricing a miscoordination equilibrium is played, then explicit discrimination will increase consumer welfare, but if under uniform pricing an entry equilibrium is played, then explicit discrimination leads to exclusion and higher prices. We would then find again the same tension between exclusion and low prices that we have stressed in the main Section above, although it is to be noticed that - apart from very specific cases ($c_E > 1 - k$) - exclusion can be achieved by the incumbent only by decreasing equilibrium prices.

	Uniform pricing	Explicit discrim. (neg. prices)
$c_I > \max\left\{\frac{1+c_E}{2}, k + c_E\right\}$	I serves: $p_I^l = p_I^s = 1$ $\Longrightarrow CS = 0$ E serves: $p_E^l = p_E^s = c_I$ $\Longrightarrow CS = 1 - c_I$	E serves: $\widehat{p}_E^l \leq c_I; \widehat{p}_E^s \leq c_I$ $\Longrightarrow CS \geq 1 - c_I$
$c_I < \min\left\{\frac{1+c_E}{2}, k + c_E\right\}$	I serves: $p_I^l = p_I^s = 1$ $\Longrightarrow CS = 0$ E serves: $p_E^l = p_E^s = c_I$ $\Longrightarrow CS = 1 - c_I$	I serves: $\widehat{p}_I^l = 1; \widehat{p}_I^s \lesseqgtr c_I$ $\Longrightarrow CS \leq 1 - c_I$

5.1.3 Implicit price discrimination (rebates)

It would be tedious to characterize all the equilibrium solutions for the case of rebates as well. Like for the case of explicit discrimination, the possibility to set negative prices allows the incumbent to make more aggressive offers, eliminating entry equilibria which would have existed under uniform prices; also, and again like for explicit discrimination, it allows the entrant to subsidize a group of buyers and induce them to use the product independently of what other buyers do, thus leading to the disruption of miscoordination equilibria. The fact that the self-selection constraint needs to be satisfied does not therefore eliminate the possibility to disrupt some of the equilibria;[24] however, it does imply that

[24] At first sight, one may wonder why a buyer may want to buy at positive prices when it could mimic a buyer who is offered a negative price. But recall that a large buyer may get

competition is softer under rebates than under explicit discrimination. Even in this case, therefore, we find the result that rebates are less exclusionary than explicit discrimination, but lead to higher prices when similar equilibrium market structures are compared.

5.2 Elastic demands

Here we relax the assumption that demands are inelastic, by assuming a simple linear demand function for the buyers. We briefly deal with two cases:

5.2.1 Linear prices

It turns out that working with elastic demands allows us to uncover an interesting feature of rebates when linear prices are considered. By incorporating a quantity threshold (a certain price is offered for demand up to a certain number of units), a rebate scheme contains a de facto rationing scheme which limits the number of units that a firm has to sell at a given price. Therefore, when offering below-cost prices, a rebate allows a firm to limit losses or, which is the same, for a given amount of losses that it can sustain, it can afford offering lower prices than under an explicit discrimination scheme. This points to an interesting comparison between the relative aggressiveness of rebates v. explicit discrimination: on the one hand, the necessity to satisfy the self-selection constraints limits the aggressiveness of rebates, but on the other hand, the presence of an inherent rationing device (the quantity thresholds) allows a rebate scheme to make more aggressive offers. Therefore, the result obtained in Corollary 6, that the parameter space where entry equilibria exist under explicit discrimination is fully included in the corresponding space under rebates, does no longer hold: it is possible to find parameter values for which prices are lower under rebates than under explicit price discrimination, and other values for which the opposite holds.

5.2.2 Two-part tariffs

While the analysis becomes more tedious when working with linear demands and two-part tariffs, the results are very similar in spirit to the ones of our benchmark case, i.e. of inelastic demands and linear prices. Given linear demands, if the firms can use explicitly discriminatory two-part tariffs, the firms will set the variable component of the price at marginal cost (thus maximizing total surplus), and use the fixed fee to transfer rents between buyers (e.g. the incumbent could break an entry equilibrium by extracting all consumer surplus from the large buyer, and sharing it among the small buyers through a negative fixed fee).

more surplus from buying $1-k$ units at a positive price than a smaller number of units k/m at a negative price. However, we have seen in Section 3.3 that small buyers will never be willing to buy at positive price if they have the chance to buy more units than they need at zero price. A fortiori, this is true when the price offered for a large number of units is negative.

Again, the only difference between rebates and explicit discrimination is given by the presence of the self-selection constraints under the former scheme. Under two-part tariffs, these self-selection constraints lead to the well-known usage price distortions: at the miscoordination equilibrium, for instance, the incumbent will charge an above-cost unit price to the small buyers, thus reducing the quantity threshold, and making it less attractive for the large buyer to behave as a small buyer. This allows the incumbent to extract more rent from the large buyer without violating his self-selection constraint.

Unlike the linear-price case, under two-part tariffs, firms will never want to ration buyers when they can explicitly discriminate among them, and so there is no sense in which rebates are superior relative to explicit discrimination. Instead, whenever they lead to usage price distortions, rebates reduce total surplus, and hence the rent that can be appropriated by firms. Thus, under two-part tariffs, rebates are always less aggressive (and therefore less exclusionary) than explicit discrimination.

5.3 Full Discrimination

In this section, we consider the case where firms can discriminate among buyers of the same type. To see why this reinforces the exclusionary potential of discriminatory pricing, consider the following example: there are $m = 5$ small buyers, and to reach the minimum size, the entrant must serve at least $m = 4$ of these small buyers, plus of course the large buyer. Then, to break an entry equilibrium, it is sufficient for the incumbent to "steal" two of the small buyers: this leaves the entrant with only 3 of them, and so it will fall short of the minimum size. This in turn would allow the incumbent to charge the monopoly price not only to the large buyer, but also to the 3 small buyers who are forced to switch to the incumbent once the first two left the entrant. Now, this means that the price offers the incumbent can make to the first two will be a lot more generous than if it had to simultaneously steal all five small buyers, with only the revenue from the large buyer left to compensate for the losses made on the small buyers. Thus, under full discrimination, the incumbent's price offers to the small buyers are more aggressive than under "explicit discrimination" (as discussed in this paper), i.e. when buyers of the same size must be offered the same price. Note that the entrant will have to offer the incumbent's lowest price to all 5 small buyers, and therefore the efficiency threshold for entry equilibria to exist will be even higher than under explicit discrimination. This confirms again the trade-off between exclusionary potential and lower prices given market structure which was emphasized in this paper.

6 Concluding remarks

The purpose of this paper was to demonstrate the exclusionary potential of rebate tariffs in the presence of network externalities. Our findings are particularly interesting insofar as, in our model, the entrant is in a fairly good initial position compared to other papers on exclusionary practices: it does not have to pay any fixed cost to start operating in the industry, entrant and incumbent can approach all buyers simultaneously (i.e. the incumbent has no first-mover advantage in offering contracts to the buyers before the entrant can do so), and the entrant has the same pricing instruments at its disposal.

In the base model, we assume that firms can only charge non-negative linear prices. First of all, we find that exclusionary equilibria exist for all parameter values, and that even monopoly prices can be sustained in these exclusionary equilibria under each price regime (uniform pricing, explicit discrimination, and rebates).

As for entry equilibria, we find that the more aggressive the price regime the smaller the region of the parameter space where they exist: under uniform pricing, entry equilibria always exist, whereas under rebates and explicit price discrimination they exist only if the entrant is sufficiently more efficient than the incumbent (and the condition is the tightest under explicit price discrimination). On the other hand, if we look at regions where entry equilibria exist under all three regimes, we find that consumers would be better off under explicit discrimination, followed by rebates (or implicit discrimination) and finally uniform pricing. This trade-off between exclusionary potential and (for given market structure) lower equilibrium prices is one of the main themes of this paper.

Allowing for subsidies (i.e., negative prices) does not change the main insight of our analysis: more aggressive pricing allows the incumbent to exclude the entrant for an even wider region of parameter values, while reducing even further the highest prices that can be sustained in any entry equilibrium. In addition, the possibility of subsidizing buyers for *using* the product, gives an important tool to the entrant to disrupt miscoordination equilibria. If the gap between incumbent's and entrant's costs is sufficiently large, miscoordination equilibria do not exist, and if they exist they can be sustained only by lower than monopoly prices. Overall, usage subsidies (i) make exclusion most likely, but (ii) given market structure, results in the lowest prices.

Finally, the same trade-off appears again if one allows for full price discrimination (under which the same type of buyer can be offered different prices).

Interesting extensions of our model could be to allow for buyers to compete against each other downstream, to see whether fierce downstream competition may eliminate miscoordination problems (as showed by Fumagalli and Motta (2006) in the context of exclusive dealing). Another issue of interest could be to allow for partial compatibility between I's and E's network, and to introduce compatibility as a strategic choice variable.[25]

In this paper, we have chosen to model scale effects as a demand-side variable, by using network effects and by considering a network's installed base as the incumbency advantage. However, our results would be identical if we assumed

[25] If networks were fully compatible, no issue of miscoordination would arise, but assuming costly interoperability and a demand which rises continuously in the size of the network might render the study of interoperability decisions in our model interesting (see Cremer et al, 2000).

there are scale economies, and that there is a firm which has already paid its sunk costs, as the incumbency advantage.

Consider the following game. At time 1, firms I and E simultaneously set prices (according to the different price regimes, prices can be uniform or differentiated); at time 2, all buyers decide which firm they want to buy from and make firm orders; at time 3, firm E decides on entry (if it does enter, it has to pay sunk cost $f > 0$); at time 4, payoffs are realized. Like in Section 2, continue to assume that there are m small buyers and 1 large buyer, and let the sunk cost f be large enough so that entry is profitable only if firm E serves the large buyer plus at least one small buyer.[26] With these modifications, results will be of the same nature as those obtained in this paper, and even the calculations will be to a large extent the same. Fumagalli and Motta (2001) set up a model with similar features (economies of scale in production, timing of the game) as those just described and show that miscoordination can indeed prevent an efficient firm from entering the market. However, they have symmetric buyers and do not consider the impact of price discrimination and rebates.

Finally, one may wonder how the existence of switching costs (which play an important role in shaping entry in the real world) would change our model. First of all, consider our basic model with network effects. One simple way to take switching costs into account would be to assume that (equivalently to the 'old' buyers in the basic model) there are buyers who have arbitrarily large switching costs and therefore would never buy from the entrant, and (equivalently to the 'new' buyers in the basic model) buyers who have switching costs σ which are small enough, so that the entrant's effective marginal cost, $c_E + \sigma \equiv \widetilde{c}_E$, is still lower than the incumbent's: $c_E + \sigma < c_I$. Provided that there are both large and small buyers among the latter category of buyers, and after replacing c_E with the effective marginal cost $\widetilde{c}_E$, the analysis would be the same as in our model, and the comparative statics on the switching costs would be straightforward. An increase in switching costs σ would be equivalent to an increase in the marginal cost of the entrant, $\widetilde{c}_E$, and would thus lead to more likely exclusionary equilibria.

Similarly, one could easily incorporate switching costs into a model where scale effects are due to a minimum efficient scale of production: it would be enough to assume again that all buyers have a small (as defined above) switching cost.

Of course, one could find more sophisticated and interesting ways to incorporate switching costs in the analysis, but it is clear that the basic mechanisms illustrated in this paper would still take place and would be exacerbated by the existence of switching costs. Both under consumption externalities and under economies of scale, switching costs would add to the incumbency advantage provided by the installed base and the sunk cost, respectively. Note, however, that in our framework, switching costs alone (i.e. without installed base or sunk cost) would not be sufficient to obtain the results.

[26] Similarly, in our base model, we have chosen the threshold network size $\overline{s}$ so that the minimum scale was reached only if the large and at least one small buyer were consuming the entrant's network product.

References

[1] Aghion, Philippe and Patrick Bolton. 1987. "Contracts as a Barrier to Entry". *American Economic Review,* 77(3): 388 - 401.

[2] Armstrong, Mark and John Vickers. 1993. "Price Discrimination, Competition and Regulation". *Journal of Industrial Economics*, 41(4): 335-359.

[3] Bernheim, B. Douglas and Michael D. Whinston. 1998. "Exclusive Dealing", *Journal of Political Economy,* 106(1): 64 - 103.

[4] Crémer, Jacques, Patrick Rey, and Jean Tirole. 2000. "Connectivity in the Commercial Internet", *Journal of Industrial Economics*, 48(4): 433-472.

[5] Farrell, Joe and Paul Klemperer. Forthcoming. "Coordination and Lock-In: Competition with Switching Costs and Network Effects". Mark Armstrong and Robert Porter (eds.) *Handbook of Industrial Organization Volume 3.* Amsterdam: Elsevier.

[6] Fumagalli, Chiara and Massimo Motta. 2001. "Buyers' Coordination and Entry", CEPR Discussion Paper No. 2908.

[7] Fumagalli, Chiara and Massimo Motta. 2006. "Exclusive Dealing and Entry When Buyers Compete", *American Economic Review*, 96(3): 785-795.

[8] Gyselen, Luc. 2003. "Rebates: Competition on the Merits or Exclusionary Practice?", in Claus-Dieter Ehlermann and Isabela Atanasiu (eds.) *European Competition Law Annual 2003: What Is an Abuse of a Dominant Position?*. Oxford/Portland Oregon: Hart Publishing.

[9] Innes, Robert and Richard J. Sexton. 1993. "Customer coalitions, monopoly price discrimination and generic entry deterrence", *European Economic Review* 37: 1569 - 1597.

[10] Innes, Robert and Richard J. Sexton. 1994. "Strategic Buyers and Exclusionary Contracts". *American Economic Review,* 84(3): 566 - 584.

[11] Kobayashi, Bruce H. 2005. "The Economics of Loyalty Discounts and Antitrust Law in the United States". *Competition Policy International*, 1(115): Autumn 2005.

[12] Rasmusen, Eric B., J. Mark Ramseyer and John S. Wiley. 1991. "Naked Exclusion". *American Economic Review,* 81(5): 1137 - 45.

[13] Rey, Patrick, Jordi Gual, Martin Hellwig, Anne Perrot, Michele Polo, Klaus Schmidt and Rune Stenbacka. 2005. "An Economic Approach to Article 82". Brussels: European Commission, Competition Directorate General.

[14] Segal, Ilya R. and Michael D. Whinston. 2000. "Naked Exclusion: Comment". *American Economic Review,* 90(1): 296 - 309.

[15] Shapiro, Carl. 1999. "Exclusivity in network industries". *George Mason Law Review*, 7(3): 1-11.

[16] Stole, Lars. Forthcoming. "Price Discrimination and Imperfect Competition." in Mark Armstrong and Robert Porter (eds.) *Handbook of Industrial Organization Volume 3*. Amsterdam: Elsevier.

[17] Thisse, Jacques-François, and Xavier Vives. 1988. "On the Strategic Choice of Spatial Price Policy". *American Economic Review*, 78(1): 122 - 37.

[18] Vickers, John. 2005. "The Abuse of Market Power". *Economic Journal,* 115(504): F244-261.

7 Appendix A - Proofs

Proof of Proposition 1:

(i) With all buyers buying from E at $p_E = c_I$, total demand is $mq_E^s(p_E) + q_E^l(p_E) = 1 \geq \bar{s}$, and so E will reach the minimum size. Thus, E's product has the same value to the buyers as I's, and it sells at the same price. Given that buyers coordinate on the entrant whenever E's offer is at least as good as I's, no buyer has an incentive to deviate and buy from I instead. I will not want to deviate either: To attract the buyers, I would have to set a price $p_I < c_I$, i.e. sell at a loss; and increasing p_I above c_I will not attract any buyers. E has no incentive to change its price either: increasing p_E would imply losing the buyers to I, and decreasing p_E will just reduce profits.

There can be no equilibrium where E serves all buyers at a price $p_E > c_I$: In this case, I could profitably undercut E, and all buyers would switch to I.

(ii) Suppose that all buyers buy from I. Then, recall that $\bar{s} > \max\{1-k, k\}$, implying that none of the individual buyers alone is sufficient for E to reach the minimum size. Thus, E's product has zero value for any single buyer, and so no buyer will want to deviate and buy from E, even if p_E were strictly lower than p_I. I sets $p_I = p_I^m$, which is the highest among all prices under which buyers will miscoordinate on the incumbent (at a price strictly above the prohibitive price, buyers would stop buying altogether). Thus, I has no incentive to increase or decrease its price. Since buyers will not switch to E even if the price difference between the two firms is maximal, i.e. even if E charges $p_E = c_E$, E has no incentive to decrease its price either.□

Proof of Proposition 3:

First, the best offer the incumbent can make to the small buyers is given by the solution of the following program:

$$\begin{array}{l} \max_{p_I^s, p_I^l} CS^s(p_I^s) = \frac{k(1-p_I^s)}{m}, \quad \text{s.to:} \\ \text{(i) } (p_I^s - c_I)k + (p_I^l - c_I)(1-k) \geq 0 \\ \text{(ii) } p_I^s \in [0,1], p_I^l \in [0,1], \end{array} \tag{10}$$

where (i) is the profitability constraint of the incumbent.

Next, note that the best offer the incumbent can make to the large buyer is given by the solution of:

$$\begin{array}{l} \max_{p_I^s, p_I^l} CS^l(p_I^l) = (1-k)(1-p_I^l), \quad \text{s.to:} \\ \text{(i) } (p_I^s - c_I)k + (p_I^l - c_I)(1-k) \geq 0 \\ \text{(ii) } p_I^s \in [0,1], p_I^l \in [0,1] \end{array} \tag{11}$$

We see that the best offer the incumbent can make to the small buyers is to set $p_I^l = p_I^M = 1$ and lower p_I^s as much as possible while still satisfying the profitability constraint (i); likewise, the best offer to the large buyer is obtained by setting $p_I^s = p_I^M$ and lowering p_I^l as much as allowed by (i).

The offer (p_I^s, p_I^M) to the small buyers is feasible as long as the incumbent breaks even (i.e., constraint (i) must be satisfied):

$$m\frac{k}{m}(-c_I + p_I^s) + (1-k)(1-c_I) \geq 0, \tag{12}$$

The offer (p_I^M, p_I^l) to the large buyer is feasible as long as:

$$(1-k)(-c_I + p_I^l) + m\frac{k}{m}(1-c_I) \geq 0. \tag{13}$$

Call $\widehat{p}_I^s$ and $\widehat{p}_I^l$ respectively the prices that solve the equations associated with inequalities (12) and (13) above. The lowest possible deviation prices of the incumbent are identified by respectively:

$$p_I^s = \max(\widehat{p}_I^s, 0) \quad \text{and} \quad p_I^l = \max(\widehat{p}_I^l, 0),$$

since we limit attention to non-negative prices (see below for the case where prices can be negative).

The entrant can match the incumbent's deviations if it is able to offer (weakly) more surplus to the buyers, while still making profits. In other words, the entrant will be able to profitably enter at equilibrium if it can set prices (p_E^s, p_E^l) such that:

$$CS^s(p_E^s) = \frac{k}{m}(1-p_E^s) \geq CS^s(p_I^s) = \frac{k}{m}(1-p_I^s) \tag{14}$$

$$CS^l(p_E^l) = (1-k)(1-p_E^l) \geq CS^l(p_I^l) = (1-k)(1-p_I^l) \tag{15}$$

$$\pi_E(p_E^s, p_E^l) \geq 0. \tag{16}$$

Optimality requires the entrant offering the highest among all prices that satisfy these conditions, so at equilibrium they will be binding:

$$p_E^s = p_I^s;\ p_E^l = p_I^l,$$

from which condition (16) becomes:

$$k(p_E^s - c_E) + (1-k)(p_E^l - c_E) \geq 0,$$

or:

$$\pi_E(p_I^s, p_I^l) : k(p_I^s - c_E) + (1-k)(p_I^l - c_E) \geq 0. \tag{17}$$

We therefore have to find $\left(p_I^s, p_I^l\right)$. By solving the equalities associated with (12) and (13) above, we obtain:

$$\widehat{p}_I^s = \frac{c_I - (1-k)}{k}; \quad \widehat{p}_I^l = \frac{c_I - k}{1-k}.$$

Note that $\widehat{p}_I^s < c_I$ and $\widehat{p}_I^l < c_I$; also:

$$\widehat{p}_I^s \geq 0 \quad \text{if} \quad c_I \geq 1-k; \text{ and } \widehat{p}_I^l \geq 0 \quad \text{if} \quad c_I \geq k.$$

Therefore, the incumbent's optimal offer will be:

$$p_I^s = \begin{cases} \widehat{p}_I^s & \text{if } c_I \geq 1-k \\ 0 & \text{if } c_I < 1-k \end{cases} \quad p_I^l = \begin{cases} \widehat{p}_I^l & \text{if } c_I \geq k \\ 0 & \text{if } c_I < k \end{cases}$$

This identifies four regions, and for each of them we have to verify whether (16) holds or not:

$$\begin{array}{rcl} \text{if } c_I & \in & [1-k, k] \text{ and } k \geq 1/2\text{:} \quad \pi_E(\widehat{p}_I^s, 0) \geq 0 \\ \text{if } c_I & \in & [k, 1-k] \text{ and } k < 1/2\text{:} \quad \pi_E(0, \widehat{p}_I^l) \geq 0 \\ \text{if } c_I < \min & & \{k, 1-k\} : \pi_E(0,0) \geq 0 \\ else & : & \pi_E(\widehat{p}_I^s, \widehat{p}_I^l) \geq 0 \end{array}$$

After replacing, we can then find that:
(1) $\pi_E(\widehat{p}_I^s, \widehat{p}_I^l) = k\left(\frac{c_I-(1-k)}{k} - c_E\right) + (1-k)\left(\frac{c_I-k}{1-k} - c_E\right) \geq 0$,
which is satisfied for:

$$c_I \geq \frac{1+c_E}{2} \equiv \overline{c}_{I1}$$

(2) $\pi_E(0, \widehat{p}_I^l) = -c_E k + (1-k)\left(\frac{c_I-k}{1-k} - c_E\right) \geq 0$
which holds for:

$$c_I \geq k + c_E \equiv \overline{c}_{I2}$$

(3) $\pi_E(\widehat{p}_I^s, 0) = k\left(\frac{c_I-(1-k)}{k} - c_E\right) - c_E(1-k) \geq 0$
which holds for:

$$c_I \geq 1 + c_E - k \equiv \overline{c}_{I3}.$$

(4) $\pi_E(0,0) = -c_E \geq 0$,
which never holds, apart from the knife-edge case where $c_E = 0$. (Since prices cannot go below zero in this basic model, the best that the incumbent can offer to buyers is to give them the good for free; but when $c_E = 0$, the entrant could match that offer without making losses, and entry equilibria would always exist. Clearly, though, this is a very special case.)

Finally, straightforward algebra shows that if $c_I \geq \max\{k, 1-k\}$, so that threshold $\overline{c}_{I1} = \frac{1+c_E}{2}$ applies, we have that $\overline{c}_{I1} = \min\{\overline{c}_{I1}, \overline{c}_{I2}, \overline{c}_{I3}\}$, and the analogous relation holds for the other two threshold values of c_I: in the parameter region where $\overline{c}_{Ii}$ applies, $\overline{c}_{Ii} = \min\{\overline{c}_{I1}, \overline{c}_{I2}, \overline{c}_{I3}\}$.□

Proof of Proposition 5:
In order to find the conditions under which entry equilibria exist, we proceed in three steps.

First, we look for the best possible offer p_I^s that the incumbent can make to the small buyers; second, we look for the best possible offer p_I^l that the incumbent can make to the large buyer; third, we see whether the entrant is able to make a profitable offer (p^s_E,p^l_E, $\bar{q}_E$) to the small and the large buyer such that they are at least as well off as if they bought from the incumbent.

The incumbent's best offer to the small buyer, $\left(\widetilde{p}^s_I, \overline{p}^l_I, \bar{q}^s_I\right)$, solves Program (18):

$$\begin{array}{ll} \max_{p^s_I, p^l_I, \bar{q}^s_I} CS^s(p^s_I, \bar{q}^s_I) = (1-p^s_I)\frac{k}{m}, \quad \text{s.to:} & \\ \text{(i) } (p^s_I - c_I)k + (p^l_I - c_I)(1-k) \geq 0 & \\ \text{(ii) } p^s_I \in [0,1], p^l_I \in [0,1], \bar{q}^s_I \leq \frac{k}{m} & \\ \text{(iii) } CS^l\left(p^l_I, 1-k\right) = (1-p^l_I)(1-k) \geq \bar{q}^s_I\left(1-p^s_I\right), & \\ \text{(iv) } CS^s\left(p^s_I, \frac{k}{m}\right) = (1-p^s_I)\frac{k}{m} \geq \left(\bar{q}^s_I + \varepsilon\right)\left(1-p^l_I\right) & \end{array} \tag{18}$$

where $k \in \left[0, \frac{m}{m+1}\right]$, and p^s_I applies to all purchases $q^j \leq \bar{q}^s_I$, while p^l_I applies whenever $q^j > \bar{q}^s_I$.

Constraints (i) to (iv) fully determine the solution. Wlog, we can set $\bar{q}^s_I = \frac{k}{m}$, and search for the prices $\left(p^s_I, p^l_I\right)$ that satisfy the remaining constraints. Note that the incumbent would like to set p^s_I as low as possible, while charging the highest possible price to the large buyer. However, I can no longer set $p^l_I = 1$ (as under explicit discrimination), because at this price the large buyer is left with zero surplus, and so his self-sorting condition can never be satisfied (he would prefer to buy even a very small quantity, k/m, at a price $p^s_I < 1$, than a large quantity $1-k$ at the prohibitive price). Hence, the large buyer's self-selection constraint will always be binding under any solution of program (18):

$$p^l_I = 1 - \frac{k(1-p^s_I)}{m(1-k)}$$

In order to satisfy the profitability constraint, the following must hold as well:

$$p^l_I \geq \frac{c_I - kp^s_I}{1-k}$$

At any solution to the program, we must have $p^s_I < p^l_I$, so the small buyers' self-sorting condition, constraint (iv), is never binding. Then, either the break-even constraint (i), or the non-negativity constraint on p^s_I (ii), are binding along with constraint (iii). This gives us the following solutions of the program:

$$\begin{array}{rcl} \widetilde{p}^s_I & = & 1 - \dfrac{m(1-c_I)}{k(m+1)}; \quad \overline{p}^l_I = 1 - \dfrac{(1-c_I)}{(1-k)(m+1)}, \quad \text{if } c_I \geq 1-k-k/m \\ \widetilde{p}^s_I & = & 0; \quad \overline{p}^l_I = 1 - \dfrac{k}{m(1-k)}, \quad \text{if } c_I < 1-k-k/m. \end{array}$$

The incumbent's best offer to the large buyer $(\overline{p}^s_I, \widetilde{p}^l_I, \bar{q}^l_I)$ solves Program (19):

$$
\begin{array}{l}
\max_{p_I^s, p_I^l, \bar{q}_I^l} CS^l(p_I^l) = (1-k)(1-p_I^l), \quad \text{s.to:} \\
\text{(i) } (p_I^s - c_I)k + (p_I^l - c_I)(1-k) \geq 0 \\
\text{(ii) } p_I^s \in [0,1], p_I^l \in [0,1], \bar{q}_I^l \leq (1-k) \\
\text{(iii) } CS^l\left(p_I^l, 1-k\right) \geq CS^l\left(p_I^s, \bar{q}_I^l - \varepsilon\right), \\
\text{(iv) } CS_I^s\left(p_I^s\right) = \frac{k}{m}\left(1-p_I^s\right) \geq \frac{k}{m} - p_I^l \bar{q}_I^l
\end{array} \tag{19}
$$

where $k \in \left[0, \frac{m}{m+1}\right]$, and p_I^s applies to all $q_I^j < \bar{q}_I^l$, while p_I^l applies to all $q_I^j \geq \bar{q}_I^l$. Note that the two quantity thresholds $\bar{q}_I^s$ and $\bar{q}_I^l$ are indexed by s and l to make it clear to which of the two programs they belong.

Now, we can set $\bar{q}_I^l = (1-k)$ wlog. The incumbent would like to set p_I^l as low as possible. (But recall that $p_I^l = 0$ can never satisfy the self-selection constraint of the small buyers, who would always prefer to buy a quantity $(1-k)$ at zero price - and throw away $1-k-k/m$ units - than a smaller quantity k/m at positive price.) In order to satisfy the profitability constraint and the small buyers' self-selection constraint respectively, the following must hold:

$$
p_I^s \geq \frac{c_I - (1-k)p_I^l}{k},
$$

and respectively:

$$
p_I^s \leq \frac{m(1-k)p_I^l}{k}.
$$

Under any solution to Program (19), we have that $p_I^s > p_I^l$ (the reason being analogous to Program (18). It follows immediately that constraint (iii) will never be binding at any solution to Program (19). Then, either self-sorting condition (iv), or the $p_I^s \leq 1$ constraint (ii), are binding along with the break-even constraint (i). This gives us the following solutions of the program:

$$
\begin{aligned}
\widetilde{p}_I^l &= \frac{c_I}{(1-k)(m+1)}; \quad \overline{p}_I^s = \frac{mc_I}{k(m+1)}, \quad \text{if } c_I < \frac{k(1+m)}{m} \\
\widetilde{p}_I^l &= \frac{c_I - k}{1-k}; \quad \overline{p}_I^s = 1, \quad \text{if } c_I \geq \frac{k(1+m)}{m}.
\end{aligned}
$$

We can now summarize the incumbent's optimal offers as follows:

$$
\widetilde{p}_I^s = \begin{cases} 1 - \frac{m(1-c_I)}{k(m+1)} & \text{if } c_I \geq 1-k-k/m \\ 0 & \text{if } c_I < 1-k-k/m \end{cases} \quad \widetilde{p}_I^l = \begin{cases} \frac{c_I - k}{1-k} & \text{if } c_I \geq \frac{k(1+m)}{m} \\ \frac{c_I}{(1-k)(m+1)} & \text{if } c_I < \frac{k(1+m)}{m} \end{cases}
$$

Again, these are the highest prices that the entrant can charge in any entry equilibrium. For entry to be feasible, $\left(\widetilde{p}_I^s, \widetilde{p}_I^l\right)$ must be high enough to allow the entrant to break even. The functions $\left(\widetilde{p}_I^s, \widetilde{p}_I^l\right)$ identifies four regions, and for each of them we have to verify whether (16) holds or not:

(i) if $c_I \in \left[1-k-k/m, \frac{k(1+m)}{m}\right]$ and $k \geq \frac{m}{2(1+m)}$: $\quad \pi_E(1-\frac{m(1-c_I)}{k(m+1)}, \frac{c_I}{(1-k)(m+1)}) \geq 0$
(ii) if $c_I \in \left[\frac{k(1+m)}{m}, 1-k-k/m\right]$ and $k < \frac{m}{2(1+m)}$: $\quad \pi_E(0, \frac{c_I-k}{1-k}) \geq 0$
(iii) if $c_I < \min\left\{\frac{k(1+m)}{m}, 1-k-k/m\right\}$: $\pi_E(0, \frac{c_I}{(1-k)(m+1)}) \geq 0$
(iv) else: $\pi_E(1-\frac{m(1-c_I)}{k(m+1)}, \frac{c_I-k}{1-k}) \geq 0$

After replacing, we can then find that:
(i) $\pi_E(1-\frac{m(1-c_I)}{k(m+1)}, \frac{c_I}{(1-k)(m+1)}) \geq 0$ holds for:

$$c_I \geq \frac{m}{1+m} + c_E - k$$

(ii) $\pi_E(0, \frac{c_I-k}{1-k}) \geq 0$ holds for:

$$c_I \geq k + c_E$$

(iii) $\pi_E(0, \frac{c_I}{(1-k)(m+1)}) \geq 0$ holds for:

$$c_I \geq c_E(1+m)$$

(iv) $\pi_E(1-\frac{m(1-c_I)}{k(m+1)}, \frac{c_I-k}{1-k}) \geq 0$ is satisfied for:

$$c_I \geq \frac{m+(1+m)c_E}{1+2m}$$

If $c_E < \frac{1}{2(m+1)}$, then we have that $c_E(1+m) < \frac{m+(1+m)c_E}{1+2m} < \frac{1}{2}$. Tedious algebra shows that in this case, $c_I \geq \frac{m+(1+m)c_E}{1+2m}$ is redundant, and that each of the remaining thresholds is the minimum of all thresholds in the parameter region where it applies. Conversely, if $c_E \geq \frac{1}{2(m+1)}$, then we have that $\frac{m+(1+m)c_E}{1+2m} < c_E(1+m)$ and $c_E(1+m) \geq \frac{1}{2}$. In this case, $c_I \geq c_E(1+m)$ is redundant, and each of the remaining thresholds is the minimum of all thresholds in the parameter region where it applies.

Finally, note that - unlike the case of explicit discrimination - *in principle* it may not be enough if the entrant simply matches the incumbent's offer, because at prices $p_E^s = \widetilde{p}_I^s$ and $p_E^l = \widetilde{p}_I^l$, it may be that one of the self-selection constraints is violated. In other words, the self-selection conditions on the one hand affect the incumbent by obliging it to set (weakly) higher prices but on the other hand also affect the entrant by obliging it to set (weakly) higher prices as well. Thus, we have to verify if each of the possible price pairs identified by $\left(\widetilde{p}_I^s, \widetilde{p}_I^l\right)$ will also satisfy the large and small buyers' self-sorting conditions, so that they can actually sustain an entry equilibrium:

(i) Suppose that $\widetilde{p}_I^s < \widetilde{p}_I^l$, and that $\bar{q}_E = \frac{k}{m}$. Then, only the large buyer's self-selection constraint could be violated (but not the small buyers'). But recall that $\widetilde{p}_I^s$ derives from Program (18), i.e. the incumbent's best offer to the *small* buyers, and that this offer $\left(\widetilde{p}_I^s, \overline{p}_I^l, \overline{q}_I^s\right)$ satisfies the large buyer's self-selection constraint by construction. Now, we also have that

$$\overline{p}_I^l > \widetilde{p}_I^l = p_E^l$$

i.e. the price that the large buyer is charged under solution $\left(\widetilde{p}_I^s, \overline{p}_I^l, \overline{q}_I^s\right)$ is always higher than the price under I's best offer to the *large* buyer, $\widetilde{p}_I^l$, which is also the highest price that the entrant can charge the large buyer. Otherwise, $(\overline{p}_I^s, \widetilde{p}_I^l, \overline{q}_I^l)$ cannot be a solution to Program (19). But that implies that $\left(\widetilde{p}_I^s, \widetilde{p}_I^l\right)$ must also satisfy the large buyer's self-selection condition.

(ii) For the complementary case $\widetilde{p}_I^s > \widetilde{p}_I^l$, and $\overline{q}_E = 1 - k$, only the small buyers' self-selection constraint could be violated. But now, we have that $(\overline{p}_I^s, \widetilde{p}_I^l, \overline{q}_I^l)$, which solves Program (19), satisfies the small buyers' self-selection constraint by construction, and that $\overline{p}_I^s > \widetilde{p}_I^s$, so that $\left(\widetilde{p}_I^s, \widetilde{p}_I^l\right)$ must satisfy the small buyers' self-selection constraint as well.□

Proof of Corollary 6:

Under explicit discrimination, the lower bound on c_I for entry equilibria to exist is $\min\left\{\frac{1+c_E}{2}, k + c_E, 1 - k + c_E\right\}$. Now, if $c_E < \frac{1}{2(m+1)}$, the corresponding condition under rebates reads $c_I \geq \min\left\{c_E(1+m), k + c_E, \frac{m}{1+m} + c_E - k\right\}$. Comparing the components of the two sets, we see that the second component is the same, $k + c_E = k + c_E$. The third component is lower under rebates, $\frac{m}{1+m} + c_E - k < 1 - k + c_E$. Finally, $c_E < \frac{1}{2(m+1)}$ implies that $c_E(1+m) < \frac{1+c_E}{2}$, i.e. the first component is lower under rebates as well. If instead $c_E \geq \frac{1}{2(m+1)}$, the first component under rebates is $\frac{m+(1+m)c_E}{1+2m}$, which is always smaller than $\frac{1+c_E}{2}$. Thus, we can conclude that the parameter space for which entry equilibria exist under rebates fully includes the corresponding parameter space under explicit discrimination.□

Proof of Proposition 7:

To make it a dominant strategy for the small buyers to buy from E, E must offer a price p_E^s that yields a (weakly) higher net surplus as I's offer to the small buyers:

$$-p_E^s \frac{k}{m} \geq \frac{k}{m}(1 - p_I^s)$$

We see immediately that $p_E^s \leq -(1 - p_I^s) < 0$ (E subsidizes small buyers' consumption of its product). If the small buyers consume E's product for sure, then the large buyer will switch to E whenever $p_E^l \leq p_I^l$. Will E be able to break-even under this optimal deviation? Inserting $p_E^s = -(1 - p_I^s)$ and $p_E^l = p_I^l$ into the profit function we have that

$$-k(1 - p_I^s) - c_E + p_I^l (1 - k) \geq 0$$

Rearranging this break-even constraint, we obtain

$$p_I^s \geq 1 - \frac{1}{k}\left[p_I^l (1 - k) - c_E\right]$$

Looking at it from the point of view of the incumbent, this means that given p_I^l, p_I^s must *not* exceed $1 - \frac{1}{k}\left[p_I^l(1-k) - c_E\right]$, or else I becomes vulnerable to the deviation described above. Hence, I's problem reads

$\max_{p_I^s, p_I^l} \pi_I = (p_I^s - c_I)\,k + \left(p_I^l - c_I\right)(1-k)$
s.t. (i) $p_I^l \le 1$
(ii) $p_I^s \le \min\left\{1 - \frac{1}{k}\left[p_I^l(1-k) - c_E\right], 1\right\}$

If $(1-k) - c_E < 0$, the problem is trivially solved by

$$p_I^s = p_I^l = 1$$

If instead $(1-k) - c_E \ge 0$, we can insert $p_I^s = 1 - \frac{1}{k}\left[p_I^l(1-k) - c_E\right]$ into the objective function to see that the choice variables drop out, so that the objective function reduces to:

$$\pi_I = k + c_E - c_I$$

Thus, I will be able to break even iff

$$c_I \le k + c_E$$

(i) Let $c_E \le 1-k$. If the incumbent raises p_I^l above 1 (the prohibitive price), the large buyer will not buy anything. Reducing p_I^l below 1 would only reduce profits. Note that $c_E \le 1-k$ implies that $p_I^s \le 1$. If the incumbent raises p_I^s above $1 - \frac{1}{k}[1-k-c_E]$, the small buyers will find it individually rational to buy from E:

$$-p_E^s \frac{k}{m} = \frac{1}{m}(1-k-c_E) > \frac{k}{m}(1-\tilde{p}_I^s)$$

Reducing p_I^s below $1 - \frac{1}{k}[1-k-c_E]$ would only reduce profits.

Under this equilibrium, all buyers buy from the incumbent, so that the entrant's profits are zero. We argued before that the entrant's optimal deviation is to set $p_E^l = p_I^l = 1$, and to reduce p_E^s below $-\frac{1-k-c_E}{k}$ to attract the small buyers. But such an offer would violate the entrant's break-even condition:

$$\tilde{p}_E^s k - c_E + p_I^l(1-k) < -k(1-p_I^s) - c_E + p_I^l(1-k) = 0$$

The entrant has no incentive either to increase p_E^s above $-\frac{1-k-c_E}{k}$, as it does not make any sales in equilibrium.

Finally, no individual buyer has any incentive to deviate and buy from the entrant instead: each of the small buyers is indifferent between I's and E's offer, and the large buyer strictly prefers to buy from I than being the only buyer to buy from E.

Can there be any other miscoordination equilibrium, where I charges a lower p_I^l, namely $p_I^l < 1$, and an accordingly higher $p_I^s = 1 - \frac{1}{k}\left[p_I^l(1-k) - c_E\right]$? No, because no matter which prices E sets, I would want to increase p_I^l to 1 without changing p_I^s, thereby increasing profits without losing the large buyer to E. Therefore, such a price pair cannot sustain an equilibrium.

(ii) Let $c_E > 1-k$, and let $p_I^s = p_I^l = 1$. Clearly, the incumbent has no incentive to change its prices. Recall that under E's optimal deviation, E's break-even condition reads

$$-k(1-p_I^s) - c_E + p_I^l\,(1-k) \geq 0$$

Inserting $p_I^s = p_I^l = 1$, we get

$$-c_E + (1-k) \geq 0$$

This condition is always violated if $c_E > 1-k$. In other words, business-stealing by the entrant is impossible even if the incumbent charges monopoly prices to both groups of buyers. Therefore, the entrant is indifferent among all the prices it can set such that I serves the buyers: $p_E^s = p_E^l = 1$ dominates all others. The rest of the proof is analogous to the reasoning above.□

Proof of Proposition 8:
The best offer the incumbent can make to the small buyers is given by the solution of the following program:

$$\begin{array}{l} \max_{p_I^s, p_I^l} CS^s(p_I^s) = \frac{k}{m}(1-p_I^s), \quad \text{s.to:} \\ \text{(i) } (p_I^s - c_I)k + (p_I^l - c_I)(1-k) \geq 0 \\ \text{(ii) } p_I^s \leq 1, p_I^l \leq 1. \end{array} \tag{20}$$

The best offer the incumbent can make to the large buyer is given by the solution of:

$$\begin{array}{l} \max_{p_I^s, p_I^l} CS^l(p_I^l) = (1-k)\,(1-p_I^l), \quad \text{s.to:} \\ \text{(i) } (p_I^s - c_I)k + (p_I^l - c_I)(1-k) \geq 0 \\ \text{(ii) } p_I^s \leq 1, p_I^l \leq 1. \end{array} \tag{21}$$

By following the same steps as in Section 3.2 one can check that the incumbent's best offers are

$$\widehat{p}_I^s = \frac{c_I - (1-k)}{k}; \quad \widehat{p}_I^l = \frac{c_I - k}{1-k}.$$

An entry equilibrium will exist only if the entrant is able to profitably match simultaneously both best offers, i.e. $p_E^s = \widehat{p}_I^s$, and $p_E^l = \widehat{p}_I^l$. Therefore, such an equilibrium exists if and only if:

$$\pi_E(\widehat{p}_I^s, \widehat{p}_I^l) = k\left(\frac{c_I - (1-k)}{k} - c_E\right) + (1-k)\left(\frac{c_I - k}{1-k} - c_E\right) \geq 0,$$

which is satisfied for:

$$c_I \geq \frac{1 + c_E}{2}.$$

□

An empirical analysis of broadband diffusion in the New Member States of the European Union

María Rosalía Vicente (University of Oviedo)*

Ana Jesús López (University of Oviedo)

Abstract

Recent studies stress the crucial role of broadband diffusion to enhance economic growth and performance. Therefore, the analysis of the factors shaping its diffusion becomes a matter of special importance. In this context, this paper examines the impact of several socioeconomic factors on the likelihood of broadband subscription, using a bivariate probit model, and making use of a unique data survey among Eastern European households. This is an area where broadband diffusion is just starting to take off and there is a lack of empirical evidence. In this context, a first interesting point is the fact that certain demographic variables (income and education) appear to be more correlated with Internet access as such than with the type of access. Moreover, the substantial differences in cross-country penetration rates are mainly explained by investments in information and communication technologies together with the level of competition in telecommunications markets.

Key Words: broadband, Internet, telecommunications competition, New Member States, European Union, bivariate probit

* Corresponding author: María Rosalía Vicente, Applied Economics, University of Oviedo, Campus del Cristo s/n, Oviedo - Asturias 33006-Spain.

Tel. +34 985 10 50 53

Fax. +34 985 10 50 50

Email: mrosalia@uniovi.es

I. Introduction

In recent years, the access to high-speed Internet through broadband connections has been recognized to open up huge possibilities for the development of new interactive and advanced applications, with potential benefits for businesses, the public sector, and consumers (Gillet et al., 2006). In particular, broadband technologies enable new services, and enhance the capacity and quality of the existing ones such as business transactions, education, health care, and government services. Therefore, broadband availability and affordability has become an element of strategic importance to all countries.

In particular, the European Commission (2006) highlights that "widespread broadband access is a key condition for the development of modern economies and is an important aspect of the Lisbon agenda. The European Union must step up its efforts to encourage take-up of broadband services and stimulate further deployment, in particular in the less developed areas of the Union". In this line, the Riga Ministerial Declaration, signed by ministers from European Union Member States in June 2006, set the specific target for broadband coverage in Europe to reach at least 90% by 2010.

Nonetheless, there is little empirical evidence on broadband diffusion at the European Union level, especially with regard to the New Member States. Within this context, the aim of this paper is to remedy this deficiency, trying to throw some light on the drivers of broadband adoption in the New Member States of the European Union. This work may come at an opportune time: since the broadband market is just starting to develop in these countries, our empirical evidence may help in the design of adequate strategies.

II. Broadband drivers

In the last few years broadband access has recorded an outstanding growth. Thus, broadband subscribers in the OECD area reached 137 million by mid-year 2005, adding 18 million broadband subscribers since January (OECD, 2006). In the case of the European Union, broadband access lines have almost doubled between 2004 and 2005 (European Commission, 2006).

In spite of such remarkable growth, recent figures show large differences both between and within countries. Hence, countries with higher income and higher level of competition in the telecommunications market tend to show higher rates of broadband access (Cava-Ferreruela and Alabau-Muñoz, 2006; Wallsten, 2006). As the OECD (2004) points out effective competition is one of the principles that have been demonstrated to assist the development of broadband markets.

Moreover broadband diffusion has affected socio-economic and demographic groups to different extents (Madden et al., 1996; Madden and Simpson, 1997; Savage and Waldman, 2005). In fact, income, educational attainment, and age are acknowledged to be the most important determinants of broadband diffusion: low income groups are less likely to purchase broadband services because their price represents a larger proportion of their budget; the lack of higher education reduces the awareness of benefits from greater access to information and the ability to manage it; while the elderly are less familiar with information technologies.

In this sense, a pioneer study by Madden et al. (1996) show that the less well educated and the elderly are the less willing to subscribe to broadband services. Madden and Simpson’s (1997) estimates indicate that household income and the installation fees are the principal determinants of take-up, together with age and household size.

Moreover, Savage and Waldman (2005) show that broadband users tend to have more years of online experience compared to dial up users.

Nonetheless, the decision to take up broadband is rather a complicated matter involving end users in conjunction with the opportunity costs of time (Rappoport et al., 2002) and the availability of services. Varian (2002) estimated the willingness to pay for various levels of speed connections across a sample of American households and found that the problem with broadband were not access but applications: Internet users were not willing to pay a premium for high speed given the applications available at that moment.

All this research on broadband diffusion refers mostly to the United States, Australia, and OECD countries; meanwhile the references to the European Union are limited and even scarcer to the New Member States. As said before, this paper will try to remedy this deficiency.

III. Model specification and explanatory variables

In the economic analysis of broadband a standard neoclassical utility maximization model framework is adopted, whereby the demand for broadband is determined by the size of consumer surplus associated with high-speed access and its costs. In this context, we postulate a model of access choice between broadband and narrowband access.

Assume that broadband access is determined by an unobserved latent variable, $Y_i^* = X_i^{'}\beta + u_i$ (1), for household i, i=1, …, N. Only Y_i is observed, which equals 1 if $Y_i^* \geq 0$, implying that household i chooses to take up broadband; Y_i^* equals zero if household i chooses to take up narrowband. X_i is a vector of family and geographical area characteristics, and u_i is the error term. Assuming that u_i is normally distributed, the data are described by the following probit model:

$\text{Prob}(Y_i = 1) = \phi(X_i'\beta)$ (2),

where ϕ is the cumulative normal distribution function.

The choice between broadband and narrowband only makes sense for those households that already have access. Since our data comes from a sample of the full population, if we restrict our analysis only to those with Internet access, sample selection bias will be introduced. To avoid this, a two-stage estimation procedure is adopted (Heckman, 1979; Greene, 1992, Yoo, 2004): we estimate a first equation to determine whether a household is connected or not, and a second equation to explain the type of access, given that the household is connected.

Similar to (1), assume that home Internet access is determined by an unobserved latent variable,

$C_i^* = Z_i'\gamma + \varepsilon_i$ (3),

where only C_i equal to 0 or 1 is observed, Z_i is a vector of explanatory variables and ε_i is the error term. Lastly, we assume that the random errors u_i and ε_i follow a bivariate standard normal distribution with correlation $\rho_{\varepsilon,u}$. The bivariate probit model is appropriate when $\rho_{\varepsilon,u} \neq 0$.

Within this framework, we try to identify the factors that shape the decision of households to have Internet connection and, in particular, broadband access. The factors examined include household income, age, education, and family size. Moreover, we have included some variables related to household technological attributes in order to check if there is some evidence of “technology clusters”, in the sense that the familiarity with one technology increases the receptivity towards others. Therefore, we have considered the presence of fixed or mobile phones at home, as well as a fax.

As already noted, large differences in the uptake of information and communication technologies persist across countries, especially with regard to broadband. Cost differentials and structural differences are among the factors explaining such disparities. Hence, some macroeconomic variables have been included in order to take account of cross-country differences and, in particular, income per capita (Gross Domestic Product, GDP), a measure of trade openness (imports and exports of goods as a percentage of GDP), a measure of the efforts on research and development activities (R&D expenditure as a percentage of GDP), and some measures of telecommunications costs.

In the absence of data on broadband prices for the area of analysis, we have included some other variables to take account of the situation of the broadband market in those countries. Since broadband and narrowband are substitutes, we have included the prices of narrowband access. Following Goel et al. (2006) we have also included investment in information and communication technologies (ICT, as a percentage of GDP) to factor out the effect of supply that is not reflected in the price. Finally, we have explored the role of competition using as a proxy the number of Internet service providers per 1,000 inhabitants.

IV. Main results

Our study covers the area of Eastern Europe, including nine of the New Member States of the European Union (Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Slovakia, Slovenia and Romania). The micro data used comes from a survey, with 9,379 interviews successfully completed, from the European project SIBIS (Empirica-SIBIS, 2003). Information on macroeconomic variables is derived from Eurostat (2006).

Table 1 presents the results for the first-stage equation on the availability of Internet access at home. As is to be expected, the probability of having Internet at home is positively related to income and negatively to age. Interestingly, results show that only the highest levels of education make a difference. Thus, those households where at least one occupant has a university degree are more likely to have Internet access. We also find that family size positively influences the likelihood of home access.

There is also support for the existence of technology clusters. In particular, the probability of having Internet access is noticeably higher for those households having a fax.

With regard to country characteristics, results confirm that the economic structure matters in order to explain Internet access. We find that income, openness and R&D have a positive significant effect on the likelihood of Internet take up. Interestingly, R&D expenditure comes as the most powerful determinant among these three variables. The effect of telecommunications costs on Internet adoption is also significant, but quite moderate (note that it gets a coefficient of -0.02)[1].

[1] These costs are given by the price of a composite basket including both national fixed and mobile calls, and international calls. We also tried several baskets of Internet access prices (with different levels of usage per month) but they were not significant.

Since the estimated models correctly predict 87% of sample observations, the goodness of fit seems to be quite good.

Table 2 shows the results for broadband access, given that the household is connected. A first interesting point is the fact that certain demographic variables appear to be more correlated with Internet access as such than with the type of access. Thus, once we have controlled for educational attainment in the access equation, it is not statistically significant any more in the broadband equation[2].

Moreover, results show that ICT investment exerts a positive and statistically significant effect on broadband take up, whereas dial-up access cost does not affect it. Finally, it is worth highlighting the role of competition as the most powerful determinant of broadband diffusion. Thus, model estimates show that the number of Internet service providers is associated with a higher probability of broadband subscription. It is notable the size of the effect as it gets a coefficient of 1.87.

Overall our modelling framework in two stages seems to be appropriate according to the results of the likelihood ratio test of independency of the equations. The correlation between the error terms significantly differs from zero, suggesting the need to correct the section bias, as previously explained in section II.

V. Concluding Remarks

The objective of this paper is to identify the factors that shape broadband diffusion, using cross-sectional data on nine of the New Member States of the European Union. We estimate bivariate probit models for the joint probability of Internet access and the type of connection. Results show that the probability of home Internet access is

[2] Although we do not report results here, the same happens with income.

primarily influenced by income, age, and education. There is also some evidence of the existence of technology clusters. In addition, the decision to take up broadband appears to be associated to younger households. Furthermore, ICT investment and competition are found to have notable positive effects on broadband diffusion. Based on these results, policy strategies should focus on both increasing ICT investment and promoting market competition in the provision of high-speed access. Such measures might result in both improvements in the quality of services and lower access prices, fostering broadband demand.

References

- Aron, D.J. and Burnstein, D.E. (2003) Broadband adoption in the United States: An empirical analysis, Paper presented at the 31st Research Conference on Communication, Information and Internet Policy (USA).
- Cava-Ferreruela, I and Alabau-Muñoz, A. (2006) Broadband policy assessment: A cross-national empirical analysis, *Telecommunications Policy*, 30, 445-463.
- Distaso, W., Lupi, P. and Manenti, F.M. (2006) Platform competition and broadband uptake: Theory and empirical evidence from the European Union, *Information Economics and Policy*, in press.
- Empirica-SIBIS (2003) *General Population Survey* (http://www.sibis-eu.org).
- European Commission (2006) *Bridging the Broadband Gap*, European Commission, Luxemburg.
- Eurostat (2006) *Structural Indicators* (http://europa.eu.in/comm/eurostat).
- Gillett S.E., Lehr W.H., Osorio, C.A. and Sirbu M.A. (2006) *Measuring Broadband's Economic Impact*, Report for the US Department of Commerce.

- Goel, R.K., Hsieh, E.T., Nelson, M.A. and Ram, R.(2006) Demand elasticities for Internet services, *Applied Economics*, 38, 975-980.
- Greene, W.H. (1992) A statistical model for credit scoring, *Department of Economics Working Paper*, 92-29, Leonard N. Stern School of Business, New York.
- Hausman, J.A., Sidak, J.G. and Singer, H.J. (2001) Cable Modem and DSL: Broadband internet access for residential customers, *American Economic Review*, 91, 302-307.
- Heckman, J. (1979) Sample selection bias as a specification error, *Econometrica*, 47, 153-161.
- Madden G. and Simpson, M. (1997) Residential broadband subscription demand: an econometric analysis of Australian choice experiment data, *Applied Economics*, 29, 1079-1078.
- Madden, G., Savage, S. and Simpson, M. (1996) Information inequality and broadband network access: An analysis of Australian household data, *Industrial and Corporate Change*, 5, 1049-1066.
- OECD (2004) *Broadband driving growth: Policy responses*, OECD, Paris.
- Rappoport, P.N., Kridel, D.J. and Taylor, L.D. (2002) The demand for broadband: Access, content, and the value of time, in *Broadband: Should we regulate high-speed access?* (Ed.) C. Robert and A. James, AEI-Brookings Joint Center for Regulatory Studies, Washington, 57-82.
- Savage, S.J. and Waldman, D. (2006) Broadband Internet access, awareness, and use: Analysis of United States household data, *Telecommunications Policy*, 29, 615-633.
- US Department of Commerce (2004) *A nation online: Entering the broadband age*, US Department of Commerce.

- Varian, H.R. (2002) The demand for bandwidth: Evidence from the INDEX project, in *Broadband: Should we regulate high-speed access?* (Ed.) C. Robert and A. James, AEI-Brookings Joint Center for Regulatory Studies, Washington, 39-56.
- Wallsten, S. (2006) Broadband and unbundling regulation in OECD countries, *Working Paper 06-16*, AEI-Brookings Joint Center for Regulatory Studies, Washington.
- Yoo, S.H. (2004) Determinants of subscription time for portable internet service, *Applied Economic Letters*, 11, 931-934.

Table 1. Home Access Equation. Probit Estimates

Variables	Model 1	Model 2	Model 3
2^{nd} Income Quartile	0.32***	0.28***	0.29***
3^{rd} Income Quartile	0.46***	0.34***	0.48***
4^{th} Income Quartile	0.96***	0.76***	0.85***
Age: 25-49	-0.14***	-0.12**	-0.10*
Age: 50-64	-0.38***	-0.39***	-0.36***
Age: 65 and more	-0.74***	-0.81***	-0.79***
Education: High School	0.13		
Education: University	0.54***	0.45***	0.42***
Family Size	0.12***	0.12***	0.14***
Mobile Phone	0.43***	0.41***	0.36***
Fixed Phone	0.83***	0.83***	0.69***
Fax	1.16***	1.18***	1.22***
GDP per capita		0.009***	
Openness			0.05***
R&D Expenditure			0.59***
Telecommunications Costs			-0.02***
Intercept	-3.19***	-3.41***	-3.61***
Pseudo R^2	0.25	0.25	0.28
Classification rate (%)	86.56	86.76	86.91

Note: For the estimation of the model we have considered the following reference groups: the first income quartile; those who are less than 24 years old; primary education. *** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 2. Estimated Broadband equation joint with Home Access Equation

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Education: High School	0.30				
Education: University	0.32				
Age: 25-49		-0.12	-0.12		
Age: 50-64		-0.31**	-0.32**		-0.23*
Age: 65 and more		-1.26***	-1.29***		-1.18***
ICT Investment			0.22**		
Dial-up Internet Access Cost				0.004	
Competition					1.87**
Intercept	-1.28***	-0.90***	-1.44***	-0.96***	-1.12***
$\rho_{\varepsilon,u}$	0.19	0.26	0.32	0.26	0.28
LR test indep. of equations					
Chi-squared	4.79**	8.53***	12.31***	2.73*	10***

Note: The equations have been estimated using Model 3 from Table 1.
*** Significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

The Impact of Internet on the Market for Daily Newspapers in Italy

Lapo Filistrucchi[•][°]

First draft: June 2001
This version: February 2007
Under revision, comments welcome!

ABSTRACT

Recent years have seen a surge in websites that provide news for free and, up to the end of 2001, daily newspapers in Italy have shown a growing trend towards making available online for free the exact articles published on paper.
To assess whether on-line news and traditional daily newspapers are substitute, complement or independent goods, I model the choice between different daily newspapers as a discrete choice among differentiated products. Considering the availability of a website as a newspaper characteristic and controlling for other observable and unobservable characteristics of newspapers and of the outside good, I estimate a logit model of demand on market level data from 1976 to 2001 for the main national daily newspapers in Italy.
Results suggest that opening a website had a negative impact both on the sales of the newspaper who opened it and on those of its rivals. I calculate the implied short-run and approximated long-run losses in both sales and profits and provide some evidence of the additional negative effect stemming from the general availability of Internet and on-line news.
Results also contribute to explaining why, starting from the end of 2001, many publishers introduced a fee to read on-line the paper edition of the newspaper.

Keywords: *daily newspapers, Internet, websites, substitution, discrete choice models, product differentiation, dynamics, market level data.*

[•] I am grateful to Massimo Motta and Andrea Ichino for their trust and comments. I also wish to thank Fabian Bornhorst, Timothy Bresnahan, Paul Chwelos, Stephan Fahr, Bronwyn Hall, Helmut Lutkepohl, Johan Stennek, Frank Vella and participants in the Microeconometrics Working Lunch, Second Year Forum 2001/02 and Third Year Forum 2002/03 at the EUI as well as participants in the Economics of Industry Work in Progress Seminar 2002/03 at the LSE and participants in ESSID-2001, the 4th Kiel-Munich Workshop on the Economics of Information and Network Industries, 31st EARIE Conference, the 2004 IDEI conference on Economics of Electronic Communication Markets and the 5th ZEW Conference on the Economics of Information and Communication Technologies. Finally, I am indebted for many suggestions to my colleague Elena Argentesi, with whom I have been building the database on the market for daily newspapers in Italy. All remaining errors in the paper are however mine. Financial support from CESIFIN "Alberto Predieri" Foundation for my visiting period at LSE is gratefully acknowledged.
[°] Tilburg University and University of Siena. Address: Tilburg University, Department of Economics, P.O.Box 90153, 5000 LE Tilburg. Email: L.Filistrucchi@uvt.nl; filistrucchi@unisi.it

JEL Classification: C23, D12, L86, O39.

1. INTRODUCTION

Internet has undoubtedly been the major economic innovation of the last two decades. Among its main effects was the introduction of new products, such as music, videos, e-books and news. Yet the question of whether these new products are substitutes, complements or independent goods has only recently received some attention in the literature.

This paper looks at the market for daily newspapers in Italy, which in the last ten years has witnessed a surge in the number of websites which provide news and other information for free and, up to the end of 2001, a growing trend by daily newspapers publishers towards putting online the exact articles published on paper.

Table 1 below reports the results from a survey of the Italian research centre Censis, carried out in the year 2000. Almost 26% of the people interviewed reported "using less" newspapers and magazines since they started surfing on Internet[1].

Table 1 - Changes in the use of old media because of Internet (% of population above 14 years of age)

"Since I use Internet I use less ..."	
Books	29.2
Newspapers and magazines	25.8
TV	50.5
Radio	21.6

Censis, 2001
(Note that it was possible to give more than one answer.)

Unfortunately the question did not distinguish between magazines and newspapers and it is also not clear whether the respondents were just spending less time reading the newspaper they bought or they were actually buying fewer newspapers[2]. This paper therefore investigates whether less people buy daily newspapers because of Internet.

[1] Similarly a survey of U.S. adult Internet users carried out in 1997, cited in Barsh et al.(1999), reports that in 1995 13% spent less time reading newspapers and 12% of them spent less time reading magazines, while in 1997 both numbers had grown to 16%.
[2] It is also not clear whether those 74% of people who did not read less in fact read as much as before or read more, that is whether, although for 26% people there was some kind of substitution, for others there was complementarity, in which case at the aggregate there might be smaller substitutability, independence or even complementarity between the new and the old product.

In fact, following the appearance of Internet and on-line news, a priori three substitution effects can be expected: one from the general availability of Internet, as people allocate less time to reading (and thus do not buy newspapers) because they prefer surfing on the net (effect 1), a second one from the general availability of news online, whereby people do not buy newspapers as they prefer to read news via Internet (effect 2)[3], and a third one from the opening of the website of the newspaper itself and the availability of the exact articles of the printed edition, whereby people do not buy a newspaper as they prefer to read it on the Internet (effect 3)[4].

I focus here on the third of these three conceivable substitution effects, and I therefore look for the effect of the appearance of a daily newspaper website on its market share on paper and on those of its rivals. In addition I look for some evidence on the first two effects jointly, that is of the general availability of Internet and news on-line.

There is one simple reason why online news and daily newspapers might be substitutes: the news appearing on websites are easily the same of those on the daily newspapers[5]. Even more, they are usually fresher. The substitutability might be stronger between a daily newspaper and its website because of a brand effect if a traditional reader likes[6] or trusts his newspaper more than other information sources. In addition, when the paper edition is available online, a surfer finds on-line the exact content of the paper edition (which is exactly the case for daily newspaper in Italy in the period I consider).

But there are also reasons to expect daily newspapers and online news to be complements. One might for instance expect online readers to get interested in a piece of news they read online and buy a newspaper to read more about it or vice versa a

---

[3] The market for newspaper is usually believed to have already experienced the first two kinds of substitution effects as the appearance of TV changed the reading habits of the Italian population and the appearance of TV news put an end to the second daily edition of most Italian newspapers. See FIEG(1982-2002), Censis(1961-2002).

[4] Since, as reported below, a growing number of daily newspaper websites has developed an original online edition in addition to making available online the paper edition, one could further distinguish between whether people do not buy the paper edition of the newspaper because they prefer to read it online (effect 3a) or whether they do so because they prefer to read the online edition (effect 3b).

[5] Notice that there is no copyright on news themselves but only on the way they are written. It is therefore impossible by both nature and law to stop news from spreading. This is why it has always been so important for traditional newspaper to be the first to write a piece of news. But even this advantage has lost its importance on the Internet, as websites can be updated much more frequently than daily newspapers.

[6] A reader might like a newspaper more than others which provide the same news because of the newspaper's layout or because of the way these news are presented or commented, the leading journalists or, more generally, the editorial line.

newspaper reader might want to go on-line to get fresh updates on some news or find older references to the issue by searching the news archives. Often a website might provide additional content and services which have found no space on the traditional newspaper, such as discussion forums, blogs, smaller cities weather forecasts, audio and video content and so on. Again this complementarity might be stronger between a daily newspaper and its website because of a brand effect if the traditional reader likes or even trusts more his newspaper or an on-line reader likes a daily newspaper website (or the preview of its paper content) and decides to buy the paper edition. In addition, a website might promote subscriptions to the traditional edition by simplifying the process of subscription itself.

Finally, the place (work, home, pub…) and the support (on paper, on screen) where news are read might substantially contribute to diversify the products. At the extreme, one could expect the two goods to be independent if for instance reading a traditional newspaper at home or in a bar is perceived to be so different than reading it on-line[7]. Yet the on-line and the paper product might also be independent at the aggregate level if the website only attracts highly price sensitive consumers who would not otherwise buy the newspaper or attracts only readers from abroad.

To assess whether traditional daily newspapers and their on-line websites are substitute, complement or independent goods, I model consumer choice among different daily newspapers as a choice for a differentiated product, consider the existence of a website as a product characteristic of the traditional newspaper and estimate a logit model of demand using market level data for the period 1976-2001 on the four main national newspapers in Italy, namely *Corriere della Sera*, *La Repubblica*, *La Stampa* and *Il Giornale,* of which all except *Il Giornale* launched and maintained a website in the period under consideration. I then calculate the effect of the decision to go online on the number of copies sold.

Results suggest that website provision had a negative impact on the sales of those who opened it and on their rivals. The estimated average short-run loss in sales per issue due to the joint openings of the websites is 23,350 for *Corriere della Sera*, 30,765 for *La Repubblica*, 24,810 for *La Stampa* and 9,055 for *Il Giornale*, respectively

[7] Of course, these two experiences might also be judged to be complementary, although in this case, if we don't want to assume that reading a second time the same piece of news has any marginal utility, we have to assume some constraint on online or on-paper reading.

around 3.6%, 5.1%, 6.9% and 3.9%. The average short-run loss from an own website is instead estimated to be 3.1%, that from a rival website 1.5%. The approximated[8] average long-run losses due to the joint appearance of their websites are instead estimated to be respectively 197,965 , 260,836 , 210,348 and 76,775, approximately 30.5%, 43.2%, 58.5% and 33.1%. The average approximated long-run[9] loss from an own website is instead estimated to be 26.4%, that from a rival website 12.8%. Finally, there also seems to be some evidence of an additional negative effect due to the general availability of Internet and on-line news.

The loss in sales is estimated to have lead to substantial losses in profits due to a decline in both sales and advertising revenues. Given the average per copy profit margin for our sample of newspapers obtained from balance-sheet data, the estimated average short-run loss per day over the period 1997-2001 due to the joint openings of the websites is 76,292,139 Italian lire (base:1995) for *Corriere della Sera*, 93,070,701 for *La Repubblica*, 80,206,593 for *La Stampa* and 29,699,231 for *Il Giornale*, while the long-run ones are respectively 646,830,507 , 789,084,819 , 680,018,566 and 251,800,102.

Unsurprisingly, starting from the end of 2001, many Italian daily newspapers introduced a fee to read on-line the paper edition of the newspaper. And so did in particular *La Repubblica* in January 2002 and *La Stampa* in 2003.[10]

After discussing in the next section the existing literature, in section 3 I give a brief description of the traditional newspaper market in Italy, define the relevant market and describe the sample of newspapers on which I carry out my analysis. In section 4 I describe the development of the on-line market for news in Italian and the history and features of the websites of the four newspapers in my sample. Section 5 introduces the structural model of demand whereas section 6 provides a general overview of the dataset and section 7 discusses the estimation. Section 8 reports and

---

[8] See below and Filistrucchi(2005) for the necessity to calculate an approximation and, more generally, a discussion of dynamics in a discrete choice logit model.

[9] The long-run effect is defined as the limit for $t \rightarrow \infty$ of the effect of a permanent change in the characteristic. All other things being equal it therefore measures the effect of the decision to go on line if that decision is not reversed.

[10] *Corriere della Sera* just set up a fee to access the pdf version and the archive of the paper edition. The non pdf online version of the paper edition is still freely accessible after registration. Il Giornale opened its website much later than the others and still provides for free access to the full pdf version of the traditional newspaper.

comments on the results. Finally, in section 9, I conclude and discuss possible rationales for the decision to open a daily newspapers website.

2. THE LITERATURE

The question of whether on-line news and news on-paper are substitutes, complements or independent goods had not initially received much attention in the literature.

A few papers in the business literature have dealt with newspapers on-line (Mings & White (2000), Cameron, Curtin & al. (1996) and Cameron, Hollander et al. (1997)). But only very recently, contemporaneously to my work, some papers have dealt with the issue of on-line and traditional newspapers.[11]

Kaiser (2003) has analysed the effect of website provision on the demand for German women's magazines. Estimates from a static and a dynamic[12] nested logit demand model on quarterly market level data from 1996 to 2001 for 41 magazines suggest that website provision did not significantly affect magazine's market shares, which would imply that either they are independent goods or at the aggregate level the complementarity and the substitution effects balance out. The latter seems more likely as there appears however to be evidence of positive spillover effects[13] from the presence of competitor's websites, which suggests that across newspapers the complementary effect is higher than the substitution one. A distinctive feature of the German women's magazine is however, as reported by Kaiser, that the content of the paper edition is not available online. More recently, Kaiser & Kongsted(2005) analyze the relationship between website visits, magazine demand and the demand for advertising pages using Granger non-causality tests on quarterly data for the German magazine market between 1998 and 2004. They find evidence for positive effects from

[11] Most of the previous econometric work on the traditional market for daily newspapers had attempted to estimate either the price elasticity of demand, as Reekie (1976) and Blair & Romano (1993), to identify the main features of the pricing decisions by a newspaper publisher, as Booth et al. (1991) and Fisher & Konieczy (2000), or to assess the cause of the observed trend towards monopolisation as Dertouzos & Trautman (1990). Cecchetti (1986) has instead studied the frequency of price adjustment whereas Willis (2000) has estimated price adjustment costs in the US market for magazines. Some other studies, such as Hakfoort & Weigand (2000), have dealt with the market for magazines, which of course shares many features with the one for daily newspapers.

[12] The dynamic model is obtained by adding lags of the dependent variable to the aggregate demand equation obtained from the discrete choice model. For a discussion of the ambiguities involved, see below and Filistrucchi(2005).

website visits to circulation, suggesting that news online and on paper are complement goods. They do not however allow for cross effects of the other magazines' websites. They also report that none of the magazines makes the printed version fully available online.

Gentzkow (2005) has instead tested for substitutability, independence or complementarity between the Washington Times, the Washington Post and the washingtonpost.com[14] in a model that allows multiple choices and complementarity using consumer level data. Estimates of his model on pooled data from a survey conducted twice a year between March 2000 and February 2003, controlling for consumer characteristics provides evidence for moderate substitutability between the Washington Post and its on-line edition. The loss in readership of the paper edition due to the presence of the website is estimated to be approximately 1.7%[15]. Given the nature of the data, which at each point in time are observations of different samples of individuals and anyway start after the introduction of the website, there is little identification of the website effect coming from the time dimension.

Finally, Simon (2005) analyses the effect on magazines circulation in the US of the different levels[16] of digital content of their website and using annual market level data on 556 magazines from 1996 to 2001 finds substantial evidence of substitution, with the greatest cannibalization of paper sales taking place when digital access to the entire on paper content is offered on-line. The loss in circulation due to the presence of a website is estimated to be approximately 2.5%, which rises to 10% if the full content of the paper edition is available on-line for free[17]. In addition, the effect of all traditional rivals having[18] a website is estimated to be -6.8%[19]. He also finds that

[13] Kaiser calls it “awareness” spillover.

[14] Given the small number of observations in the sample, Gentzkow is forced to omit the washingtontimes.com from the analysis, therefore restricting the estimation of the website cross-effect to that of the post.com on the Washington Times.

[15] Gentzkow does not unfortunately report the cross effect of the post.com on the Washington Times sales (nor the elasticity of the post.com with respect to the Washington Times’ price), although it is probably small (given that the previous elasticity is).

[16] Simon ranks available online content in four categories: related content (that which does not appear on paper), preview content, selected articles and all articles of the paper edition.

[17] Simon also finds that this effect is due to the effect on subscriptions rather than on sales at the newsstand which seem instead to suffer relatively more from the availability of related content. In Italy however, subscriptions are only a very small proportion of total sales, the opposite of what Simon reports for US magazines, where one can explain the finding by saying that either subscribers are more price sensitive or that non subscribers are more interested in the news themselves than in the articles.

[18] Unfortunately, Simon does not distinguish also between the different level of content provided by rival websites.

[19] This effect is instead higher on subscription than on newsstand sales.

both having full online access to the paper edition[20] and having rivals with a website reduce international sales more than US sales. Also the negative effect of a rival website is stronger on foreign sales than on US sales. Competition between traditional newspapers is however modelled only by allowing circulation to depend on the number of competitors.

My results are therefore qualitatively in line with those of both Gentzkow (2005) and Simon (2005), although I additionally distinguish between a short and a long-run effect. They are also not inconsistent with those of Kaiser (2003) and Kaiser & Kongsted(2005), given Simon (2005) finding that substitutability is the highest when there is full on-line access to the magazine content and the fact that, in the period I am considering, all the three newspapers who had a website made available on-line for free the exact content they had on paper.

3. NATIONAL DAILY NEWSPAPERS

In order to investigate if there is any evidence of a substitution or complementarity between the product on paper and the one on-line, I choose to carry out the analysis on the four main national daily newspapers, namely *Corriere della Sera*, *La Repubblica*, *La Stampa* and *Il Giornale*. I follow here the definition of national newspapers market used by the Italian Federation of Newspapers Publishers (FIEG)[21], as done also by the Italian Antitrust Authority in a complaint investigated some years ago[22]. According to FIEG, the four newspapers above belong to the

[20] As Simon suggests, the stronger effect could be due to the greater reduction in transaction costs for non US based readers, who could access the newspaper earlier and not pay the additional mail costs they likely face with respect to the national subscribers.

[21] The Italian Federation of Newspapers Publishers (FIEG) traditionally classifies daily newspapers according to their geographic diffusion, to their content and to whether they are owned by a political party or not. It therefore distinguishes between a) provincial b) regional c) multiregional d) national e) political f) financial g) sport and other) daily newspapers. But the market for daily newspapers has also been evolving substantially in the 26 years under consideration. Many national newspapers have been adding local chronicles through the years, while others made agreements with local newspapers which allowed the two to be sold together at a lower price. Moreover, there has been a growing trend of both national and local newspapers to become generalist, by adding richer business and sport sections. Last but not least, the '90s saw the introduction of all kinds of weekly supplements and of a growing number of promotions which resulted in the bundling of the copy of the daily newspapers with books, videotapes, cassettes, audio CDs, CD-ROMs and, more recently, DVDs. Although in most cases bundling left the consumer free to buy the newspaper alone or together with the bundled product (mixed bundling), in the case of some weekly magazines bundled to the newspaper, if the reader wanted to read the newspaper (the weekly magazine), he had to buy the supplement (the newspaper) and was thus forced to pay the higher price for the bundle (pure bundling). In the econometric analysis which follows I try to control for these evolving characteristics of daily newspapers.

[22] See Decision 3354/95 Ballarino vs. Grandi Quotidiani

national market together with *L'Avvenire, Il Giorno* and *Il Foglio* but even alone account for most of the sales in that market. For instance, in 2001 *Corriere della Sera*, *La Repubblica*, *La Stampa* and *Il Giornale* alone accounted for 91% of the average daily sales of the national daily newspapers, whereas the national daily newspapers accounted for almost 36% of all the average daily sales of daily newspapers in Italy.[23]

Figure 1 - Corriere della Sera, La Repubblica, La Stampa & Il Giornale on paper

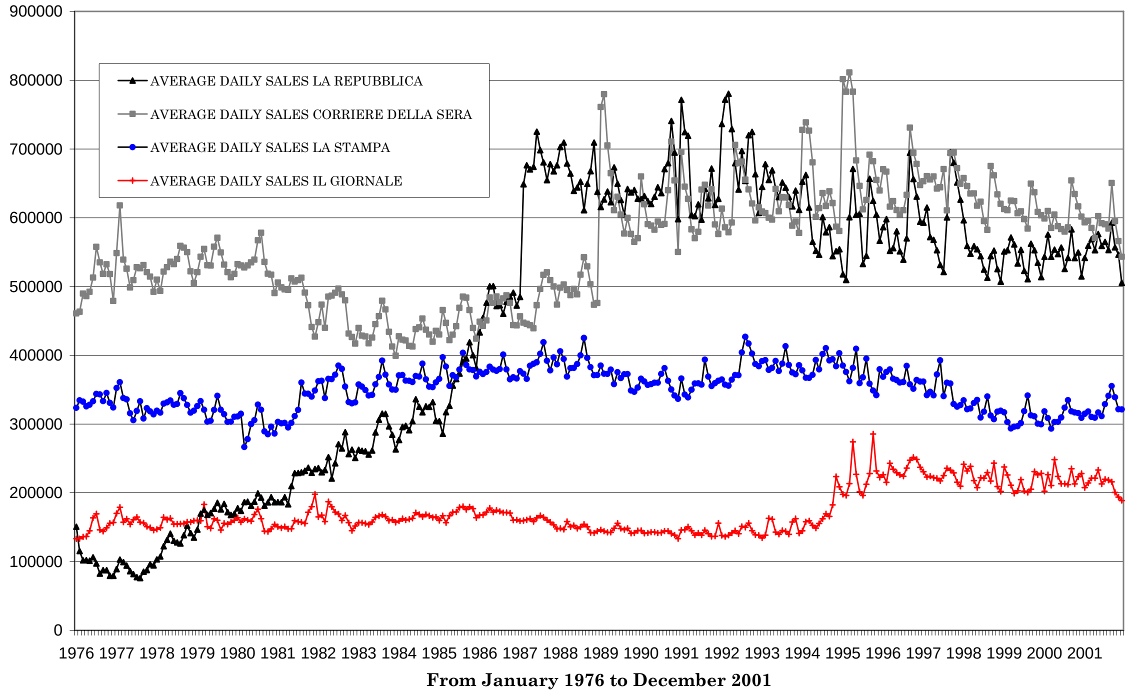


Figure 1 reports the average daily sales in each month from January 1976 to December 2001 for these four daily newspapers. *Corriere della Sera,* founded in March 1876, has been, in the last few years, the one selling more copies, fiercely competing with *La Repubblica,* which was born exactly in January 1976[24], while *La Stampa*, which having been founded in Febbraio 1867 is the oldest, shows consistently lower sales in the period considered here. Finally *Il Giornale* was instead launched in June

[23] Today in Italy there are well above 100 registered daily newspapers. However, many of them have a very limited circulation and some are not even sold at the newsstand. For instance, only 64 publishers of 70 daily newspapers are today members of FIEG (thought they constitute by far the most of the market in terms of circulation) and only 52 daily newspapers certified for advertising purposes by Accertamenti Diffusione Stampa (ADS). Interestingly, tabloids are in practice not existent in the daily market (though they are quite widespread in the weekly and especially in the monthly ones) while free daily newspapers appeared only in 2000.

1974 and is the one with the smallest circulation among the four. The opening of the website is an event which took place near the end of the period considered here, as we will discuss in the next section.[25]

A particular feature of the Italian newspaper market has always been the lack of price competition. Up to the end of 1987 the price was regulated. From the 1st of January 1988 the price was officially liberalized. However, up to today suspicious coordination of price changes appears to be common practice, at least among the main national newspapers[26]. Only through pure bundling to weekly magazines a limited variability of prices across newspapers has appeared[27]. Table 2 shows the nominal price of the four daily newspapers in my sample from January 1976 to December 2001 in a day of the week when none of them issues a supplement. It is evident at first sight that prices have always changed almost simultaneously even after their liberalization. It is because of this particular feature that in the econometric analysis I will claim to be allowed to consider prices as econometrically exogenous, or at least pre-determined, to the single publisher decision.

[24] Unsurprisingly the graph of its average daily sales seems to follow the usual S-shape well-known in the literature on product diffusion, which already suggests the necessity to use a dynamic model of demand.

[25] Looking at the graphs it is also possible to notice that a strong monthly seasonality affects the data. The timing of the spikes which can be observed in January 1989 for *Corriere della Sera* and in January 1987 for *La Repubblica* coincide respectively with *Portfolio* and *Replay*, two games of the lotto kind which could be played only and simply by buying a copy of the newspaper (at the normal price). The jump in sales for *Il Giornale* which took place in 1994 coincides in time with the appearance into politics of its owner Silvio Berlusconi.

[26] See the Italian Antitrust Authority communication to the Italian Authority for Broadcasting and the Publishing Industry in January 1996. More in Argentesi&Filistrucchi(2005).

[27] For a discussion on identification of the price effect when it varies only with bundling see Argentesi&Filistrucchi(2005).

Table 2 – Nominal Prices without supplements

Since	Corriere della Sera	La Repubblica	La Stampa	Il Giornale
01/06/74	150	150	150	150
01/05/77	200	200	200	200
11/03/79	250	250	250	250
01/08/79	300	300	300	300
17/08/80	400	400	400	400
01/08/82	500	500	500	500
01/02/83	500	500	500	500
01/07/84	600	600	600	600
20/10/85	650	650		650
05/01/86			650	
01/08/86	700	700	700	700
14/06/87	800	800	800	800
01/03/88	900	900	900	900
01/08/88	1,000	1,000	1,000	1,000
01/08/90	1,200	1,200	1,200	1,200
28/06/93	1,300	1,300	1,300	1,300
02/01/95	1,400	1,400	1,400	1,400
10/04/95	1,500	1,500	1,500	1,500
01/06/01				1,700
01/12/01	1,700	1,700	1,700	

unit of measure: Italian lire

4. NEWS ON THE WEB

The last few years have witnessed a surge in the number of websites providing news and other information, as reported in Table 3 below. Many of them were opened by publishers already present in the traditional markets, but a growing number of them is constituted by webzines, that is publications which are available on-line only.

Most traditional publishers are by now on-line, albeit with much different products and not always with a product specially designed for the web. In particular 106 Italian daily newspapers were present on-line in December 2001.[28]

Table 3 - Number of news and information sites by type and date

Site type	Dec. 1997	Dec. 1998	Dec. 1999	Dec. 2000	Dec. 2001
Daily newspapers	44	54	62	76	106
Periodicals	226	326	501	559	1,051
Webzines	128	197	412	772	1,141
Total	398	577	975	1408	2,298

Source: Webtime[29]

The increase in the number of news and information sites corresponded also to an increase in the quality of the sites themselves. Among daily newspapers in particular there appears to have been a growing trend towards putting online the exact articles published on paper, but also an increasing tendency towards developing an original on-line edition, as reported in Table 4 below.

[28] The first newspaper to appear online was the local *L'Unione Sarda* at the beginning of 1995, followed by the political *L'Unità* in August and the national *La Stampa* in September.
[29] *Webtime* is an online observatory on news on the web. Its address is www.ipse.com

Table 4 - Number of daily newspaper websites by maximum on-line content and date

On line content	Dec. 1997	Dec. 1998	Dec. 1999	Dec. 2000	Dec. 2001
All articles	11	25	33	54	73
Some articles	6	6	8	7	10
First page	6	7	8	1	3
Presentation	6	8	7	4	12
Under construction	5	3	3	6	7
Not updated	6	2	0	2	0
Newspapers review	4	3	3	2	1
Total	44	54	62	76	106
of which					
with editorial portal	0	0	4	23	23
with local portal	0	3	3	10	17
with on-line newspaper edition	1	1	1	9	12

Source: Webtime

Up to the end of 2001 all daily newspaper websites did not charge any fee to access to the information and news they provided, even for reading the articles which appeared also on the traditional newspaper. Since the end of 2001 some of them started to charge a fee, usually to access on-line the paper edition and/or to consult its catalogue.[30]

Among the four national newspapers in my sample the first one to go on-line was *La Stampa*, which opened its website *www.lastampa.it* in September 1995 and since the beginning made available on the web the traditional newspaper. Only in December 1999 however it started to provide an original on-line edition. *Corriere della Sera* opened its website in December 1996 and it too put on-line the exact articles of the printed edition, but initially the website had to be found inside the website of the publisher holding company RCS. Only in January 1998 was the new website *www.corriere.it* opened and in June 2000 it started to provide in addition to all the articles of its printed edition an original on-line edition. The official website of *La Repubblica* was instead opened in January 1997 at the address *www.repubblica.it*[31]. From its very beginning it started to provide an original on-line newspaper. Finally *Il Giornale* opened a website www.*ilgiornale.it* in May 2005 only, well outside our

[30] Most of them offered subscriptions for a given period of time, others also for a given number of issues. It appears in any case that a new business model is believed necessary for the future. This provides weak evidence in favour of product substitutability. But an alternative or additional explanation could be their inability to cover costs through online advertising following the end of the .com bubble on the stock exchange and the economic downturn that followed.

[31] There was a brief but successful trial in April 1996 when La Repubblica opened a specialised website on the occasion of the Italian elections at the address www.repubblica.interbusiness.it

sample period, and chose to provide both an original on-line edition and full access to the pdf version of the newspaper.

After a two month trial registration period, in January 2002 *La Repubblica* introduced fees to read on-line the traditional newspaper and search its archive, while maintaining free access to its on-line edition. From the same date it became also possible to download the whole newspaper in pdf format. In January 2003 *La Stampa* adopted a similar business model. *Corriere della Sera*, which had already required registration since August 2000, chose instead to charge a fee only to access the pdf of the paper edition. Maybe because its site is relatively young Il Giornale still provides fpr free full access to the content of the paper edition.

Unfortunately data on on-line readership for newspapers websites, for news sites in general and evidence on Internet use are impossible to find for the first years of life of news websites and more generally of Internet. Even when available they are not easily comparable to each other because standards in measuring Internet audience have been varying a lot[32]. As a result I cannot use them in the estimation process to identify the three candidate Internet effects mentioned above.[33] They offer however a descriptive, albeit incomplete, picture and some qualitative findings.

According to all sources, *La Repubblica* has up to now been by far the most popular newspaper website and also one of the most popular news and information sites for connections from Italy. Figure 2 shows average daily page views per month for *La Repubblica*[34], *Corriere della Sera* and *La Stampa*, from January 1999 to December

[32] Internet diffusion and use can be measured by number of hosts, by number of Internet accounts or through surveys or panels of individuals (or households). Internet traffic to a website can be measured either through surveys, panels of surfers, website log files and/or through website cookies. Only in March 2002 Audiweb has started to operate, providing complete measures of Internet audience in Italy for advertising purposes. Still it chose to provide two kinds of audience measures: those deriving from a panel of individuals and those deriving from website log files. It set however rules for the collection of each of these data. See www.audiweb.it

[33] Another difficulty when one wants to use data on online news readership in the estimation is that, unless registration and a password is required, it is not easy to identify how many different individuals (or households) chose to access a website in a given day, which would be the equivalent of how many people (or households) bought a copy of the newspaper. So that one has to trust, if at all available, the representativity of panels of individuals (or households) or of those who did not delete cookies. Downloaded pages (also called page views), although the most easily available measures of Internet traffic, do not in general allow to distinguish between the number of individuals (or households) who chose the product and the intensity of use (pages read). In addition data on page views have to be cleaned, as for instance downloading a page twice might be due to the first attempt not being completely successful or automatic refresh of an open window.

[34] On the first day of its life, the 14th of January 1997, the website of *La Repubblica* recorded approximately 500,000 page views. During the 19 days trial period in April 1996 it had enjoyed an average of approximately 300,000 page views a day.

2002. Although the panel of observations is very unbalanced and quite incomplete, average daily page views for *Corriere della Sera* appear to be lower than those for *La Repubblica*[35] but higher than those for *La Stampa*.

Figure 3 shows instead average daily page views for *La Repubblica*, in each month from January 1999 to January 2001, disaggregated by sectors of the website. In this period, page views relating to the traditional edition of the newspaper grew from 66,447 to 218,813 and were, after the homepage itself, the main component of overall page views (around 20%), with the on-line edition being the third most visited[36]. So that the possibility of a full access to the content of the paper edition appears to be a very appreciated feature of a daily newspaper's website. In addition, on-line news reading appears to have been growing considerably and to have become a substantial phenomenon well before the year 2001, the end of our sample period.

Figure 2 - Corriere della Sera, La Repubblica & La Stampa online

[35] Also other sources suggest the same ranking. In May 2001, for instance, using a panel of online surfers, Nielsen-NetRatings estimated 724,473 unique visitors and 15,804,751 page views for La Repubblica while respectively 384,351 and 7,676,761 for *Corriere della Sera*. Exactly one year before through 12,000 random phone-calls to Italians above 14 years of age Between had estimated 1,140,000 online readers for La Repubblica and only 540,000 for *Corriere della Sera*.

[36] However, it is impossible to say how much of the observed growth in the number of downloaded pages is due to an increase in the number of online readers and how much is instead due to an increased number of pages read by each of them.

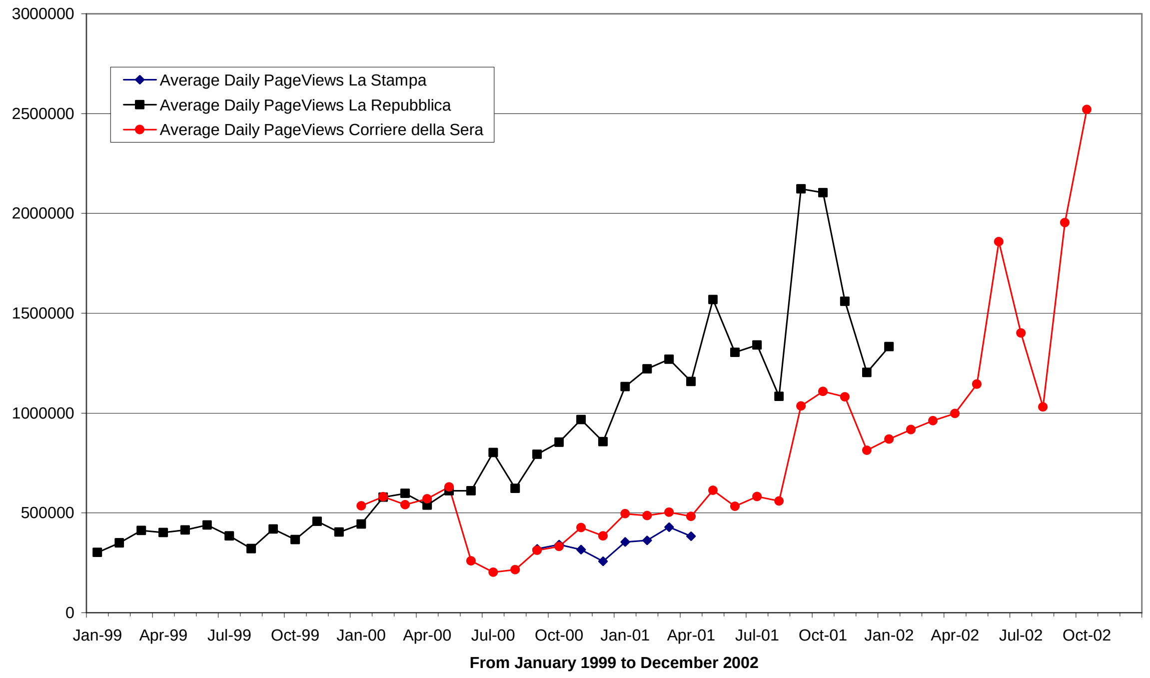
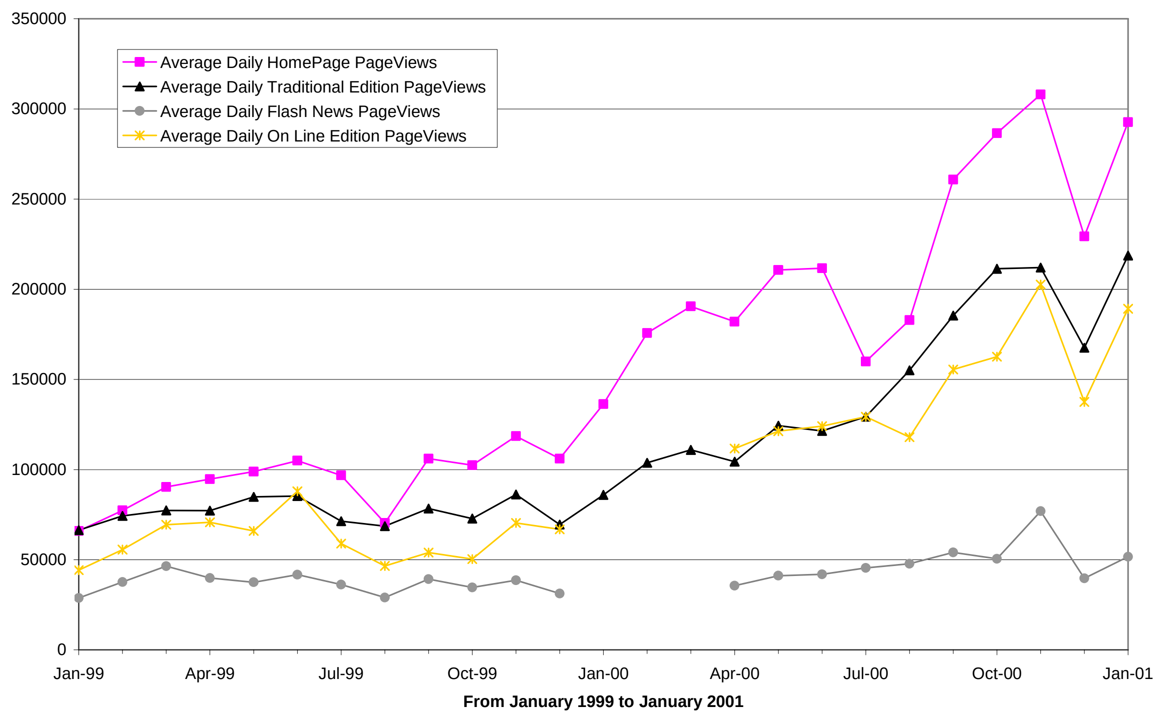


Figure 3 - La Repubblica on line



5. THE MODEL

In order to test for product substitutability or complementarity, I follow Kaiser(2003) and model consumer choice among different newspapers as a choice for a differentiated product, consider website availability as a product characteristic and derive a logit model of demand.

The starting assumption, which is common to fixed coefficients models of product differentiation in general, is the following functional form of consumer *i* indirect utility from reading newspaper *j* at time *t* on day *d* of the week:

$$u_{ijtd} = \alpha(y_{it} - p_{jtd}) + \bar{x}_{jtd}\bar{\beta} + \xi_{jtd} + \varepsilon_{ijtd} \qquad (1)$$

where y_{it} is the income of consumer *i* at time *t*, p_{jtd} is the price of newspaper *j* at time *t* on day *d* of the week, $\bar{x}_{jtd}$ is a vector of observed characteristics (including havng a website), ξ_{jtd} is an unobserved (by the econometrician) characteristic, ε_{ijtd} is a mean-zero stochastic term, α is consumers marginal utility from income and $\bar{\beta}$ is a vector of taste coefficients.

Such an indirect utility specification assumes a quasi-linear utility function, free therefore of wealth effects, which sounds plausible for daily newspapers. It also assumes that both observed and unobserved product characteristics are the same across all individuals in the market and thus rules out both the possibility of different supplements, promotions and chronicles in different regions and of different prices for different consumers[37]. Finally, the marginal utility from income and the taste parameters are assumed fixed across consumers and, as a result, consumers heterogeneity enters only through the separable additive random shock ε_{ijtd}.

As consumers may decide not to read any of the daily newspapers considered (or any newspaper at all), an outside good is introduced[38], consuming which yields to consumer *i* at time *t* on day *d* of the week the indirect utility:

[37] Whereas there are in general no discounts in the price to consumers, not even for subscriptions, and local supplements and promotions are not frequent, local chronicles and bundling to local daily newspapers at an unchanged price are quite common.

[38] In the absence of an outside good, the model would assume consumers to be forced to choose one of the newspapers. Therefore, it would assume that, if all newspapers opened a website at the same point in time, this would not influence the aggregate sales of all the newspapers, which would be unfortunate, since it would in fact amount to assuming that no substitution takes place between the paper and the online product.

$$u_{iotd} = \alpha y_{it} + \xi_{otd} + \varepsilon_{iotd} \,. \tag{2}$$

Since the outside good is a composite one, its price and its characteristics are not defined. The price of the outside good is then assumed to be equal to zero[39] and all the characteristics are assumed to be unobservable[40]. But as ξ_{otd} is not identified, the standard practice is to set it equal to 0, which, as the term αy_{it} eventually vanishes because it is common to all products, is equivalent to normalizing the mean utility from the outside good to zero[41].

Consumers mean utility, δ_{jtd} from reading newspaper *j* at time *t* on day *d* of the week is instead given by:

$$\delta_{jtd} = E_i[u_{ijtd}] = \alpha(y_i - p_{jtd}) + \bar{x}_{jtd}\bar{\beta} + \xi_{jtd} \tag{3}$$

Consumers are then assumed to purchase the newspaper which gives them the highest utility. For convenience, they are assumed to be never indifferent between buying one or another newspaper and never to choose more than one newspaper. The latter assumption is common to most empirical studies on differentiated products markets, the usual justification being that assuming otherwise is econometrically very cumbersome and the assumption is instead, at worst, a reasonable approximation. That is because multiple purchases, though by no means uncommon, are not the rule and in any case even if two products are bought together they are then often consumed at different times, so that the multiple purchase is just an organisational device.[42] Furthermore, if the potential market size is defined large enough, we might also claim that observed multiple purchases by the same individual are the result not only of his choice but also of somebody else's decision. In particular, if potential market size is defined as total population instead of number of households, the observation that an individual buys two newspapers might, at least to a certain

[39] In other words, the consumer is assumed to be choosing between buying one of the above newspapers or not buying it, not between buying one of the newspapers above or buying something else. The decision of whether to buy something is not simultaneous.
[40] Or equivalently, both the price and the characteristics are assumed to be unobservable and therefore included in ξ_{otd}.
[41] So that neither the market shares of the outside good nor those of the inside goods respond to changes in the characteristics or in the price of the outside good, unless time fixed effects are used, as discussed below.
[42] In this market, the latter justification is, however, weaker than usual, as newspapers become quickly old and buying two newspapers in the morning may indeed be equivalent to buying one newspaper in the morning and one in the evening, but very different from buying one today and one tomorrow.

extent, reflect the fact that he is buying one newspaper for himself and another for another member of his household who asked him to.

Decomposing $\xi_{jtd} = \zeta_{jd} + \zeta_t + \eta_{jtd}$, with η_{jtd} a random shock independent of ε_{ijtd}, allows me to model newspaper-day and time specific unobserved characteristics. Given that the assumption of no correlation between the observed product characteristics and the unobserved product characteristics, which lies at the basis of the random effect specification, does not appear plausible, I consider ζ_{jd} as an unknown parameter specific to each product *j* on day *d* of the week, thus leading to a fixed effects model. This choice also allows me to better estimate product differentiation, as in this model the product fixed effects are usually believed to capture also the vertical component.[43] In addition, using product-day fixed effects instead of both product fixed effects and day fixed effects has the advantage of leaving room to a different ranking of newspapers for each day of the week, thus allowing vertical product differentiation to vary across days of the week. Moreover, given that unobserved (to me) newspapers' characteristics vary also across day of the week, product-day fixed effects help to identify the Internet effect.

The inclusion of time fixed effects is instead justified by the necessity to control for the change through time in the utility of the outside good. There are many reasons why the latter may change in time. As a result of the appearance of TV, video games, CDs, DVDs, Internet and, more generally, of alternatives to reading a newspaper[44], the characteristics and price of the outside good might change. Also, changes in the average consumer taste may change the relative utility of the choice to buy the newspaper with respect to any of the activities included in the composite outside good. As its utility is by construction normalised to zero, the absence of time fixed effects or some equivalent control[45] would raise questions of identification for the estimated coefficients, particularly so in the case of the website effect, which in my model is identified through a dummy variable only.

However, controlling for changes in the utility of the outside good also defines the substitution effect measured by the website dummy variable. In as much as it is

[43] See Nevo (2001).
[44] See Censis (1961-2002) and Censis (2001).
[45] An alternative to time fixed effects is the use of a polynomial trend, as in Argentesi (2004) and Argentesi&Filistrucchi(2005) or the use of year and month fixed effects, as Kaiser (2003). The time fixed effects are however more flexible in controlling for changes in the utility of the outside good.

possible to assume that the general availability of news on line has the same negative percentage impact on the market shares of all the newspapers in the sample, the substitution effects 1) and 2) discussed above will be captured, among many other things, by the time dummy variables. As a result, the coefficient on the website dummy variable will measure only the substitution effect 3), that is will capture the loss in market share of a traditional newspaper due to people who shift to reading the same newspaper website.

I then assume that ε_{ijtd} is i.i.d. across consumers and products and that it is distributed according to a type I extreme value distribution. Assuming ε_{ijtd} to be i.i.d. across consumers rules out, in particular, the possibility that individual specific random shocks are correlated across products or equivalently only allows shocks to demand to be correlated across products if they are not individual specific

All the assumptions above lead to a logit model, where the market share of product *j* at time *t* on day *d* of the week are given by:

$$S_{jtd} = \text{Prob}\{u_{ijtd} \geq u_{iktd} \ \forall_{k \neq j}\}$$

which implies

$$s_{jtd} = \frac{\exp(\delta_{jtd})}{1 + \sum_{k \neq 0} \exp(\delta_{ktd})} \text{ for any newspaper } j \tag{4}$$

and

$$s_{0td} = \frac{1}{1 + \sum_{k \neq 0} \exp(\delta_{ktd})} \text{ for the outside option[46].} \tag{5}$$

It should be noted that the presence of an outside good with market share s_{0td} means that observations of newspapers sales are not sufficient to calculate market shares. As a result it is necessary to introduce the concept of potential market size as distinct from the observed market size which would simply be the sum of national newspapers sales. Thus the definitions of market size and market shares are different from the ones commonly used. Potential market size can either be assumed or estimated by parameterising it as depending on some market level data (such as population) which

[46] Note that the term y_{it} drops out as it is common to all options.

vary across time. Here, as discussed below, I assume potential market size is equal to population above 14 years of age[47].

For any given characteristic which is expressed by a continuous variable x its own and cross marginal effects on market shares are:

$$\frac{\partial s_{jtd}}{\partial x_{jtd}} = \beta(1 - s_{jtd}) s_{jtd} \tag{6}$$

and

$$\frac{\partial s_{jtd}}{\partial x_{ktd}} = -\beta s_{ktd} s_{jtd} \quad \text{with } k \neq j. \tag{7}$$

So that the own and cross elasticities of the market shares with respect to that characteristic are respectively:

$$\eta_{jjtd} = \frac{\partial s_{jtd}}{\partial x_{jtd}} \frac{x_{jtd}}{s_{jtd}} = \beta x_{jtd}(1 - s_{jtd}) \tag{8}$$

and

$$\eta_{jktd} = \frac{\partial s_{jtd}}{\partial x_{ktd}} \frac{x_{ktd}}{s_{jtd}} = -\beta x_{ktd} s_{ktd} \quad \text{with } k \neq j. \tag{9}$$

The model thus predicts a different demand, different market shares and therefore different marginal effects and elasticities for each time t and each day d of the week.

If the characteristic is instead a dummy variable the derivatives and elasticities above are not defined. However, defining $\delta_{jtd}(1)$ as δ_{jtd} when $x=1$ and defining $\delta_{jtd}(0)$ as δ_{jtd} when $x=0$, the own effect of the characteristic can be calculated as

$$\frac{\Delta s_{jtd}}{\Delta x_{jtd}} = \frac{\exp(\delta_{jtd}(1))}{1 + \sum_{n \neq 0,J} \exp(\delta_{ntd}(0)) + \exp(\delta_{jtd}(1))} - \frac{\exp(\delta_{jtd}(0))}{1 + \sum_{n \neq 0} \exp(\delta_{ntd}(0))} \tag{10}$$

for any newspaper j , while the cross effect of the characteristic is

$$\frac{\Delta s_{jtd}}{\Delta x_{ktd}} = \frac{\exp(\delta_{jtd}(0))}{1 + \sum_{n \neq 0,k} \exp(\delta_{ntd}(0)) + \exp(\delta_{ktd}(1))} - \frac{\exp(\delta_{jtd}(0))}{1 + \sum_{n \neq 0} \exp(\delta_{ntd}(0))} \tag{11}$$

for any newspaper j and

[47] An alternative would be using the number of households.

$$\frac{\Delta s_{otd}}{\Delta x_{ktd}} = \frac{1}{1+\sum_{n\neq 0,k}\exp(\delta_{ntd}(0))+\exp(\delta_{ktd}(1))} - \frac{1}{1+\sum_{n\neq 0}\exp(\delta_{ntd}(0))} \tag{12}$$

for the outside option.

Finally, dividing each newspaper market share by the outside good market share, simplifying and taking natural logarithms leads to the following market shares equation to estimate:

$$\delta_{jtd} \equiv \ln(s_{jtd}) - \ln(s_{otd}) = \bar{x}_{jtd}\bar{\beta} + \alpha p_{jtd} + \zeta_{jd} + \zeta_t + \eta_{jtd} \tag{13}$$

As it is well-known to the empirical industrial organization literature[48], the use of a logit model to estimate demand places restrictive assumptions on own and cross price elasticities or equivalently on own and cross marginal effects of price. The same restrictions are placed on the marginal effects and elasticities with respect to any characteristic which is measured as a continuous variable. In particular, two newspapers with the same market shares will have the same own derivative and also the same cross derivative with respect to any third newspaper. In addition cross derivatives are symmetric. So that conditional on market shares, own and cross elasticities depend only on the characteristic which changes, while in addition all newspapers have the same elasticity of demand with respect to any given newspaper. That is because additive separability together with the i.i.d. structure of the random shocks, when the amount of a positive (negative) characteristic of one newspaper is raised (decreased), requires consumers to substitute towards other newspapers in proportion to market shares, regardless of the other newspapers characteristics[49]. The same restrictions extend of course to the case of a characteristic measured as a dummy variable.[50]

[48] See for instance Berry (1994), Berry et al. (1995) and Nevo (2000)
[49] See Berry (1994) and Nevo (2000)
[50] Although traditionally the most frequently used model specifications have been logit and nested logit, recent research often uses also the more flexible mixed logit models (Berry & al (1995), Nevo (2001), McFadden & Train (2000), Petrin (2002)). Such a model is more general and allows for substitution between products to depend on product characteristics through observable consumers demographics. But it is not solvable analytically and requires estimation by simulation (Nevo (2000)). So that logit or nested logit models not only offer interesting benchmark cases but are still widely used for their computational simplicity or when the implied restrictions on price elasticities and marginal effects are not considered crucial (Brenkers & Verboven (2002), Kaiser (2003), Rysman (2004)), which I believe is the case here.

Whereas the restrictions on own and cross price derivatives and elasticities do not appear to be too much of a problem in this case as I am not directly interested in the effect of price, those on the effects of product characteristics are potentially more problematic. However, I analyse a sample of newspapers, which although differentiated, is quite homogeneous, so that assuming substitution to take place on the basis of market shares is likely to be a good approximation.

Yet even in this context some concerns might be raised with respect to the substitution towards the outside good. In fact, when using, as potential market size, the population above 14 year of age, the highest market share is the one of the outside good[51]. As a result, for any increase in a product characteristic which provides positive (negative) utility, most of the consumers are assumed to substitute from (towards) the outside good. However, the characteristic I am interested in is the availability of a website. Then, if a daily newspaper publisher opens a website and there is substitution away from the paper edition of that newspaper, the assumption that most people do not substitute towards other national newspapers but rather substitute towards the composite outside good is not at all restrictive, as in this case the outside good includes the choice of reading news on-line.

Last but not least, considering the availability of a website as a newspaper characteristic in a logit model, introduces a restriction on the sign of the cross marginal effects, in that it assumes the introduction of the website to have a positive effect on the competitors sales. As reported below, to release this assumption I let the utility of a given newspaper depend also on the number of competitors' websites, which I include in the vector of characteristics $\bar{x}_{jtd}\bar{\beta}$. Although this does not release the other assumptions of the logit model, it allows for the sign of the cross marginal effect to be decided by the estimation.

6. THE DATA

The dataset I use in the analysis mainly draws from the data collected every year, from 1976 onwards, by the association Accertamenti Diffusione Stampa (ADS) for advertising purposes. Newspapers are free to choose whether to have their data

[51] This would be true even if I used the number of households as potential market size.

certified by ADS or not, but if they choose so they are obliged to provide all the information required and the truthfulness of the reported information is verified by the ADS. Most of the Italian newspapers chose certification, some of them did not, or at least did so only discontinuously.

The information available for each newspaper includes, at various levels of disaggregation, data on sales, prints, gift copies, free subscriptions and paid subscriptions. In particular, they include data on average daily sales in each month and average daily prints in each month and for each different day of the week in each month[52].

We then added to the database other relevant information, mainly obtained by newspaper publishers themselves, such as the nominal prices of the newspapers, the dates of editor changes and their names, the dates the different supplements first appeared, the list of all promotions with the corresponding periods and the dates the different local chronicles were added to some of the national newspapers or the national newspaper was bundled to a local one.

For the purpose of this study, I also collected the dates of the opening of the newspapers web-sites, the period the traditional newspaper was available on-line for free and, more generally, the dates of major changes to the characteristics of the websites.

I choose to estimate the model on sales data at the newsstand, thus leaving aside the number of paid subscriptions. As in my sample period Internet news are a relatively recent phenomenon and subscriptions are bound to react with lags to any external shock, the effect of news on-line on traditional newspapers would be more difficult to detect if subscription were considered. In doing so I implicitly assume that an observed decline in the average daily sales would not be matched by a counter increase in paid subscription[53]. In fact, the latter are only a very small part of most Italian daily newspapers circulation[54], mainly because traditionally copies to subscribers have always been delivered through the general mail service, which

---

[52] Thus, for instance, for the Mondays of July 1979, the Tuesdays of that month and so on.
[53] This could happen for instance if having a web page makes it easier to subscribe to the paper edition. But subscribers might also,on the contrary, switch to purchase if they learn they can read on the net any issue the might miss.
[54] In 2001 for instance paid subscriptions amounted to only 6,8 % of total circulation for national daily newspapers and to 9% for all daily newspapers certified by ADS . See FIEG (2002).

implies they reach destination in the middle of the morning, too late for people going to work[55].

By using monthly observations, I would be unable to control adequately for confounding factors when identifying the effect of the website, as I would not, for instance, exploit the fact that weekly supplements are bundled to a given newspaper only on given days of the week and that on that day the price is higher compared to same newspaper in other days and to other newspapers on that day. As I have data on average daily prints for each different day of the week in each month and I can calculate the average daily ratio of sales to prints in each month, I derive data on average daily sales by day of the week multiplying prints by this ratio. The plausibility of the results clearly relies on the assumption that the ratio is constant across days of the weekin a given month. This is an assumption that, of course, I cannot test but that I believe reasonable and that allows me to enjoy the greater information provided by more disaggregated data[56].

Potential total market size was defined as total Italian population above 14 years. This is the usual potential readers definition in studies on newspapers consumption[57]. As estimates of population were available only for the beginning or the end of each year, the data were interpolated linearly to get monthly observations[58].

Market shares s_{jtd} were thus calculated as the number of sales of newspaper j at time t on day d of the week over total Italian population above 14 years of age at time t[59]. As discussed above, average daily sales on each day of the week in each month were calculated multiplying the average daily prints on that day of the week by the ratio of average daily sales to the average daily prints in that month. Since for some years and for some daily newspapers the prints (and sales) of the daily newspaper when bundled to the supplement were recorded separately by ADS, average daily sales and average daily prints by month were calculated by averaging with weights given by the effective number of issue reported by ADS, whereas average daily prints by day of the

[55] See FIEG(2002).
[56] Alternatively one could use prints as a proxy for sales, assuming the latter to be just a constant proportion, as in Argentesi (2004) and Argentesi&Filistrucchi(2005). Results do not change substantially if average daily prints for each day of the week in a month are used as the dependent variable.
[57] See FIEG (1982-2002), Censis (1961-2002) and also Kaiser (2003).
[58] Results are however qualitatively robust to using a constant potential market size during each year.
[59] Total market size was assumed constant across different days of the week in a month.

week in each month were obtained with weights given by the number of each day of the week in the month[60].

The outside good market share was calculated as $s_{otd} = 1 - \sum_j s_{jtd}$. Clearly it enjoyed the highest market share.

Average nominal prices for each day of the week in each month were obtained by averaging over the official nominal prices of the newspaper with weights given by the number of each day of the week in the month[61]. Average real prices were then obtained dividing average nominal prices by the Italian monthly CPI.

Most characteristics included in x_{jtd} were dichotomous and had to be introduced as dummy variables. All of them however changed across time and product. They included dummies for supplements[62] (both of generalist and women's kind[63]), for having a Monday issue[64], for newspaper editors in chief as a proxy for editorial line (some of them switched from one newspaper to the other during the period under consideration), and for games of the "lotto" kind played simply and only by buying the newspaper. I also included the number of competitors' websites as an additional regressor.

Finally, as already mentioned, newspaper-day fixed effects were used as well as time fixed effects.

7. ESTIMATION

Estimating equation (13) as it is leads to substantial autocorrelation in the residuals. The main source of shocks to demand in the market for daily newspapers can be expected to be news itself. Since some events take place either in more than one

[60] That is they were calculated disregarding strikes, as if the newspapers had always been sold when they were supposed to be. Official days in which newspapers are not sold due to holidays were instead considered.
[61] See note 60.
[62] In particular, I control for the effect of the supplement on the day it is issued but also for the promotional effect on the other days of the week. See Argentesi (2004)
[63] Women's supplements are *Io Donna* for *Corriere della Sera* and *D-Donna della Repubblica* for *La Repubblica*. Generalist supplements are instead *Il Venerdì* of *La Repubblica*, *Sette* of *Corriere della Sera* and *Specchio* of *La Stampa*. The day in which they are issued has for some of them changed through time.
[64] La Repubblica started to have a Monday issue in January 1994, *La Stampa* in January 1992 (though there was a Monday issue of *La Stampa* Sera up to December 1991, when this evening edition of *La Stampa* ceased to be published) and Il Giornale in January 1980.

month or in between two of them, the error term η_{jtd} might be auto correlated. If so, the autocorrelation due to news could probably be assumed to be of order one and common to all newspapers.

Yet, there is also another potential source of autocorrelation: the omission of dynamics. Figure 1 above clearly shows that dynamics is relevant, at least in order to explain the diffusion of *La Repubblica*. Dynamics in our model can be due to the presence of consumer habits[65] and/or the existence of consumption externalities, the latter affecting either consumers' evaluation of the product (that is the indirect utility of consuming it) or consumers' knowledge about the product (that is the choice set). In any case aggregate market shares today would not only be a function of newspapers characteristic and random shocks today, but also of market shares yesterday. Similarly for the ratio of any newspaper market share with respect to the outside good.

I thus estimated the equation above also with the inclusion of lags of the dependent variable, instead of or in addition to an auto correlated random shock, as adding lags of the dependent variable is the usual way to take dynamics into account when estimating a structural model of demand which starts from the specification of an aggregate demand equation. This specification is however not fully consistent with the utility maximization framework at the basis of the aggregate logit model.[66]

If one lag of the dependent variable is introduced, the estimated equation, becomes

$$\ln(s_{jtd} / s_{otd}) = \rho \ln(s_{jt-1d} / s_{ot-1d}) + \overline{x}_{jtd}\overline{\beta} + \alpha p_{jtd} + \varsigma_{jd} + \varsigma_t + \eta_{jtd} \qquad (14)$$

Then (10), (11) and (12) are short-run effects, whereas the long-run effects of the characteristics are unfortunately not defined as including a lag in the equation above only imposes the inter-period marginal effects to be such that

[65] Dewenter (2002) finds evidence of myopic habits formation while rejecting rational addiction in the market for newspapers in Germany.

[66] Usually, the problem of dynamics is not recognised in the empirical literature on discrete choice models of product differentiation, which treats observations on the same market at different points in time as observations of different markets. See for instance Nevo (2001) and Brenkers & Verboven (2002). Whereas the issue might not be relevant in a market for durable goods such as automobiles, where those who buy in period t are not likely to buy again in period t+1, it certainly is an issue in a market for non durables, such as magazines, newspapers or cereals, where multiple purchases in time by the same consumer can be expected to be the rule rather than the exception.

$$\frac{\partial s_{jdt+T}}{\partial x_{jtd}} = \rho^T \beta\, s_{jdt+T} + \frac{s_{jdt+T}}{s_{0dt+T}} \frac{\partial s_{0dt+T}}{\partial x_{jdt}} \quad \forall T > 0 \tag{15}$$

[67]

and therefore the long-run effects of a permanent change in x to be such that

$$\lim_{T\to\infty}\left(\sum_{s=o}^{T} \frac{\partial s_{jdt+T}}{\partial x_{jdt+T-s}}\right) = \lim_{T\to\infty}\left(\beta \frac{1-\rho^{T+1}}{1-\rho} s_{jdt+T} + \frac{s_{jdt+T}}{s_{0dt+T}} \sum_{s=0}^{T} \frac{\partial s_{0dt+T}}{\partial x_{jdt-s}}\right) \tag{16}$$

There is therefore more than one structural random utility model which satisfies the condition above. As a result it is not possible to calculate the long-run effect unless we impose some structure on the sources of dynamics. And even so, as the model is nonlinear, assumptions are needed on the behaviour extra sample of the price, the other characteristics, and all the other explanatory variables. [68]

However, if we assume market-shares to be constant through time, then the condition above is satisfied by

$$\frac{\partial s_{jdt+T}}{\partial x_{jtd}} = \rho^T \beta\, s_{jdt}(1 - s_{jdt}) \tag{17}$$

and

$$\frac{\partial s_{kdt+T}}{\partial x_{jtd}} = -\rho^T \beta\, s_{jdt} s_{0dt} \;\forall k \neq j \tag{18}$$

in which case the long-run effects are

---

[67] That is because

$$\frac{\partial(\ln s_{jdt+T} - \ln s_{0dt+T})}{\partial x_{jtd}} = \frac{\partial(\ln s_{jdt+T} - \ln s_{0dt+T})}{\partial(\ln s_{jdt} - \ln s_{0dt})} \frac{\partial(\ln s_{jdt} - \ln s_{0dt})}{\partial x_{jtd}} = \rho^T \beta$$

but

$$\frac{\partial(\ln s_{jdt+T} - \ln s_{0dt+T})}{\partial x_{jtd}} = \frac{\partial \ln s_{jdt+T}}{\partial x_{jtd}} - \frac{\partial \ln s_{0dt+T}}{\partial x_{jtd}},$$

so that

$$\frac{\partial \ln s_{jdt+T}}{\partial x_{jtd}} - \frac{\partial \ln s_{0dt+t}}{\partial x_{jtd}} = \rho^T \beta$$

$$\frac{1}{s_{jdt+T}} \frac{\partial s_{jdt+T}}{\partial x_{jtd}} - \frac{1}{s_{0dt+T}} \frac{\partial s_{0dt+T}}{\partial x_{jtd}} = \rho^T \beta$$

$$\frac{\partial s_{jdt+T}}{\partial x_{jtd}} = \rho^T \beta\, s_{jdt+T} + \frac{s_{jdt+T}}{s_{0dt+T}} \frac{\partial s_{0dt+T}}{\partial x_{jtd}}$$

[68] See Filistrucchi(2005).

$$\sum_{s=0}^{\infty}\frac{\partial s_{jdt+s}}{\partial x_{jtd}}=\frac{1}{1-\rho}\beta s_{jdt}(1-s_{jdt}) \tag{19}$$

and

$$\sum_{s=0}^{\infty}\frac{\partial s_{kdt+s}}{\partial x_{jtd}}=-\frac{1}{1-\rho}\beta s_{jdt}s_{kdt} \quad \forall k\neq j\,. \tag{20}$$

Similar problems arise when calculating the long-run effects of a change in a dummy variable. But under the assumption of constant market shares the long-run own effect is

$$\sum_{s=t}^{+\infty}\frac{\Delta s_{jsd}}{\Delta x_{jtd}}=\left(\frac{\exp(\delta_{jtd}(1))}{1+\sum_{n\neq 0,J}\exp(\delta_{ntd}(0))+\exp(\delta_{jtd}(1))}-\frac{\exp(\delta_{jtd}(0))}{1+\sum_{n\neq 0}\exp(\delta_{ntd}(0))}\right)\frac{1}{(1-\rho)} \tag{21}$$

for any newspaper j , while the long-run cross effect is

$$\sum_{s=t}^{+\infty}\frac{\Delta s_{jsd}}{\Delta x_{ktd}}=\left(\frac{\exp(\delta_{jtd}(0))}{1+\sum_{n\neq 0,k}\exp(\delta_{ntd}(0))+\exp(\delta_{ktd}(1))}-\frac{\exp(\delta_{jtd}(0))}{1+\sum_{n\neq 0}\exp(\delta_{ntd}(0))}\right)\frac{1}{(1-\rho)} \tag{22}$$

for any newspaper j and

$$\sum_{s=t}^{+\infty}\frac{\Delta s_{osd}}{\Delta x_{ktd}}=\left(\frac{1}{1+\sum_{n\neq 0,k}\exp(\delta_{ntd}(0))+\exp(\delta_{ktd}(1))}-\frac{1}{1+\sum_{n\neq 0}\exp(\delta_{ntd}(0))}\right)\frac{1}{(1-\rho)} \tag{23}$$

for the outside good.

I estimate the dynamic model above by OLS. As discussed by Nickell (1981), including lags of the dependent variable, or more generally predetermined variables, in a fixed effects model leads to estimates which are inconsistent for n $\rightarrow\infty$ but consistent for $t \rightarrow \infty$. Given that in our case n=28 & t=312 the relevant asymptotics is that for $t \rightarrow \infty$ and OLS estimates are consistent.

Although they are usually recognised as endogenous and instrumented, I did not instrument prices. Given the already discussed lack of price competition among newspapers in Italy, I claim they can be considered exogenous or at least predetermined. It is a common assumption in discrete choice models of product differentiation to assume that product characteristics are exogenous or predetermined[69]. Given the evidence provided above, there is no reason to believe prices too cannot be considered econometrically exogenous or predetermined in this case.

There is finally a potential endogeneity issue to be discussed about the decision to open a website. As it is very difficult to find an instrument for the opening of a website, I cannot do much about it, except pointing out that it is anyway a one-time decision, not one that is repeated each period, so that econometrically it is in many periods predetermined.[70] Again not instrumenting for the website dummy is therefore as much right as not instrumenting for characteristics in any discrete choice model of product differentiation. In addition, as I discuss in Section 9, it is possibly an unavoidable decision, one that is mainly driven by the appearance of a new technology, which would have in any case been adopted, as shown by the data on news websites in Table 3.

8. RESULTS

Results from the OLS estimation of equation (14) are reported below in Table 5 below. In an aggregate logit model the estimated parameters are the taste parameters in the indirect utility function, which are assumed not to vary across consumers. Therefore a negative (positive) coefficient indicates that a given characteristic of the newspaper on average reduces (increases) readers' utility from reading it. But of course it also implies that the characteristic taken into consideration had a negative (positive) impact on the newspaper market shares.

The coefficient for the availability of a website has a negative sign and is significant. Therefore the availability of a website decreases mean consumer utility from reading that newspaper on paper[71], as it appears to have had, in general, a negative impact on its market shares. So that daily newspapers and their websites are to be understood,

---

[69] See for instance Nevo (2001) or Brenkers & Verboven (2002).

[70] Regarding the direction of the possible bias, if those who opened a website did it because their sales were already declining more (less) or where expected to decline more (less) than those who didn't, then the negative effect is probably overestimated (underestimated) and might not even be there (might anyway be there).

[71] This is however different from saying that the introduction of websites decreased consumers' welfare because in order to measure the change in consumers' welfare one would need to measure consumers' mean utility from reading the online newspapers.

at least at the aggregate level, as substitutes rather than complements or independent goods.

Also the estimated coefficient for the number of rivals' websites has a negative sign, though significant only at a 90% confidence level. Thus, the presence of another newspaper's website decreases mean readers' utility from reading a given newspaper. So that there is also a stealing effect from other daily newspapers' websites. Overall, the effect of opening a website is a substitution from all newspapers in the market to that newspaper's website.

The coefficient for real price, though small in size, is negative and significant at a 95% confidence level. The bundling of weekly generalist magazines to some of the newspapers appears to have had a positive impact on market shares both on the day of the week of issue and on the other days of the week, whereas the bundling of weekly women magazines does not appear to have had any effect[72].

Finally, the possibility to play games such as Lotto simply and only by buying a copy of the daily newspaper appears to yield a positive utility to consumers and therefore to have had a significant positive impact on newspapers market shares.

[72] See Argentesi(2004).

Table 5 – Logit estimates of taste parameters[73]

Explanatory variable	**Coefficient (standard error)**
Website	-0.0323*** (0.0082)
N° of rival newspapers' websites	-0.0162* (0.0085)
Real price	$-3.34*10^{-15}$** $(1.53*10^{-5})$
Games	+0.0325*** (0.0047)
Generalist supplement (day of issue)	+0.0373*** (0.0096)
Women's Supplement (day of issue)	+0.0212* (0.0116)
Generalist supplement (all days)	+0.0853*** (0.0063)
Women's Supplement (all days)	-0.0075 (0.0127)
ρ	+0.8820*** (0.0051)

dependent variable $\ln(s_{jtd}/s_{otd})$ *; *** 1% **5% * 10%;*
number of observations: 8160; number of regressors: 366 ; R^2 *: 0.9859*

From the estimated taste parameters, using the formulas reported above, I calculate the own and cross effect in number of copies sold of the decisions to go on-line of *Corriere della Sera*, *La Repubblica* and *La Stampa*.

[73] Estimated coefficients and standard errors for editors dummies, newspaper-day and time fixed-effects are not reported.

Table 6 – Average short-run own and cross effects of a website (number of copies lost in a day)

Effect of (row) On (column)	Corriere della Sera	La Repubblica	La Stampa	Il Giornale	Outside Good
corriere.it	-19,830	-8,743	-5,041	-3,415	+37,029
	(6,140)	(5,766)	(3,318)	(2,258)	(15,172)
repubblica.it	-9,627	-17,988	-5,042	-3,416	+36,073
	(6,050)	(5,555)	(3,308)	(2,254)	(14,966)
lastampa.it	-7,567	-8,979	-10,707	-3,486	+30,738
	(5,017)	(5,591)	(3,315)	(2,173)	14,514

bootstrapped standard errors in parenthesis

Table 7 – Average approximated long-run own and cross effects of a website (number of copies lost in a day)

Effect of (row) On (column)	Corriere della Sera	La Repubblica	La Stampa	Il Giornale	Outside Good
corriere.it	-168,123	-74,123	-42,744	-28,957	+313,947
	(53,022)	(50,087)	(28,814)	(19,609)	(132,555)
repubblica.it	-81,623	-152,509	-42,751	-28,959	+305,841
	(52,525)	(47,963)	(28,721)	(19,579)	(130,702)
lastampa.it	-64,152	-76,131	-90,774	-29,553	+260,610
	(43,510)	(48,491)	(28,616)	(18,876)	(126,434)

bootstrapped standard errors in parenthesis

Table 6 reports the average short-run effects on sales, whereas the average approximated long-run effects are reported in Table 7. Also reported are bootstrapped standard errors for these averages obtained by 1,000 draws with repetition from the data[74]. The opening of an own website is estimated to have caused a short-run loss in sales of 19,830 to *Corriere della Sera*, of 17,988 to *La Repubblica* and of 10,707 to *La Stampa*. These losses are on average approximately 3.1% of their sales (with a bootstrapped standard error of.0.9%). The estimated long-run losses are instead 168,123, 152,509 and 90,774 respectively, on average approximately 26.4% of their sales. Also losses due to a rival's website are substantial: around 1.5% (s.e.0. 95-0.99%) in the short-run and 12.8% in the long-run.

[74] Given that for each product the model predicts a different effect at each time t on each day of the week d the reported estimates are averages over more than 400 observations. The latter is one of the reasons why the bootstrapped standard errors are so high.

Clearly, losses in sales imply losses in sales revenues. However, as first pointed out by Corden(1952) and Reddaway(1963), a newspaper publisher maximizes profits obtained by selling both newspapers and advertising slots on them, taking into account possible indirect externalities between the two markets, particularly the one whereby the more readers a newspaper has the highest the demand for advertising space[75]. Consistently Chapter 2 find a significant coefficient for circulation in the aggregate demand equation for advertising. As a result, the estimated loss in sales is likely to have lead ceteris paribus to a loss in advertising revenues on paper. Table 8 below reports the average profit margin per printed copy for each year in the period 1997-2001. The margin is calculated by dividing the total annual revenues from sales and advertising of each newspaper[76] in my sample by the number of copies sold, subtracting from the revenue per copy of each newspaper the average cost per copy in paper and other non durable material[77] for daily newspapers with an average circulation higher than 200,000 a day[78] and then averaging across the four newspapers.

Table 8 – Average profit margin (per printed copy)

Year	**Average profit margin**
1997	1,640
1998	1,695
1999	1,884
2000	2,135
2001	1,775

Source: FIEG, ADS and Deloitte&Touche&FIEG
unit of measure: 1995 Italian lire

[75] The daily newspaper market is therefore a two-sided market. See Argentesi&Filistrucchi(2005).
[76] These are reported in FIEG (1986-2002).
[77] As in Argentesi&Filistrucchi(2005) I choose costs net of both labour costs (which include journalists and editors) and other services costs (which include also freelance journalists) as they should not affect the marginal cost of a copy.
[78] These are reported in Deloitte & Touche & FIEG (1993-2002). They also report the average cost across national newspapers. However, given that the other national newspapers are very small and might thus have very different costs, I prefer to use the average cost for daily newspapers with circulation higher than 200,000. The latter however also include sport newspapers such as *La Gazzetta dello Sport* and business ones such as *Il Sole24 Ore*.

An approximation to the short and long-run losses in profits on the traditional market can then be calculated by multiplying the estimated average daily losses in sales by the average profit per copy. The estimated losses in profits per newspaper issue are the ones reported in Tables 9 and 10.

Table 9 – Average short-run own and cross effects of a website from 1997 to 2001 (profits lost in a day)

Effect of (row) on (column)	Corriere della Sera	La Repubblica	La Stampa	Il Giornale
corriere.it	-51,505,421	-22,412,560	-15,218,082	-9,709,789
repubblica.it	-25,029,742	-46,206,001	-15,236,930	-9,721,968
lastampa.it	-25,161,181	-22,558,397	-31,448,533	-9,772,163

unit of measure: 1995 Italian lire

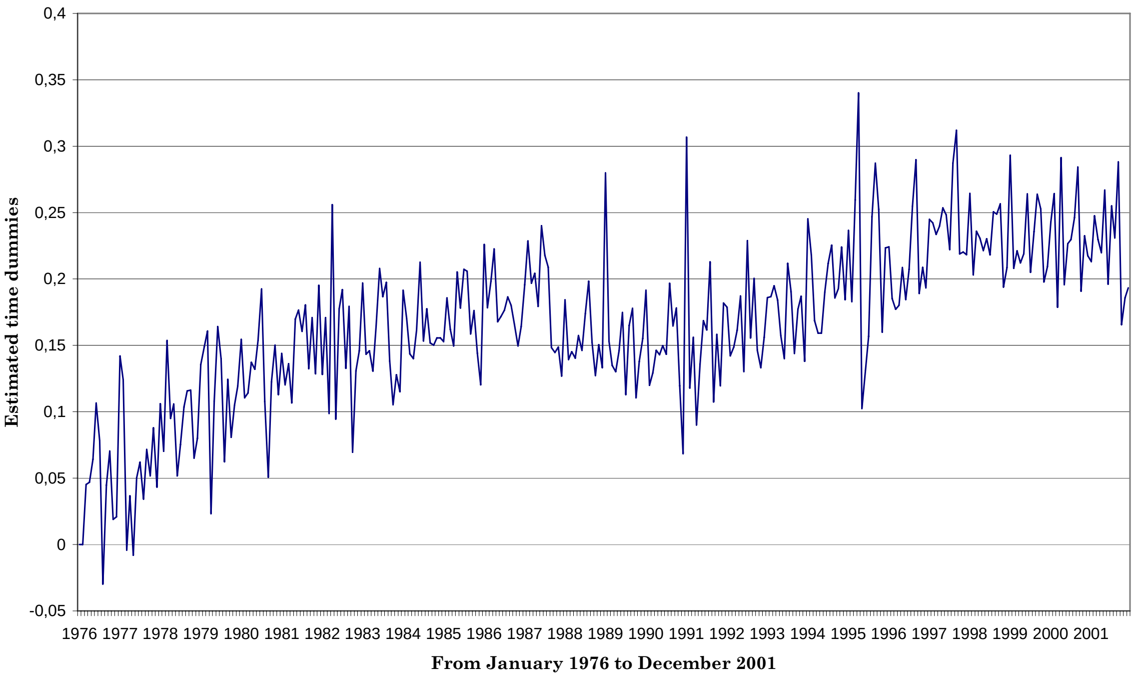
Table 10 – Average approximated long-run own and cross effects of a website from 1997 to 2001 (profits lost in a day)

Effect of (row) on (column)	Corriere della Sera	La Repubblica	La Stampa	Il Giornale
corriere.it	-436,680,350	-190,021,251	-129,024,034	-82,322,866
repubblica.it	-212,210,598	-391,750,071	-129,183,838	-82,426,127
lastampa.it	-213,324,988	-191,257,703	-266,631,281	-82,851,700

unit of measure: 1995 Italian lire

As discussed above, thanks to the use of time fixed-effects, in as much as it is possible to assume that the general availability of news on line has the same negative percentage impact on the market shares of all the newspapers in the sample, the ones reported in Tables 6 and 7 are the estimated losses in sales of a traditional newspaper due to people who shift to reading the same newspaper website or a competitor's website (effect 3 discussed in section 1). I cannot instead identify the other substitution effects (effects 1 and 2), that is those due either to people allocating less time to reading (and thus not buying newspapers) in order to surf the web or to people not buying newspapers as they prefer to read news via Internet. These are however captured, among many other things, by the time dummy variables. Figure 4 shows a graph through time of the estimated time fixed effects. Although in the whole sample period they are not a monotonic function of time, they appear to have been slowly declining since the second half of 1997[79], just as internet news have been growing considerably (see Table 3 above). The implied increase in the utility of the outside good might thus be due also to the appearance of Internet.

Figure 4 - Estimated changes in the utility of the outside good



9. CONCLUSIONS

All in all, Internet appears to have had a negative impact on the sales of the four main national daily newspapers in Italy. So that news on line and daily newspapers appear to be substitute goods at least at the aggregate level. But of course it is not possible to rule out that for some people they are not only independent goods but even complements, in which case my analysis simply suggests that the number of those who consider them substitutes is higher than those for whom they are complements. Also the estimated long-run, albeit approximated, effect does not support the idea that they will disappear, but rather predicts a substantial drop in sales, all other things being equal (including the absence of on-line fees).

[79] This might be due also to the fact that in November 1997 costs of phonecalls to connect to the Internet were cut substantially as a consequence of a specific government decree which aimed at increasing Internet use in the country. More or les in the same period many portals started to offer free internet access, e-mail and web space.

These findings are qualitatively in line with those of Simon (2005) for the magazine market in the US and those by Gentzkow (2005) for the daily newspaper market in Washington DC. They are also not inconsistent with Kaiser (2003) who analyses the women's magazine market in Germany. I believe the reason why I find such a significant negative effect lies in the particular feature of the Italian daily newspaper market in the period I consider, as publishers were making available online for free the exact articles published on paper. As a result the degree of substitutability between the traditional newspapers and its website was at its maximum, as suggested also by Simon (2005). For women's magazines in Germany instead the features of the websites seem to suggest a lower degree of substitutability and possibly leave more room to complementarity between the two products. The reason why my estimates are higher than Gentzkow (2005) might instead be, in addition to analysing a different market, that he cannot fully exploit the time dimension in order to identify the effect of websites.

Some caution is needed when discussing the implication of the finding that for a daily newspaper opening a website leads to a loss in sales. As discussed in the previous section such a loss will surely lead not only to a loss in sales revenues but also to a loss in advertising revenues. Yet opening a website might still have been a profitable choice or at least a rational one. If so, possible explanations for why traditional newspapers publishers opened websites and even put online for free the exact content of the paper edition could therefore be a) the prediction (or at least the expectation) that they would be able to raise on-line advertising revenues that compensate or more than compensate the costs of on-line publishing, the loss in paper sales and that in advertising revenues of the paper edition[80] or b) the attempt to establish a position in the new market or to build up consumers loyalty, especially among young people, in the expectation of a shift of news consumption from paper to the web and with the objective to set a fee for newspapers on-line once they succeeded in this. Even if predicted (or expected) not to be profitable, because overall on-line revenues do not cover the costs of on-line publishing, the loss in paper sales and in advertising, the choice to open a website might still have been a rational one, if c) publishers expected in any case a shift of consumption to the new media and therefore chose to minimize loss by attracting to their website part of the readers who would anyway have

[80] There might also be an additional loss in advertising sales if also online and on-paper advertising are substitutes, an issue which has not been analysed yet in the empirical literature.

switched to on-line reading. The latter explanation is in particular consistent with the estimation of a negative effect of rival newspapers' websites.

These reasons, though not necessarily only one of them, might explain why *La Repubblica* set up a fee for reading online the paper edition from January 2002, five years after the first opening of the website, and why its example was soon followed by many other daily newspapers, including *La Stampa* and, in part, *Corriere della Sera*. More generally, this choice appears part of a new business strategy of daily newspapers in Italy after 2001[81], a strategy which seems to build on a) attempting to reduce substitutability, while increasing complementarity by adding additional related content and services to the website, and b) trying to get additional on-line revenues, not only through fees but also by enjoying the higher advertising value of those on-line readers who register in order to access on-line the paper edition of the newspaper.

[81]This tendency is not only a feature of the Italian market. From September 2005 the New York Times started to charge visitors to access part of its online content: the editorials and the archive. The announcement set the end of the New York Times exception in the U.S.

REFERENCES

Anderson Simon P., de Palma André & Thisse Jacques-François, Discrete Choice Theory of Product Differentiation, Cambridge (Massachusetts)-London (UK): The MIT Press, 1992.

Argentesi Elena, Demand Estimation for Italian Newspapers: The Impact of Weekly Supplements, European University Institute Economics Working Paper n. 28, 2004.

Argentesi Elena & Filistrucchi Lapo, Estimating Market Power in a Two-Sided Market: the Case of Newspapers, European University Institute Economics Working Paper n.7, 2005.

Barsh J., Shao-Chi L. & Miles A., Beyond print: A future for magazines, *The McKinsey Quarterly*, 3, 1999.

Berry S., Estimating Discrete-choice Models of Product Differentiation, The *RAND Journal of Economics*, 1994, 25(2), 242-262.

Berry S., Levinsohn J. & Pakes A., Automobile Prices in Market Equilibrium, *Econometrica*, 1995, 63, 841-890

Blair Roger D. & Romano Richard E., Pricing Decision of the Newspaper Monopolist, *Southern Economic Journal*, 1993, 59 (4), 721-32.

Booth, D. L., Kanetkar V., Vertinsky I. & Whistler D., An Empirical Model of Capacity Expansion and Pricing in an Oligopoly with Barometric Price Leadership: A Case Study of the Newsprint Industry in North America, The Journal of Industrial Economics, 1991, 39(4),255-276.

Brenkers Randy & Verboven Frank, Liberalizing a Distribution System: the European Car Market , *CEPR Discussion Paper n◦ 3622*, September 2002.

Cameron G.T, Hollander B.A., Nowak G.J. & Shamp S.A., Assessing the potential of a full-featured electronic newspaper for the young adult market, *Media Management Review*, 1997,15-28.

Cameron, G. T., Curtin, P. A., Hollander, B. A., Nowak, G. J., & Shamp, S. A., Electronic Newspapers: Toward a Research Agenda, *Journal of Mediated Communication,* 1996, *11*, 4-53.

Canoy Marcel & de Bijl Paul, Publishers Caught in the Web?, *CPB Netherlands Bureau for Economic Policy Analysis Working Paper n°119*, March 2000.

Cecchetti Stephen, The Frequency of Price Adjustment: A Study of the Newsstand Prices of Magazines, *Journal of Econometrics*, August 1986, 31, 255-274.

Censis, Primo rapporto annuale sulla comunicazione in Italia. Offerta di informazione e uso dei media nelle famiglie italiane, Censis, 2001

Censis, Comunicazione e Cultura in Censis, Rapporto sulla Situazione Sociale del Paese, Censis, 1961-2002

Choi Soon-Young, Stahl Dale O. & Whinston Andrew B., The Economics of Electronic Commerce, Indianapolis, Indiana: Macmillan Technical Publishing, 1997.

Deloitte & Touche & FIEG, Indagine sui bilanci delle imprese editrici di quotidiani, FIEG, 1981-2002.

Dertouzos James N. & Trautman William B., Economic Effects of Media Concentration: Estimates from a Model of the Newspaper Firm, *Journal of Industrial Economics*, September 1990, 39 (1), 1-14.

Dewenter Ralf, Rational Addiction to News? Habit Formation and Print Media Usage, mimeo, University of Essen, February 2002.

FIEG, *La Stampa* in Italia, FIEG, 1982-2002.

Filistrucchi Lapo, On Dynamics in Discrete Choice Models of Product Differentiation with Market Level Data, mimeo, European University Institute, August 2005.

Fisher Timothy C.G. & Konieczny Jerzy D., Synchronization of price changes by multiproduct firms: evidence from Canadian newspaper prices, Economics Letters, 2000, 68, 271-277.

Gentzkow Matthew, Valuing New Good in a Model with Complementarity: Online Newspapers, mimeo, Chicago Graduate School of Business, May 2005.

Hakfoort Jacco & Welgand Jürgen, Magazine Publishing-A Quiet Life? The Dutch Market for Consumer Magazines, *CPB Netherlands Bureau for Economic Policy Analysis Working Paper n°120*, March 2000.

Kaiser Ulrich., The Effects of website provision on the demand for German women's magazines, Centre for Economic and Business Research Discussion Paper n° 2003-01, 2003.

Kaiser Ulrich & Kongsted Hans Christian, Do Magazines' "Companion Websites" Cannibalize the Demand for the Print Version?, CAM Working Papers n° 7, 2005.

McFadden Daniel & Train Kenneth, Mixed MNL Models for Discrete Response, mimeo, University of California at Berkeley, May 2000.

Mings Susan M. & White Peter B., Profiting from Online News: The Search for Viable Business Models in *Kahin Brian & Varian Hal R., Internet Publishing and Beyond-The Economics of Digital Information and Intellectual Property, Cambridge, USA: MIT, 2000.*

Mosconi Franco, Economia dei Quotidiani, Bologna: Il Mulino, 1998.

Nevo Aviv, A Practitioner's Guide to the Estimation of Random-Coefficients Logit Models of Demand, Journal of Economics & Management Strategy, 2000,9 (4), 513-548.

Nevo Aviv, Measuring Market Power in the Ready-to-eat Cereal Industry, *Econometrica*, 69(2), March 2001, 307-342.

Nickell S., Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1981, 1417-1426.

Petrin Amil & Train Kenneth, Omitted Product Attributes in Discrete Choice Models, mimeo, University of Chicago & University of California at Berkeley, December 2002.

Petrin Amil, Quantifying the Benefits of New Products: the Case of the Minivan, Journal of Political Economy,1110(4),2002,705-729.

Reekie W. Duncan, The price elasticity of demand for evening newspapers, *Applied Economics*, 1976, 8, 69-79.

Rosse James N., Estimating Cost Functions Parameters without Using Cost Data: Illustrated Methodology, *Econometrica*, 38, 1970, 255-275.

Rosse James N., Owen B.M. & Dertouzos James N., Trends in the Daily Newspaper Industry, 1923-1973, *Stanford University Studies in Industry Economics* n° 99, 1975.

Rosse James N., The daily newspaper firm: a 24 equation reduced form model, *Stanford University Studies in Industry Economics* n° 76, 1977.

Rosse James N., The evolution of one newspapers cities, *Stanford University Studies in Industry Economics* n° 95, 1978.

Rysman Marc, Competition Between Networks: A Study on the Market for Yellow Pages, *Review of Economic Studies*, 71 (2), 483-512.

Simon, Daniel H., The Effect of a Magazine's Digital Content on its Print Circulation: Cannibalization or Complementarity?, Cornell University mimeo, 2005.

Train Kenneth, Qualitative Choice Analysis, Cambridge (Massachusetts)-London (UK): The MIT Press,1993.

Train Kenneth, Discrete Choice Methods with Simulation, Cambridge (Massachusetts); New York (NY) : Cambridge University Press, 2003.

Willis Jonathan, Estimation of Adjustment Costs in a Model of State-Dependent Pricing, *Federal Reserve Bank of Kansas City Research Working Paper n°7*, December 2000.

Unbundling and Incumbent Investment in Quality Upgrades and Cost Reduction*

João Vareda†
Faculdade de Economia
Universidade Nova de Lisboa

March, 2007

Abstract

We study the investment of a telecommunications incumbent in quality and in cost reduction when an entrant can use its network through unbundling of the local loop. We find that unbundling may lower incentives for quality improvements, but raises incentives for cost reduction. Therefore, it is not true that all types of investment are crowded out with unbundling. If the regulator can commit to a socially optimal unbundling price before investment, the incumbent makes both types of investment. In the absence of commitment, the incumbent will not invest, so that unbundling regulation may lower welfare as compared to no regulation.

JEL classificafion: L41, L43, L51, L96

Keywords: Telecommunications Regulation; Unbundling; Investments; Commitment.

---

*I gratefully acknowledge the helpful comments from my PhD adviser Steffen Hoernig. I would also like to thank the comments and discussions at the CEPR Applied IO School conference (Funchal, May 2006), at the QED meeting (Vienna, May 2006) and at the EARIE conference (Amsterdam, August 2006). Financial support from Fundação para a Ciência e Tecnologia and Fundo Social Europeu within the III Quadro Comunitário de Apoio BD/10280/2002, and project POCTI/ECO/44146/2002 of FCT and FEDER.

†Address: Faculdade de Economia, Universidade Nova de Lisboa, Campus de Campolide, 1099-032 Lisboa, Portugal; Tel: +351 21 380 16 00; Fax: +351 21 387 09 33; E-mail: jvareda@fe.unl.pt

1 Introduction

In the last 10 years, mandatory unbundling has become a standard remedy proposal for solving the bottleneck problem in fixed telecoms competition. Since there are high entry barriers in the telecommunications market, because of scale economies, sunk costs and first-mover advantages, it is hard for a new operator to enter the market as a full-facility competitor. In particular, the building of local access networks, which are composed of circuits connecting end users to switches located in central offices, requires large investments in terms of money and time.

Under mandatory unbundling an incumbent firm has to share the use of some of its facilities with its competitors. This implies that an essential input is, at the wholesale level, separated from the incumbent's overall facilities, in order to allow for commercial wholesale supply of this input. In the particular case of local loop unbundling, this means that a new operator can directly plug into the incumbent's network by creating a connection from its switch to the incumbent's local access network (Figure 1). This policy is supposed to generate entry in the market, and to encourage entrants to build their own network in the future when their stock of costumers is large enough.

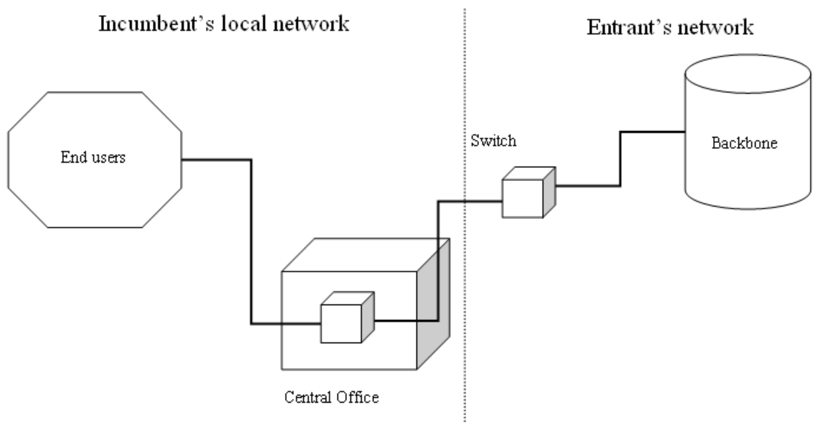


Figure 1

Mandatory unbundling is promoted both in the United States and the European Union. In the US, the Telecommunications Act of 1996 requires that incumbents unbundle their networks.[1] Incumbents and entrants are then required to negotiate a price for the entrant's use of an unbundled element.

[1] In its Local Competition Order in August 1996, the FCC specified seven unbundled elements: local loops, network interface devices, local and tandem switching, interoffice transmission facilities, signaling networks, operations support systems and operator services and assistance.

If they are unable to reach an agreement, the price is determined by the regulator. The calculation of the regulated prices is guided by a framework called Total Element Long Run Incremental Cost (TELRIC). This follows a forward looking methodology based on the assumption that an efficient and modern network is in place.[2]

In the EU, the New Regulatory Framework (NRF) of 2003 embraces the view that regulation should be used to actively promote service-based competition by facilitating access to existing infrastructure.[3] But even before the NRF, European legislation mandated the provision of unbundled access.[4] While recognizing that infrastructure competition is the primary means to attain sustainable competition in telecommunications because it increases the pressure to minimize costs and induces a higher scope for innovation, the European Commission sustains that service competition is a necessary pre-requisite for infrastructure competition. According to the Commission, competition would never be able to develop in the short term if entrants were not able to gain access to the incumbent's network.

Service-based competition promoted by unbundling has been criticized on the basis that it only promotes static efficiency. The main argument is that incumbents would not have incentives to invest if they had to share the benefits of their investments with rivals. Moreover, if access to the incumbents' network was allowed too easily, this would create inefficiencies in the long run since an entrant would not have incentives to build competing facilities (see Jorde et al., 2000).

Partially as a response to these arguments, several empirical studies analyzing the effect of unbundling on incumbent firms' investment have emerged. For instance, Willig et al. (2002) examine the relationship between unbundling prices and Bell companies' investments. They test two opposite hypotheses. The first is the investment deterrence hypothesis, according to which a low unbundling price encourages utilization by the entrants and, as a consequence, the incumbents invest less. The second one is the competitive stimulus hypothesis, according to which a low unbundling price encourage entry, and this increased competition strengths the incumbents' incentive to invest. Their results support the second hypothesis, and therefore they conclude that lower unbundling prices stimulate incumbents' investment. A

[2]See more about the Long Run Incremental Cost methodology in Laffont and Tirole (2000).

[3]Directive 2002/19/EC of the European Parliament and of the Council of 7 March 2002 on access to, and interconnection of, electronic communications networks and services.

[4]Regulation 2887/2000 of the European Parliament and of the Council of 18 December 2000 on unbundled access to the local loop.

study by Hassett et al. (2003) obtains similar conclusions.

However, the Willig et al. (2002) result is not without controversy. Haring et al. (2002) criticize Willig's estimation methodology and develop their own econometric model. They obtain the opposite relationship, i.e., low unbundling prices reduce the profitability of incumbents' investment leading to a reduction in that investment. Hausman and Sidak (2005) corroborate this opinion in their case study about the unbundling experience in the US, New Zealand, Canada, United Kingdom, and Germany. Gabel and Huang's (2003) econometric results indicate that in the US the higher the unbundling price, the more likely is the introduction of new services by the incumbents. Ingraham and Sidak (2003) show that mandatory unbundling increases the volatility of the incumbents' stock returns, which increases their equity cost.

In 2003 a new controversy has emerged after the publication of the Phoenix Center Policy Bulletin n^{o}5, which shows that the rise in unbundling lines has increased investment by incumbents. This gave origin to two replies, one by Hazlett et al. (2003) on behalf of Verizon, and another by Hill (2003) on behalf of Z-Tell-Communications, both contesting the empirical estimation and arguing that the rise in unbundling lines has led to a decline in incumbents' investment. As a response, the Phoenix Center published its Policy Bulletin n^{o}6 which, by incorporating the comments of the two replies, shows that its previous result was robust.

Finally, there is also a study by Chang et al. (2003) that finds, using US data, that lower access rates have spurred investment in digital systems by incumbent local carriers. Even so, the same study points in the opposite direction for Europe.

We can conclude that there is an unresolved controversy about the true effects of unbundling prices in incumbent's investments and, following from this, what the regulated unbundling price should be. In this paper we will focus on these two points distinguishing between investment in cost reduction and in quality upgrades.

In contrast with the large amount of research on static access pricing (Armstrong, 2002), the dynamic study of optimal access pricing is still in its early stages. Valletti (2003) reviews the existing literature about the relationship between access pricing and investment, and provides a discussion about investment incentives by relating them with questions common to R&D.

Foros (2004) shows that under some conditions the investment level in quality is lower with price regulation, since the access price is set equal to marginal cost. Kotakorpi (2006) considers a similar model with vertical differentiation, and obtains similar results. Cambini and Valletti (2004) study the impact of access charges on the incentives to invest, but in a context of two-way access. They derive the result that firms would choose a price above marginal cost in order to diminish each other's incentives to invest. In addition, there are some papers that consider cost-reducing investments, as Biglaiser and Ma (1999), Cabral and Riordan (1989) and Sappington (2002), the first in a context of an incumbent firm and the other two in monopoly. Vareda and Hoernig (2006) study both the incumbent and entrant's investment in the building of a new network.

In our paper, we develop a theoretical model with two operators that offer differentiated services, and try to explain the relationship between the unbundling price and the investment made by the incumbent. Since it is a partial consumer participation model, it portrays non-mature markets, such as the broadband market. Bourreau and Dogan (2005) assume full consumer participation represented by a Hotelling model. In this model profits are insensitive to the unbundling price for a large interval of unbundling prices, which does not seem to be reasonable in the context of investment choice.

The main contribution of our model is the comparison of the incumbent's incentives for two different types of investment: quality-upgrades and cost-reduction. We show that, although these investments are complements, the direct effect of the unbundling price on each one differs. Indeed, a lower unbundling price decreases incentives for quality improvements, but raises incentives for cost reduction. This follows from the fact that, for a lower unbundling price, the incumbent wants to maintain its competitive advantage. Thus, it has more incentives to invest in cost reduction increasing its cost-advantage. On the other hand, it has less incentives to invest in quality upgrades because this benefits both operators. In equilibrium, we always have a higher investment in cost reduction for a lower unbundling price, while investment in quality can be higher or lower due to the complementarity relationship.

We also determine the socially optimal unbundling price. First, we assume that the decision about the unbundling price is taken before investment and that the regulator commits to it until the end of the game. We show that the regulator sets a higher unbundling price when the cost of improving quality is relatively low, in order to give incentives for this type of investment. When cost reduction is less expensive, then the unbundling price the regulator should set is lower. We contrast these results with a context where the regulator cannot commit to his decisions and revises the unbundling price after the investment has been made. In this case, the incumbent does not invest since the regulator sets a price such that it always earns zero profits. Social welfare is then lower in a no-commitment context.

We compare both contexts with an unregulated market. We show that the incumbent has always incentives to unbundle its infrastructure in order to attract new consumers to the market.[5] This is always worse than the context where the regulator sets the unbundling price before investment as the price set by the incumbent is too high, but it can be better than a no-commitment context since there is some investment. Therefore, we conclude that the unbundling problem raised by some authors is more a problem of commitment rather than unbundling as such.

Finally, we provide a short analysis of the case of mature markets. In these markets, the investment in quality upgrades increases with the unbundling price, while the investment in cost reduction is independent of it. A relevant example is fixed telephony.

The remainder of the paper is organized as follows. We describe the model in Section 2. In Section 3 we obtain the equilibrium prices and quantities, and in Section 4 we find the profit-maximizing investments. Then, in Section 5, we solve the regulator's problem. In Section 6 we find the unregulated market equilibrium and compare it with the regulated contexts, and in Section 7 we analyze a mature market. Finally, in Section 8 we conclude.

2 The Model

We introduce a model of a telecommunications market, where two firms compete on subscription prices and supply horizontally differentiated services. The operators on this market are: one vertically integrated network (denoted as incumbent) which owns the local loop, and one non-integrated network (denoted as entrant) which only owns a backbone and switches, and needs access to the incumbent's local loop.

We introduce a third party, a regulator, who sets the unbundling price in order to maximize social welfare. We assume that the unbundling price is the only instrument available for the regulator. This corresponds closely to the current European practice. First, we consider a context where the regulator fixes the unbundling price at the beginning of the game. Later, we consider a context where the regulator only takes the final decision about the unbundling price after the investment stage. We adopt the simplifying assumption of complete information, i.e., the regulator is supposed to have full information about the incumbent's technology and costs.

[5] Some incumbents have voluntarily entered into agreements with entrants for unbundled access. For example, Verizon and Covad in the US.

We assume that the incumbent can invest in its infrastructure both to increase quality and to reduce cost. We also assume that there is no uncertainty about returns on investment, i.e., for a given amount of expenditures on investment a given effect is obtained for sure.

After observing the price set by the regulator and the investment made by the incumbent, the entrant decides if it asks for access to the local loop. In this paper we exclude the possibility of entering as a facility-based competitor.

Demand side

In a telecommunications market consumers usually subscribe services from only one operator, thereby, they face a discrete decision problem of which operator to subscribe to. However, if we aggregate the demand of all consumers and divide by their number, we obtain the demand of a representative consumer who subscribes services from both operators. We can then use a quasi-linear consumer surplus function similar to Bowley (1924):

$$U\left(q_I, q_E\right) = a_I q_I + a_E q_E - \frac{1}{2} b \left(q_I^2 + 2\theta q_I q_E + q_E^2\right) - p_I q_I - p_E q_E, \tag{1}$$

where a_i is the reservation price for service i, and θ indicates the degree of substitutability. When θ is higher the services are stronger substitutes. In the extreme case, when $\theta = 0$ we have independent services, and when $\theta = 1$ we have perfect substitutes. (q_I, p_I) is the number of subscribers and the subscription price of the incumbent, while (q_E, p_E) is the number of subscribers and the subscription price of the entrant. For simplicity, we assume that $b = 1$.

If we solve the representative consumer's problem:

$$\max_{q_I, q_E} U\left(q_I, q_E\right), \tag{2}$$

we obtain the following demand functions:

$$q_I = \frac{a_I - \theta a_E - p_I + \theta p_E}{1 - \theta^2} \tag{3}$$

$$q_E = \frac{a_E - \theta a_I - p_E + \theta p_I}{1 - \theta^2} \tag{4}$$

(these expressions are valid provided that $\theta < 1$ and that both $q_I, q_E \geq 0$).

We assume partial consumer participation. Contrary to the Hotelling model often used in literature, consumer participation depends on price, which creates the opportunity to consider welfare effects neglected by it.

Supply side

Regarding the incumbent's cost structure, we assume that cost per subscriber is just a constant marginal cost c. For simplicity, irrespective of being the incumbent or the entrant that sells the services to subscribers, we assume the incumbent's marginal cost per subscriber to be the same.

If the entrant decides to ask for access to the incumbent's local loop, the incumbent receives from the entrant a price r per rented line (unbundling price). For its own retail services it receives a subscription price p_I per consumer. Given these, in the absence of investment, incumbent's profit is equal to:

$$\pi_I = (p_I - c)\, q_I + (r - c)\, q_E. \tag{5}$$

The entrant receives a subscription price p_E from its customers and pays the correspondent unbundling price to the incumbent. Hence, its profit is:

$$\pi_E = (p_E - r)\, q_E. \tag{6}$$

The regulator maximizes social welfare, which is the following:

$$W = \pi_I + \pi_E + CS, \tag{7}$$

where CS is consumer surplus.

Throughout, we make the following assumptions:

Assumption 1: Reservation prices are equal for both operators and higher than marginal cost:

$$a_I = a_E = a > c.$$

Assumption 2: Firms only operate in the market if they have non-negative profit:

$$\pi_I \geq 0 \ , \ \pi_E \geq 0.$$

According to Assumption 1, if it asks for access, the entrant is restricted to providing services with a quality equal to the incumbent's. This happens because it depends on the incumbent's infrastructure, thus it is not able to supply services that the incumbent could not supply, too.

As we are not considering in our model questions related with foreclosure, we assume that the entrant has already incurred in a sunk cost of entry. Thus, it asks for access if it is able to obtain non-negative retail profits.

Investments

As we have said, the incumbent has the possibility to invest in its network. We will consider two types of investment: quality-upgrades and cost reduction.

In the first case, we assume that the investment increases the reservation price by g, which implies a parallel shift in both demand functions. In fact, as the entrant supplies its services through the incumbent's local network, it also benefits from this investment, consequently, the reservation price for its services also increases in g. We assume that spillovers are complete, contrary to Foros (2004) and Kotakorpi (2006).[6] An example of this kind of investment is an upgrade of the switching equipment or the installation of new fibre optic cables, which allows to increase the velocity of the transmission or the capacity to deliver voice and data traffic.

The investment cost function is quadratic and given by:

$$C_q(g) = \frac{\alpha}{2} g^2. \tag{8}$$

The second type is an investment to decrease the cost of providing the services by turning the local network more efficient and reliable. Since the entrant uses the incumbent's lines to supply the services to its subscribers, it is the incumbent that supports all the operating costs. Consequently, if the incumbent invests in cost reduction, the marginal cost of supplying all consumers is reduced, no matter whether they are the incumbent's or the entrant's. Since the incumbent has constant marginal cost c, the innovation represents a decrease of h in marginal cost.

The investment cost function is also quadratic:

$$C_c(h) = \frac{\beta}{2} h^2. \tag{9}$$

Timing of the game

1. The regulator fixes the unbundling price.
2. The incumbent decides how much to invest in its infrastructure.
3. The entrant decides if it asks for access.
4. Firms compete in prices.

When the regulator cannot commit to his decision the order of the first two moves is reversed.

We now find the Subgame-Perfect Equilibrium using backward induction.

[6] In these papers, an investment increases the willingness to pay for services, but the dimension of the effect depends on the ability of each operator to transform input to output. Thus, we can have the incumbent offering higher quality services, and vice-versa.

3 Price competition stage

For given reservation prices (and therefore quality levels) and cost levels, using demand functions (3) and (4), and maximizing profit with respect to subscription price, we obtain the following Nash-equilibrium prices and quantities of the price competition stage (see Appendix A):

$$p_I - c - h = \frac{1-\theta}{2-\theta}(x_0 + g + h) + \frac{3\theta}{4-\theta^2}(y_0 + h) \tag{10}$$

$$p_E - r = \frac{1-\theta}{2-\theta}(x_0 + g + h) - 2\frac{1-\theta^2}{4-\theta^2}(y_0 + h) \tag{11}$$

$$q_I = \frac{1}{2-\theta^2+\theta}(x_0 + g + h) - \frac{\theta}{4-\theta^2}(y_0 + h) \tag{12}$$

$$q_E = \frac{1}{2-\theta^2+\theta}(x_0 + g + h) - \frac{2}{4-\theta^2}(y_0 + h), \tag{13}$$

where $x_0 = a - c$ and $y_0 = r - c$.

Parameter y_0 represents the incumbent's *ex ante* access margin. It can also be interpreted as the entrant's cost disadvantage, since the entrant has to pay r for each line while the incumbent only incurs a cost of c. In the future we will work with y_0 when we want to find the optimal unbundling price.

Firms' profit and welfare become:

$$\pi_I = \frac{1}{(2-\theta)^2}\left(\frac{1-\theta}{1+\theta}(x_0+g+h)^2 - \frac{8+\theta^2}{(2+\theta)^2}(y_0+h)^2 \right. \tag{14}$$
$$\left. + \frac{4-2\theta+\theta^2}{2+\theta}(x_0+g+h)(y_0+h)\right) - \frac{\alpha}{2}g^2 - \frac{\beta}{2}h^2.$$

$$\pi_E = \frac{1-\theta}{(2-\theta)^2}\left(\frac{1}{1+\theta}(x_0+g+h)^2 + 4\frac{1+\theta}{(2+\theta)^2}(y_0+h)^2 \right. \tag{15}$$
$$\left. - \frac{4}{2+\theta}(x_0+g+h)(y_0+h)\right)$$

$$W = \frac{1}{(2-\theta)^2}\left(\frac{3-2\theta}{1+\theta}(x_0+g+h)^2 - \frac{1}{2}\frac{4+5\theta^2}{(2+\theta^2)^2}(y_0+h)^2 \right. \tag{16}$$
$$\left. - (1-\theta)(x_0+g+h)(y_0+h)\right) - \frac{\alpha}{2}g^2 - \frac{\beta}{2}h^2.$$

These functions are only valid if the entrant asks for access.

Lemma 1 *The entrant only asks for access if:*[7]

$$y_0 \leq \overline{y}_0 \equiv \frac{1}{2}\frac{2+\theta}{1+\theta}\left(x_0+g+h\right)-h. \tag{17}$$

Proof. Equivalent to $\pi_E \geq 0$. ■

Note that when services are more substitutable, i.e., when θ is close to 1, the entrant's equilibrium profit tends to zero and the incumbent's equilibrium profit tends to the monopolist's profit at $p_I = r$.

In monopoly, profit and welfare become:

$$\pi_I^M = \frac{1}{4}\left(x_0+g+h\right)^2 - \frac{\alpha}{2}g^2 - \frac{\beta}{2}h^2 \tag{18}$$

$$W^M = \frac{1}{2}\left(x_0+g+h\right)^2 - \frac{\alpha}{2}g^2 - \frac{\beta}{2}h^2. \tag{19}$$

4 Investment stage

We start to solve the incumbent's problem about how much to invest in quality and in cost reduction when the regulator acts as a first-mover.

Proposition 2 *The investments in quality upgrades and cost reduction are complements. Moreover, the marginal revenue of investing in quality upgrades (cost reduction) is increasing (decreasing) in the unbundling price.*

Proof. From the derivatives of (14) we easily find that:

$$\frac{\partial^2 \pi_I}{\partial g \partial h} > 0,\ \frac{\partial^2 \pi_I}{\partial g \partial y_0} > 0,\ \frac{\partial^2 \pi_I}{\partial h \partial y_0} < 0.$$

■

The higher is the investment in quality upgrades, the higher is the marginal benefit of investing in cost reduction, since it decreases the cost of serving a higher number of subscribers. Thus, investments are complements. However, they are affected differently by the unbundling price. When the unbundling price is higher, the incumbent earns more profit with the entrant's subscribers. As a consequence, it has a higher incentive to invest in quality upgrades since this increases the entrant's number of subscribers. On

[7] This condition is also sufficient to guarantee that $q_I, q_E \geq 0$ and $p_E \geq r$ at equilibrium prices.

the other hand, it has less incentives to invest in cost reduction since this decreases its rival's number of subscribers. This follows from the fact that this investment only reduces incumbent's cost per line, while the entrant's cost per line, which is given by r, remains the same. Consequently, the incumbent's cost advantage over its rival increases. That is why, despite the complementarity between the two investments, we may observe each investment going in a different direction after a change in the unbundling price

If the regulator set an access margin instead of an unbundling price, so that r decreased when c decreased, the investment in cost reduction would be equivalent to an investment in quality upgrades, since there would not be any gains in terms of cost advantage after an investment in cost reduction.

Given the unbundling price set by the regulator, the incumbent maximizes its profit function (14) with respect to g and h. The profit-maximizing investments in quality upgrades and in cost reduction are:

$$\begin{aligned} g^* = & \frac{\left(8+\theta^3\right)(1+\theta)\beta-(2+\theta)\left(6-\theta+\theta^2\right)}{V(\theta,\alpha,\beta)}y_0 \\ & + \frac{\left(6-\theta+\theta^2\right)(2+\theta)+2(1-\theta)(2+\theta)^2\beta}{V(\theta,\alpha,\beta)}x_0 \end{aligned} \tag{20}$$

$$\begin{aligned} h^* = & -\frac{(1+\theta)\left(8+2\theta^2-\theta^3\right)\alpha-(2+\theta)\left(6-\theta+\theta^2\right)}{V(\theta,\alpha,\beta)}y_0 \\ & + \frac{(2+\theta)\left(8-3\theta^2+\theta^3\right)}{V(\theta,\alpha,\beta)}\alpha x_0, \end{aligned} \tag{21}$$

where:

$$\begin{aligned} V(\theta,\alpha,\beta) = & (1+\theta)\left(4-\theta^2\right)^2\alpha\beta-(2+\theta)\left(6-\theta+\theta^2\right) \\ & -2(1-\theta)\left((2+\theta)^2\beta+\left(4+4\theta-\theta^3\right)\alpha\right). \end{aligned} \tag{22}$$

Assumption 3: The socially optimal quality upgrades and cost reduction investments are finite, which is equivalent to have:[8]

$$\alpha > \underline{\alpha} \equiv \frac{6-4\theta}{4-3\theta^2+\theta^3} \tag{23}$$

[8] These are obtained by solving the problem of a social planner who takes the decisions regarding investments and unbundling price.

$$\beta > \underline{\beta} \equiv \frac{(1-\theta)\left(12+8\theta-\theta^2-2\theta^3\right)\alpha+\left(7+3\theta+7\theta^2+\theta^3\right)}{(2+\theta)^2\left((2-\theta)^2(1+\theta)\alpha-(6-4\theta)\right)}. \tag{24}$$

If both these conditions are verified we have $V(\theta, \alpha, \beta) > 0$, and both h^* and g^* are a maxima.[9]

Note that (20) and (21) are the profit-maximizing investments given that $h^* < c$. However, if c is low enough, we may have $h^* > c$, and in this case the best the incumbent can do is to invest $h^{**} = c$. The profit-maximizing investment in quality upgrades is then:

$$\begin{aligned} g^{**} = {} & \frac{(1+\theta)\left(4-2\theta+\theta^2\right)}{\left((1+\theta)(2-\theta)^2\alpha-2(1-\theta)\right)(2+\theta)}(y_0+c) \\ & + \frac{2(1-\theta)}{(1+\theta)(2-\theta)^2\alpha-2(1-\theta)}(x_0+c) \end{aligned} \tag{25}$$

Having found the equilibrium investments in quality upgrades and in cost reduction, we are now able to determine the effect of unbundling on each type:

Proposition 3 *In the presence of both types of investment and for $c > h^*$, the profit-maximizing investment in cost reduction is decreasing in the unbundling price.*

Proof. Taking the derivative of h^* with respect to r, we obtain:

$$\frac{\partial h^*}{\partial r} = \frac{\partial h^*}{\partial y_0}\frac{\partial y_0}{\partial r} = -\frac{(1+\theta)\left(8+2\theta^2-\theta^3\right)\alpha-(2+\theta)\left(6-\theta+\theta^2\right)}{V(\theta,\alpha,\beta)} < 0,$$

which is always positive for $\alpha > \underline{\alpha}$. ■

Proposition 4 *Define $\widetilde{\beta} = \frac{6-\theta+\theta^2}{(1+\theta)\left(4-2\theta+\theta^2\right)}$. In the presence of both types of investment and for $c > h^*$, the profit-maximizing investment in quality upgrades is increasing in the unbundling price if $\beta > \widetilde{\beta}$ and decreasing if $\beta < \widetilde{\beta}$.*

Proof. Taking the derivative of g^* with respect to r, we obtain:

$$\frac{\partial g^*}{\partial r} = \frac{\partial g^*}{\partial y_0}\frac{\partial y_0}{\partial r} = \frac{\left(8+\theta^3\right)(1+\theta)\beta-(2+\theta)\left(6-\theta+\theta^2\right)}{V(\theta,\alpha,\beta)},$$

which is positive for $\beta > \widetilde{\beta}$ and negative for $\beta < \widetilde{\beta}$. ■

[9] These conditions are sufficient to guarantee that the Hessian of the incumbent's problem verifies the maxima conditions.

As we have seen in Proposition 2, when the unbundling price is lower, the marginal revenue of investing in cost reduction is higher. Thus, as expected, we have a higher equilibrium investment in this type.

In contrast, the effect of the unbundling price in the equilibrium quality improvements does not follow immediately from Proposition 2, as only for a high β we obtain a positive relationship. This results from the complementarity relationship between the two investments. Indeed, if we take into account the indirect effect of a higher unbundling price through cost reduction, we observe that this has a negative impact on the marginal revenue of quality improvements. Consequently, when this indirect effect is relatively higher, we obtain a negative relationship between investment in quality upgrades and the unbundling price. This happens for a low β, i.e., when the reaction of cost reduction to an increase in the unbundling price is high.

The same indirect effect is present in the cost reduction equilibrium investment, but in this case the effect is weaker for all $\alpha > \underline{\alpha}$, and consequently, we always obtain a negative relationship. In fact, as the marginal revenue of cost reduction reacts more to changes in r than the marginal revenue of quality improvements, the direct effect of an increase in the unbundling price on cost reduction is higher.

For $c < h^*$ cost reduction is independent of the unbundling price, while quality improvements is always increasing in r, since the indirect effect does not exist in equilibrium.

If the incumbent has a restriction of funds to spend in investment the complementarity result also disappears. In fact, if the incumbent wants to increase its investment in one component it must reduce its investment in the other. In this case investments become substitutes. Hence, when the unbundling price is higher, the investment in quality upgrades is always higher and the investment in cost reduction is always lower.

For the rest of the paper we will assume that $c > h^*$ and there is no restriction of funds, so that we always have (20) and (21) as the investment choices of the incumbent. In this case, we obtain the following result:

Proposition 5 *Let* $\Phi = \left(\frac{8+2\theta^2-\theta^3}{8+\theta^3}\right)$. *When* $c > h^*$, *if* $\frac{\beta}{\alpha} > \Phi$, *the profit-maximizing total investment is increasing in the unbundling price and if* $\frac{\beta}{\alpha} < \Phi$, *it is decreasing.*

Proof. Summing g^* and h^*, and then taking the derivative with respect to r, we obtain:

$$\frac{\partial\left(h^*+g^*\right)}{\partial r} = (1+\theta)\frac{\left(8+\theta^3\right)\beta - \left(8+2\theta^2-\theta^3\right)\alpha}{V\left(\theta,\alpha,\beta\right)},$$

and thus $\frac{\partial(h^*+g^*)}{\partial r} > 0$ if and only if $\beta > \left(\frac{8+2\theta^2-\theta^3}{8+\theta^3}\right)\alpha$. ■

According to this Proposition, if cost reduction is sufficiently expensive as compared with quality improvements, the higher is the unbundling price, the higher is the total amount of investment we expect the incumbent to do. Otherwise, we expect total investment to be lower when the unbundling price is higher.

Note that $\Phi > 1$ for $\theta \in (0, 1)$, i.e., we can have a $\beta > \alpha$ and even so the relationship is negative. This is a consequence of the stronger direct effect of a higher unbundling price on cost reduction.

The existent empirical studies do not distinguish between these two types of investment. By this way, it is natural that we observe some contradictory results about the relationship between the unbundling price and the incumbent's investment. In fact, if $\frac{\beta}{\alpha} > \Phi$, a more intense utilization generated by a lower unbundling price leads an incumbent to invest less. Therefore, we expect to see more investment when the unbundling price is higher, which confirms the results of Haring et al. (2002) and Gabel and Huang (2003). If $\frac{\beta}{\alpha} < \Phi$, we obtain a negative relationship between investment and unbundling price, which is according to the result by Willig et al. (2002), Hassett et al. (2003) and the Phoenix Center Studies (2003), which state that a lower unbundling price increases the intensity of competition, and this increases the incentives of an incumbent to invest in order to gain a competitive advantage.

5 Regulation stage

5.1 A no-investment benchmark

Let us consider first the absence of an investment stage. In this case, a regulator maximizes social welfare over r without having to take into account the incumbent's investment incentives. Thus, given our assumption that incumbent must make non-negative profits, we obtain the following result:

Proposition 6 *In the absence of an investment stage, the second-best socially optimal unbundling price is such that the incumbent earns zero profits.*

Proof. See Appendix B. ■

The regulator wants the incumbent to subsidize the entrant's activity through a negative access margin. In fact, when the unbundling price is lower than marginal cost, competition between operators is more intense. The incumbent wants the entrant to have fewer subscribers in order to lose less money with unbundled lines, and the entrant wants to have more subscribers because profit per subscriber is higher. As a result, the subscription price of both operators decreases, increasing social welfare.

5.2 Commitment to unbundling price before investment

In a commitment context the regulator sets the unbundling price before the incumbent takes its decision about investment and commits to it until the end of the game. Hence, when he decides, the regulator takes into account how the incumbent will invest given the unbundling price. This implies that he has three objectives: He wants to increase the intensity of competition, to give incentives for an investment in cost reduction and to give incentives for an investment in quality upgrades. When $\beta \leq \widetilde{\beta}(\theta)$ the three objectives are all favored by a low r, and therefore the first-best unbundling price (y_0^{c1}) is so low that the incumbent would earn *ex post* negative profits. In this case, the second-best unbundling price (y_0^{c2}) set by the regulator is such that the incumbent earns *ex post* zero profits. Only for $\beta > \widetilde{\beta}(\theta)$ quality improvements become increasing in r, and the regulator has incentives to set a higher r. However, if β is lower than a threshold $\overline{\beta}(\theta, \alpha)$, the first-best unbundling price is still such that the incumbent would earn *ex post* negative profits. Therefore, the regulator continues to set a second-best unbundling price. Only when $\beta > \overline{\beta}(\theta, \alpha)$ quality improvements become sufficiently important so that the first-best r allows the incumbent to earn *ex post* positive profits, and thus it can be implemented by the regulator.

We then have the following results:

Proposition 7 *Define* $\overline{\beta}(\theta, \alpha)$ *by* $\pi_I^*\left(y_0^{c1}\left(\theta, \alpha, \overline{\beta}, x_0\right), \theta, \alpha, \overline{\beta}, x_0\right) = 0$. *When the regulator sets the unbundling price before the investment decision:*

(a) At the socially optimal the incumbent earns ex post positive profits for $\beta > \overline{\beta}(\theta, \alpha)$, *and ex post zero profits for* $\beta \leq \overline{\beta}(\theta, \alpha)$.

(b) The socially optimal unbundling price is increasing in β *and decreasing in* α *when* $\beta > \overline{\beta}(\theta, \alpha)$, *and when* $\beta \leq \overline{\beta}(\theta, \alpha)$ *it is increasing both in* α *and* β.

Proof. See Appendix C. ■

The first-best unbundling price is decreasing in β and increasing in α, as when cost reduction is relatively less expensive, the regulator wants the incumbent to invest relatively more in this type of investment than in quality upgrades. Therefore, he sets a lower unbundling price. However, when it hits the non-negativity condition in the incumbent's profit, the unbundling price set is the second-best one, and this is increasing in both β and α, as when an investment becomes more expensive, the restriction in incumbent's profit becomes tighter, and therefore the unbundling price must be higher.

De Bijl and Peitz (2004) argue that the regulator can give stronger incentives for an incumbent to invest in the quality of its network by increasing the sensitivity of the unbundling price to the quality level. In fact, if the regulator could set an unbundling price dependent on investment, it should increase in both investment types. In this case, the incumbent would have a higher incentive to invest both in quality and cost reduction, in order to receive a higher unbundling price.

5.3 No commitment to unbundling price before investment

In the previous section we assumed that the regulator acts as a first-mover and sets the unbundling price before the incumbent invests. This commitment may, however, not be credible if the regulator can change price at will later on. In this case, he has all the incentives to revise his decision after observing the investment made by the incumbent. Knowing this, the incumbent takes into account how the regulator will change his decision about unbundling price when it invests.

We start by solving the regulator's problem. Given the value of g and h chosen by the incumbent in the first stage, the regulator maximizes social welfare with respect to r.

Proposition 8 *When the regulator sets the unbundling price after the investment decision, it is such that the incumbent earns zero profits.*

Proof. See Appendix B, but instead of x_0 and y_0 consider $x_0 + g + h$ and $y_0 + h$. ∎

Corollary 9 *The incumbent does not invest in its network if the regulator only sets the unbundling price ex post. Therefore, welfare is lower as compared with the commitment context.*

Proof. See Appendix D. ∎

When the regulator acts as a second-mover, he only cares for low equilibrium prices, which are favoured by a low unbundling price. The incumbent foresees this behavior by the regulator, and thus it does not invest. In fact, every gain from its investment is expropriated later by a low unbundling price. Note that, in this case, we cannot observe any relationship between the unbundling price and investment. What we observe is that under unbundling there is no investment.

If a regulator cannot commit to his decisions, unbundling policies affect welfare negatively. This result supports the criticisms of service-based competition, namely of its impact on dynamic efficiency. Indeed, if a regulator is implementing an unbundling policy he must show to the market participants that he has the ability to commit to his decisions. If he cannot commit, it may be better to leave the market unregulated as we will see next.

6 Unregulated market

When there is no regulator in the market, the incumbent takes all decisions regarding investment and unbundling price. In this case, a high r is equivalent to no unbundling.

As we have a simultaneous decision over (r, g, h), by the envelope theorem, we just need to substitute the optimal investment functions (20) and (21) into the incumbent's profit function (14), and then maximize it with respect to r. The profit-maximizing *ex ante* access margin becomes:

$$y_0^{ur} = (2+\theta)\frac{(1+\theta)\left(4-2\theta+\theta^2\right)\beta - \left(6-\theta+\theta^2\right)}{T(\theta,\alpha,\beta)}\alpha x_0, \tag{26}$$

where

$$T(\theta,\alpha,\beta) = 2(1+\theta)\left(8+\theta^2\right)\beta\alpha - (2+\theta)\left(6-\theta+\theta^2\right)(\beta+\alpha), \tag{27}$$

which is positive for $\beta > \underline{\beta}^{ur} \equiv \frac{(2+\theta)\left(6-\theta+\theta^2\right)\alpha}{2(1+\theta)\left(8+\theta^2\right)\alpha-(2+\theta)\left(6-\theta+\theta^2\right)}$.[10]

Therefore, profit-maximizing investments are:

$$g^{ur} = \frac{(2+\theta)\left(6-\theta+\theta^2\right)\beta}{T(\theta,\alpha,\beta)}x_0 \tag{28}$$

$$h^{ur} = \frac{(2+\theta)\left(6-\theta+\theta^2\right)\alpha}{T(\theta,\alpha,\beta)}x_0. \tag{29}$$

Proposition 10 *The incumbent prefers to rent out its loops to remaining on as a monopolist.*

[10] This condition is also necessary for a negatively defined Hessian.

Proof. See Appendix E. ■

The incumbent unbundles its access network in order to attract more consumers to the market. Then, it sets a high unbundling price to absorb part of the profit the entrant earns with these new subscribers, increasing its own profit.

Proposition 11 *The profit-maximizing unbundling price is higher than the socially optimal one in a commitment context. Therefore, for $\beta < \widetilde{\beta}(\theta)$ the incumbent invests less in both types of investment in an unregulated market, while for $\beta > \widetilde{\beta}(\theta)$ it invests more in quality upgrades and less in cost reduction. Social welfare is always lower as compared to a commitment context.*

Proof. We have $y_0^{ur} > y_0^c$. The second part follows from Propositions 3 and 4. For the welfare result, we just need to see that the regulator could always have chosen y_0^{ur} when he set the socially optimal unbundling price. ■

Proposition 12 *Define $\beta^*(\theta, \alpha)$ by $W^{nc}(\theta, x_0) = W^{ur}(\theta, \alpha, \beta^*, x_0)$, where W^{nc} is welfare in a no-commitment context and W^{ur} is welfare in an unregulated context. The profit-maximizing unbundling price is higher than the socially optimal one in a no-commitment context, and the incumbent invests more in both types of investment. Social welfare is lower as compared to a no-commitment context if $\beta > \beta^*(\theta, \alpha)$, and higher if $\beta < \beta^*(\theta, \alpha)$.*

Proof. See Appendix F. ■

We find that if the regulator can commit to his decisions, social welfare is higher when he intervenes in the market *ex ante* as compared to no intervention since the unbundling price the incumbent would set is too high as compared to the one set by the regulator. Hence, in this case, it is better to have the regulator intervening.

On the other hand, if we compare the unregulated market with the context where the regulator cannot commit, the profit-maximizing unbundling price is still higher, but now welfare can be lower or higher depending on the level of investment costs. Indeed, when investment costs are low, it is preferable to leave the market unregulated since the incumbent unbundles and invests, while with *ex post* regulation we obtain a zero investment by the incumbent. For $\beta > \beta^*(\theta, \alpha)$ the investment objectives become less important as their socially optimal values are low, and thus regulation *ex post* becomes preferable since it assures low subscription prices.

7 Mature markets

In mature markets, we usually have full-consumer participation, which means that total demand is perfectly inelastic with respect to price changes. A relevant example is fixed telephony.

De Bijl and Peitz (2004) show that, contrary to the context of partial consumer participation, an increase in the unbundling price is totally passed on by the entrant to consumers. Bourreau and Dogan (2005) show that this is only true when the marginal consumer obtains positive surplus, which happens for low values of r.

Bourreau and Dogan (2005) use a Hotelling model to formalize a mature market, where, in equilibrium, each operator is located at one of the extremes of the line. The incumbent's profit and social welfare are given by:

$$\pi_I = \begin{cases} \frac{1}{2} + r - c & \text{if} \quad r \in \left[0, v - \frac{5}{4}\right) \\ \frac{1}{2}\left(v - \frac{1}{4} + r\right) - c & \text{if} \quad r \in \left[v - \frac{5}{4}, v - \frac{3}{4}\right) \\ r - 1 + \sqrt{3}\sqrt{v-r} - \frac{2\sqrt{3}}{9}(v-r)^{\frac{3}{2}} - c & \text{if} \quad r \in \left[v - \frac{3}{4}, v\right) \end{cases}, \tag{30}$$

and

$$W = \begin{cases} v - \frac{1}{12} - c & \text{if} \quad r \in \left[0, v - \frac{3}{4}\right) \\ \frac{2}{3}v + \frac{1}{3}r - \frac{1}{3} - c + \frac{\sqrt{3}}{3}\sqrt{v-r} & \text{if} \quad r \in \left[v - \frac{3}{4}, v\right) \end{cases}, \tag{31}$$

where $v > 3$ is the fixed utility of consumption

Let us first consider an investment in quality. As in previous sections this investment increases the reservation price, i.e., it increases v. In this case, when the unbundling price is low ($r < v - \frac{5}{4}$) the incumbent has no incentive to invest since the increases in quality are totally passed on to consumers. For intermediate values of r the incumbent invests in quality upgrades because this increases the valuation of the marginal consumer, allowing firms to increase prices. For high values of r the incumbent invests even more, and the investment is increasing in r. In fact, as for $r > v - \frac{3}{4}$ the entrant's market share increases when the incumbent invests, the higher is r the higher is the incentive to invest.

As concerns cost reduction, the incumbent always invests, but the size of the investment is independent of the unbundling price. This is because the number of subscribers is fixed, and translates into the lack of interaction between r and c in π_I.

We can conclude that for a full-consumer participation model, the socially optimal unbundling price must be high enough to give incentives for the incumbent's investment in quality.

8 Conclusions

The main objective of this paper is trying to bring some light into the contradictory results in the empirical literature about the effects of unbundling on incumbent firms' investment. Hence, we develop a model with two telecommunications operators, an incumbent and an entrant, that offer differentiated services in a market with partial consumer participation. The incumbent can invest in quality upgrades and in cost reduction, which are complements but have different impacts on both firms. We conclude that both empirically observed relationships are possible. In fact, a low unbundling price increases the intensity of competition, which gives incentives for an incumbent to invest in cost reduction in order to gain a cost advantage for a given unbundling price. On the other hand, it decreases the incumbent's return from investing in quality upgrades. Thus, although one should expect to have both investments moving together due to their complementarity, it is not obvious what the equilibrium effect of a lower unbundling price will be.

Secondly, we compare social welfare when the regulator can commit to an unbundling price set *ex ante* and when he cannot. We show that in the latter case the incumbent does not invest since it does not retain any gain from its investments. As a consequence, social welfare is lower. Here, it may be better to let the market unregulated since the incumbent firm will unbundle its local loop and invest. Thus, for the welfare effects of unbundling policies it is decisive whether the regulator can or cannot commit to unbundling prices.

Appendix A

We start to solve price competition in the absence of investment.

The representative consumer maximizes the utility function:

$$U = a\left(q_I + q_E\right) - \frac{1}{2}\left(q_I^2 + 2\theta q_I q_E + q_E^2\right) - p_I q_I - p_E q_{E.}$$

First-order conditions are:

$$p_I = a - q_I - \theta q_E \tag{32}$$
$$p_E = a - \theta q_E - q_I. \tag{33}$$

Inverting equations (32) and (33), we obtain:

$$q_I = \frac{a\left(1-\theta\right) - p_I + \theta p_E}{1-\theta^2} \tag{34}$$
$$q_E = \frac{a\left(1-\theta\right) - p_E + \theta p_I}{1-\theta^2}. \tag{35}$$

Given (34) and (35), the incumbent and entrant's profit function become:

$$\pi_I = \left(\frac{a(1-\theta) - p_I + \theta p_E}{1-\theta^2}\right)(p_I - c) + \left(\frac{a(1-\theta) - p_E + \theta p_I}{1-\theta^2}\right)(r-c)$$

$$\pi_E = (p_E - r)\left(\frac{a(1-\theta) - p_E + \theta p_I}{1-\theta^2}\right).$$

If we maximize each profit function with respect to the price of the respective operator, we obtain the following best response functions:

$$\begin{aligned} p_I &= \frac{1}{2}(a+c)(1-\theta) + \frac{1}{2}\theta p_E + \frac{1}{2}\theta r \\ p_E &= \frac{1}{2}a(1-\theta) + \frac{1}{2}\theta p_I + \frac{1}{2}r. \end{aligned}$$

Solving these, equilibrium prices become:

$$\begin{aligned} p_I &= \frac{1-\theta}{2-\theta}(a-c) + \frac{3\theta}{4-\theta^2}(r-c) + c \\ p_E &= \frac{1-\theta}{2-\theta}(a-c) - 2\frac{1-\theta^2}{4-\theta^2}(r-c) + r. \end{aligned}$$

Substituting both in (34) and (35) we find:

$$\begin{aligned} q_I &= \frac{1}{2-\theta^2+\theta}(a-c) - \frac{\theta}{4-\theta^2}(r-c) \\ q_E &= \frac{1}{2-\theta^2+\theta}(a-c) - \frac{2}{4-\theta^2}(r-c). \end{aligned}$$

Therefore, equilibrium profits and welfare are:

$$\begin{aligned} \pi_I = {} & \frac{1-\theta}{(1+\theta)(2-\theta)^2}(a-c)^2 - \frac{8+\theta^2}{\left(4-\theta^2\right)^2}(r-c)^2 \\ & + \frac{4-2\theta+\theta^2}{(2-\theta)\left(4-\theta^2\right)}(a-c)(r-c) \end{aligned}$$

$$\begin{aligned} \pi_E = {} & \frac{1-\theta}{(1+\theta)(2-\theta)^2}(a-c)^2 + \frac{4\left(1-\theta^2\right)}{(2+\theta)^2(2-\theta)^2}(r-c)^2 \\ & - \frac{4(1-\theta)}{(2-\theta)^2(2+\theta)}(a-c)(r-c) \end{aligned}$$

$$W = \frac{3-2\theta}{\left(2+\theta-\theta^2\right)(2-\theta)}(a-c)^2 - \frac{4+5\theta^2}{2\left(4-\theta^2\right)^2}(r-c)^2 - \frac{1-\theta}{(2-\theta)^2}(a-c)(r-c).$$

Finally, we introduce the two investments into the equilibrium and obtain $a+g$ instead of a, and $c-h$ instead of c. We also have to introduce the investment cost functions (8) and (9) into welfare and incumbent's profit.

Appendix B. Proof of Proposition 6

The regulator maximizes social welfare subject to the non negativity condition in the incumbent's profit. Thus, we have the following Lagrangian:

$$\pounds = W(y, x, \theta) + \lambda_1 \pi_I.$$

First-order conditions are:

$$-2\frac{(1-\theta)(2+\theta)^2 x_0 + \left(4+5\theta^2\right) y_0}{\left(4-\theta^2\right)^2} + \lambda_1 \frac{\left(8+\theta^3\right) x_0 - 2y_0\left(8+\theta^2\right)}{\left(4-\theta^2\right)^2} = 0$$

$$\pi_I \geq 0 \quad , \quad \lambda_1 \geq 0 \quad , \quad \pi_I \lambda_1 = 0. \tag{36}$$

If $\lambda_1 = 0$, we obtain:

$$y_0 = -\frac{(1-\theta)(2+\theta)^2}{4+5\theta^2} x_0,$$

but this violates $\pi_I \geq 0$ restriction.

If $\lambda_1 > 0$, we obtain $\pi_I = 0$, and thus (36) becomes:

$$\lambda_1 = 2\frac{x_0(1-\theta)(2+\theta)^2 + \left(4+5\theta^2\right) y_0}{\left(8+\theta^3\right) x_0 - 2y_0\left(8+\theta^2\right)} > 0.$$

Therefore, the second-best socially optimal access margin is

$$y_0^{sb} = (2+\theta)\frac{(1+\theta)\left(4-2\theta+\theta^2\right) - (2-\theta)\sqrt{(2+\theta)(1+\theta)\left(6-\theta+\theta^2\right)}}{2(1+\theta)\left(8+\theta^2\right)} x_0. \tag{37}$$

This access margin is lower than $\overline{y}_0$, and thus the entrant asks for access.

Appendix C. Proof of Proposition 7

First, we substitute the profit-maximizing investments (20) and (21) into the welfare function (16) and then maximize it with respect to y_0 to obtain the first-best access margin $y_0^{c1} = Z(\theta, \alpha, \beta)\, x_0$.

Second, we define the incumbent's profit at the profit-maximizing investments by

$$\pi_I^*(\theta, \alpha, \beta, x_0, y_0) = \pi_I(\theta, \alpha, \beta, x_0, y_0, g^*(\theta, \alpha, \beta, x_0, y_0), h^*(\theta, \alpha, \beta, x_0, y_0)).$$

Then we introduce the first-best access margin and find that:

$$\pi_I^*\left(\theta, \alpha, \beta, x_0, y_0^{c1}(\theta, \alpha, \beta, x_0)\right) \geq 0,$$

if and only if, $\beta \geq \overline{\beta}(\theta, \alpha)$ (this is independent of x_0). Thus, for $\beta \geq \overline{\beta}(\theta, \alpha)$, the socially optimal *ex ante* access margin is $y_0^{c1} = Z(\theta, \alpha, \beta)\, x_0$, and we have:

$$\frac{\partial Z(\theta, \alpha, \beta)}{\partial \alpha} < 0, \frac{\partial Z(\theta, \alpha, \beta)}{\partial \beta} > 0.$$

When $\beta < \overline{\beta}(\theta, \alpha)$, we have

$$\pi_I^*\left(\theta, \alpha, \beta, x_0, y_0^{c1}(\theta, \alpha, \beta, x_0)\right) < 0,$$

and therefore, the second-best socially optimal *ex ante* access margin is obtained by solving $\pi_I^*(\theta, \alpha, \beta, x_{0,} y_0^{c2}) = 0$, which gives $y_0^{c2} = U(\theta, \alpha, \beta)\, x_0$ where:

$$\frac{\partial U(\theta, \alpha, \beta)}{\partial \alpha} > 0, \frac{\partial U(\theta, \alpha, \beta)}{\partial \beta} > 0.$$

The socially optimal *ex ante* access margin under commitment is then $y_0^c = \max\{y_0^{c1}, y_0^{c2}\}$ and, after tedious calculations, it is possible to prove that $y_0^c < \overline{y}_0$.

Appendix D. Proof of Proposition 9

The first part of the Proposition is a natural consequence of Proposition 8, since no matter what the incumbent invests it earns zero profits. Hence, it will never invest. The *ex ante* access margin set by the regulator is then given by (37), and social welfare becomes:

$$W^{nc} = \left(\frac{\left(128 + 16\theta + 22\theta^2 - 4\theta^3 + 3\theta^4 - 3\theta^5\right)}{4(1+\theta)\left(8+\theta^2\right)^2} - \frac{\left(24 + 6\theta^2 - 3\theta^3\right)\sqrt{(2+\theta)(1+\theta)\left(6 - \theta + \theta^2\right)}}{4(1+\theta)\left(8+\theta^2\right)^2} \right) x_0^2.$$

Comparing W^{nc} with welfare in a commitment context (W^c) it is possible to show, after tedious calculations, that $W^c > W^{nc}$.

Appendix E. Proof of Proposition 10

y_0^{ur} is such that the entrant asks for access since $y_0^{ur} < \overline{y}$. Thus, we can introduce y^{ur}, g^{ur} and h^{ur} into the incumbent's profit function (14) and obtain:

$$\pi_I^{ur} = \frac{(2+\theta)\left(6-\theta+\theta^2\right)\alpha\beta}{2T(\theta,\alpha,\beta)}x_0^2.$$

A monopolist incumbent solves the following problem:

$$\max_{g,h} \frac{1}{4}(x_0+h+g)^2 - \frac{\alpha}{2}g^2 - \frac{\beta}{2}h^2,$$

which gives optimal investments under monopoly:

$$g^M = \frac{\beta}{2\alpha\beta-\alpha-\beta}x_0$$

$$h^M = \frac{\alpha}{2\alpha\beta-\alpha-\beta}x_0.$$

Thus, incumbent's monopoly profit is:

$$\pi_I^M = \frac{1}{2}\left(\frac{\alpha\beta}{2\alpha\beta-\alpha-\beta}\right)x_0^2.$$

If we compare with profit under unbundling, we find that $\pi_I^{ur} \geq \pi_I^M$.

Appendix F. Proof of Proposition 12

If we substitute (26), (28), (29) into (16), we obtain welfare in an unregulated market:

$$W^{ur} = \left(\frac{(1+\theta)\left(304+48\theta+108\theta^2+16\theta^3+11\theta^4-\theta^5\right)\beta\alpha}{T(\theta,\alpha,\beta)^2}\right. \\ \left. -\frac{(\beta+\alpha)(2+\theta)^2\left(6-\theta+\theta^2\right)^2}{T(\theta,\alpha,\beta)^2}\right)\frac{1}{2}\alpha\beta x_0^2.$$

Comparing with W^{nc} we obtain that $W^{nc} > W^{ur}$ for $\beta > \beta^*(\theta,\alpha)$.

References

Armstrong, M., 2002. The Theory of Access Pricing and Interconnection. In: Cave, M., Majumdar, S., Vogelsang, I. (Eds.), Handbook of Telecommunications Economics, vol. 1. North-Holland, Amsterdam, pp. 295-384.

Biglaiser, G., Ma, C-T, 1999. Investment Incentives of a Regulated Dominant Firm. Journal of Regulatory Economics 16, 215-235.

Bourreau, M., Dogan, P., 2005. Unbundling the Local Loop. European Economic Review 49, 173-199.

Bowley, A.L., 1924. The Mathematical Groundwork of Economics. Oxford University Press.

Cabral, L., Riordan, R., 1989. Incentives for Cost-Reduction under Price-Cap Regulation. Journal of Regulatory Economics 1, 93-102.

Cambini, C., Valletti, T., 2004. Access Charges and Quality Choice in Competing Networks. Information Economics and Policy 16, 391-409.

Chang, H., Koski, H., Majumdar, S.K., 2003. Regulation and Investment Behavior in the Telecommunications Sector: Policies and Patterns in US and Europe. Telecommunications Policy 27, 677-699.

De Bijl, P., Peitz, M., 2004. Unbundling the Local Loop: one way access and imperfect competition. In: Dewenter, R., Haucap, J. (Eds.), Access Pricing: Theory and Practice. Elsevier Science.

Foros, O., 2004. Strategic Investments with Spillovers, Vertical Integration and Foreclosure in the Broadband Access Market. International Journal of Industrial Organization 22, 1-24.

Gabel, D.J., G-L Huang, 2003. Promoting Innovation: Impact of Local Competition and Regulation of Deployment of Advanced Telecommunications Services for Businesses. draft, Internet and Telecom Convergence Program. MIT, Cambridge, MA.

Haring, J., Rettle, M.L., Rohlfs, J.H., Shooshan III, H.M., 2002. UNE Prices and Telecommunications Investment. Strategic Policy Research 2002.

Hassett, K.A., Ivanova, Z., Kotlikoff, L.J., 2003. Increased Investment, Lower Prices - The Fruits of Past and Future Telecom Competition. mimeo.

Hausman, J., Sidak, J., 2005. Did mandatory unbundling achieve its purpose? Empirical evidence from five countries. Journal of Competition Law and Economics 1, 173-245.

Hazlett, T.W., Havenner, A.M., Bazelon, C., 2003. Declaration of Thomas W. Hazlett, Arthur M. Havenner, and Coleman Bazelon: In the Matter of Petition for Expedited Forbearance from the Current Pricing Rules for the Unbundled Network Element Platform. U.S. F.C.C. W.C. Docket N^o 03-157.

Hill, C., 2003. Declaration of R.Carter Hill: In the Matter of Petition for Forbearance from the Current Pricing Rules for the Unbundled Network Element Platform, U.S. F.C.C. WC Docket N^o 03-157.

Ingraham, S., Sidak, J., 2003. Mandatory Unbundling, UNE-P, and Cost of Equity: Does TELRIC Pricing Increase Risk for Incumbent Local Exchange Carriers?. Criterion Economics. Cambridge, MA.

Jorde, T.M., Sidak, J.G., Teece,D.J, 2000. Innovation, Investments, and Unbundling. Yale Journal of Regulation 17, 1-37.

Kotakorpi, K., 2006. Access Price Regulation, Investment and Entry in Telecommunications. International Journal of Industrial Organization 24, 1013-1020.

Laffont, J.J., Tirole, J., 2000. Competition in Telecommunications. MIT Press.

Phoenix Center Policy Bulletin, 2003. Competition and Bell Company Investment in Telecommunications Plant: The Effects of UNE-P, n^{o}5.

Phoenix Center Policy Bulletin, 2003. UNE-P Drives Bell Investment: A Synthesis Model, n^{o}6.

Sappington, D., 2002. Price Regulation and Incentives. In: Cave, M., Majumdar, S., Vogelsang, I. (Eds.), Handbook of Telecommunications Economics, vol. 1. North-Holland, Amsterdam, pp.227-293.

Valletti, T., 2003. The Theory of Access Pricing and its Linkage with Investment Incentives. Telecommunications Policy 27, 659-675.

Vareda, J., Hoernig, S., 2006. The race for telecoms infrastructure investment with bypass: Can access regulation achieve the first-best?. mimeo, FEUNL.

Willig, R., Lehr, W., Bigelow, J., Levinson, S., 2002. Stimulating Investment and the Telecommunications Act of 1996. report filed by AT&T in FCC Docket, 1-338.

Dynamic Competition in Telecom Networks under the Receiver Pays Principle

Ángel Luis López[1]

Version: March 2007

This paper analyzes dynamic network competition under the Receiver Pays Principle. We investigate how the implementation of the Receiver Pays Principle affects the networks operators' pricing strategies in a model of dynamic competition. We characterize the equilibrium and provide sufficient conditions under which it exists and is unique. In the region where the equilibrium exists we show that networks price calls at their off-net cost. We further show that, even in this dynamic setting, the off-net cost pricing equilibrium neutralizes the potential role that future reciprocal access charges could play as an instrument to soften retail competition. Last, we argue that while the level of the access price does not affect networks' profits, it clearly distorts consumer welfare.

Keywords: dynamic competition, interconnection, receiver pays principle, access pricing.

JEL Classifications: D43,D92,K21,L51,L96.

[1]GREMAQ, University Toulouse 1. Email: angel.lopez@univ-tlse1.fr. The author thanks Patrick Rey for his guidance and helpful advice, and gratefully acknowledge the comments of Jean Tirole and Javier Campos. All remaining errors are my own.

1. INTRODUCTION

Since the introduction of competition in telecommunications, network interconnection has been one of the most controversial issues of telecoms regulation. The need for interconnection stems from the fact that networks need to connect their subscribers with those on other networks; indeed it is one of the keys, but not the only, to achieving effective competition in the market. This involves "two-way" access agreements whereby networks provide call origination, transit and termination services to each other. It then raises the non-trivial question of how accepting traffic from or delivering traffic to other networks should be priced.

"One-way" access refers to the case where an incumbent monopolizes the local network and must provide a bottleneck input to new entrants that compete with it in a downstream market. Since the incumbent could use the bottleneck to expel competitors from the market, there is a wide consensus in the literature that regulation is socially desirable. In the case of two-way access, however, networks operate at the same level of network hierarchy, that is, they do not only compete for subscribers in the retail market but do depend on each other to supply the retail service. Thus one may at first sight be induced to think that regulation is unnecessary. In practice, access charges are frequently set cooperatively, while cooperation over retail prices is in general considered to be illegal. Some might wonder whether networks could not agree on setting a specific access charge that softens competition in the retail market.

In order to develop an optimal policy it is key to determine whether unconstrained interconnection negotiations over access charges can undermine retail competition or on the contrary are socially optimal, in which case no regulation is needed. Indeed, this question has been studied by the seminal papers of Armstrong (1998) and Laffont, Rey and Tirole (1998a,b). Assuming symmetric networks, reciprocal access charges and linear retail pricing these papers show that competition in the retail market can be undermined by collusion over the access charge. This result stems from the fact that if a network lowers its retail price, then its subscribers will make more calls, which provokes an access deficit whenever the access charge is above the cost. Then, by agreeing to high access charges, networks reduce the incentive to undercut each other. More surprisingly, Laffont, Rey and Tirole (1998a) show that in the same setting but under two-part pricing, the collusive power of the access charge vanishes, that is, networks' equilibrium profits do not depend on the level of the access charge. This result comes about because of an intense waterbed effect. Intuitively, an increase in the access charge leads to an increase in the usage price, which makes it more desirable for networks to build market share. In the linear pricing case, networks cannot build market share without incurring an access deficit, but under two-part tariffs they can by lowering their fixed fees while keeping usage prices constant.

This waterbed effect occurs in the limit where networks find it worthwhile to spend the full revenue from access fees in order to attract subscribers. That is, higher usage prices are offset by lower fixed fees such that networks' profits remain independent of the reciprocal access charge.

This striking result has become the focus of much research,[2] and also has been proved to hold when customers are heterogeneous.[3] Indeed, this neutrality result depends crucially on three assumptions: full-participation, no termination-based price discrimination and symmetry.[4] Carter and Wright (2003) allow asymmetric networks by providing for brand loyalty and show that the incumbent strictly prefers the access charge to be set at marginal cost, and that both networks prefer cost-based access charges when there is a sufficient degree of asymmetry. Intuitively, the larger network or incumbent faces a higher proportion of intra-network calls, whereas the smaller network faces a higher proportion of inter-network calls. Then, since networks price calls at the *perceived* marginal cost, a reciprocal access charge above cost increases the *perceived* marginal cost of the smaller network (because of most of its calls are inter-network) and hence its call price also increases. This, consequently, implies that the larger network will face a net outflow of calls with an above-cost access charge and hence a deficit in the wholesale market. We show below how this last result partially explains our non-neutrality result in a dynamic model of competition even though networks are symmetric. To sum up, established telecoms networks under nonlinear pricing and no termination-based price discrimination cannot use reciprocal access charges as an instrument of collusion as long as there is full participation or an exogenous participation rate, and thereby unconstrained interconnection negotiations over reciprocal access charges might be a socially optimal policy.

So far we have only considered the literature that studies competition in a static model. What about dynamic competition? Does it alter our conclusion? De Bijl and Peitz (2000, 2002) study dynamic network competition but focusing only on myopic behaviour or, in other words, on the per-period profit maximizing equilibria. They study the asymmetric case and find in the short term a similar result to that of Carter and Wright (2003), and in the long term a result that is very close to neutrality.[5] Our previous work (López, 2005) however depicts

[2] See Armstrong (2002) and Vogelsang (2003) for a survey of this literature.

[3] Dessein (2003) and Hahn (2004) introduce heterogeneity in volume demand and shows that equilibrium profits are still independent of the reciprocal access charge under second-degree price discrimination. De Bijl and Peitz (2000, chpt. 7) allow for third-degree price discrimination and still find the same result.

[4] Firstly, Poletti and Wright (2004) by allowing customers' participation constraint to be binding in equilibrium show that access charges above cost can play a collusive role. In addition, Schiff (2002) show that under partial consumer participation and some other assumptions, as for instance an exogenous participation rate, networks prefer the access charge equal to the marginal cost, but when these assumptions are relaxed, networks instead prefer either cost-based or below-cost access prices depending on the case that is under consideration. Secondly, Gans and King (2001) show that networks prefer access charges below cost when they can price discriminate.

[5] It is worth to remark, however, that they make numerical analyses of a wide range of

the competition between two differentiated networks in a two-period model and under the subgame-perfect equilibrium concept. We show that even symmetric networks with full participation can use reciprocal access charges to soften competition when they compete in a dynamic setting. In particular, the networks' overall profits are neutral with respect to the first-period access charge but increase when the second-period access charge departs away from the marginal cost. A robust economic argument supports this result: in the second period the profits of the larger firm decrease when the access charge departs away from the marginal cost, which in turn decreases the incentives to fight for market share in the first period. This result holds both when consumer expectations are naive and when they are rational. Thus regulation might be needed in order to prevent anticompetitive behaviour since cost-based access charges maximize the full-period welfare surplus. Price controls of course is a draconian policy that regulators normally avoid if others alternatives are available, in particular because it is not clear whether regulation costs are lower than the potential benefits derived from price controls. A possible solution that avoids direct intervention in the market is moving towards a Receiver Party Pays system.

The Receiver Pays Principle (RPP) is already applied to mobile call pricing in U.S and some Asian countries, and it has also been widely adopted in international roaming, although in both cases for non-economic reasons.[6] An important economic argument that may support its implementation is the existence of call externalities, which occur because both callers and receivers may benefit from a phone call. In practice, RPP has been recently invoked as an instrument to reduce mobile termination charges.[7] Despite the spectacular growth of mobile telephony in recent years, mobile termination charges have remained high in Europe, where the Caller Pays Principle (CPP) applies. These high termination rates are from fixed to mobile calls, and have become a serious concern in most European countries; they do not only affect negatively the consumer welfare but are also perceived to be damaging the fixed telecoms sector's ability to innovate and invest in new technologies.[8] In this respect, some observers see RPP as a good alternative to price controls and predict that its implementation in the telecoms industry would exert downward pressure on mobile termination charges.[9]

interesting scenarios that are not considered here, as for instance the non-reciprocal access price case and the process of entry (De Bijl and Peitz, 2004.)

[6] In the former case it has been so mainly because of technological reasons: the access codes of the mobile service providers are not distinct to those of the fixed network.

[7] Those charges that mobile operators levy on each other and on fixed network operators for terminating calls on their networks.

[8] In France, Germany and the UK, the total transfer of funds for fixed to mobile calls (computing the excess of termination charges paid over costs and including a normal return on capital employed) from fixed networks to the mobile sector is estimated to be €19 billion between 1998 and 2002 (see Cave et al. 2003.)

[9] Intuitively, under the receiver pays regime, if a mobile network sets high termination charges it will decrease the utility of its own subscribers and so its attraction. Consequently, competition in reception charges should result in lower termination charges.

We are primarily interested in determining whether future reciprocal access charges can still soften first-period competition when networks compete in fixed fees, call prices and reception charges. Obviously, adopting RPP will significantly change the networks' incentives. This in turn makes important to develop a conceptual framework in which the resulting industry can be analyzed. We build on previous literature to propose such a framework, and aim to investigate how networks' pricing strategies react to the adoption of the receiver pays regime when they compete in a multi-period setting. Our starting point is that callers and call receivers derive utility from making and receiving calls. Moreover, networks are allowed to charge customers for receiving calls. The analysis faces the problem of *sovereignty:* who decides to end the call? It will be argued that in a deterministic framework allowing receivers to hang up makes the model discontinuous. We thus generalize this setup by assuming that both the caller's and receiver's utilities are subject to a random noise, the purpose of this is to smooth the demand, in fact this makes the model even more realistic.

The receiver pays principle has been studied under different settings (all of them focusing on static competition) by Berger (2001), Fabrizi (2005), Hermalin and Katz (2001, 2005), Kim and Lim (2001), Jeon, Laffont and Tirole (2004) and Laffont et al. (2003).[10] Nevertheless, the most related papers to the problem we study are the last two papers. Laffont et al. analyze Internet backbone competition. In their framework there are two types of customers: senders or websites and receivers or consumers. In our model however every consumers both send and receive traffic, and get surplus from and are charged for making and receiving calls. On the other hand, Jeon, Laffont and Tirole (2004) and our paper analyze three-part tariff competition in a telecommunications environment where the volume of traffic between each caller and receiver is endogenously determined by one of them though subject to a random noise. More specifically, Jeon, Laffont and Tirole assume that only the receiver's utility is subject to a noise and a certain proportionality between the receiver's and the caller's utilities. Our setup however generalizes their work by allowing a random noise in both the callers and receivers' utilities, and by removing the assumption of a given proportionality between the utility functions. Yet, the main contributions of our paper are that in this general setup we easily show that the off-net-cost pricing principle is a candidate equilibrium, and more importantly we prove that under general conditions the off-net-cost pricing equilibrium exists and is the unique possible equilibrium. Instead, Jeon, Laffont and Tirole (2004) establish only the existence (and not uniqueness) of the off-net-cost pricing equilibrium in a very specific case that will be commented later. Finally, we extend our results to a multi-period setting. In concrete, our main insights are as follows:

[10] For an overview of this literature see Jeon, Laffont and Tirole (2004).

Existence. Under linear demands, low enough substitutability between networks and a random noise with a wide enough support, there exists a unique equilibrium, which is interior and where networks choose the same call and reception prices over the time.

Pricing. In equilibrium, networks price calls at their off-net cost, whatever the sizes of the installed bases. Fixed fees and full-period profits are neutral with respect to the level of the *per-period* access markup.

Role of access charges. Should one ban unconstrained interconnection negotiations over reciprocal access charges? The off-net-cost pricing equilibrium neutralizes the potential anticompetitive role that future reciprocal access charges could play. In other words, under RPP networks cannot use access charges as an instrument to soften retail competition, whereas under CPP they can increase profits and decrease consumers surplus by setting future reciprocal access charges different from marginal costs.

Welfare. Should one set cost-bases access charges? As already noted, full-period profits do not depend on the level of the access markups, nevertheless an increase in the access markup raises the call price and decreases the reception charge. These two effects introduce a clear distortion in the consumer welfare. Given we have assumed a random noise in the utility functions, we look for the level of the access markup that maximizes the *expected* social welfare. We conclude that the optimal value of the access markup depends on the characteristics of each market in particular. Although, we can demonstrate that if the caller's and receiver's utility functions are identical, then starting from zero access markup, a decrease in the access charge raises the expected social welfare. Indeed, we show that 'bill and keep' might be optimal in this situation.

The article is organized as follows. Section 2 presents the main insights of our previous work. Section 3 describes the model and makes the main assumptions. Section 4 analyzes the two-period game, characterizes the equilibrium, studies the role of the access charge and extends the basic model to a multi-period setting. Section 5 investigates how the access charge affects the social welfare and studies its optimal level. Section 6 summarizes the main insights.

2. HOW TO SOFTEN NETWORK COMPETITION UNDER THE CPP

To provide a motivation for our analysis, it is convenient to introduce briefly the main insights of our previous work, where CPP is assumed throughout. In particular, we show below that under CPP networks can soften retail competition by setting access charges different from marginal costs. To that end, let $\widehat{\pi}_2^i$ denote the equilibrium second-period profits, which depends on the second-period access markup m_2, and the network $i's$ first-period market share α_1^i provided that switching costs exist. In the second period the model is similar to the traditional static model in which the symmetric equilibrium profits are neutral with respect to the access markup[11]. Moreover, the equilibrium second-period call price p_2^i is equal to the cost of an average call originating on network i, that is,

$$p_2^i = c + \alpha_2^j m_2,$$

where c is the industry's marginal cost of a call. Recall that in the first period, each network sets prices taking into account its first-period profitability, but also the effect that its first-period market share will have on its second-period profits. In particular, network i chooses the first-period call price p_1^i and the first-period fixed fee F_1^i so as to maximize its total discounted profits, taking network $j's$ first-period call price and fixed fee as given. As already pointed out in Laffont et al. (1998a), it is analytically convenient to view network competition as one in which the networks pick usage fees and net surpluses rather than usage fees and fixed fees, so that market shares are determined directly by net surpluses. The net surplus that a network $i's$ subscriber derives in the first period is: $w_1^i \equiv v(p_1^i) - F_1^i$, where $v(p_1^i)$ is the subscriber's indirect utility function who faces a call price of p_1^i. Thus network i solves:

$$\max_{p_1^i, w_1^i} \Pi \equiv \pi_1^i(p_1^i, p_1^j, w_1^i, w_1^j) + \delta \widehat{\pi}_2^i(m_2, \alpha_1^i(w_1^i, w_1^j)),$$

where π_1^i denotes the network $i's$ first-period profits and δ the discount factor. The first-order condition with respect to p_1^i yields $p_1^i = c + \alpha_1^j m_1$, that is, networks choose their call prices in the same way as they do in the second period. Further, the first-order condition with respect to w_1^i is

$$0 = \frac{\partial \pi_1^i}{\partial w_1^i} + \delta \frac{\partial \widehat{\pi}_2^i}{\partial \alpha_1^i}(m_2, \alpha_1^i) \frac{\partial \alpha_1^i}{\partial w_1^i}. \tag{1}$$

That is, the equilibrium first-period fixed fees are given as a function of m_2 through the term $\partial \widehat{\pi}_2^i / \partial \alpha_1^i$. In addition, we can show that in a symmetric equilibrium the full-period profits are equal to

$$\Pi(m_2) = \frac{1+\delta}{4\sigma} - \frac{\delta}{2} \frac{\partial \widehat{\pi}_2^i}{\partial \alpha_1^i}(m_2, 1/2),$$

where σ is the degree of substitutability between the two networks. Moreover, $\partial \widehat{\pi}_2^i / \partial \alpha_1^i(0, 1/2) > 0$, thus so as to satisfy (1) it must hold that $\partial \pi_1^i / \partial w_1^i < 0$, that is, in the neighborhood of $m_2 = 0$, networks compete more aggressively in the first period than they would do in a market without switching costs. More

[11] Indeed, the symmetric equilibrium profits are equal to the profits that networks would obtain under unit demands.

importantly,

$$\frac{\partial \widehat{\pi}_2^i}{\partial m_2 \partial \alpha_1^i}(0, 1/2) = 0, \qquad \frac{\partial}{\partial m_2}\left(\frac{\partial \widehat{\pi}_2^i}{\partial m_2 \partial \alpha_1^i}(0, 1/2)\right) < 0.$$

That is, slightly moving m_2 away from zero reduces the value of having a higher market share in the second period: $\partial \widehat{\pi}_2^i / \partial m_2 \partial \alpha_1^i$; this in turn softens competition for market share in the first period and increases full-period profits. An explanation for this result can be found in the Proposition 1 of Carter and Wright (2003), which proves that the profits of the larger network decrease when the access charge departs away from the marginal cost. Intuitively, as higher or lower is the second-period access markup with respect to the marginal costs, lower the second-period profits for the larger firm will be, and consequently the competition for first-period market share is disincentived. Notice that equilibrium first-period profits are independent of the reciprocal access charges, thus m_1 does not have to be different from m_2 to undermine network competition, that is, $m_1 = m_2 = m \neq c_0$ will also increase networks profits and decrease consumer welfare. The analysis below shows that under RPP networks can no longer increase full-period profits by colluding over the level of the future reciprocal access charges. Indeed, under RPP: $\partial \widehat{\pi}_2^i / \partial m_2 \partial \alpha_1^i = 0 \; \forall m_2, \alpha_1^i$; in effect, competition in call prices, fixed fees and reception prices neutralizes the effects that access charges have on equilibrium full-period profits.

3. THE MODEL

There are two networks indexed by i and j. Each has its own full coverage network and competes for a consumer set of measure 1. It is assumed that every consumer joins one of the networks, that is, there is full participation. In addition, networks are assumed to be interconnected, therefore a consumer who subscribes to one network can call any other consumer on either network. The usual balanced-traffic assumption is maintained throughout the analysis, which implies that the percentage of calls originating on a network and completed on the same network is equal to the market share of this network. Networks compete in nonlinear prices, and offer a three-part tariff: $\{F^i, p^i, r^i\}$, where F^i denotes network $i's$ fixed fee, and p^i and r^i represent respectively the per-unit call and reception charge. For off-net calls, the originating network must pay a reciprocal access charge a per unit of termination to the terminating network.[12] Moreover, networks are not allowed to price discriminate between calls that terminate on- and off-net.

Cost structure. Symmetric costs are assumed for simplicity. The cost of

[12] Reciprocity means that a network pays as much for termination of a call on the rival network as it receives for completing a call originated on the rival network.

serving a customer is $f \geq 0$, which reflects the cost of connecting the customer's home to the network and of billing and servicing the customer. The marginal cost of terminating or originating a call is denoted by c_0, and the marginal trunk cost of a call by c_1. The total cost is c. The access mark-up is thus

$$m \equiv a - c_0.$$

Demand structure. Networks sell a differentiated but substitutable product, they are differentiated à la Hotelling. Consumers are uniformly located on the segment $[0, 1]$ and the two networks are located at the two ends of the interval. Thus, consumers' tastes for networks are represented by their position on the segment and taken into account through the transportation costs τ. Given income y a consumer located at x and joining network i has utility

$$y + v_0 - \tau |x - x_i| + w^i,$$

where v_0 represents a fixed surplus from being connected to either network,[13] $\tau |x - x_i|$ is the cost of subscribing to a network located at x_i, and w^i is the net surplus of a network i subscriber from making and receiving calls on that network.

Timing. We consider three stages. In the first stage or period zero, reciprocal access charges are set by a regulator or negotiated between carriers; a flexible regulation is allowed, so that access charges may differ over time. In the first and second periods, which are indexed by $t \in \{1, 2\}$, networks compete in retail prices, taking as given the access charges.

Dynamics. Every customer incurs a cost $s > 0$ when switching networks.[14] Note that if $s > \tau$ every consumer remains with the same network in a symmetric equilibrium. We assume instead that $s < \tau$, so that at least some consumers switch. In addition, we shall make the following two assumptions:

A.1. *Preferences are independent across periods.*

A.2. *Consumers have naive expectations.*

The first assumption only says that preferences may change over time.[15] On the other hand, A.2. imposes a strong condition on the consumers' behavior.

[13] y and v_0 are assumend to be large enough such that the full participation assumption is satisfied.

[14] Quite obviously, in the absence of switching costs, networks are per-period profit maximizing. There is however much evidence suggesting that switching costs are significant (see for instance De Bijl and Peitz, 2000)

[15] This case might also arise when the customers are different in different periods and second-period customers are exposed to the choice of first-period customers. Actually, asumming constant preferences over time introduces technical problems when the Hotelling model is used: for some variations in prices, market shares remain constant.

It will however be argued that rational consumer expectations would not affect the main insights of the paper. From now on and without any loss of generality assume network i is located at the beginning of the segment $[0,1]$ and network j at the end. Then, a consumer located at $x = \alpha_1$ is indifferent between the two networks in the first period if and only if

$$w_1^i - \tau\alpha_1 = w_1^j - \tau(1-\alpha_1).$$

Therefore, the network $i's$ market share is

$$\alpha_1^i = \frac{1}{2} + \sigma\left(w_1^i - w_1^j\right),$$

where $\sigma = 1/2\tau$ is the index of substitutability between the two networks. At the beginning of the second period there is a fraction α_1^i of consumers initially attached to network i. For these and given A.1 and A.2, a consumer located at $x \in [0,1]$ will remain associated with network i if $w_2^i - \tau x \geq w_2^j - \tau(1-x) - s$. A consumer initially attached to network j, say x, will instead switch to network i if $w_2^i - \tau x - s \geq w_2^j - \tau(1-x)$. Therefore, the network $i's$ second-period market share is

$$\begin{aligned} \alpha_2^i &= \alpha_1^i\left[\frac{1}{2} + \sigma\left(w_2^i - w_2^j + s\right)\right] + \alpha_1^j\left[\frac{1}{2} + \sigma\left(w_2^i - w_2^j - s\right)\right] \\ &= \frac{1}{2} + (2\alpha_1^i - 1)\sigma s + \sigma\left(w_2^i - w_2^j\right). \end{aligned} \tag{2}$$

Finally, networks have rational expectations and discount second-period revenues and costs by a factor δ.

Demand for traffic. Subscribers derive a surplus from making and receiving calls. The utility from placing q calls is denoted by $\mu(q)$, whereas the utility from receiving $\widetilde{q}$ calls is denoted by $\widetilde{\mu}(\widetilde{q})$; we assume that these utility functions are twice continuously differentiable, with $\mu' > 0$, $\mu'' < 0$, $\widetilde{\mu}' > 0$, and $\widetilde{\mu}'' < 0$.[16] The analysis faces the problem of *sovereignty:* who decides to end the call? The receiver's demand function $\widetilde{q}(r)$ is given by $\widetilde{\mu}'(\widetilde{q}) = r$, whereas the caller's demand function $q(p)$ is given by $\mu'(q) = p$; When receivers are allowed to hang up the volume of calls from network i to network j is thus given by $Q(p^i, r^j) = \min\{q(p^i), \widetilde{q}(r^j)\}$; In a deterministic framework, this makes the model discontinuous and complicates its analysis.[17] In order to get around this

[16] Throughout this paper the apostrophe symbol means the first derivative of the considered function with respect to its argument. In this case for instance $\mu' = d\mu/dq$ and $\mu'' = d^2\mu/(dq)^2$.

[17] When reception charges are regulated or contractually determined before networks compete in retail tariffs, an assumption that simplifies much the analysis is that the caller determines the volume of calls. However, when reception charges are set by the networks at the same time they chose call prices and fixed fees, this assumption introduces a potential problem in the analysis due to the *multiplicity of equilibria:* from the viewpoint of networks and subscribers, only the sum $\{F^i + r^i\widetilde{q}\}$ matters, not its composition; hence different combinations of F^i and

problem we assume that both the caller's and receiver's utilities are subject to a random noise, which smooths the demand. To that end, let ε and $\widetilde{\varepsilon}$ denote, respectively, the random term of the caller's and receiver's utilities, and assume that: i) they follow respectively the distribution functions $F(\cdot)$ and $\widetilde{F}(\cdot)$ with supports $[\underline{\varepsilon}, \overline{\varepsilon}]$ and $[\underline{\widetilde{\varepsilon}}, \overline{\widetilde{\varepsilon}}]$ where $\overline{\varepsilon} - \underline{\varepsilon} > 0$ and $\overline{\widetilde{\varepsilon}} - \underline{\widetilde{\varepsilon}} > 0$, and strictly positive density functions $f(\cdot)$ and $\widetilde{f}(\cdot)$; ii) they are identically and independently distributed for each caller-receiver pair. We then make the following assumption:

A.3. *The caller's utility is given by:* $u = \mu(q) + \varepsilon q$, *whereas the receiver's utility is given by:* $\widetilde{u} = \widetilde{\mu}(\widetilde{q}) + \widetilde{\varepsilon}\widetilde{q}$.

Assumption A.3. allows the willingness to stay on the phone to be state-contingent for both callers and receivers. In addition, demands q and $\widetilde{q}$ are assumed to be bounded, hence for a given $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$ there exist price levels $\underline{p}$ and $\overline{p}$ such that if $p \leq \underline{p}$ then $q = \overline{q}$, where $0 < \overline{q} < \infty$, and if $p \geq \overline{p}$ then $q = 0$. Similarly, for a given $\widetilde{\varepsilon} \in [\underline{\widetilde{\varepsilon}}, \overline{\widetilde{\varepsilon}}]$ there exist price levels $\underline{r}$ and $\overline{r}$ such that if $r \leq \underline{r}$ then $\widetilde{q} = \overline{\widetilde{q}}$, where $0 < \overline{\widetilde{q}} < \infty$, and if $r \geq \overline{r}$ then $\widetilde{q} = 0$. Therefore, $p^i \in [\underline{p}, \overline{p}]$, $r^i \in [\underline{r}, \overline{r}]$, and since $\alpha^i \in [0, 1]$ the networks' profit functions are also bounded.

4. ANALYSIS

Under A.3., and for a given pair of prices (p_t^i, r_t^j), the *expected* volume of calls from a network i subscriber to a network j subscriber at period t is given by:[18]

$$\begin{aligned} Q(p_t^i, r_t^j) &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} [q(p_t^i, \varepsilon) \Im_{\text{if } q(p_t^i, \varepsilon) \leq \widetilde{q}(r_t^j, \widetilde{\varepsilon})} \\ &\quad + \widetilde{q}(r_t^j, \widetilde{\varepsilon}) \Im_{\text{if } q(p_t^i, \varepsilon) > \widetilde{q}(r_t^j, \widetilde{\varepsilon})}] f(\varepsilon) \widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon}. \end{aligned}$$

Further, the *expected* utility that a network i subscriber derives from calling a network j subscriber at period t is

$$\begin{aligned} U(p_t^i, r_t^j) &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} [u(q(p_t^i, \varepsilon)) \Im_{\text{if } q(p_t^i, \varepsilon) \leq \widetilde{q}(r_t^j, \widetilde{\varepsilon})} \\ &\quad + u(\widetilde{q}(r_t^j, \widetilde{\varepsilon})) \Im_{\text{if } q(p_t^i, \varepsilon) > \widetilde{q}(r_t^j, \widetilde{\varepsilon})}] f(\varepsilon) \widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon}, \end{aligned}$$

r^i are feasible equilibria but nonequivalent since each combination may affect differently the rival network.

[18] Throughout the analysis the symbol $\Im_{if}$ means that the double integral of the term that is located at its left side is defined if and only if the condition that is located at its right side is satisfied.

while the *expected* utility that a subscriber from network j derives from receiving calls from a network i subscriber at period t is given by:

$$\begin{aligned}\widetilde{U}(p_t^i, r_t^j) &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\underline{\widetilde{\varepsilon}}_r}^{\overline{\widetilde{\varepsilon}}} [\widetilde{u}(q(p_t^i, \varepsilon))\Im_{\text{if } q(p_t^i,\varepsilon)\leq \widetilde{q}(r_t^j,\widetilde{\varepsilon})} \\ &\quad + \widetilde{u}(\widetilde{q}(r_t^j, \widetilde{\varepsilon}))\Im_{\text{if } q(p_t^i,\varepsilon) > \widetilde{q}(r_t^j,\widetilde{\varepsilon})}] f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon}.\end{aligned}$$

Therefore, the volume of traffic from network i to network j depends on two usage prices and is sometimes determined by the caller and at other times by the receiver. In this framework we still find the following standard results:

$$\frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i} = \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \frac{\partial q(p_t^i - \varepsilon)}{\partial p_t^i} \Im_{\text{if } q(p_t^i,\varepsilon)\leq \widetilde{q}(r_t^j,\widetilde{\varepsilon})} f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon},$$

$$\frac{\partial U(p_t^i, r_t^j)}{\partial p_t^i} = p_t^i \frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i}, \tag{3}$$

where we have used $\partial u(q(p_t^i - \varepsilon))/\partial q = \mu'(\mu'^{-1}(p_t^i - \varepsilon)) + \varepsilon = p_t^i$. And,

$$\frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i} = \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \frac{\partial \widetilde{q}(r_t^i - \widetilde{\varepsilon})}{\partial r_t^i} \Im_{\text{if } q(p_t^j,\varepsilon)\leq \widetilde{q}(r_t^i,\widetilde{\varepsilon})} f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\widetilde{\varepsilon} d\varepsilon,$$

$$\frac{\partial \widetilde{U}(p_t^j, r_t^i)}{\partial r_t^i} = r_t^i \frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i}, \tag{4}$$

where we have used $\partial \widetilde{u}(\widetilde{q}(r_t^i - \widetilde{\varepsilon}))/\partial \widetilde{q} = \widetilde{\mu}'(\widetilde{\mu}'^{-1}(r_t^i - \widetilde{\varepsilon})) + \widetilde{\varepsilon} = r_t^i$. For the sake of the presentation, we will write $Q_t^{ij} = Q(p_t^j, r_t^j)$, $U_t^{ij} = U(p_t^i, r_t^j)$ and $\widetilde{U}_t^{ij} = \widetilde{U}(p_t^i, r_t^j)$ $\forall i, j$. Recall that network $i's$ second-period market share is

$$\alpha_2^i = \frac{1}{2} + (2\alpha_1^i - 1)\sigma s + \sigma(w_2^i - w_2^j), \tag{5}$$

where the *expected* net surplus of a network i consumer at period t is defined as

$$w_t^i = \phi_t^i - F_t^i, \tag{6}$$

with

$$\begin{aligned}\phi_t^i(\alpha_t^i, p_t^i, p_t^j, r_t^i, r_t^j) &= \alpha_t^i U_t^{ii} + \alpha_t^j U_t^{ij} + \alpha_t^i \widetilde{U}_t^{ii} + \alpha_t^j \widetilde{U}_t^{ji} \\ &\quad - p_t^i \left(\alpha_t^i Q_t^{ii} + \alpha_t^j Q_t^{ij}\right) - r_t^i \left(\alpha_t^i Q_t^{ii} + \alpha_t^j Q_t^{ji}\right).\end{aligned} \tag{7}$$

4.1. THE SECOND-PERIOD CASE

In the second period networks maximize profits with respect to call prices, reception charges and fixed fees; thus, any network i solves:

$$\max_{p_2^i, r_2^i, F_2^i} \pi_2^i \equiv \alpha_2^i\{\alpha_2^i(p_2^i - c)Q_2^{ii} + \alpha_2^j(p_2^i - c - m_2)Q_2^{ij} + \alpha_2^j m_2 Q_2^{ji} + r_2^i(\alpha_2^i Q_2^{ii} + \alpha_2^j Q_2^{ji}) + F_2^i - f\}. \quad (8)$$

We can solve (8) by maximizing it with respect to p_2^i and r_2^i for a given α_2^i, adapting fixed fees so that market shares remain constant. For this to hold, net surpluses must satisfy $w_2^i - w_2^j = (1/\sigma)(\alpha_2^i - 1/2) - (2\alpha_1^i - 1)s$; using (6) it follows that the fixed fee must be equal to

$$F_2^i = \phi_2^i - \phi_2^j + F_2^j - \frac{1}{\sigma}\left(\alpha_2^i - \frac{1}{2}\right) + (2\alpha_1^i - 1)s.$$

By substituting this last expression into the profit function we have:

$$\overline{\pi}_2^i(p_2^i, r_2^i) = \alpha_2^i\{\alpha_2^i(p_2^i - c)Q_2^{ii} + \alpha_2^j(p_2^i - c - m_2)Q_2^{ij} + \alpha_2^j m_2 Q_2^{ji} + r_2^i(\alpha_2^i Q_2^{ii} + \alpha_2^j Q_2^{ji}) + \phi_2^i - \phi_2^j + F_2^j - \frac{1}{\sigma}\left(\alpha_2^i - \frac{1}{2}\right) + (2\alpha_1^i - 1)s - f\}. \quad (9)$$

For given $r_2^i = r_2^j = r_2$ and p_2^j, the call price p_2^i determines the volume of calls generated by network i when callers are sovereign on average, network i incurs a unit cost $c + \alpha_2^j m_2$ from delivering these calls to network i and network j. However, since the call price affects subscribers' net surplus as well, fixed fees must be adapted in order to maintain markets shares constant; more precisely, a decrease in the call price p_2^i :

- affects network $i's$ revenue, but at the expense of consumers; hence to keep market shares constant fixed fees must be adapted so as to neutralize this transfer.
- allows network i to increase its fixed fee by $U(p_2^i, r_2)$, which is the utility that network $i's$ subscribers obtain from making calls (call prices also affect the utility from receiving calls but they affect both networks' consumers in the same way, so that fixed fees do not need to be adapted to maintain market shares.)
- affects the quantity of money that network $j's$ subscribers pay for calls received from network i : this effect is called *pecuniary externality* in Jeon et al. (2004), and allows network i to increase its fixed fee by $r_2\alpha_2^i$ and keep market shares constant.

We can summarize in the following expression the terms that are affected by the level of the network $i's$ call price, when adjusting the fixed fee so as to maintain market shares constant:

$$\alpha_2^i\{[-(c+\alpha_2^j m_2)+r_2\alpha_2^i]Q(p_2^i,r_2)+U(p_2^i,r_2)\}. \tag{10}$$

For given $p_2^i = p_2^j = p_2$ and r_2^j, setting the reception charge r_2^i similarly determines the volume of calls generated by network i when receivers are sovereign. For this volume of calls, network i incurs a cost $\alpha_2^i c$, but earns $\alpha_2^j m_2$ again from off-net calls. The reception charge r_2^i also affects subscribers' net surpluses, which requires fixed fees to be adapted so as to maintain market shares constant:

- First, network i gains revenue from reception charges, but its fixed fee must be altered by the same amount to keep market shares constant.
- Keeping market shares constant, network i can increase its fixed fee to reflect the utility obtained from receiving calls: $\widetilde{U}(p_2, r_2^i)$ (reception charges affect similarly both networks' subscribers for the calls they place to network $i's$ subscribers: $\alpha_2^i U(p_2, r_2^i)$; this therefore does not require fixed fees to be adapted).
- Finally, we find a new sort of *pecuniary externality:* the reception charge r_2^i determines how much network $j's$ consumers must pay for the calls they make to network $i's$ consumers. This externality allows network i to increase its fixed fee by $p_2\alpha_2^i$ while keeping market shares constant.

The following expression summarizes the terms that are affected by the level of network $i's$ reception charge, when adjusting the fixed fee so as to maintain market shares constant:

$$\alpha_2^i\{[-\alpha_2^i c+\alpha_2^j m_2+p_2\alpha_2^i]Q(p_2,r_2^i)+\widetilde{U}(p_2,r_2^i)\}. \tag{11}$$

By differentiating (10) with respect to p_2^i and (11) with respect to r_2^i, and using (3) and (4) we obtain the first-order conditions:

$$p_2^i = c+\alpha_2^j m_2 - \alpha_2^i r_2, \tag{12}$$

$$r_2^i = \alpha_2^i c - \alpha_2^j m_2 - \alpha_2^i p_2. \tag{13}$$

Essentially, we see that networks price calls and call receptions at their *strategic marginal cost:*[19] network $i's$ equilibrium call prices are equal to the average unit cost of a call originating on network i, minus the pecuniary externality imposed on network $j's$ subscribers; likewise, network $i's$ equilibrium reception charges are equal to the average cost of receiving calls on network i, minus the pecuniary externality imposed on network $j's$ consumers. Using $p_2^i = p_2$ and $r_2^i = r_2$, we

[19] Using the terminology of Jeon et al. (2004).

obtain the equilibrium call and reception prices:

$$p_2 = c + m_2, \tag{14}$$

$$r_2 = -m_2. \tag{15}$$

We shall emphasize that this symmetric solution is valid for any given level of market shares: hence (14) and (15) characterize the equilibrium second-period usage prices, which are symmetric whatever the sizes of customer installed bases. Now, setting call and reception prices at the equilibrium level, we can rewrite network $i's$ second-period profits as follows:

$$\pi_2^i = \left(\frac{1}{2} + (2\alpha_1^i - 1)\sigma s - \sigma(F_2^i - F_2^j)\right)(F_2^i - f). \tag{16}$$

By differentiating this last expression with respect to F_2^i we obtain the following first-order condition:

$$F_2^i = \frac{1}{2}\left[f + \frac{1}{2\sigma} + (2\alpha_1^i - 1)s + F_2^j\right]. \tag{17}$$

Similarly, we can obtain network $j's$ first-order condition with respect to its fixed fee, and by solving that system of two equations we obtain the equilibrium second-period fixed fees as a function of the first-period market shares:

$$\widehat{F}_2^i(\alpha_1^i) = f + \frac{1}{2\sigma} + \frac{(2\alpha_1^i - 1)s}{3}. \tag{18}$$

By substituting $\widehat{F}_2^i$ and $\widehat{F}_2^j$ into (16), we then obtain the equilibrium second-period profits as a function of first-period market shares:

$$\widehat{\pi}_2^i(\alpha_1^i) = \frac{1}{4\sigma} + (2\alpha_1^i - 1)\frac{s}{3} + (2\alpha_1^i - 1)^2\frac{\sigma s^2}{9}. \tag{19}$$

Notice that equilibrium second-period profits do not depend on m_2. Moreover, note from (19) that if $\alpha_1^i = 1/2$ the equilibrium second-period profits are equal to the profits that networks would obtain under unit demands, that is, $\pi_2^i(1/2) = 1/4\sigma$. In order to prove the existence and uniqueness of this equilibrium we will have to be more specific about the noise and the caller's and receiver's demand. We then make the following assumption:

A.4. $\mu(q) = aq - (b/2)q^2$ and $\widetilde{\mu}(\widetilde{q}) = d\widetilde{q} - (e/2)\widetilde{q}^2$, where $a, b, d, e > 0$. Moreover, $\varepsilon, \widetilde{\varepsilon} \in [\underline{\varepsilon}, \overline{\varepsilon}]$, where $\underline{\varepsilon} < 0 < \overline{\varepsilon}$, $E(\varepsilon) = E(\widetilde{\varepsilon}) = 0$, and both random terms follow a uniform distribution with density function: $f(\varepsilon) = \widetilde{f}(\widetilde{\varepsilon}) = 1/\Delta$, where $\Delta = \overline{\varepsilon} - \underline{\varepsilon}$.

Notice that A.3. and A.4. implies linear demand functions: $q = (a - p + \varepsilon)/b$

and $\widetilde{q} = (d - r + \widetilde{\varepsilon})/e$. Then we have the following proposition:

Proposition 1. *(Existence and Uniqueness) Under A.1, A.2, A.3 and A.4, for a small enough σ and a large enough Δ there exists a unique second-period equilibrium, which is interior and where networks choose:*

$$p_2^i = c + m_2,$$

$$r_2^i = -m_2,$$

$$F_2^i = f + \frac{1}{2\sigma} + \frac{(2\alpha_1^i - 1)s}{3}.$$

Proof. See Appendix. ∎

In summary, networks price calls at their off-net cost, that is, each network sets prices for making and receiving calls equal to the marginal cost that it could incur if all other subscribers belonged to the rival network. The off-net-cost pricing principle dates back to Laffont et al. (2003), which found this pricing rule in a framework for Internet backbone competition. In contrast, Jeon, Laffont and Tirole (2004) and our paper analyze three-part tariff competition in a telecommunications environment. At the expense of assuming linear demands, our setup however generalizes their work by allowing a random noise in both the callers and receivers' utilities, and by removing the assumption of a given proportionality between the utility functions. Moreover, Jeon, Laffont and Tirole only establish the existence of the off-net-cost pricing equilibrium when the noise on the receiver side converges to zero, so that the volume is determined by callers with probability converging to one. Instead, we have showed that the off-net-cost pricing equilibrium exists and is unique for a small enough σ and a large enough Δ. Indeed, a small (enough) σ (i.e., networks are relatively poor substitutes) is a standard assumption in the "two-way" access literature; and a large (enough) Δ is not a too restrictive assumption since extreme situations might happen in reality. For example, there exist many situations in which a person may not want to receive or make a call even though it is free. Finally, it is worth to remark that in the second period networks do not have incentives to corner the market by choosing a strategy different to that of the off-net-cost pricing one.

4.2. THE FIRST PERIOD

Recall that networks are assumed to be initially symmetric; thus, first-period market shares are given by

$$\alpha_1^i = \frac{1}{2} + \sigma(w_1^i - w_1^j),$$

where $w_1^i = \phi_1^i - F_1^i$. In the first period, network i chooses first-period usage price, reception charge and net surplus in order to maximize its total discounted profits:

$$\Pi^i(p_1^i, r_1^i, F_1^i) = \pi_1^i(p_1^i, r_1^i, F_1^i) + \delta\widehat{\pi}_2^i(\alpha_1^i(p_1^i, r_1^i, F_1^i)), \tag{20}$$

$$\begin{aligned} \text{with } \pi_1^i &= \alpha_1^i\{\alpha_1^i(p_1^i - c)Q_1^{ii} + \alpha_1^j\left[p_1^i - c - m_1\right]Q_1^{ij} + \alpha_1^j m_1 Q_1^{ji} \\ &\quad + r_1^i\left[\alpha_1^i Q_1^{ii} + \alpha_1^j Q_1^{ji}\right] + F_1^i - f\}, \end{aligned}$$

and where $\widehat{\pi}_2^i$ is given by (19). As above, we can maximize first Π^i with respect to p_1^i and r_1^i for a given α_1^i, adjusting fixed fees so as to keep α_1^i constant. Then, $\partial\Pi^i/\partial p_1^i = \partial\pi_1^i/\partial p_1^i$ and $\partial\Pi^i/\partial r_1^i = \partial\pi_1^i/\partial r_1^i$; therefore, networks choose their retail prices and reception charges in the same way as they do in the second period, that is, $p_1^i = c + m_1$ and $r_1^i = -m_1$. Now, we may proceed similarly to the analysis of the previous section, assume that first-period call and reception prices are at the equilibrium level, and rewrite the full-period profits as follows:

$$\Pi^i = \left(\frac{1}{2} - \sigma(F_1^i - F_1^j)\right)\{F_1^i - f\} + \delta\widehat{\pi}_2^i(\alpha_1^i(F_1^i, F_1^j)). \tag{21}$$

By differentiating this last expression with respect to F_1^i we obtain the following first-order condition:

$$0 = -\sigma\{F_1^i - f\} + \left(\frac{1}{2} - \sigma(F_1^i - F_1^j)\right) - \delta\sigma\frac{d\widehat{\pi}_2^i(\alpha_1^i)}{d\alpha_1^i}. \tag{22}$$

From (19) we have that $d\widehat{\pi}_2^i(\alpha_1^i)/d\alpha_1^i = (2s/3) + (4/9)(\alpha_1^i - \alpha_1^j)\sigma s^2$, therefore:

$$F_1^i = \frac{f}{2} + \frac{1}{4\sigma} + \frac{F_1^j}{2} - \frac{\delta s}{3} - \frac{4\sigma^2 s^2\delta}{9}(F_1^j - F_1^i). \tag{23}$$

Given the symmetry of the game in the first period, we may look for a symmetric solution where $F_1^i = F_1^j$, then it is easy to see from (23) that in equilibrium network i chooses:

$$F_1^i = f + \frac{1}{2\sigma} - \frac{2\delta s}{3}. \tag{24}$$

The following proposition gives the conditions for the existence and uniqueness of the first-period equilibrium:

Proposition 2. *Under A.1, A.2, A.3, and A.4, for a small enough σ and a large enough Δ : i) there exists a unique interior equilibrium where networks choose their first-period call and reception prices in the same way as they do in the second period:*

$$p_1 = c + m_1, \quad r_1 = -m_1,$$

ii) the equilibrium first-period fixed fees and full-period profits do not depend on *the level of the first or second-period access markup:*

$$F_1 = f + \frac{1}{2\sigma} - \frac{2s\delta}{3}, \qquad \Pi = \frac{1+\delta}{4\sigma} - \frac{s\delta}{3},$$

iii) there exists no "cornered-market" equilibrium if switching costs are small enough.

Proof. See Appendix. ∎

We may then conclude that networks can no longer use future reciprocal access charges as an instrument to soften first-period competition. Notice that as long as $\partial\widehat{\pi}_2^i/\partial\alpha_1^i > 0$ networks compete more aggressively in the first period, so as to build market share that is profitable in the second period. From López (2005) we know that when networks only compete in call prices and fixed fees, $\partial\widehat{\pi}_2^i/\partial\alpha_1^i$ depends on both α_1^i and m_2, and in a symmetric equilibrium slightly moving m_2 away from zero can reduce the value of having a higher market share in the second-period, $\partial\widehat{\pi}_2^i/\partial\alpha_1^i$ is strictly concave in m_2 at $m_2 = 0$, and therefore increase their full-period profits by softening first-period competition for market share. In contrast, when networks compete also in reception charges, $\widehat{\pi}_2^i$ depends only on first-period market shares, implying that $\partial^2\widehat{\pi}_2^i/\partial m_2\partial\alpha_1^i = 0 \ \forall m_2, \alpha_1^i$; hence, first-period competition does not depend on m_2, and neither do the full-period profits. In the rational consumer expectations case the expressions for the second-period equilibrium are the same as with naive expectations: (14), (15) and (18). In the first period, however, consumers recognize that a network with higher market share will charge higher prices in the second-period whenever switching costs are positive. Nevertheless, since the value of having a higher second-period market share is neutral with respect to the level of m_2, first-period prices are also neutral, and hence m_2 does not affect the subscribers' first-period net surpluses. In summary, with both naive and rational consumers expectations, networks cannot increase their full-period profits by departing m_2 away from zero when competition is in call prices, fixed fees and reception charges.

4.3. THE MULTI-PERIOD CASE

Assume networks compete in (finite) T discrete periods of time. Our setup is as follows: in each period $t = 1...T$, networks can condition their play at time t on the history of play until that date h_{t-1} (closed-loop or feedback strategies). Let $V_t^i(\cdot)$ denote the value function for network i at time t, with $V_{T+1}^i = 0$. We will provide sufficient conditions under which there exists a unique subgame-perfect equilibrium.

CLAIM 1: *Suppose networks compete in finite $T > 1$ discrete periods of time, and assume A.1, A.2, A.3 and A.4 holds, then for a small enough σ and a large enough Δ there exists an interior subgame-perfect equilibrium such that in any* *continuation equilibria (even off the equilibrium path): (i) networks price calls at their off-net cost, (ii) the fixed fees and per-period profits depend on h_t only through α^i_{t-1}, and moreover (iii) do not depend on the access markup levels.*

The proof of Claim 1 will proceed in several steps. First of all, note that the analysis of the game in period T is the same as in the two-period case, thus from proposition 1 we know that in period T, under A.1, A.2, A.3 and A.4, for a small enough σ and a large enough Δ there exists a unique equilibrium, which is interior and where networks price calls at the off-net cost. Moreover, we know that V^i_T exists, depends on h_{T-1} only through α^i_{T-1}, and is quadratic. Consider now period $T-1$ where α^i_{T-2} is given, network i knows h_{T-2}, and solves:

$$\begin{aligned}\max_{p^i_{T-1},r^i_{T-1},F^i_{T-1}} \Pi^i_{T-1} &\equiv \pi^i_{T-1}(p^i_{T-1},r^i_{T-1},F^i_{T-1},p^j_{T-1},r^j_{T-1},F^j_{T-1},\alpha^i_{T-2})\\ &\quad +\delta V^i_T(\alpha^i_{T-1}),\end{aligned}$$

where π^i_{T-1} is given by (8) and

$$\begin{aligned}\alpha^i_{T-1} &= 1/2+(2\alpha^i_{T-2}-1)\sigma s+\sigma(\phi^i_{T-1}(p^i_{T-1},r^i_{T-1},p^j_{T-1},r^j_{T-1},\alpha^i_{T-1})\\ &\quad -\phi^j_{T-1}(p^j_{T-1},r^j_{T-1},p^i_{T-1},r^i_{T-1},\alpha^i_{T-1})+F^j_{T-1}-F^i_{T-1}).\end{aligned}$$

The analysis can again be simplified by invoking the one-to-one relationship between F^i_t and α^i_t : network i choosing a tariff (p^i,r^i,F^i) given network $j's$ tariff (p^j,r^j,F^j), is equivalent to choosing (p^i,r^i,α^i). We can thus rewrite network $i's$ problem as follows:

$$\begin{aligned}\max_{p^i_{T-1},r^i_{T-1},\alpha^i_{T-1}} \overline{\Pi}^i_{T-1} &\equiv \overline{\pi}^i_{T-1}(p^i_{T-1},r^i_{T-1},\alpha^i_{T-1},p^j_{T-1},r^j_{T-1},F^j_{T-1},\alpha^i_{T-2})\\ &\quad +\delta V^i_T(\alpha^i_{T-1}),\end{aligned}$$

where $\overline{\pi}^i_{T-1}$ is given by (9). It then follows that $\partial\overline{\Pi}^i_{T-1}/\partial p^i_{T-1}=\partial\overline{\pi}^i_{T-1}/\partial p^i_{T-1}$ and $\partial\overline{\Pi}^i_{T-1}/\partial r^i_{T-1}=\partial\overline{\pi}^i_{T-1}/\partial r^i_{T-1}$; thus a candidate solution for the four first-order equilibrium conditions with respect to usage prices in period $T-1$ is $p^{i\,*}_{T-1}=p^{j\,*}_{T-1}=c+m_{T-1}$ and $r^{i\,*}_{T-1}=r^{j\,*}_{T-1}=-m_{T-1}$. Replacing these expressions into $\overline{\Pi}^i_{T-1}$, we can derive the corresponding candidate equilibrium fixed fees $F^{i\,*}_{T-1}$, which solve

$$\max_{F^i_{T-1}} \alpha^i_{T-1}\left\{F^i_{T-1}-f\right\}+\delta V^i_T(\alpha^i_{T-1}), \tag{25}$$

subject to

$$\alpha^i_{T-1}=(1/2)+(2\alpha^i_{T-2}-1)\sigma s-\sigma(F^i_{T-1}-F^j_{T-1}).$$

Note that (25) is a quadratic optimization problem, which implies that $F^{i\,*}_{T-1}$

is a linear function of α^i_{T-2}, and hence

$$\widetilde{V}^i_{T-1} \equiv \overline{\Pi}^i_{T-1}(p^{i\ *}_{T-1}, p^{i\ *}_{T-1}, F^{i\ *}_{T-1}, p^{j\ *}_{T-1}, r^{j\ *}_{T-1}, F^{j\ *}_{T-1}, \alpha^i_{T-2})$$

is a quadratic function of α^i_{T-2}. Consequently, if V^i_T exists, depends on h_{T-1} only through α^i_{T-1}, and is quadratic, there exists in period $T-1$ a candidate equilibrium where networks price calls at the off-net cost and where fixed fees depend on h_{T-2} only through α^i_{T-2}, and do so linearly, so that for this candidate equilibrium the valuation function $\widetilde{V}^i_{T-1}$ is also quadratic and depends on h_{T-2} only through α^i_{T-2}. Therefore, by mathematical induction we can derive a sequence of candidate equilibria for all t where networks price calls at the off-net cost and $F^{i\ *}_t$ depends on h_{t-1} only through α^i_{t-1}, and so a sequence of candidate valuation functions $\widetilde{V}^i_t$ that are quadratic and depends on h_{t-1} only through α^i_{t-1}.

Since we exhibit a candidate equilibrium by solving the first-order conditions, this candidate equilibrium is indeed an equilibrium if the Hessian of $\overline{\Pi}^i_{T-1}$ is definite negative since in that case second-order conditions are satisfied. That is, if the Hessian of $\overline{\Pi}^i_{T-1} \equiv \overline{\pi}^i_{T-1} + \delta V^i_T(\alpha^i_{T-1})$ is definite negative, then our candidate equilibrium is an equilibrium from $t = T-1$ onwards and $\widetilde{V}^i_{T-1}$ is well defined and the valuation function. It then follows that if the Hessian of $\overline{\Pi}^i_{T-2} \equiv \overline{\pi}^i_{T-2} + \delta\widetilde{V}^i_{T-1}$ is also definite negative, our candidate equilibrium is an equilibrium from $t = T-2$ onwards and $\widetilde{V}^i_{T-2}$ is also well defined and the valuation function, and so on. Before providing conditions under which this is so we first derive the sequence of candidate equilibrium fixed fees and valuation functions. In each period, the two networks each solve a linear-quadratic dynamic programming problem, and thus the candidate value functions $\widetilde{V}^i_t(\alpha^i_{t-1})$ are quadratic and characterized by coupled Ricatti equations that can be solved recursively. Let us define $y_t = (F^1_t, F^2_t, \alpha^1_t, 1)'$ and $x_t = (F^1_t, F^2_t)'$; the optimization problem for networks 1 and 2 in period t can be formulated, respectively, as follows:

$$\max_{F^1_t} \frac{1}{2} y'_t I y_t + \delta\widetilde{V}^1_{t+1}(y_t),$$

$$\max_{F^2_t} \frac{1}{2} y'_t J y_t + \delta\widetilde{V}^2_{t+1}(y_t),$$

subject to $y_t = Ay_{t-1} + Bx_t$, where

$$I = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -f \\ 0 & 0 & -f & 0 \end{bmatrix}, \ J = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & f \\ 0 & 1 & f & -2f \end{bmatrix},$$

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2\sigma s & 1/2 - \sigma s \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -\sigma & \sigma \\ 0 & 0 \end{bmatrix}.$$

Moreover, since $\widetilde{V}_t^i(y_{t-1})$ is quadratic for any $t < T$ we can write

$$\widetilde{V}_t^i(y_{t-1}) = \frac{1}{2} y'_{t-1} S_t^i y_{t-1}. \tag{26}$$

The matrix S_t^i can be obtained as follows[20]: define $B = [b_1, b_2]$, let

$$\Phi_t = \begin{bmatrix} b'_1 \Sigma_t^1 \\ b'_2 \Sigma_t^2 \end{bmatrix},$$

where

$$\Sigma_t^1 = I + \delta S_{t+1}^1,$$

$$\Sigma_t^2 = J + \delta S_{t+1}^2.$$

Note that network 1 and 2 solve, respectively,

$$\max_{F_t^1} \left\{ \frac{1}{2} (A y_{t-1} + b_1 F_t^1 + b_2 F_t^2)' \Sigma_t^1 (A y_{t-1} + b_1 F_t^1 + b_2 F_t^2) \right\}$$

$$\max_{F_t^2} \left\{ \frac{1}{2} (A y_{t-1} + b_1 F_t^1 + b_2 F_t^2)' \Sigma_t^2 (A y_{t-1} + b_1 F_t^1 + b_2 F_t^2) \right\}$$

Consequently, the couple of first-order conditions are

$$b'_1 \Sigma_t^1 (A y_{t-1} + b_1 F_t^1 + b_2 F_t^2) = 0,$$

$$b'_2 \Sigma_t^2 (A y_{t-1} + b_1 F_t^1 + b_2 F_t^2) = 0.$$

Finally, by solving this system of two linear equations we might find the rule for the *candidate* equilibrium fixed fees, given by

$$x_t = E_t y_{t-1}, \tag{27}$$

where $E_t = -(\Phi_t B)^{-1} \Phi_t A y_{t-1}$. In addition, S_t^1 and S_t^2 are determined by

$$S_t^1 = (A + BE_t)'(I + \delta S_{t+1}^1)(A + BE_t),$$

$$S_t^2 = (A + BE_t)'(J + \delta S_{t+1}^2)(A + BE_t).$$

[20] We follow here the same procedure as in Kydland (1975), although we use the trick of including the constant 1 in the list of state variables so as to express networks' profits in a simple quadratic form.

By iterating the above process one can compute the candidate value function for both networks at any $t \in [1, T-1]$. In fact, we can compute recursively the unique solution in fixed fees for any period t and hence guarantee the existence of a well-defined $\widetilde{V}_t^1(y_{t-1})$ only if $|\Phi_t B| \neq 0$ for $t = 1, ...T$, which is satisfied here since networks compete for market share. That is, we would have $|\Phi_t B| = 0$ only if reaction functions had the same slope, in which case there would be infinitely many solutions, or no solution (see Kydland, 1975.) Note that both $\widetilde{V}_t^1$ and $F_t^{1\ *} \equiv x(1)$ do not depend on m_t. Recall that if the Hessians of $\overline{\Pi}_t^1$ and $\overline{\Pi}_t^2$ are definite negative in own strategies for all t, our candidate equilibrium will be a subgame-perfect equilibrium. Let H_t^i denote the Hessian matrix of $\overline{\Pi}_t^i$ under the candidate equilibrium, and let $(H_t^i)_k$ denote the $k-th$ principal minor of the Hessian matrix H_t^i. To prove Claim 1, it suffices to apply the following proposition,

Proposition 3. *Under A.1, A.2, A.3 and A.4, for a small enough σ and a large enough Δ, $\left|\left(H_t^i\right)_1\right| < 0$, $\left|\left(H_t^i\right)_2\right| > 0$, and $\left|\left(H_t^i\right)_3\right| < 0$ $\forall t$.*

Proof. See Appendix. ∎

A couple of remarks are in order:

Remark 1. (Uniqueness) To prove uniqueness we can follow a similar reasoning to that of proof of proposition 2: by assuming a large enough Δ we can reduce the set of candidate equilibria in usage prices to a singleton where usage prices are set at their off-net cost: $p_t^{i\ *} = p_t^{j\ *} = c + m_t$ and $r_t^{i\ *} = r_t^{j\ *} = -m_t$. Moreover, at this level we have that $[F_t^{i\ *}(\alpha_{t-1}^i), F_t^{j\ *}(\alpha_{t-1}^i)]$ are uniquely determined and given by (27). Finally we know from above that this unique candidate equilibrium is indeed an equilibrium for a (positive) small enough σ.

Remark 2. (Corner Equilibrium) We now show that no cornered-market equilibrium exists when switching costs are not too high. Suppose there exists an equilibrium where network i corners the market in any period t by setting $(p_t^{i\ *}, r_t^{i\ *}, F_t^{i\ *})$ given that network j sets $(p_t^{j\ *}, r_t^{j\ *}, F_t^{j\ *})$. Then, $\pi_t^{j\ *} = 0$ and $\Pi_t^{j\ *} = \delta V_{t+1}^j(0)$. And $\Pi_t^{i\ *} = \pi_t^{i\ *} + \delta V_{t+1}^i(1)$, where $\pi_t^{i\ *} = (p_t^{i\ *} - c + r_t^{i\ *})Q(p_t^{i\ *}, r_t^{i\ *})$. But in order to corner the market network i must sacrifice present profits so as to attract consumers. It means that $\pi_t^{i\ *}$ is lower than the static equilibrium profits, which is always interior. Moreover, as switching costs decrease, the link between the present and the future vanishes, that is, $\lim_{s\to 0} V_{t+1}^i(1) = V_{t+1}^i(1/2)$. Therefore, $\lim_{s\to 0} \Pi_t^{i\ *} < \lim_{s\to 0} \widehat{\Pi}_t^i = \widehat{\pi}_t^i(\alpha_{t-1}^i) + \delta V_{t+1}^i(1/2)$, where as before $\widehat{\pi}_t^i(\alpha_{t-1}^i)$ denote the equilibrium profits of network i in period t as a function of α_{t-1}^i. Thus, a (positive) small enough s is a sufficient condition under which no "cornered-market" equilibrium exists.

5. SOCIAL OPTIMUM

Jeon et al. (2004) already pointed out that efficiency cannot be achieved in the presence of noise since marginal utilities have a random term, which in turn requires price instruments to be contingent on the realization of this term. We address this problem in a different way, and we look for the level of the access markup that maximizes the *expected* social welfare.

We begin by considering symmetric networks, that is, $\alpha_1^i = \alpha_1^j = 1/2$. It follows from proposition 1 that the symmetric equilibrium is the unique possible one so the market is again equally divided in the second period. This symmetric solution minimizes the average consumer's disutility from not being able to join to his preferred network, and hence allow us to rule out this social cost from the analysis for the moment. Since payments are only transfers from one agent to another, from a social-welfare viewpoint what matters is the utility that consumers derive from making and receiving calls, and the costs of these calls. Consider a call from a network i consumer to a network j consumer, its length is given by Q^{ij}, and the total utility derived by both the caller and the receiver from this call is: $U^{ij} + \widetilde{U}^{ij}$. Let $W(m_t)$ denote the expected welfare arising from this call, in equilibrium

$$W(m_t) = U(c + m_t, -m_t) + \widetilde{U}(c + m_t, -m_t) - cQ(c + m_t, -m_t). \tag{28}$$

The first-order condition is

$$\begin{aligned} \frac{dW}{dm_t}(m_t) &= \left(\frac{\partial U}{\partial p_t}(m_t) - \frac{\partial U}{\partial r_t}(m_t)\right) + \left(\frac{\partial \widetilde{U}}{\partial p_t}(m_t) - \frac{\partial \widetilde{U}}{\partial r_t}(m_t)\right) \\ &\quad - c\left(\frac{\partial Q}{\partial p_t}(m_t) - \frac{\partial Q}{\partial r_t}(m_t)\right) \\ &= 0. \end{aligned} \tag{29}$$

A small increase in m_t implies two opposite effects: it increases p_t^i but also decreases r_t^i. Moreover, a small increase in p_t^i reduces the callers' willingness to stay on the phone, and consequently it decreases both the callers' utility $\partial U/\partial p_t^i$ and the receivers' utility $\partial\widetilde{U}/\partial r_t^i$. On the other hand, a small decrease in r_t^i increases the receivers' willingness to stay on the phone, which in turn increases the utility of both callers and receivers: $-(\partial U/\partial r_t + \partial\widetilde{U}/\partial r_t)$. Thus, on one hand, it decreases the volume of traffic in which callers are sovereign and hence the costs incurred in these calls; this social gain is given by: $-c(\partial Q/\partial p_t)$. At the same time, however, it increases the volume of traffic in which receivers are sovereign, which implies a social cost equal to $c(\partial Q/\partial r)$. Using (3) and (4) yields, in equilibrium, $\partial U/\partial p_t = (c + m_t)(\partial Q/\partial p_t)$ and $\partial\widetilde{U}/\partial r_t = -m_t(\partial Q/\partial r_t)$; so equation (29) boils down to:

$$\begin{aligned}\frac{dW}{dm_t}(m_t) &= m_t\left(\frac{\partial Q}{\partial p_t}(m_t)+\frac{\partial Q}{\partial r_t}(m_t)\right)+c\frac{\partial Q}{\partial r_t}(m_t) \\ &\quad+\left(\frac{\partial \widetilde{U}}{\partial p_t}(m_t)-\frac{\partial U}{\partial r_t}(m_t)\right) \\ &= 0.\end{aligned} \tag{30}$$

Moreover, under the existence and uniqueness conditions of proposition 1:

$$d^2W/(dm_t)^2 \simeq -(1/2)(1/b+1/e+e/b^2+b/e^2)<0.$$

Letting m_t^* denote the optimal access markup we thus have that any m_t^* such that $(dW/dm_t)(m_t^*)=0$ is socially optimal. In order to be more precise let us state the following proposition,

Proposition 4. *Under A.1, A.2, A.3, A.4, for a small enough σ and a large enough Δ, in equilibrium:*

$$\frac{\partial \widetilde{U}}{\partial p}(m_t) \simeq -\left(\frac{1}{b}\right)\left[\frac{e}{2b}(c+m_t-a)+\frac{d}{2}+\frac{\Delta}{8}\left(\frac{e}{b}+\frac{b}{3e}\right)\right],$$

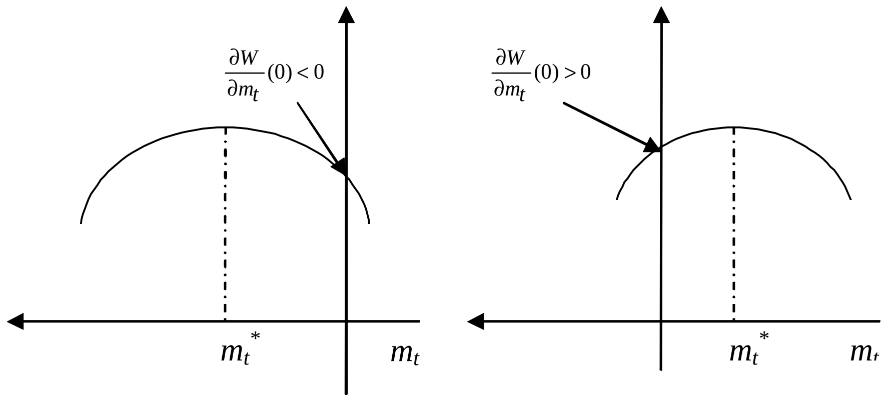
$$\frac{\partial U}{\partial r}(m_t) \simeq -\left(\frac{1}{e}\right)\left[-\frac{b}{2e}(m_t+d)+\frac{a}{2}+\frac{\Delta}{8}\left(\frac{b}{e}+\frac{e}{3b}\right)\right].$$

Proof. See Appendix. ∎

Clearly, the above proposition points out that the optimal value of the access markup depends on the characteristics of each market in particular. Consider now a small increase in the access markup starting from $m_t=0$, it slightly increases call prices and slightly decreases the reception price. More precisely,

$$\frac{dW}{dm_t}(0)=\left(c\frac{\partial Q}{\partial r_t}(0)+\frac{\partial \widetilde{U}}{\partial p_t}(0)\right)-\frac{\partial U}{\partial r_t}(0). \tag{31}$$

This expression is in general different from zero. Roughly speaking, $m_t=0$ is (generically) never optimal. Indeed, if $(dW/dm_t)(0)<0$ it follows that $m_t^*<0$, and conversely $m_t^*>0$ if $(dW/dm_t)(0)>0$ (see figure below.) Assume that $u(x)=\widetilde{u}(x)$, that is, $a=d$ and $b=e$. Now, making use of Proposition 4 yields: $(\partial\widetilde{U}/\partial p)(0)-(\partial U/\partial r)(0)\simeq -c/2b<0$. Consequently, the consumers' surplus decreases: the small decrease in r_t increases the callers' utility less than the decrease in the receivers' utility that is driven by the small increase in p_t. This social cost together with the cost incurred by the increase in the average length of calls yield $(dW/dm_t)(0)\simeq -c/b<0$. Conversely, a small decrease in m_t will decrease p_t and increase r_t such that the receivers' utility increases more than the loss in the callers' utility, moreover the average length of calls decreases since



r_t increases, which indeed decreases costs in $|c\partial Q/\partial r_t|$. Then, it is optimal to decrease m_t, that is, $m_t^* < 0$. Given this, we have proved the following proposition:

PROPOSITION 5. *If $u(x) = \widetilde{u}(x)$, then $m_t^* < 0$ and is given by (30) if $a_t^* = m_t^* + c_0 > 0$. Otherwise, 'bill and keep' is socially optimal and $m_t^* = -c_0$.*

So far we have assumed symmetric networks; let us now turn to the asymmetric case. The utility that any network $i's$ subscriber derives from calls is $\alpha_t^i U^{ii} + \alpha_t^j U^{ij} + \alpha_t^i \widetilde{U}^{ii} + \alpha_t^j \widetilde{U}^{ji}$, and the costs incurred by his calls are $(\alpha_t^i Q^{ii} + \alpha_t^j Q^{ij})c$. Since there are α_t^i consumers attached to network i and α_t^j consumers attached to network j, the total utility that consumers derive is:

$$\begin{aligned} &\alpha_t^i(\alpha_t^i U^{ii} + \alpha_t^j U^{ij} + \alpha_t^i \widetilde{U}^{ii} + \alpha_t^j \widetilde{U}^{ji}) + \alpha_t^j(\alpha_t^j U^{jj} + \alpha_t^i U^{ji} \\ &+\alpha_t^j \widetilde{U}^{jj} + \alpha_t^i \widetilde{U}^{ij}) - \alpha_t^i(\alpha_t^i Q^{ii} + \alpha_t^j Q^{ij})c - \alpha_t^j(\alpha_t^j Q^{jj} + \alpha_t^i Q^{ji})c. \end{aligned} \tag{32}$$

But in equilibrium expression (32) boils down to (28). Therefore, the above analysis remain valid in the asymmetric case. The intuition is very simple: since usage prices are identical in both networks whatever the market shares are, we have that in equilibrium $U^{ii} = U^{ij}$, $\widetilde{U}^{ij} = \widetilde{U}^{ji}$ and $Q^{ij} = Q^{ji}$; it then follows that consumers derive from calls the same utility in both networks. Let us now turn back to the consumer's disutility from not being able to join to his preferred network and the switching costs issue. Given first-period market shares α_1^i and α_1^j, the socially optimal configuration of market shares $(\alpha_2^{i\,*}, \alpha_2^{j\,*})$ minimizes both social costs. Suppose that $s = 0$ and the market is initially unequal divided between the two competitors (i.e., $\alpha_1^i \neq \alpha_1^j$), then $\alpha_2^{i\,*} = \alpha_2^{j\,*} = 1/2$ will still minimize the average consumer's disutility since preferences are assumed to be independent across periods. Nevertheless, if every subscriber incurs a cost when switching networks, then $\alpha_2^i = \alpha_2^j = 1/2$ is not necessarily optimal if $\alpha_1^i \neq \alpha_1^j$.

Note however that in equilibrium $\phi_t^i = \phi_t^j$, which amounts to

$$w_t^i - w_t^j = \widehat{F}_t^j(\alpha_{t-1}^j) - \widehat{F}_t^i(\alpha_{t-1}^i).$$

That is, net surpluses in the equilibrium do not depend on the access markup, so neither do the market shares. We thus need one more instrument or a direct regulation of fixed fees so as to achieve $(\alpha_2^{i\,*}, \alpha_2^{j\,*})$.

6. CONCLUSION

This article has studied the implications of adopting the receiver pays regime when networks compete in a dynamic framework. We allowed callers and call receivers to derive utility from making and receiving calls, and networks to price calls and charge customers for receiving calls. Assuming the existence of a random noise in the caller's and receiver's utility, we first showed that the off-net-cost pricing principle is a candidate equilibrium.

Second, we showed that under linear demands q and $\widetilde{q}$, this candidate equilibrium is indeed the unique equilibrium provided that the degree of substitutability between networks is low enough and the random noise has a wide enough support. Other insights were derived. In the region where the equilibrium exists, an increase in the access charge raises the call price and decreases the reception charge, but does not affect the networks' full-period profits. Instead, the access charge level clearly affects the consumer welfare; indeed its optimal level from the social welfare viewpoint depends on the characteristics of each market. In the particular case where the linear demand functions q and $\widetilde{q}$ are the same, starting from zero access markup, a small decrease in the access charge decreases the call price and raises the reception charge. As a result, the receivers' utility increases more than the loss in the callers' utility, and the average length of calls decreases, which in turn decreases costs. Consequently, we find optimal to decrease access charges so that either a interior solution is reached, or 'bill and keep' might be socially optimal.

Third and finally, in our previous work (López, 2005) we showed that networks are able to soften present competition by departing away future reciprocal access charges from marginal costs. Under the receiver pays regime we showed however that in a multi-period setting the off-net-cost pricing equilibrium neutralizes the potential anticompetitive role that reciprocal access charges could play.

Our article is a further step in the research agenda; it has characterized the equilibrium that arises in dynamic network competition under the receiver pays regime, and has studied how networks operators' pricing strategies might react to the adoption of such regime. We expect further research extending our analysis. Three key directions are noteworthy. Firstly, as already pointed out by Jeon et al. (2004), the "noncooperative volume setting" assumption should be extended to allow more cooperative behaviors, as for instance the maximization of joint surplus over the call length. Secondly, asymmetric calling patterns should be analyzed. It is not difficult to find cases in which the calling pattern is unbalanced, which might affect the incentives of the networks in the industry. Thirdly, it would be interesting to check whether the off-net-cost pricing principle still applies to the case of multiple networks competing for market share.

7. APPENDIX

LEMMA 1. *Under A.1, A.2, A.3 and A.4:*

$$\frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i} = -\frac{1}{2b} - \frac{1}{\Delta}\left[\frac{d - r_t^j}{e} + \frac{p_t^i - a}{b}\right]$$

$$\frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i} = -\frac{1}{2e} - \frac{1}{\Delta}\left[\frac{a - p_t^j}{b} + \frac{r_t^i - d}{e}\right]$$

$$\frac{\partial^2 U(p_t^j, r_t^i)}{(\partial r_t^i)^2} = -\left(\frac{b}{2e^2}\right) - \left(\frac{1}{e\Delta}\right)\left[a - \frac{b}{e}(d - r_t^i)\right]$$

$$\frac{\partial^2 \widetilde{U}(p_t^i, r_t^j)}{(\partial p_t^i)^2} = -\left(\frac{e}{2b^2}\right) - \left(\frac{1}{b\Delta}\right)\left[d - \frac{e}{b}(a - p_t^i)\right]$$

$$\frac{\partial^2 U(p_t^j, r_t^i)}{\partial p_t^j \partial r_t^i} = \frac{p_t^j}{b\Delta}$$

$$\frac{\partial^2 \widetilde{U}(p_t^i, r_t^i)}{\partial r_t^i \partial p_t^i} = \frac{r_t^i}{\Delta e}$$

Proof. Let us construct $Q(p_t^i, r_t^j)$ by means of several illustrative steps. First of all, notice that for a given pair of prices (p_t^i, r_t^j) and a given pair of realized values $(\varepsilon, \widetilde{\varepsilon})$, the length of a call from a network i consumer to a network j consumer is given by $Q(p_t^i, r_t^j, \varepsilon, \widetilde{\varepsilon}) = \min[q(p_t^i - \varepsilon), \widetilde{q}(r_t^j - \widetilde{\varepsilon})]$, where $q = \mu'^{-1}(p_t^i - \varepsilon)$ and $\widetilde{q} = \widetilde{\mu}'^{-1}(r_t^j - \widetilde{\varepsilon})$, that is,

$$q = \frac{a - (p_t^i - \varepsilon)}{b}, \qquad \widetilde{q} = \frac{d - (r_t^j - \widetilde{\varepsilon})}{e}.$$

Step 1. Assume for the moment $\widetilde{\varepsilon}$ is exogenous and takes value $\widetilde{\varepsilon}'$.

Step 2. Note that $q(p^i - \varepsilon)$ is strictly increasing in ε, which means that it will exist an ε^* such that $q(p_t^i - \varepsilon^*) = \widetilde{q}(r_t^j - \widetilde{\varepsilon}')$, namely

$$\varepsilon^* = p_t^i - q^{-1}(\widetilde{q}(r_t^j - \widetilde{\varepsilon}')).$$

Moreover, if $\varepsilon^* \notin [\underline{\varepsilon}, \overline{\varepsilon}]$ then $f(\varepsilon^*) = 0$.

Step 3. For any $\varepsilon \leq \varepsilon^*$, the caller will be sovereign, whereas the receiver will be sovereign provided that $\varepsilon > \varepsilon^*$. Therefore, we can write the demand as follows:

$$\begin{aligned} d(p_t^i, r_t^j, \widetilde{\varepsilon}') &= \int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} q(p_t^i - \varepsilon) f(\varepsilon) d\varepsilon + \int_{\varepsilon^*(\cdot)}^{\overline{\varepsilon}} \widetilde{q}(r_t^j - \widetilde{\varepsilon}') f(\varepsilon) d\varepsilon \\ &= \int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} q(p_t^i - \varepsilon) f(\varepsilon) d\varepsilon + \widetilde{q}(r_t^j - \widetilde{\varepsilon}')[F(\overline{\varepsilon}) - F(\varepsilon^*(\cdot))], \end{aligned}$$

which can be rewritten for any value of $\widetilde{\varepsilon} : d = d(p_t^i, r_t^j, \widetilde{\varepsilon})$.

Step 4. Therefore, under A.3., for a given pair of prices (p_t^i, r_t^j), the volume of calls from a network i consumer to a network j consumer at period t is given by:

$$\begin{aligned} Q(p_t^i, r_t^j) &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} d(p_t^i, r_t^j, \widetilde{\varepsilon}) \widetilde{f}(\widetilde{\varepsilon}) d\widetilde{\varepsilon} \\ &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} (\int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} q(p_t^i - \varepsilon) f(\varepsilon) d\varepsilon \\ &\quad + \widetilde{q}(r_t^j - \widetilde{\varepsilon})[F(\overline{\varepsilon}) - F(\varepsilon^*(\cdot))]) \widetilde{f}(\widetilde{\varepsilon}) d\widetilde{\varepsilon} \end{aligned}$$

Now, for a given r_t^j we can differentiate $Q(p_t^i, r_t^j)$ with respect to p_t^i :

$$\begin{aligned} \frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i} &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \left(\int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} \frac{\partial q(p_t^i - \varepsilon)}{\partial p_t^i} f(\varepsilon) d\varepsilon \right) \widetilde{f}(\widetilde{\varepsilon}) d\widetilde{\varepsilon} \\ &= -\left(\frac{1}{b\Delta^2} \right) \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} (\varepsilon^*(\cdot) - \underline{\varepsilon}) d\widetilde{\varepsilon}, \end{aligned}$$

where $\varepsilon^* = (b/e)(d - r_t^j + \widetilde{\varepsilon}) + p_t^i - a$. Then,

$$\frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i} = -\frac{1}{2b} - \frac{1}{\Delta} \left[\frac{d - r_t^j}{e} + \frac{p_t^i - a}{b} \right]$$

In a similar way, we can assume p_t^i as given and differentiate $Q(p_t^i, r_t^i)$ with respect to r_t^i. To that end, we can rewrite the demand as follows:

$$\begin{aligned} Q(p_t^j, r_t^i) &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} (\int_{\underline{\varepsilon}}^{\widetilde{\varepsilon}^*(\cdot)} \widetilde{q}(r_t^i - \widetilde{\varepsilon}) \widetilde{f}(\widetilde{\varepsilon}) d\widetilde{\varepsilon} \\ &\quad + q(p_t^i - \varepsilon)[\widetilde{F}(\overline{\varepsilon}) - \widetilde{F}(\widetilde{\varepsilon}^*(\cdot))]) f(\varepsilon) d\varepsilon, \end{aligned}$$

, where $\widetilde{\varepsilon}^* = (e/b)(a - p_t^j + \varepsilon) + r_t^i - d$. Then,

$$\begin{aligned} \frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i} &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \left(\int_{\underline{\varepsilon}}^{\widetilde{\varepsilon}^*(\cdot)} \frac{\partial \widetilde{q}(r_t^i - \widetilde{\varepsilon})}{\partial r_t^i} \widetilde{f}(\widetilde{\varepsilon}) d\widetilde{\varepsilon} \right) f(\varepsilon) d\varepsilon \\ &= -\left(\frac{1}{e\Delta^2} \right) \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} (\widetilde{\varepsilon}^*(\cdot) - \underline{\varepsilon}) d\varepsilon \end{aligned}$$

Thus,

$$\frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i} = -\frac{1}{2e} - \frac{1}{\Delta}\left[\frac{a - p_t^j}{b} + \frac{r_t^i - d}{e}\right]$$

Assume a given p_t^i and rewrite $U(p_t^i, r_t^i)$ as follows:

$$\begin{aligned} U(p_t^i, r_t^i) &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} (\int_{\underline{\varepsilon}}^{\widetilde{\varepsilon}^*(\cdot)} u(\widetilde{q}(r_t^i - \widetilde{\varepsilon}))\widetilde{f}(\widetilde{\varepsilon})d\widetilde{\varepsilon} \\ &\quad + u(q(p_t^i - \varepsilon))[\widetilde{F}(\overline{\varepsilon}) - \widetilde{F}(\widetilde{\varepsilon}^*(\cdot))])f(\varepsilon)d\varepsilon, \end{aligned}$$

Then,

$$\frac{\partial U(p_t^i, r_t^i)}{\partial r_t^i} = \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} (\int_{\underline{\varepsilon}}^{\widetilde{\varepsilon}^*(\cdot)} u'(\widetilde{q}(r_t^i - \widetilde{\varepsilon}))\frac{\partial \widetilde{q}(r_t^i - \widetilde{\varepsilon})}{\partial r_t^i}\widetilde{f}(\widetilde{\varepsilon})d\widetilde{\varepsilon})f(\varepsilon)d\varepsilon$$

Note that $u'(\widetilde{q}(r_t^i - \widetilde{\varepsilon})) = a - b\widetilde{q}(r_t^i - \widetilde{\varepsilon}) + \varepsilon$. Thus,

$$\begin{aligned} \frac{\partial U(p_t^i, r_t^i)}{\partial r_t^i} &= -\left(\frac{1}{e\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} (\int_{\underline{\varepsilon}}^{\widetilde{\varepsilon}^*(\cdot)} a - b\left[\frac{d - (r_t^i - \widetilde{\varepsilon})}{e}\right] + \varepsilon d\widetilde{\varepsilon})d\varepsilon \\ &= -\left(\frac{1}{e\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \left[a - \frac{b}{e}(d - r_t^i) + \varepsilon - \frac{b}{2e}(\widetilde{\varepsilon}^*(\cdot) + \underline{\varepsilon})\right](\widetilde{\varepsilon}^*(\cdot) - \underline{\varepsilon})d\varepsilon \end{aligned}$$

It follows that:

$$\begin{aligned} \frac{\partial^2 U(p_t^i, r_t^i)}{(\partial r_t^i)^2} &= -\left(\frac{1}{e\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} a - \frac{b}{e}(d - r_t^i + \underline{\varepsilon}) + \varepsilon d\varepsilon \\ &= -\left(\frac{b}{2e^2}\right) - \left(\frac{1}{e\Delta}\right)\left[a - \frac{b}{e}(d - r_t^i)\right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 U(p_t^i, r_t^i)}{\partial p_t^i \partial r_t^i} &= -\left(\frac{1}{e\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \widetilde{\varepsilon}^*(\cdot) - \frac{e}{b}(a + \varepsilon) + (d - r_t^i)d\varepsilon \\ &= \frac{p_t^i}{b\Delta} \end{aligned}$$

For a given p_t^i we can write $\widetilde{U}(p_t^i, r_t^i)$ as follows:

$$\begin{aligned} \widetilde{U}(p_t^i, r_t^i) &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \left(\int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} \widetilde{u}(q(p_t^i - \varepsilon))f(\varepsilon)d\varepsilon\right) \\ &\quad + \left(\widetilde{u}(\widetilde{q}(r_t^i - \widetilde{\varepsilon}))\left[F(\overline{\varepsilon}) - F(\varepsilon^*(\cdot))\right]\right)\widetilde{f}(\widetilde{\varepsilon})d\widetilde{\varepsilon} \end{aligned}$$

Then,

$$\begin{aligned}\frac{\partial\widetilde{U}(p_t^i,r_t^i)}{\partial p_t^i} &= -\left(\frac{1}{b\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}}(\int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} d-e\left[\frac{a-(p_t^i-\varepsilon)}{b}\right]+\widetilde{\varepsilon}d\varepsilon)d\widetilde{\varepsilon} \\ &= -\left(\frac{1}{b\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}}\left[d-\frac{e}{b}(a-p_t^i)+\widetilde{\varepsilon}-\frac{e}{2b}(\varepsilon^*(\cdot)+\underline{\varepsilon})\right](\varepsilon^*(\cdot)-\underline{\varepsilon})d\widetilde{\varepsilon}\end{aligned}$$

It follows that:

$$\begin{aligned}\frac{\partial^2\widetilde{U}(p_t^i,r_t^i)}{(\partial p_t^i)^2} &= -\left(\frac{1}{b\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} d-\frac{e}{b}(a-p_t^i+\underline{\varepsilon})+\widetilde{\varepsilon}d\widetilde{\varepsilon} \\ &= -\left(\frac{e}{2b^2}\right)-\left(\frac{1}{b\Delta}\right)\left[d-\frac{e}{b}(a-p_t^i)\right]\end{aligned}$$

$$\begin{aligned}\frac{\partial^2\widetilde{U}(p_t^i,r_t^i)}{\partial r_t^i\partial p_t^i} &= -\left(\frac{1}{b\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}}\varepsilon^*(\cdot)-\frac{b}{e}(d+\widetilde{\varepsilon})+(a-p_t^i)d\widetilde{\varepsilon} \\ &= \frac{r_t^i}{e\Delta}\end{aligned}$$

∎

Lemma 2. *Under A.1, A.2, A.3, A.4., and for a large enough* Δ :

$$\partial^2\overline{\pi}_t^i/(\partial p_t^i)^2 \simeq -\alpha_t^i/2b$$

$$\partial^2\overline{\pi}_t^i/(\partial r_t^i)^2 \simeq -\alpha_t^i/2e$$

$$\frac{\partial^2\overline{\pi}_t^i}{(\partial\alpha_t^i)^2} = 2\lambda_{\alpha,t}(p_t^i,r_t^i,p_t^j,r_t^j)-2/\sigma,$$

$$\partial^2\overline{\pi}_t^i/\partial r_t^i\partial p_t^i \simeq 0$$

$$\partial^2\pi_t^i/\partial p_t^i\partial r_t^i \simeq 0$$

$$\frac{\partial\overline{\pi}_t^i}{\partial p_t^i\partial\alpha_t^i} \simeq [2(p_t^i-c)\alpha_t^i+((p_t^i-c-m_t)(\alpha_t^j-\alpha_t^i)+2\alpha_t^i r_t^j)]\left(-\frac{1}{2b}\right) \equiv \lambda_{p^i,t}(p_t^i,r_t^j,\alpha_t^i)$$

$$\frac{\partial\overline{\pi}_t^i}{\partial r_t^i\partial\alpha_t^i} \simeq [2\alpha_t^i(r_t^i-c)+((\alpha_t^j-\alpha_t^i)(r_t^i+m_t)+2p_t^j\alpha_t^i)]\left(-\frac{1}{2e}\right) \equiv \lambda_{r^i,t}(p_t^j,r_t^i,\alpha_t^i),$$

where $\lambda_{\alpha,t} = -cQ_t^{ii}+(c+m_t)Q_t^{ij}-m_tQ_t^{ji}+p_t^j(-Q_t^{jj}+Q_t^{ji})+r_t^j(-Q_t^{jj}+Q_t^{ij})+(U_t^{ii}+\widetilde{U}_t^{ii}-U_t^{ji}-\widetilde{U}_t^{ij})-(U_t^{ij}+\widetilde{U}_t^{ji}-U_t^{jj}-\widetilde{U}_t^{jj})$ *is a bounded function.*

Proof. Using the market share definition, we can rewrite the second-period profits in terms of p^i_2, r^i_2 and α^i_2 :

$$\begin{aligned}\overline{\pi}^i_t(p^i_t, r^i_t, \alpha^i_t) \quad = \quad & \alpha^i_t\{-c\alpha^i_t Q^{ii}_t - (c+m_2)\alpha^j_t Q^{ij}_t + \alpha^j_t m_t Q^{ji}_t + p^j_t(\alpha^j_t Q^{jj}_t \\ & +\alpha^i_t Q^{ji}_t) + r^j_t(\alpha^j_t Q^{jj}_t + \alpha^i_t Q^{ij}_t) + \alpha^i_t(U^{ii}_t + \widetilde{U}^{ii}_t - U^{ji}_t \\ & -\widetilde{U}^{ij}_t) + \alpha^j_t(U^{ij}_t + \widetilde{U}^{ji}_t - U^{jj}_t - \widetilde{U}^{jj}_t) + F^j_t - \frac{1}{\sigma}\left(\alpha^i_t - \frac{1}{2}\right) \\ & +(2\alpha^i_{t-1} - 1)s - f\}\end{aligned} \tag{33}$$

From expression (9) we have that:

$$\begin{aligned}\frac{\partial \overline{\pi}^i_t}{\partial p^i_t} \quad = \quad & \alpha^i_t\{(-\alpha^j_t(c+m_t) + \alpha^j_t p^i_t + \alpha^i_t r^j_t)\frac{\partial Q(p^i_t, r^j_t)}{\partial p^i_t} \\ & +\alpha^i_t\left(-c + p^i_t\right)\frac{\partial Q(p^i_t, r^i_t)}{\partial p^i_t} + \alpha^i_t\left(\frac{\partial \widetilde{U}(p^i_t, r^i_t)}{\partial p^i_t} - \frac{\partial \widetilde{U}(p^i_t, r^j_t)}{\partial p^i_t}\right)\}\end{aligned} \tag{34}$$

$$\begin{aligned}\frac{\partial^2 \overline{\pi}^i_t}{(\partial p^i_t)^2} \quad = \quad & \alpha^i_t\{(-\alpha^j_t(c+m_2) + \alpha^j_t p^i_t + \alpha^i_t r^j_t)\frac{\partial Q^2(p^i_t, r^j_t)}{\left(\partial p^i_t\right)^2} + \alpha^j_t\frac{\partial Q(p^i_t, r^j_t)}{\partial p^i_t} \\ & +\alpha^i_t\frac{\partial Q(p^i_t, r^i_t)}{\partial p^i_t} + \alpha^i_t(-c + p^i_t)\frac{\partial Q^2(p^i_t, r^i_t)}{(\partial p^i_t)^2}\}\end{aligned}$$

$$\begin{aligned}\frac{\partial \overline{\pi}^i_t}{\partial r^i_t} \quad = \quad & \alpha^i_t\{\alpha^i_t\left(-c + r^i_t\right)\frac{\partial Q(p^i_t, r^i_t)}{\partial r^i_t} + (\alpha^j_t(m_t + r^i_t) + \alpha^i_t p^j_t)\frac{\partial Q(p^j_t, r^i_t)}{\partial r^i_t} \\ & +\alpha^i_t\left(\frac{\partial U(p^i_t, r^i_t)}{\partial r^i_t} - \frac{\partial U(p^j_t, r^i_t)}{\partial r^i_t}\right)\}\end{aligned} \tag{35}$$

$$\begin{aligned}\frac{\partial^2 \overline{\pi}^i_t}{(\partial r^i_t)^2} \quad = \quad & \alpha^i_t\{\alpha^i_t\frac{\partial Q(p^i_t, r^i_t)}{\partial r^i_t} + \alpha^i_t\left(-c + r^i_t\right)\frac{\partial Q^2(p^i_t, r^i_t)}{(\partial r^i_t)^2} + \alpha^j_t\frac{\partial Q(p^j_t, r^i_t)}{\partial r^i_t} \\ & +(\alpha^j_t(m_t + r^i_t) + \alpha^i_t p^j_t)\frac{\partial^2 Q(p^j_t, r^i_t)}{(\partial r^i_t)^2}\}\end{aligned}$$

Then, using Lemma 1:

$$\frac{\partial^2 \overline{\pi}^i_t}{(\partial p^i_t)^2} = \alpha^i_t\{\alpha^j_t\frac{\partial Q(p^i_t, r^j_t)}{\partial p^i_t} + \alpha^i_t\frac{\partial Q(p^i_t, r^i_t)}{\partial p^i_t} - \left(\frac{1}{\Delta b}\right)(p^i_t - c - \alpha^j_t m_t + \alpha^i_t r^j_t)\}$$

$$\frac{\partial^2 \overline{\pi}^i_t}{(\partial r^i_t)^2} = \alpha^i_t\{\alpha^i_t\frac{\partial Q(p^i_t, r^i_t)}{\partial r^i_t} + \alpha^j_t\frac{\partial Q(p^j_t, r^i_t)}{\partial r^i_t} - \left(\frac{1}{\Delta e}\right)(r^i_t - \alpha^i_t c + \alpha^j_t m_t + \alpha^i_t p^j_t)\}$$

Moreover,

$$\begin{aligned}\frac{\partial^2 \overline{\pi}_t^i}{\partial r_t^i \partial p_t^i} &= \alpha_t^i \{ \alpha_t^i \left(-c + p_t^i \right) \frac{\partial Q^2(p_t^i, r_t^i)}{\partial r_t^i \partial p_t^i} + \alpha_t^i \frac{\partial^2 \widetilde{U}(p_t^i, r_t^i)}{\partial r_t^i \partial p_t^i} \} \\ &= \frac{(\alpha_t^i)^2}{\Delta e} (p_t^i + r_t^i - c)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \overline{\pi}_t^i}{\partial r_t^j \partial p_t^i} &= \alpha_t^i \{ \alpha_t^i \frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i} + (-\alpha_t^j (c + m_t) + \alpha_t^j p_t^i + \alpha_t^i r_t^j) \frac{\partial Q(p_t^i, r_t^j)}{\partial r_t^j \partial p_t^i} \\ &\quad - \alpha_t^i \frac{\partial \widetilde{U}(p_t^i, r_t^j)}{\partial r_t^j \partial p_t^i} \} \\ &= \alpha_t^i \{ \alpha_t^i \frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i} + \frac{(-\alpha_t^j (c + m_t) + \alpha_t^j p_t^i)}{\Delta e} \}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \overline{\pi}_t^i}{\partial p_t^i \partial r_t^i} &= \alpha_t^i \{ \alpha_t^i \left(-c + r_t^i \right) \frac{\partial^2 Q(p_t^i, r_t^i)}{\partial p_t^i \partial r_t^i} + \alpha_t^i \frac{\partial^2 U(p_t^i, r_t^i)}{\partial p_t^i \partial r_t^i} \} \\ &= \frac{(\alpha_t^i)^2}{\Delta b} (p_t^i + r_t^i - c)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \overline{\pi}_t^i}{\partial p_t^j \partial r_t^i} &= \alpha_t^i \{ \alpha_t^i \frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i} + (\alpha_t^j (m_t + r_t^i) + \alpha_t^i p_t^j) \frac{\partial Q(p_t^j, r_t^i)}{\partial p_t^j \partial r_t^i} \\ &\quad - \alpha_t^i \frac{\partial U(p_t^j, r_t^i)}{\partial p_t^j \partial r_t^i} \} \\ &= \alpha_t^i \{ \alpha_t^i \frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i} + \frac{\alpha_t^j (m_t + r_t^i)}{\Delta b} \}\end{aligned}$$

On the other hand,

$$\frac{\partial \overline{\pi}_t^i}{\partial \alpha_t^i} = \frac{\overline{\pi}_t^i}{\alpha_t^i} + \alpha_t^i \{ \lambda_{\alpha,t}(p_t^i, r_t^i, p_t^j, r_t^j) - \frac{1}{\sigma} \},$$

where

$$\begin{aligned}\lambda_{\alpha,t} &= -cQ_t^{ii} + (c + m_t) Q_t^{ij} - m_t Q_t^{ji} + p_t^j (-Q_t^{jj} + Q_t^{ji}) + r_t^j (-Q_t^{jj} + Q_t^{ij}) \\ &\quad + (U_t^{ii} + \widetilde{U}_t^{ii} - U_t^{ji} - \widetilde{U}_t^{ij}) - (U_t^{ij} + \widetilde{U}_t^{ji} - U_t^{jj} - \widetilde{U}_t^{jj})\end{aligned}$$

And,

$$\frac{\partial^2 \overline{\pi}_t^i}{(\partial \alpha_t^i)^2} = 2\lambda_{\alpha,t}(p_t^i, r_t^i, p_t^j, r_t^j, \alpha_t^i) - 2/\sigma,$$

Moreover,

$$
\begin{aligned}
\frac{\partial \pi_t^i}{\partial p_t^i \partial \alpha_t^i} &= 2(p_t^i - c)\alpha_t^i \frac{\partial Q(p_t^i, r_t^i)}{\partial p_t^i} + ((p_t^i - c - m_t)(\alpha_t^j - \alpha_t^i) + 2\alpha_t^i r_t^j)\frac{\partial Q(p_t^i, r_t^j)}{\partial p_t^i} \\
&\quad + 2\alpha_t^i \left(\frac{\partial \widetilde{U}(p_t^i, r_t^i)}{\partial p_t^i} - \frac{\partial \widetilde{U}(p_t^i, r_t^j)}{\partial p_t^i} \right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \pi_t^i}{\partial r_t^i \partial \alpha_t^i} &= 2\alpha_t^i (r_t^i - c)\frac{\partial Q(p_t^i, r_t^i)}{\partial r_t^i} + ((\alpha_t^j - \alpha_t^i)(r_t^i + m_t) + 2p_t^j \alpha_t^i)\frac{\partial Q(p_t^j, r_t^i)}{\partial r_t^i} \\
&\quad + 2\alpha_t^i \left(\frac{\partial U(p_t^i, r_t^i)}{\partial r_t^i} - \frac{\partial U(p_t^j, r_t^i)}{\partial r_t^i} \right),
\end{aligned}
$$

where:

$$
\left(\frac{\partial \widetilde{U}(p_t^i, r_t^j)}{\partial p_t^i} - \frac{\partial \widetilde{U}(p_t^i, r_t^i)}{\partial p_t^i} \right) = \frac{1}{2e\Delta}[(r_t^j)^2 - (r_t^i)^2] \tag{36}
$$

$$
\left(\frac{\partial U(p_t^i, r_t^j)}{\partial r_t^i} - \frac{\partial U(p_t^i, r_t^i)}{\partial r_t^i} \right) = \frac{1}{2b\Delta}[(p_t^j)^2 - (p_t^i)^2] \tag{37}
$$

Thus, for a large enough Δ and using Lemma 1 it follows the stated results. ∎

Proof. Proposition 1.

We first focus on network $i's$ best response to given prices of the rival: p_t^j, r_t^j and F_t^j. Note first that, for given p_t^i and r_t^i, $\Delta\phi_t \equiv \phi_t^i - \phi_t^j : [0,1] \to R$ is an affine function of the market share at period $t : \Delta\phi_t(\alpha_t^i) = \kappa\alpha_t^i + y$, where κ and y are real numbers. Note further that relevant fixed fees are bounded: given the pair (p_2^i, r_2^i) there exists an upper bound $\overline{F}$ such that $\overline{\alpha}_t^i(\overline{F}) = 0$ and thus $F_t^i > \overline{F}$ cannot be a best response; similarly there exists a lower bound $\underline{F}$ such that $\overline{\alpha}_t^i(\underline{F}) = 1$ and hence for any $F_t^i < \underline{F}$ we still have that $\overline{\alpha}_t^i = 1$ but lower network $i's$ profits, thus $F_t^i < \underline{F}$ cannot be a best response. Therefore, the F_t^i that can be a best response to the triple (p_t^j, r_t^j, F_t^j), for given (p_t^i, r_t^i), belongs to the interval $[\underline{F}, \overline{F}]$ (see figure below.) Accordingly, for given (p_t^i, r_t^i), $\overline{\alpha}_t^i : [\underline{F}, \overline{F}] \longrightarrow [0,1]$ is one-to-one or injective in F_t^i iff $\kappa \neq 1/\sigma$:

$$
\overline{\alpha}_t^i = \frac{1}{1 - \sigma\kappa}\left(\frac{1}{2} + (2\alpha_1^i - 1)\sigma s + \sigma(F_t^j - F_t^i + y) \right),
$$

that is, $\overline{\alpha}_t^i$ is well-defined and monotonically increasing or monotonically decreasing in F_t^i iff $\kappa \neq 1/\sigma$. The *degenerate case* where $\kappa = 1/\sigma$ could exist for given usage prices and σ, however as long as q and $\widetilde{q}$ are bounded, which is assumed, there will always be a small enough σ such that this *degenerate case* cannot occur. Consequently, for a small enough σ, $\overline{\alpha}_t^i$ is well-defined and injective, and thus invertible on its domain; its inverse $\overline{\alpha}_t^{i-1} = \overline{F}_t^i$ is then uniquely defined.

Each network i maximizes π_2^i with respect to p_2^i, r_2^i and F_2^i, for given p_2^j, r_2^j

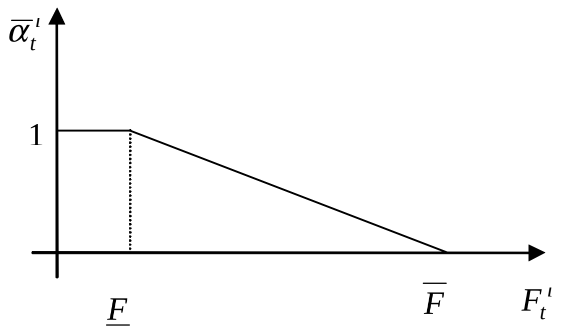


FIG. 1 .

and F_2^j, subject to $\alpha_2^i = (1/2) + (2\alpha_1^i - 1)\sigma s + \sigma(\phi_2^i - F_2^i - \phi_2^j + F_2^j)$, where π_2^i is given by (8) and ϕ_2^i is given by (7). Using the market share definition, we can rewrite the second-period profits in terms of p_2^i, r_2^i and α_2^i : $\overline{\pi}_2^i(p_2^i, r_2^i, \alpha_2^i)$, which is given in (33). Moreover, since for any (p_t^i, r_t^i) $\overline{\alpha}_2^i$ is one-to-one, for given p_2^j, r_2^j and F_2^j, maximizing π_2^i with respect to p_2^i, r_2^i and F_2^i is equivalent to maximizing $\overline{\pi}_2^i$ with respect to p_2^i, r_2^i and α_2^i; that is, there exists a one-to-one correspondence between both best response correspondences (p_2^i, r_2^i, F_2^i) and $(p_2^i, r_2^i, \alpha_2^i)$, to a given triple (p_2^j, r_2^j, F_2^j). Now, we check whether such a best response correspondence $(p_2^i, r_2^i, \alpha_2^i)$ is well-defined, in other words whether the Hessian of the network $i's$ profit function $\overline{\pi}_2^i$ is negative definite:

$$H^i = \begin{bmatrix} \frac{\partial^2 \overline{\pi}_2^i}{(\partial p_2^i)^2} & \frac{\partial^2 \overline{\pi}_2^i}{\partial r_2^i \partial p_2^i} & \frac{\partial^2 \overline{\pi}_2^i}{\partial \alpha_2^i \partial p_2^i} \\ \frac{\partial^2 \overline{\pi}_2^i}{\partial p_2^i \partial r_2^i} & \frac{\partial^2 \overline{\pi}_2^i}{(\partial r_2^i)^2} & \frac{\partial^2 \overline{\pi}_2^i}{\partial \alpha_2^i \partial r_2^i} \\ \frac{\partial^2 \overline{\pi}_2^i}{\partial p_2^i \partial \alpha_2^i} & \frac{\partial^2 \overline{\pi}_2^i}{\partial r_2^i \partial \alpha_2^i} & \frac{\partial^2 \overline{\pi}_2^i}{(\partial \alpha_2^i)^2} \end{bmatrix}$$

Let H_k^i denote the $k-th$ principal minor of the Hessian matrix H^i. Using Lemma 2, for a large enough Δ, we have that $\left|H_1^i\right| \simeq -\alpha_2^i/2b$ and $\left|H_2^i\right| \simeq (\alpha_2^i)^2/4be$, moreover

$$\left|H_3^i\right| \simeq \lambda_{\alpha,2}\frac{(\alpha_2^i)^2}{2be} - \frac{1}{\sigma}\frac{(\alpha_2^i)^2}{2be} + \frac{\alpha_2^i}{2e}(\lambda_{p^i,2})^2 + \frac{\alpha_2^i}{2b}\left(\lambda_{r^i,2}\right)^2$$

Then, for any $\alpha_2^i \in (0,1]$ and a large enough Δ : $\left|H_1^i\right| < 0$ and $\left|H_2^i\right| > 0$, moreover since demands are bounded by assumption, λ_α, λ_{p^i} and λ_{r^i} are also bounded functions, and hence there exists a small enough σ such that $\left|H_3^i\right| < 0$. Let us now show that no cornered-market equilibrium exists. Suppose that network i corners the market by setting $(p_2^{i\,*}, r_2^{i\,*}, F_2^{i\,*})$. Then, $\pi_2^j = 0$ and using (8): $\pi_2^{i\,*} = [(p_2^{i\,*} - c + r_2^{i\,*})Q(p_2^{i\,*}, r_2^{i\,*}) + F_2^{i\,*} - f]$, with $\pi_2^{i\,*} \geq 0$, otherwise cornering the market would not be an optimal strategy. But network j could charge $p_2^{j\,*} = p_2^{i\,*}$, $r_2^{j\,*} = r_2^{i\,*}$

and $F_2^{j\,*} = F_2^{i\,*} + \epsilon$, where $\epsilon > 0$. It follows that $\alpha_2^{j\,*} = (1/2) + (2\alpha_1^j - 1)\sigma s - \sigma\epsilon$, and if $\alpha_1^j = 0$ we have that $\alpha_2^{j\,*} = (1/2)(1 - s/\tau) - \sigma\epsilon$, then since $s < \tau$ it exists a small enough and positive ϵ such that $\alpha_2^{j\,*} > 0$ for any $\alpha_1^j \in [0,1]$. It follows that for such a small enough ϵ and using (8), the network $j's$ profits would then be

$$\begin{aligned}\pi_2^j &= \alpha_2^{j\,*}[(p_2^{j\,*} - c + r_2^{j\,*})Q(p_2^{j\,*}, r_2^{j\,*}) + F_2^{j\,*} - f] \\ &= \alpha_2^{j\,*}(\pi_2^{i\,*} + \epsilon) \geq \alpha_2^{j\,*}\epsilon > 0,\end{aligned}$$

a contradiction. In summary, for a large enough Δ there exists a small enough σ such that profit functions are strictly concave whatever the rival prices are, which means that the network $i's$ best response is a continuos function. Therefore, any candidate equilibrium must satisfy the first-order conditions, and any solution that satisfy the first-order conditions is an equilibrium. The set of first-order conditions can be written as follows:

$$\begin{array}{ll}\frac{\partial\pi_2^i}{\partial p_2^i}(p_2^i, r_2^i, \alpha_2^i, r_2^j) = 0 \quad (C.1), & \frac{\partial\pi_2^j}{\partial p_2^j}(p_2^j, r_2^j, \alpha_2^j, r_2^i) = 0 \quad (C.3), \\ \frac{\partial\pi_2^i}{\partial r_2^i}(p_2^i, r_2^i, \alpha_2^i, p_2^j) = 0 \quad (C.2), & \frac{\partial\pi_2^j}{\partial r_2^j}(p_2^j, r_2^j, \alpha_2^j, p_2^i) = 0 \quad (C.4),\end{array}$$

$$\frac{\partial\pi_2^i}{\partial\alpha_2^i}(p_2^i, r_2^i, \alpha_2^i, p_2^j, r_2^j, F_2^j) = 0, \quad \frac{\partial\pi_2^j}{\partial\alpha_2^j}(p_2^j, r_2^j, \alpha_2^j, p_2^i, r_2^i, F_2^i) = 0.$$

Together with the market share definitions, we have 8 equations and 8 unknown variables. Consider the first four first-order conditions derived from maximizing profits with respect to usage prices *(C.1-C.4),* notice that fixed fees do not enter these conditions (as can been seen from (34) and (35).) Using lemma 1 and expressions (36) and (37) we can write:

$$\begin{aligned}\frac{\partial\pi_2^i}{\partial p_2^i} &= -\alpha_2^i\left(\frac{1}{2b}\xi_p^i(p_2^i, r_2^j) + \frac{1}{\Delta}\omega_p^i(p_2^i, r_2^i, r_2^j)\right), \\ \frac{\partial\pi_2^i}{\partial r_2^i} &= -\alpha_2^i\left(\frac{1}{2e}\xi_r^i(r_2^i, p_2^j) + \frac{1}{\Delta}\omega_r^i(r_2^i, p_2^i, p_2^j)\right), \\ \frac{\partial\pi_2^j}{\partial p_2^j} &= -\alpha_2^j\left(\frac{1}{2b}\xi_p^j(p_2^j, r_2^i) + \frac{1}{\Delta}\omega_p^j(p_2^j, r_2^j, r_2^i)\right), \\ \frac{\partial\pi_2^j}{\partial r_2^j} &= -\alpha_2^j\left(\frac{1}{2e}\xi_r^j(r_2^j, p_2^i) + \frac{1}{\Delta}\omega_r^j(r_2^j, p_2^j, p_2^i)\right),\end{aligned}$$

where $\xi_p^i(p_2^i, r_2^j) = -c - \alpha_2^j m_2 + p_2^i + \alpha_2^j r_2^j$, $\xi_r^i(r_2^i, p_2^j) = -\alpha_2^i c + \alpha_2^j m_2 + r_2^i + \alpha_2^i p_2^j$, and ω_p^i and ω_r^i are nonlinear functions that do not depend on Δ. That is, each one of these equations can be written as the sum of a linear function (ξ) and a nonlinear function (ω). Moreover, this system of equations have at least one solution, which is given by $p_2^i = p_2^j = c + m_2$, and $r_2^i = r_2^j = -m_2$, and where

$\xi_p^i = \xi_p^j = 0$, $\omega_p^i = \omega_p^j = 0$, $\xi_r^i = \xi_r^j = 0$, $\omega_r^i = \omega_r^j = 0$. Let Ξ denote the set of solutions to the system (C.1-C.4), which we already know is non-empty. Note that by increasing Δ the nonlinear components of this system tend to vanish, indeed the non-linear equations tend to be linear as Δ increases. Therefore, by assuming a large enough Δ we can make vanish all those solutions that might come from nonlinearities and thereby make Ξ tend to be finite and have *at most* one element or to have infinite elements, which is/are the solution/s that would come from the linear system: $\xi_p^i(p_2^i, r_2^j) = 0$, $\xi_r^i(r_2^i, p_2^j) = 0$, $\xi_p^j(p_2^j, r_2^i) = 0$, $\xi_r^j(r_2^j, p_2^i) = 0$. Indeed, we know that for a large enough Δ the set Ξ is non-empty nor infinite but tend to a singleton since $\xi_p^i(c+m_2, -m_2) = \xi_p^j(c+m_2, -m_2) = 0$ and $\xi_r^i(-m_2, c+m_2) = \xi_r^j(-m_2, c+m_2) = 0$. Therefore, for any $\alpha_2^i \in (0,1)$ and a large enough Δ there exists a unique equilibrium in usage prices, where networks price calls at their off-net cost. Let us now return to the original formulation of the profit function that is given in (8) and where the strategic variables are p_2^i, r_2^i and F_2^i. Substituting $p_2^i = p_2^j = c + m_2$ and $r_2^i = r_2^j = -m_2$ into (8) gives us the expression (16). By maximizing this expression with respect to the network $i's$ fixed fee we obtain linear reaction functions: $F_2^i(F_2^j)$, which are given in (17). Moreover, $dF_2^i/dF_2^j = 1/2$, therefore there exists a unique equilibrium in fixed fees that is given in (18). ∎

Proof. Proposition 2.

Following Proposition 1 and using the market share definition, we can rewrite the first-period profits in terms of p_1^i, r_1^i and α_1^i. Moreover, since for a small enough σ, $\overline{\alpha}_1^i$ is one-to-one for any (p_1^i, r_1^i) and given p_2^j, r_2^j and F_2^j, maximizing π_1^i with respect to p_1^i, r_1^i and F_1^i is equivalent to maximizing $\overline{\pi}_1^i$ with respect to p_1^i, r_1^i and α_1^i; that is, there exists a one-to-one correspondence between both best response correspondences (p_1^i, r_1^i, F_1^i) and $(p_1^i, r_1^i, \alpha_1^i)$, to a given triple (p_1^j, r_1^j, F_1^j). Hence, we only need to check whether the Hessian of the network $i's$ full-period profit function: $\overline{\Pi}^i(p_1^i, r_1^i, \alpha_1^i) = \overline{\pi}_1^i(p_1^i, r_1^i, \alpha_1^i) + \delta\widehat{\pi}_2^i(\alpha_1^i)$, with $\widehat{\pi}_2^i(\alpha_1^i)$ given by (19), is negative definite. Let H_k^i denote the $k-th$ principal minor of the Hessian matrix. Using Lemma 2, for a large enough Δ, we have that $\left|H_1^i\right| \simeq -\alpha_1^i/2b$ and $\left|H_2^i\right| \simeq (\alpha_1^i)^2/4be$, moreover

$$\left|H_3^i\right| \simeq \lambda_{\alpha,1}\frac{(\alpha_1^i)^2}{2be} - \frac{1}{\sigma}\frac{(\alpha_1^i)^2}{2be} + \delta\sigma\frac{2s^2\left(\alpha_1^i\right)^2}{9be} + \frac{\alpha_1^i}{2e}(\lambda_{p^i,1})^2 + \frac{\alpha_1^i}{2b}\left(\lambda_{r^i,1}\right)^2$$

Then, for any $\alpha_1^i \in (0,1]$ and a large enough Δ : $\left|H_1^i\right| < 0$ and $\left|H_2^i\right| > 0$, and since $\lambda_{\alpha,1}$, $\lambda_{p^i,1}$ and $\lambda_{r^i,1}$ are bounded functions, there exists a small enough σ such that $\left|H_3^i\right| < 0$. We now show that in the first period no cornered-market equilibrium exists if switching costs are small enough. Suppose that network i corners the market by setting $(p_1^{i\,*}, r_1^{i\,*}, F_1^{i\,*})$, then $\Pi^{i\,*} = \pi_1^{i\,*} + \delta\widehat{\pi}_2^i(1)$, where $\widehat{\pi}_2^i(1) = 1/4\sigma - s/3 - \sigma s^2/9$. Note that in order to corner the market network i

must sacrifice present profits so as to build market share. This implies that $\pi_1^{i\ *}$ is lower than the static equilibrium profits, which is always interior. As switching costs decrease, the link between the present and the future vanishes, that is, $\lim_{s\to 0} \widehat{\pi}_2^i(1) = \widehat{\pi}_2^i(1/2) = 1/4\sigma$. Thus, $\lim_{s\to 0} \Pi^{i\ *} = \pi_1^{i\ *} + \delta/4\sigma < \lim_{s\to 0} \Pi = 1/4\sigma + \delta/4\sigma$. Finally, notice that $\partial\overline{\Pi}^i/\partial p_1^i = \partial\overline{\pi}_1^i/\partial p_1^i$ and $\partial\overline{\Pi}^i/\partial r_1^i = \partial\overline{\pi}_1^i/\partial r_1^i$, therefore we can construct a system of equations similar to the system $C.1 - C.4$ given in the proof of proposition 1 with the unique difference that the time index subscript takes now value 1. Then, following a similar reasoning to that used in the proof of proposition 1, one can show that for a large enough Δ there exists a unique equilibrium in usage prices, which is given by $p_1^i = p_1^j = c + m_1$ and $r_1^i = r_1^j = -m_1$, and hence do not depend on the level of the market shares. Given this, we can return to the original formulation of the full-period profit function that is given in (20) and where the strategic variables are p_1^i, r_1^i and F_1^i. By substituting the equilibrium usage prices into (20) we obtain the expression (21). Finally, maximizing this expression with respect to the network $i's$ fixed fee yields linear reaction functions $F_1^i(F_1^j)$ that are given in (23) and have got a unique intersection point that is given in (24). ∎

Proof. Proposition 3.

Let $s_{t+1}^i(m,n)$ denote the $(m,n).th$ entry of the matrix S_{t+1}^i. The following lemma will be needed: ∎

LEMMA 3. $\lim_{\sigma\to 0} s_t^i(3,3) = 0$ *and* $\lim_{\sigma\to 0} s_t^j(3,3) = 0\ \forall t$.

Proof. By matrix computation we can show that

$$\text{if } S_{t+1}^i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & s_{t+1}^i(3,3) & s_{t+1}^i(3,4) \\ 0 & 0 & s_{t+1}^i(4,3) & s_{t+1}^i(4,4) \end{bmatrix} \text{ then } S_t^i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & s_t^i(3,3) & s_t^i(3,4) \\ 0 & 0 & s_t^i(4,3) & s_t^i(4,4) \end{bmatrix},$$

where

$$s_t^i(3,3) = \frac{-4\sigma s^2\left[-2 + \sigma\delta s_{t+1}^i(3,3)\right]}{\left[-3 + \sigma\delta(s_{t+1}^j(3,3) + s_{t+1}^i(3,3))\right]^2}.$$

Thus if $\lim_{\sigma\to 0} s_{t+1}^i(3,3) = 0$ then $\lim_{\sigma\to 0} s_t^i(3,3) = 0$. Now, noting that $s_T^i(3,3) = (8/9)\sigma s^2$ the lemma is proved by mathematical induction. ∎

Our candidate equilibrium is $p_t^{i\ *} = p_t^{j\ *} = c + m_t$, $r_t^{i\ *} = r_t^{j\ *} = -m_t$, and $(F_t^{i\ *}, F_t^{j\ *})$, which are given by (27) if $|\Phi_t B| \neq 0$ for $t = 1, ...T$, which is satisfied since reaction functions have different slopes. Thus there exists a unique closed-loop sequence of candidate equilibria, and hence $F_t^{i\ *}(\alpha_{t-1}^i) \equiv x(1)$, which is given by (27), is uniquely determined and define $\widetilde{V}_t^i(\alpha_{t-1}^i)$, which is given by (26). Therefore, $d\widetilde{V}_{t+1}^i/d\alpha_t^i = s_{t+1}^i(3,3)\alpha_t^i + s_{t+1}^i(3,4)$. The proposition will be proved by mathematical induction. First, assume there exists an equilibrium in any period $t+1$, so that $\widetilde{V}_{t+1}$ is a true valuation function, using lemma 2 and for a large enough Δ, we have that $\left|\left(H_t^i\right)_1\right| \simeq -\alpha_t^i/2b$, $\left|\left(H_t^i\right)_2\right| \simeq (\alpha_t^i)^2/4be$, and

$$\begin{aligned} \left|\left(H_t^i\right)_3\right| &\simeq \lambda_{\alpha,t}\frac{(\alpha_t^i)^2}{2be} - \frac{1}{\sigma}\frac{(\alpha_t^i)^2}{2be} + \delta\frac{(\alpha_t^i)^2}{4be}\frac{d^2\widetilde{V}_{t+1}^i}{(d\alpha_t^i)^2} \\ &\quad + \frac{\alpha_t^i}{2e}(\lambda_{p^i,t})^2 + \frac{\alpha_{T-1}^i}{2b}\left(\lambda_{r^i,t}\right)^2, \end{aligned}$$

where $d^2\widetilde{V}_{t+1}^i/(d\alpha_t^i)^2 = s_{t+1}^i(3,3)$. Thus for any $\alpha_t^i \in (0,1]$ and a large enough Δ : $\left|\left(H_t^i\right)_1\right| < 0$ and $\left|\left(H_t^i\right)_2\right| > 0$, and since $\lambda_{\alpha,t}, \lambda_{p^i,t}$ and $\lambda_{r^i,t}$ are bounded functions, and by lemma 3 $\lim_{\sigma\to 0} s_{t+1}^i(3,3) = 0$, there exists a (positive) small enough σ such that $\left|\left(H_t^i\right)_3\right| < 0$. In short, given that there exists an interior equilibrium in period $t+1$, we can construct the Hessian matrix of the network $i's$ profit function in period t, and by assuming *i)* a large enough Δ obtain that for any $\alpha_t^i \in (0,1]$: $\left|(H_t^i)_1\right| < 0$, $\left|(H_t^i)_2\right| > 0$, and *ii)* a (positive) small enough σ obtain that $\left|\left(H_t^i\right)_3\right| < 0$ since $\lambda_{\alpha,t}$, $\lambda_{p^i,t}$ and $\lambda_{r^i,t}$ are bounded functions, and $\lim_{\sigma\to 0} s_{t+1}^i(3,3) = 0$. Therefore, the existence of this candidate equilibrium can be proved by mathematical induction as long as we prove its existence in the last period of the game. In this respect, using proposition 1 we have that for a (positive) small enough σ and a large enough Δ there exists a unique equilibrium in period T, which is interior.

Proof. Proposition 4.

Making use of the proof of Lemma 2 we can write:

$$\frac{\partial\widetilde{U}^{ij}}{\partial p_t^i} = -\left(\frac{1}{b\Delta^2}\right)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}}[\int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} d - e\left(\frac{a-p_t^i+\varepsilon}{b}\right) + \widetilde{\varepsilon}d\varepsilon]d\widetilde{\varepsilon},$$

where $\varepsilon^* = (b/e)(d - r_t^j + \widetilde{\varepsilon}) + p_t^i - a$. Note that,

$$\begin{aligned} \int_{\underline{\varepsilon}}^{\overline{\varepsilon}}\int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} \widetilde{\varepsilon}d\varepsilon d\widetilde{\varepsilon} &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \widetilde{\varepsilon}(\varepsilon^* - \underline{\varepsilon})d\widetilde{\varepsilon} \\ &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \widetilde{\varepsilon}\left[(\frac{b}{e})(d - r_t^j + \widetilde{\varepsilon}) + p_t^i - a - \underline{\varepsilon}\right]d\widetilde{\varepsilon} \\ &= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \frac{b}{e}\widetilde{\varepsilon}^2 + \widetilde{\varepsilon}\left[(\frac{b}{e})(d - r_t^j + \widetilde{\varepsilon}) + p_t^i - a - \underline{\varepsilon}\right]d\widetilde{\varepsilon} \\ &= \frac{b}{e}\frac{\overline{\varepsilon}^3 - \underline{\varepsilon}^3}{3} \\ &= \frac{2b}{24e}\Delta^3 \end{aligned}$$

On the other hand,

$$\begin{aligned}
\xi &\equiv \int_{\underline{\varepsilon}}^{\overline{\varepsilon}}[\int_{\underline{\varepsilon}}^{\varepsilon^*(\cdot)} d - e\left(\frac{a-p_t^i+\varepsilon}{b}\right)d\varepsilon]d\widetilde{\varepsilon} \\
&= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}}\left(d-\frac{e}{b}(a-p_t^i)\right)(\varepsilon^*-\underline{\varepsilon})-\frac{e}{b}\frac{(\varepsilon^*)^2-\underline{\varepsilon}^2}{2}d\widetilde{\varepsilon} \\
&= \int_{\underline{\varepsilon}}^{\overline{\varepsilon}}\left[d-\frac{e}{b}(a-p_t^i)-\frac{e}{2b}(\varepsilon^*+\underline{\varepsilon})\right](\varepsilon^*-\underline{\varepsilon})d\widetilde{\varepsilon}
\end{aligned}$$

Replacing the definition of ε^* into last expression yields

$$\xi = \int_{\underline{\varepsilon}}^{\overline{\varepsilon}}[y+\frac{e}{2b}\frac{\Delta}{2}-\frac{\widetilde{\varepsilon}}{2}][v+\frac{\Delta}{2}+\frac{b}{e}\widetilde{\varepsilon}]d\widetilde{\varepsilon},$$

where $y=(e/2b)(p_t^i-a)+(d+r_t^j)/2$ and $v=(b/e)(d-r_t^j)+p_t^i-a)$. Then,

$$\xi = yv\Delta+\frac{y}{2}\Delta^2+\frac{ev}{4b}\Delta^2+\frac{e}{8b}\Delta^3-\frac{b}{24e}\Delta^3$$

And,

$$\begin{aligned}
\frac{\partial\widetilde{U}^{ij}}{\partial p_t^i} &= -\left(\frac{1}{b\Delta^2}\right)\left[\xi+\frac{2b}{24e}\Delta^3\right] \\
&= -\left(\frac{1}{b\Delta^2}\right)\left[yv\Delta+\frac{y}{2}\Delta^2+\frac{ev}{4b}\Delta^2+\frac{\Delta^3}{8}\left(\frac{e}{b}+\frac{b}{3e}\right)\right] \\
&= -\left(\frac{1}{b}\right)\left[\frac{yv}{\Delta}+\frac{y}{2}+\frac{ev}{4b}+\frac{\Delta}{8}\left(\frac{e}{b}+\frac{b}{3e}\right)\right]
\end{aligned}$$

Thus, for a large enough Δ we can write

$$\frac{\partial\widetilde{U}^{ij}}{\partial p_t^i} \simeq -\left(\frac{1}{b}\right)\left[\frac{e}{2b}(p_t^i-a)+\frac{d}{2}+\frac{\Delta}{8}\left(\frac{e}{b}+\frac{b}{3e}\right)\right]$$

In equilibrium: $p_t^i = c+m_t$, which proves the first part of the proposition. Last, using the same steps as before one can show that for a large enough Δ :

$$\frac{\partial U^{ij}}{\partial r_t^j} \simeq -\left(\frac{1}{e}\right)\left[\frac{b}{2e}(r_t^j-d)+\frac{a}{2}+\frac{\Delta}{8}\left(\frac{b}{e}+\frac{e}{3b}\right)\right],$$

and using that in equilibrium $r_t^j = -m_t$ it is proved the second part of the proposition. ∎

REFERENCES

[1] ARMSTRONG M. "Network Interconnection in Telecommunications." *Economic Journal.* Vol. 108 (1998), pp. 545-564.

[2] ARMSTRONG M. "The Theory of Access Pricing and Interconnection." In M.E. Cave, S.K. Majumdar, and I. Vogelsang, eds., *Handbook of Telecommunications Economics.* Amsterdam: North-Holland, 2002.

[3] BERGER, U. "Network Competition and the Collusive Role of the Access Charge." Mimeo, University of Vienna, 2001.

[4] BOMSEL, O., M. CAVE, G. LE BLANC AND K-H. NEUMANN. *How Mobile Termination Charges Shape the Dynamics of the Telecom Sector,* Final Report, wik Consult, University of Warwick, July, 2003.

[5] CARTER, M. AND WRIGHT, J.“Asymmetric Network Interconnection.” *Review of Industrial Organization,* Vol. 22 (2003), pp. 27-46.

[6] DE BIJL, P. AND MARTIN, P. *Competition and Regulation in Telecommunications Markets.* CPB Netherlands Bureau for Economic Policy Analysis, November 2000.

[7] DE BIJL, P. AND MARTIN, P. *Regulation and Entry into Telecommunications Markets.* Cambridge, UK: Cambridge U. Press, 2002.

[8] DE BIJL, P. AND MARTIN, P. “Dynamic Regulation and Entry in Telecommunications Markets: A Policy Framework.” *Information Economics and Policy.* Vol. 16 (2004), pp. 411-437.

[9] DESSEIN, W. "Network Competition in Nonlinear Pricing," *RAND Journal of Economics,* Vol. 34 (2004), pp. 1-19.

[10] FABRIZI, S. "International Telecommunications Pricing: Does a Scope for Reform Exist?." *Université Toulouse 1 and Università degli Studi di Bologna,* 2005.

[11] GANS, J.S. AND KING, S.P. "Using ‘Bill and Keep’ Interconnect Arrangements to Soften Network Competition." *Economics Letters,* Vol. 71 (2001), pp. 413-420.

[12] HAHN, J. "Network Competition and Interconnection with Heterogeneous Subscribers." *International Journal of Industrial Organization,* Vol. 22 (2004), pp. 611-631.

[13] HERMALIN, B. AND KATZ, M.L. "Network Interconnection with Two-Sided User Benefits." Mimeo, University of California-Berkeley, 2001.

[14] JEON, D.S., LAFFONT, J.-J., AND TIROLE, J. "On the Receiver Pays Principle." RAND Journal of Economics. Vol. 35 (2004), pp. 85-110.

[15] KIM, J.-Y. AND LIM. Y. "An Economic Analysis of the Receiver Pays Principle." *Information Economics and Policy.* Vol. 13 (2001), pp. 231-260.

[16] KYDLAND, F., "Noncooperative and Dominant Player Solutions in Discrete Dynamic Games." *International Economic Review,* Vol. 16 (1975), pp. 321-335.

[17] LAFFONT, J.-J., REY, P. AND TIROLE, J. "Network Competition: I. Overview and Nondiscriminatory Pricing." *RAND Journal of Economics.* Vol. 29 (1998a), pp. 1-37.

[18] LAFFONT, J.-J., REY, P. AND TIROLE, J. "Network Competition II: Price Discrimination." *RAND Journal of Economics.* Vol. 29 (1998b), pp. 38-56.

[19] LAFFONT, J.-J., MARCUS, S., REY, P., AND TIROLE, J. "Internet Interconnection and the Off-Net-Cost Pricing Principle." *RAND Journal of Economics.* Vol. 34 (2003), pp. 370-390.

[20] LOPEZ, A.L. "Dynamic Network Interconnection Under Consumer Switching Costs." *Fundación de las Cajas de Ahorro.* Working Paper No. 214/2005 and *Economics of Networks - WPS.* Available at SSRN: http://ssrn.com/abstract=826387.

[21] POLETTI, S. AND WRIGHT, J. "Network Interconnection with Participation Constraints." *Information, Economics and Policy,* Vol. 16 (2004), pp. 347-373.

[22] SCHIFF, A. "Two-Way Interconnection with Partial Consumer Participation." *Networks and Spatial Economics,* Vol. 2 (2002), pp. 295-315.

[23] VOGELSANG, I. "Price Regulation of Access to Telecommunications Networks." *Journal of Economic Literature.* Vol. 41 (2003), pp. 830-862.

Mergers & Acquisitions and Innovation Performance in the Telecommunications Equipment Industry

Tseveen Gantumur[@]
European University Viadrina

Andreas Stephan
European University Viadrina and DIW Berlin

ABSTRACT

In response to global market forces such as deregulation and globalization, technological change and digital convergence, the telecommunications in the 1990s witnessed an enormous worldwide round of Mergers & Acquisitions (M&A). Given both *M&A and Innovation* a major means of today's competitive strategy development, this paper examines the innovation determinants of M&A activity and the consequences of M&A transactions on the technological potential and the innovation performance. We examine the telecommunications equipment industry over the period 1988-2002 using a newly constructed data set with firm-level data on M&A and innovation activity as well as financial characteristics. By implementing a counterfactual technique based on a matching propensity score procedure, the analysis controls for merger endogeneity and ex-ante observable firms characteristics. The study provides evidence that M&A realize significantly positive changes to the firm's post-merger innovation performance. The effects of M&A on innovation performance are in turn driven by both the success in Research and Development (R&D) activity and the deterioration in internal technological capabilities at acquiring firms prior to a merger.

Keywords: Mergers & Acquisitions, Innovation Performance, Telecommunications Equipment Industry

*JEL Classifications***:** L63, O30, L10

[@] Corresponding author: European University Viadrina, Faculty of Economics and Management Studies, Grosse Scharrnstrasse 59, D-15230 Frankfurt (Oder), E-mail: gantumur@euv-ffo.de

1 Introduction

The telecommunications industry is moving fast both on the technology front and in terms of structure. A recent surge of Mergers and Acquisitions (M&A) in the telecommunications industry is a reflection of the drastically changing environment of the market.[1] Deregulation and liberalization, technological innovation and digital convergence and the evolving requirements of the capital markets have driven dramatic changes in the telecommunications industry as a whole. The industry in turn has sparked fundamental changes in the economic landscape worldwide. As the telecommunications firms face increasing exposure to international competition, the industry has undergone a radical transformation creating exciting new opportunities and new challenges for infrastructure and service providers (Li and Whalley, 2002). Market winners are in most cases also technology leaders or highly capable of turning a base technology into a superior product that meets the customer needs (Brodt and Knoll, 2004).

The rapid technological change, growing technological complexity and the shortening of product life cycles add new dimensions to an already complex scenario and increasingly force firms to source technologies externally. Firms will often prefer M&A to other cooperative approaches of R&D network building, e.g., R&D joint ventures, because M&A provide an immediate controlling presence in the new, fast expanding market, rather than having to gradually build a new business or negotiate with a partner about developing a cooperation (Caves, 1982; Capron and Mitchell, 1997). While several analyses have stressed that the telecommunications industry has undergone major restructuring in the 1990s through intense M&A activity (e.g., Jamison, 1998; Kim, 2005; Rosenberg, 1998, Warf, 2003), we are not aware of any study which investigates the linkage between recent rises both in M&A and innovation activity. The goal of this paper is to uncover the keen reliance of the telecommunications firms on M&A as a technology sourcing strategy.

We aim at providing an answer to the following question: Why do firms in the telecommunications industry increasingly use M&A as a technology source? Does M&A affect the innovation performance of firms involved as their proponents expect? Before attempting to determine this task, however, a more basic question needs to be addressed, namely: Does the innovation activity of firms depict a significant predictor of entering the M&A market? Admittedly, technological reasons do not motivate all M&A. M&A can be motivated, for instance, by the desire to obtain financial synergies or market power, to obtain access to distribution channels, and/or to gain entry into new markets.[2] Such M&A may not be able to be

[1] Between 1996 and 2001, more than twenty M&A deals worth over $20 billion took place in the telecom sector, 14 of which were in the US. Telecom mergers amount for seven of the largest operations announced in 2000, and eight out of the ten largest of all times (Le Blanc and Shelanski, 2002)

[2] For extensive review, see Shimizu et al. (2004)

directly expected to improve the firms' innovation performance. However, in high technology industries where innovation is key to a competitive advantage, firms will incorporate the impact of M&A on technological performance even when the transaction is not innovation-driven, thus choosing the most appropriate innovation and financial strategies. Moreover, to the extent that access to technology and know-how become increasingly important to succeed in the market, factors such as the firm's size, history and equity become less and less critical requirements. This allows new challengers to realize tremendous growth rates. Furthermore, it spurs the quest for external knowledge sourcing both at the established and new firms in the market. As innovation is becoming indispensable for strategic competitiveness in the high technology industry, we ask: How do firms that choose M&A and firms that stay outside of the M&A market differ with respect to their innovation performance? The follow up question is then, what are the effects of M&A on the innovative performance of firms if we control for the differences in innovation performance prior to M&A activities?

Though occurrence of M&A has grown dramatically in the last years, academic research on the relationship between innovation and M&A has not kept pace with the changes. In spite of the vast and rapidly growing body of literature on M&A,[3] empirical evidence which has explored this relationship is rather limited and often inconclusive.[4] The literature on the technological effects of M&A shows contradictory implications. On the one hand, M&A may build up competencies and foster innovation for a number of reasons. M&A can reduce high transaction costs related to the transmission of knowledge between firms (Bresman et al., 1999). Furthermore, in fast moving markets with abbreviated product life cycles, firms may perceive that they do not have the time to develop the required skills and knowledge internally, and therefore, turn outward to M&A. In this sense, M&A may offer a quick access to knowledge assets (Warner et al., 2006). Moreover, M&A may extend the technological base of firms involved allowing them to achieve greater economies of scale and scope through more efficient deployment of knowledge resources. Also, M&A may enlarge the overall R&D budget of firms engaged, which then enables them to tackle larger R&D projects and, thereby, this spreads the risk of innovation. In addition, the integration of complementary knowledge may also increase innovation through M&A leading to more advanced technologies being developed (Gerpott, 1995). Finally, exchanging the best practices on innovation management within the combined entity, firms may employ efficient technology integration.

On the other hand, innovation-driven M&A encompass the difficulties associated with innovation as well as the obstacles faced in mergers. First of all, differences in corporate culture, processes and knowledge base may impede a smooth transition of knowledge (Lane and

[3] For review see Roeller et al. (2001) and Shimizu et al. (2004)
[4] For review see Veugelers (2005)

Lubatkin, 1998; Very, 1997). Furthermore, M&A integration process is time consuming and costly. This may divert management attention away from innovation (Hitt et al., 1996). Also, trade off payment of debt and debt costs for investment in R&D can occur due to M&A (Hitt et al., 1990). In addition, a disadvantage of M&A is that it involves entire firms whereas the advantages for knowledge exchange may be limited to only a small part of the firms involved. In M&A, knowledge beyond that required is also acquired. This may cause indigestibility: a firm may acquire more knowledge than it can use in a meaningful way (Hennart and Reddy, 1997). Finally, as the literature has shown, technologically motivated and intensive acquisitions are highly vulnerable to failure (Chakrabarti et al., 1994). One of the main reasons for this value destruction lies in the miscarried and inappropriate integration of the technology-based firm after the acquisition (Duysters and Hagedoorn, 2000). Even when the merger is successful in terms of the integration of R&D departments, in other business areas the merger may not be a success, thereby influencing a disintegration of the entire firm.

One of the main reasons for the contradictions and inconclusiveness of previous studies might be due to cross-industry investigations. Consequently, we focus on the recent increase in M&A and innovation activities of firms in the telecommunications equipment industry as the existence of industry clustering in M&A activity in the 1990s is evident (Andrade et al., 2001). Lying at the core of the telecommunications industry, the telecommunications equipment industry takes a central role in the technological transformation of the entire industry. Also, the growing market concentration observed for technology producers and the relevance of patent-based technological standards in recent years are motives for selecting this well-defined industry.This study provides empirical evidence on our research questions by examining the M&A that took place between telecommunications equipment firms during the 1988 to 2002 period. It uses a newly created data set from the following four complementary data sources: NBER Patent, Thomson One Banker-Deals, Compustat and Global Vantage databases. We adopt a treatment effect estimation approach with endogenous selection using a matching propensity score technique. In order to explore the interrelationship between M&A and R&D of firms, we capture broad dimensions on both activities. We analyze R&D input and output as well as the knowledge stock and research productivity of firms. Furthermore, we reveal the role that a firm plays in a transaction as an acquirer, a target and a pooling merger.

We find evidence consistent with the following propositions. First, the telecommunications equipment firms undertake M&A in order to strengthen their success in innovation, and thereby, their market position. Second, the equipment manufacturers in telecommunications, which experienced low research productivity from ongoing exploitation of R&D efforts in the past, are forced to explore potential future innovation trajectories in new business units by acquisitions. Third, those telecommunications equipment firms with a declining inventive portfolio, are involved in pooling mergers to offer comprehensive and integrated equipment solutions. Finally, equipment firms in telecommunications can successfully outsource R&D through M&A as a means of revitalizing a firm by enhancing and supplementing its knowledge base effectively.

The article proceeds as follows: Section 2 draws a broad picture on the developments in telecommunications equipment industry with regard to its M&A and R&D activities. Section 3 presents the theoretical underpinnings of our research questions. Data description is provided in Section 4, while Section 5 discusses measures of performance and empirical methodology. We report empirical results in Section 6 and conclude with a discussion of findings in Section 7.

2 Telecommunications Equipment Industry

The telecommunications equipment industry provides all equipment required for the use or provision of telecommunications and data services. The structure of the industry is presented in Figure 4 and described in the Appendix.

Until recently, with its regulated and very stable market structure, the telecommunications industry had correspondingly little, if any, opportunities for mergers.[5] Most national telecommunications markets were characterized by vertically integrated national monopolies where telecommunications equipment production and its demand were largely in the hands of government agencies. This picture radically changed in late 1980s with the beginning of the liberalization and deregulation process which culminated in the standardization and subsequent rapid diffusion of the GSM.[6] During this period, the vertically integrated monopolies had dissolved, while R&D and innovation increasingly shifted towards the producers of telecommunications equipment. For the telecommunications equipment producers, regulatory liberalization has both implied the opening of new markets abroad and a shift in competition from the national to the global level.

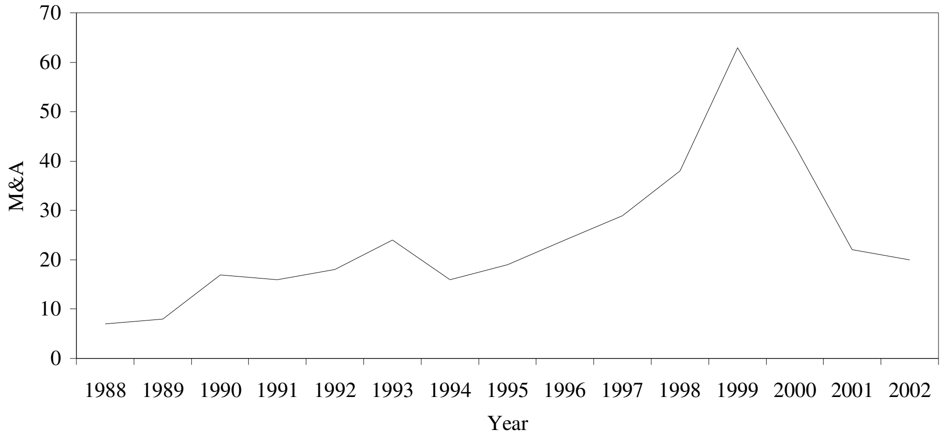
In addition, the process of convergence between telecommunications, IT and broadcasting through its technological, organizational, and market/service aspects has a far-reaching influence on the market structure of the telecommunications industry (Tadayoni and Skouby, 1999; Kim, 2005). The ICT convergence essentially has come about through two parallel technological changes. The first initial radical innovation is the application of integrate circuit technology to allow digital switching of telephone calls that change the core design concepts of telephone systems. The second development is the Internet. It is based on the

[5] We employ, hereafter, the term “merger” to define both merger and acquisition if not otherwise indicated.
[6] Global System for Mobile Communication

TCP/IP[7] protocol that standardizes the rules of packaging, transmitting and receiving data over the Internet, providing more flexibility and fitting further between information and telecommunications technologies (Minin and Palmberg, 2006).

As a result of international competition stemming from the liberalization of its market and pace of technological evolution, telecommunications firms that traditionally operated only in their home markets have expanded throughout the world, often by undertaking mergers of foreign telecommunications firms. The 1990s witnessed an enormous wave of mergers that dramatically reconfigured the market structure of global telecommunications equipment (see Figure 1). Mergers provided a means of rapidly overcoming the shortcomings of existing networks and services as well as helping address the inadequacies associated with the provision of such global services under the existing framework of correspondent relationships between telecommunications operators (Capron and Mitchell, 1997).

Figure 1. M&A in the Telecommunications Equipment Industry, 1988-2002



Source: Authors' calculations from Thomson One Banker-Deals

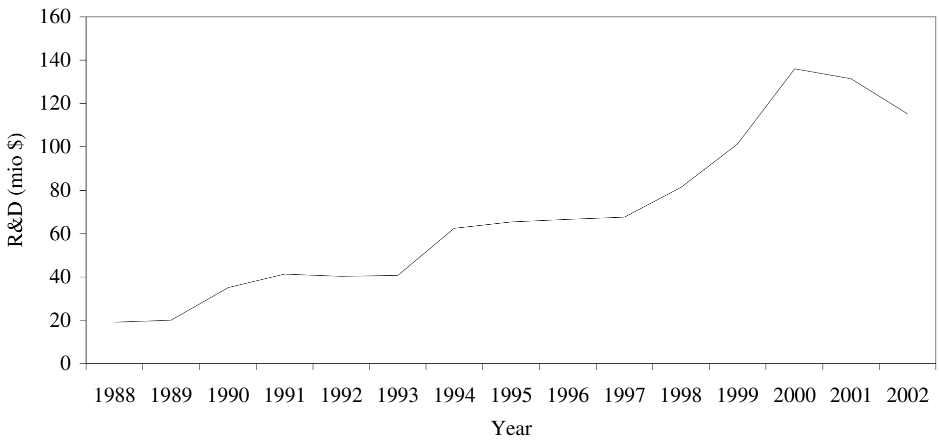
In the recent years, the existing network operators have rebuilt their telecommunications infrastructure to accommodate increased voice and data traffic. Major investments were undertaken to improve network capacity, to enable the transport of high-speed data and to enhance the intelligence of the fixed network to enable customers to benefit from advanced services. Investment in R&D plays a major role in the rise and fall of the telecommunications equipment industry. While the trade and regulatory liberalization primarily has globalized the demand for telecommunications equipment, technological change in the industry has had pervasive effects further upstream on R&D. One aspect of this is the ongoing convergence between various technology sub-fields of ICT, which require technological diversification

[7] Transmission Control Protocol/Internet Protocol

amongst the incumbents and open up multiple entry options (Minin and Palmberg, 2006) also for new players.

Figure 2 gives a broad overview of the R&D expenditures for the equipment manufacturing firms of our sample over the last decade. During the monopoly era, hardly any visible importance to R&D was given as the technology was dominated by analog systems. Digitalization was the main technological achievement in communications: hence, new technologies were introduced into the market in all areas of communications, e.g., voice, data networks, satellite, cellular, switching and routing network devices, etc. For the first time service firms had choices in terms of selecting the equipment manufactures. R&D experienced its peak during those years. However, after the telecom crash, both the sales volume and retail prices dropped. That directly affected the amount of R&D spending in 2001. Research did not make it through the product lifecycles and the launching of new technologies was delayed, e.g., third generation (3G) wireless networks.

Figure 2. Average R&D expenditures in the Telecommunications Equipment Industry, 1988-2002

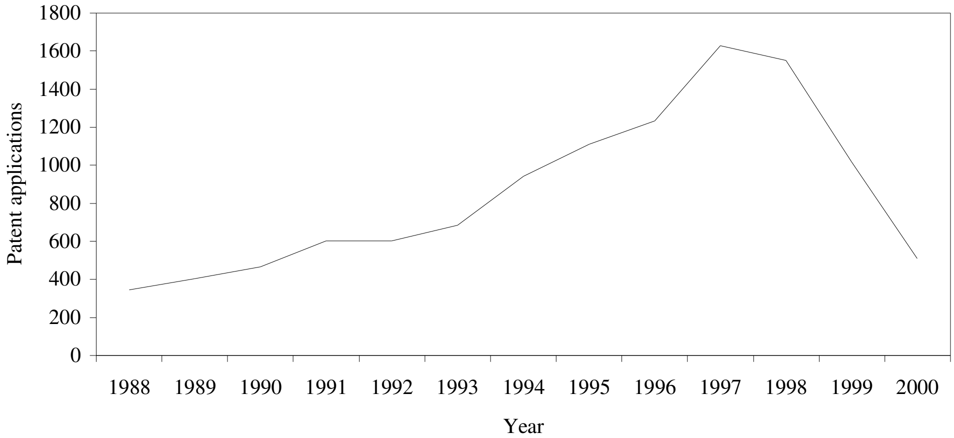


Source: Authors' calculations from Compustat and Global Vantage databases

In the digitalized markets, success is increasingly based on non-physical or intangible factors rather than merely on the telecommunications networks and other physical assets that provided the previous national telecom monopolies with a primary source of revenues (Quah, 2000). Recent intensive competition has dramatically changed the business environment of the telecommunications firms and forced them to seek new strategies. Not only technological innovations themselves but also related patent right royalties or license fees may provide telecommunications firms with an outstanding source of income. It may be increasingly costly not to be innovative because the license fees for patents relating to the communication products become a major cost for communication product manufacturers and their customers

(Kretschmer, 1997). The growth in the patenting of the worldwide telecommunications equipment industry has been tremendous (see Figure 3). From 1988 to 1998 the number of communication equipment patents applied by the UPSTO increased by more than four times. The abrupt fall in the patent applications after 1998 in Figure 3 is primarily caused by the truncation of the patent data sample. [8]

Figure 3. Patenting in the Telecommunications Equipment Industry, 1988-2000



Source: Authors' calculations from NBER Patent Database

To summarize the M&A and R&D trends, the telecommunications equipment industry in the 1990s has experienced the race of technological changes and a major restructuring, which is characterized by intense merger activity.

3 Theoretical Background

Technological change influences the ability of firms to integrate, build and reconfigure internal and external competencies in order to address altering competitive and technological challenges. Dosi (1988) described the technological changes to be continuous or incremental because they reflect a path dependent and cumulative development as a technological paradigm or pattern of inquiry. Incremental change tends to reinforce the market power of incumbent firms because it utilizes existing competencies in development and can be deployed through an established set of sales and marketing resources (Teece, 1996). Accumulated prior knowledge and heuristics constitute the related problem-solving knowledge that permits incumbents to acquire related

[8] While considering only patents that eventually granted, we date the patents in our sample by the date of application. We have all patents which were granted by 2002; thus, our patent numbers are not quite complete in the latter years. We end our analysis on patents in 2000 because, in the subsequent years, a truncation due to the grant lag appears to be more visible.

problem-solving capabilities. Learning capabilities involve the development of the capacity to assimilate existing knowledge, while problem-solving skills represent a capacity to create new knowledge; and they are mutually inclusive.

To the extent that the innovation embodies new skills or knowledge, incumbents can be hindered in responding as they may have little or no relevant development history to draw upon (Dosi, 1988). Firms must absorb the environmental information on innovation and eventually be able to exploit it through new products or processes in the market. Cohen and Levinthal (1989) elucidated the two faces of R&D activity. That is, R&D activity does not only stimulate innovation, but it also enhances the firms' ability to assimilate outside knowledge. The second face of R&D is called the absorptive capacity, and it is considered to be crucial particularly for assessing the effective contribution by spillovers from others. Defined as a set of knowledge and competencies, the firm's knowledge base remains a preliminary condition in the assimilation of spillovers from R&D efforts of environment. For Rosenberg (1990), fundamental research inside the firm has strong complementarities with external R&D. All in all, both Cohen & Levinthal and Rosenberg insist on potential synergies between the firm's own knowledge base and external flows of scientific and technical knowledge. Thus, the responsiveness of R&D activity to exploit external knowledge flows is an indication of the importance of absorptive capacity. In industries like telecommunications, this response must be quick due to highly competitive conditions caused by short product lifecycles, new technologies, frequent entry by unexpected outsiders, repositioning of incumbents and radical redefinitions of market boundaries as ICT industries converge. Highly reactive firms with highly absorptive capacities will not wait for failure to spur development. By contributing R&D to the firms' absorptive capacity, however, one should bear in mind the fact that technological performance does not necessarily depend on past or referential performance, but rather on absorptive capacity generated in the past. In other words, firms with high absorptive capacity will exploit new ideas regardless of their past performance.

Firms, especially those with high technological content, strive to overcome constraints aligned with cost, appropriation, absorptive capacity and time regarding R&D performance. Thus, firms are faced with the associated objectives of developing a response to an innovation and doing so in a timely fashion. Therefore, there is a crucial strategic choice to be made for firms that decide to conduct R&D activities. Traditionally, the economic literature on the choice of strategies for performing R&D basically considers two alternatives: either R&D is performed internally or it is acquired externally; these two alternatives could also be combined. The theoretical literature, drawing on transaction costs economics (Coase, 1937; Williamson, 1981) and property rights theory (Hart and Moore, 1986), considers the choice between external sourcing and internal development as substitutes, i.e., the classical make-or-buy decision.

Technological know-how is often tacit and can, therefore, not be easily transmitted from one firm to another (Larsson et al., 1998). In order to avoid high transaction costs, firms may be induced to engage in a merger in order to solve problems related to the transmission of tacit knowledge (Bresman et al., 1999). At the same time, internal developments may be perceived by firms because of the high risk due to the low probability of the innovation success and the length of required time for the innovations to provide adequate returns (Hundley et al., 1996). Thus, firms prefer to invest fewer resources in internal R&D when faced with resource constraints or attractive external innovation sources exist. It is argued that the acquisitions of firms with an innovative portfolio of interest often represent more certainty and lower risk of exploiting knowledge assets than new ventures do (Chakrabarti et al., 1994). Engaging in acquisitions, firms, however, may trade off payment of debt and debt costs for investment in R&D. That is, as the innovation developments embed assets that are largely non-redeployable, firms are likely to prefer the use of debt to fund acquisitions rather than to support innovation activities (Hitt et al., 1990).

However, due to the fact that the financial and innovation strategies of future-oriented firms are jointly decided, a lack financial resources may be imperative for firms pursuing a competitive strategy premised on innovation. Hence, the mutually exclusive choice between these innovation strategies is too restrictive. Moreover, R&D strategy adopted by a firm depends on its environment and on differences in the abilities of the firms to conduct R&D activities. While the difficulty in being a good ‘buyer’ when one is not also a ‘maker’ and vice versa occurs, most theories of economic organization which rely on a comparison of costs or benefits per transaction to explain the organization of economic activity have typically ignored the possibility of multiple innovation sources.

The studies inspired by the resource- and knowledge-based approaches argue that a firm can rely on a combination of different strategies to engage in innovation. Following the absorptive capacity argument, the ability to recognize the value of new information and also to integrate and deploy it is enhanced when the new knowledge is related to what is already known (Cohen and Levinthal, 1990). The combination of similar fields of technology are expected to enable firms to share technological expertise, to shorten the innovation lead time and to engage in projects larger than what would be possible within the once separated firms (Kogut and Zander, 1992). Yet, the relatedness between the knowledge bases is likely to have a non-monotonic influence on subsequent integrated innovation performance (Lubatkin, 1983; Lane and Lubatkin, 1998). In other words, with increasing relatedness, the innovation performance will initially increase reaching some optimum and then it will decrease. Ahuja and Katila (2001) showed that, from one side, when the knowledge bases of acquiring and acquired firms are very similar, then the beneficial effects of such combination would be limited, as there will not be much to learn from each other. On the other hand, when the knowledge bases are very distinct, the absence of common skills and similar cognitive structures and the lack of an adequate absorptive capacity will impede communication and knowledge transfer. Nevertheless, distinctive knowledge bases of internal and external sources can be particularly valuable under conditions of technological uncertainty (Sorenson and Sorensen, 2001) and might be of use in creating knowledge complementarities.

It is theoretically evident that the firms who invest in R&D have, in any case, a natural advantage in the exploitation of high technology, and if the results of their own R&D have not been satisfactory, those firms are inclined to procure technology externally. In this context, to justify the desideratum of the external technology source, it is essential to attend to the increasing evidence that a firm's size and position within the industry affects the nature and the type of innovation in which it is engaged (Hart and Ramanantsoa, 1992; Christensen, 1997). First, pursuing to develop the knowledge and to create a new one internally, firms might be particularly hindered by radical or significant innovations rather than by minor or incremental innovations since the latter are technological changes that are close to the current expertise. This is distinctive to established firms in the market or market leaders - mostly large firms, which tend to innovate in order to reinforce their positions or to enhance their core competencies. The improvements on their R&D can be, indeed, significant, but they are not likely to change their status quo.

Second, prior success in developing competencies may block firms from adjusting from environment. The former competencies may become rigidities or barriers to performance for developments that differ greatly from existing knowledge. To fully utilize its strengths, large firms need the path to innovation to be predictable. Having less to gain from a radically new design than a market challenger, they are less likely to pursue disruptive technologies or to embrace new innovations which would threaten their dominance.

Finally, large firms might prefer exploitative investments rather than explorative as the latter is uncertain in payoff and organizationally challenging. However, the ongoing exploitation of the existing knowledge and capabilities, even those that make an organization successful for a certain time, after a certain point hinders the creation of new knowledge and eventually leads to a technological exhaustion (March, 1991; Vermeulen and Barkema, 2001). These self-reinforcing capacities can also create competency-destroying technological change.

On the other hand, new firms or market challengers, mostly small firms, are more entrepreneurial and can respond more quickly to unexpected opportunities. By creating new fields of technology or new skills where the market leader does not have an expertise or an established position, they are looking for opportunities to upset the leader's position and to radically change the competitive situation, thus eliminating or diminishing the leader' market dominance. While they are more likely to fail, they are more willing or able to venture into completely new directions because they have less of a vested interest in the current technology and are not tied to sunk investments in obsolete technologies.

At the same time, small challengers have fewer resources to spend on R&D and because there is a lack of strong enterprise channels, they are less likely to have the resources to bring an invention to the marketplace. This lack of manufacturing and distributing activity can be filled by large firms which possess a greater ability to finance a large amount of R&D as well.

In sum, we suppose that a disruption in productive efficiency in the innovation activity of firms forces them to turn outward to technological mergers. Next, we expect that a merger between accumulated absorptive capacities of firms facilitates the merging firm's ability to understand new knowledge held by its acquisition target or partner in a pooling merger. Furthermore, we presume that relative absorptive capacity of merging firms enhances the merging firm's ability to assimilate new knowledge from its acquisition target or partner in a pooling merger. In addition, we predict that the strengthening of building absorptive capacity of merging firms shapes the merging firm's ability to apply the assimilated knowledge. Finally, we propose that a merger influences the merging firm's ability to strengthen and to build up its knowledge base. Consequently, the R&D functions of merging firms have to be effectively integrated and coordinated through the interaction emerging in such circumstances.

4 Data Description

In order to examine the interaction between merger and innovation activity, a new firm-level database is constructed which covers all firms in the telecommunications equipment industry that operated in any year over the 18 years period, 1987 to 2004 (including lagged periods). This database is created by complex matching process of information from initially four separate datasets. The first two datasets include firms' financial characteristics and the additional two data sets describe the firms' merger and innovation activities, respectively.

We define the telecommunications equipment firms as those which have primary activity in the communications equipment Standard International Codes (SIC) 3661, 3663, or 3669. The population of firms and their financial information including R&D expenditures were drawn from Compustat and Global Vantage databases. After eliminating firms with missing financial information, we could identify a sample of 638 telecommunications equipment firms for those a data on R&D expenditures, total assets, market value, cash flow, long term debt were available.

Our patent statistics for the telecommunications equipment industry are based on the database which is compiled by the National Bureau of Economic Research (NBER, Hall et al., 2001). This database comprises detailed information on all US patents granted between 1963 and 2002 and all patent citations made between 1975 and 2002. The patent and citations data were procured originally from the US Patent Office and from Derwent Information Services, respectively. Although this US data could imply a bias in favor of US firms and against non-US firms, the group of non-US firms in this sample represents a group of innovative and rather large firms that are known to patent worldwide. Our database includes information on the patent number, the application and grant dates, the detailed technology field(s) of the innovation, the name(s) of the inventors, the city and state from which the patent was filed and citations of prior patents on which the current work builds. Following the classification in Hall et al. (2001), we include the patents for which firms applied in twelve main classes of the International Patent Classification (IPC) 178, 333, 340, 342, 343, 358, 367, 370, 375, 379, 385 or 455 - in the category communication equipment. As the distribution of the value of patented innovations is extremely skewed, we also consider the number of forward citations as an indicator of the importance or the value of innovations for each patent, thereby overcoming the limitations of simple counts (Brouwer and Kleinknecht, 1999; Griliches, 1990). During the observed period, 251 firms from our sample have applied for a total of 11,226 patents in communication equipment (including multiple applications by the same firm in the same year and for the whole period); this produces a total of 86,442 citations.[9] The most active firms in the patent applications were Motorola, Siemens, Nortels Network, Qualcomm, and Alcatel Bell Telephone with 61.72% of the total patent applications.

M&A transaction data were obtained from the Thomson One Banker-Deals database. Updated daily, the database offers detailed information on merger transactions including target and acquirer profiles, deal terms, financial and legal advisor assignments, deal value and deal status. This database includes alliances with a deal value of more than 1 million USD, thus ensuring that the overwhelming majority of mergers are covered. Our final sample on merger transactions contains information on 364 completed deals (including multiple deals by the same firm in the same year and during the observed period) carried out by 178 firms and announced during the period from 1988 to 2002. Using information from the initial database, we distinguished between the role that a firm played in a M&A transaction and classified the firms in our sample in generally as an acquirer, the firm which purchased the stock or other equity interests of another entity or acquired all or a substantial portion of its assets; a target, the firm which sold a significant amount or all of itself to another firm; or a partner in a pooling merger, the firm which pooled its assets with another firm or merged with another firm of approximately

[9] The data set is truncated, which might cause a downward bias in the citation counts of recent patents.

equal size. Out of 364 M&A transactions, we could identify 217 acquirer, 25 targets and 122 partners in pooling mergers.[10] Furthermore, 59.6% of all of the mergers involve innovative firms, i.e., firms that applied for at least one patent during the observed period. While 84.8% of the merger firms took part up to three times in a merger, we can observe that the merger activity of the telecommunications equipment industry is characterized by the transactions of certain firms. For instance, the large-scale firms such as Ericsson, Siemens, ADC Telecommunications, Motorola and Alcatel carried out 17.86% of the total merger transactions. For our econometric analysis, we restrict the multiple transactions carried out by one firm in the same year to the largest transaction only.[11] Finally, the estimation sample consists of total 300 M&A transactions, which involve 186 acquirer, 22 targets and 94 partners in pooling mergers.

The databases were matched on the basis of firm names, CUSIP numbers[12] and address information provided by each database. The firms that are lacking information or have inadequate data on the matching procedure were cross-checked and completed with information reported in the Dun & Bradstreet's "Who owns whom" annual issues.

5 Econometric Methodology

In this section we describe the econometric approach. The aim of the analysis is twofold: first, to investigate the impact of success in innovation activity on the likelihood that a firm engages in a merger, and second, to analyze the effect of a merger on a firm's innovation performance.

The general perspective on mergers and the technological-related reasons of merger activity both stress the importance of understanding the conditions under which the change in ownership might have a significant effect on the innovation performance of firms. Hence, the effects of a merger must be related to the reasons and expectations behind the transaction. In order to explore the link between merger and innovation performance of firms, we estimate the innovation determinants that influence the merger and may alter following the merger.

[10] We lack financial data on the target firms for transactions that involve the acquisitions mostly of a privately held and/or a relatively small firms that are not operated in the US and not listed in Global Vantage,.

[11] The frequency of M&A transactions carried out by one firm in the same year is as follows: 294 firms with one deal, 44 firms with two deals, six firms with three deals, and three firms with four deals in a given year during the sample period.

[12] CUSIP stands for Committee on Uniform Securities Identification Procedures. A CUSIP number identifies most securities, including stocks of all registered US and Canadian companies and US government and municipal bonds. The CUSIP system facilitates the clearing and settlement process of securities. A similar system is used to identify foreign securities (CUSIP International Numbering System).

5.1 Estimating the Propensity to Merge

We start our analysis by exploring the determinants of mergers and by investigating the attractiveness of telecommunications equipment firms as merger candidates. Employing a random utility model, we consider the firm i's decision of whether to acquire, to be acquired, to have involvement in a pooling merger or to stay outside the merger market. The utilities associated with each of these choices k are modeled as a function of the firm's characteristics X_i which affect the utilities differently:

$$U_{ik} = X_i \beta_k + e_{ik} \tag{1}$$

While the level of utility is not observable, we can, however, infer from the firms' choices how they rank each of these alternatives. If we assume that the e_{ij} are distributed Weibull, the differences in the disturbances are distributed logistic and a multinomial logit can be used to estimate the differences in the parameters β.

The propensity of engaging in a merger is modeled as a function of the firm's characteristics. We base the analysis on a panel that consists of innovation-related and financial variables on both merged and non-merged firms for which data were available during the 1988 to 2002 period. The probability that firm i chooses alternative k is specified

$$Pr(\, i \text{ chooses } k\,) = \frac{exp\left(\beta_k' X_i\right)}{\sum_l^m exp\left(\beta_l' X_i\right)} = \frac{1}{\sum_l^m exp\left[\left(\beta_l - \beta_k\right)' X_i\right]} \tag{2}$$

where $\beta_1,...,\beta_m$ are m vectors of unknown regression parameters.

An important property of the multinomial logit model is that relative probabilities are independent from each other, which is the so-called independence of irrelevant alternatives (IIA) property. In order to obtain robust standard errors of estimated coefficients, appropriate tests were conducted, which are discussed in Section 5.1.

In the following, we explain the determinants of a merger captured by our analysis and assess the appropriateness and plausibility of the merger choice. Summary statistics of the variables are shown in Table 1.[13]

The innovation performance of a firm is examined with respect to its R&D input, R&D output, the stock of accumulated knowledge generated by past R&D efforts, and the research productivity. R&D input and R&D output of firms are measured by their R&D expenditure and the number of patent applications, respectively. R&D expenditures involve both current and

[13] We checked that there exists no multicollinearity among selected variables.

capital expenditures, where the current expenditures are composed of labor costs and other current costs, and the capital expenditures are the annual gross expenditures on fixed assets used in the R&D projects of firms. We analyze the firms' intellectual property rights registered as patents that are actually granted. Each patent contains highly detailed information on the innovation itself (Hall, 2001). As a strong relationship exists between the size of the firm and its R&D expenditure and total number of patents, as suggested by common innovation studies, we took the ratios of the R&D expenditures and the patent counts to the total assets; we then defined them as R&D intensity and patent intensity, respectively.

INSERT TABLE 1 ABOUT HERE

In order to account not only for the quantity but also the quality of the patented inventions, we measured the patent-based characteristics of a firm using the number of forward citations of patents. As mentioned above, the number of citations received by any given patent is truncated in time because we only know about the citations received thus far. In other words, the number of forward citations a patent received depends on the year of the application. We, therefore, normalize the citation counts by their average value calculated over all patents belonging to the same technological sub-class whose application was filed in the same year.[14] We then weight each patent of a firm by the number of normalized citations that it subsequently received (Trajtenberg, 1990).

The stock of accumulated knowledge of a firm is measured using citation-based patents and calculated by applying the perpetual inventory method by assuming a depreciation rate of 15% per annum (Hall, 1990). Hence, the individual patents in the firm's knowledge base provide the basis for comparing the firm's own knowledge base with that of other firms. R&D productivity, defined as the ratio of citation-weighted patent to R&D expenditure, accounts for the firm's research productivity. Research productivity may be interpreted as the efficiency with which R&D brings forth new and useful knowledge.

Since financial profiles of firms are likely to influence both their innovative and merger activity, we also include the firms' financial characteristics. To express all monetary values in real terms, we employ the US industry-based Producer Price Index with basis year 1999. All covariates in the regressions have been lagged by one year in order to avoid potential endogeneity problems as well as possible biases arising from different merger accounting methods and financial statement consolidation.[15]

[14] This is the *fixed-effects* approach proposed in Hall et al. (2001)

[15] Similar methods have been widely used in the literature (Hall, 1999). We are aware, however, that lagging regressors do not control for all sources of endogeneity (Wooldridge, 2002).

Firm size is proxied by the book value of the total assets. Some empirical evidence has shown that the purchase of larger companies is positively related to post-merger performance, as larger targets can benefit the buyer in terms of economies of scale, a larger resource base and a larger customer base (Seth, 1990; Loderer and Martin, 1992; Clark and Ofek, 1994; Ahuja and Katila, 2001). However, other studies have claimed that these potential benefits might not be realized if the integration of larger acquired organizations creates greater coordination problems and needs resources to be devoted to solve this at the expense of business operations, thus leading to a negative impact of a merger (Lubatkin 1983; Kusewitt 1985; Ahuja and Katila, 2001).

The economic performance of a firm is proxied by firm growth and Tobin's q. Firm growth is measured by the annual growth rate of the market value. Firms with growing market value may appear as likely acquisition targets for mature firms looking to absorb growth opportunities. We approximate Tobin's q by calculating the ratio of the market value to the book value of a firm's assets, where the former is the sum of the book value of long-term debt and the market value of common equity (Danzon et al., 2004). According to the q-theory of investment, capital should flow from low-q to high-q firms. Indeed, by knowledge flows, technology shocks cause a large variation in the firms' Tobin's q (Jovanovic and Rousseau, 2004). The interpretation of the effect of Tobin's q should be treated with some caution, because, apart from being a forward looking indicator - a firm's growth opportunities (Gugler et al., 2004), Tobin's q is also likely to reflect stock undervaluation (Mork et al., 1990), or managerial performance (Powell, 1997).

The cash flow ratio is defined as the ratio of cash flow to the total assets, and it represents the financial capabilities of the firms. The cash-flow ratio amounts for funds available to a firm for operations, investments and acquisitions. Given the argument that R&D is primarily financed by internally generated resources, the cash-flow ratio might be an important determinant of the (inclusively) choice between internal R&D or external know-how of innovative firms.

Following the practice of previous studies (e.g., Hall, 1999), a dummy variable is included which indicates missing R&D values and equals one when R&D is missing and zero otherwise. For the firm-years observations with missing R&D intensity, we then set the R&D intensity equal to zero. Moreover, to capture the difference between firms with no R&D output, we employ similarly a dummy for firms with zero (citation-weighted) patent intensity.

Table 2 depicts the t-statistics of the differences in means of the firms' characteristics separately for merged and non-merged firms. Firms that actually merged are characterized by a greater knowledge stock expressed in accumulated intellectual property rights than firms that did not merge. In terms of total assets, there is a significant size difference between merged and non-merged firms, thus showing that larger firms are more likely to merge.

INSERT TABLE 2 ABOUT HERE

The merged firms had, on average, a larger Tobin's *q* and cash-flow ratio, and they were less likely to have missing R&D values and zero (citation-weighted) patent intensity. Furthermore they had experienced more merger transactions in the past relative to firms that did not engage in a merger in a given year. The firms in our sample do not differ significantly in their R&D and (citation-weighted) patent intensity as well as research productivity prior to a merger.

5.1 Estimating the Impact of M&A on Innovation

Our analysis of the effects of mergers controls for endogeneity and ex-ante observable firms characteristics using a propensity score method (Dehejia and Wahba, 2002).

For each firm i in the sample, let M_i be a merger indicator that equals one when the firm engages in a merger and zero otherwise. We denote Y_{i1} as the innovation performance of merging and Y_{i0} as the innovation performance of non-merging firms and observe M_i and hence $Y_i = M_i \cdot Y_{i1} + (1 - M_i) \cdot Y_{i0}$. Accordingly, let $E\left[Y_{i1} \middle| M_i = 1\right]$ and $E\left[Y_{i0} \middle| M_i = 0\right]$ denote average outcomes of the technological performances of merged and non-merged firms, respectively. The effect we are interested in is that of merger on the technological performance of the merged firms, or the difference between the expected innovative performances of the merged firms and the firms that would have experienced if they did not merge:

$$\tau\Big|_{M_i=1} = E\left[Y_{i1} \middle| M_i = 1\right] - E\left[Y_{i0} \middle| M_i = 1\right] \tag{3}$$

This denotes the expected treatment effect on the treated. Since we do not have the counterfactual evidence of what would have happened if a firm had not engaged in a merger, $E\left[Y_{i0} \middle| M_i = 1\right]$ is unobservable. However, it can be estimated by $E\left[Y_{i0} \middle| M_i = 0\right]$ and the effect can be then given by the difference in the average outcome between the merged and non-merged innovative performances:

$$\tau^e = E\left[Y_{i1} \middle| M_i = 1\right] - E\left[Y_{i0} \middle| M_i = 0\right] \tag{4}$$

In fact, we have observations on the firms which did not engage in a merger, but if the merged and the non-merged firms systematically differ in their firm characteristics, (4) will be a biased estimator of (3) (Hirano et al., 2002).[16]

Rubin (1997), Rosenbaum and Rubin (1983, 1984) showed that a propensity score analysis of observational data can be used to create groups of treated and control units that have similar characteristics, whereby comparisons can be made within these matched groups. In these groups, there are firms that have been merged and firms that have not been merged; hence, the allocation of the merger can be considered to be random inside the groups of firms.

The merger propensity score is defined as the conditional probability of engaging in a merger given a set of observed covariates X_i:

$$p\left(M_i\right) = \Pr\left(M_i = 1 \middle| X_i\right) = E\left[M_i \middle| X_i\right] \tag{5}$$

The propensity score is a balancing score, meaning that conditional on the propensity score the distributions of the observed covariates are independent of the binary treatment assignment (Rosenbaum and Rubin, 1983; 1984). In other words, the propensity score matching relies on the "strong ignorability" assumption, which implies that for common values of covariates, the choice of treatment is not based on the benefits of alternative treatments.

The treatment effect of a merger is then estimated as the expectation of the conditional effects over the distribution of the propensity score in the merged sample:

$$\tau\Big|_{M_i=1} = E_{p(M_i)}\left\{E\left[Y_{i1}\middle|p\left(M_i\right), M_i = 1\right] - E\left[Y_{i0}\middle|p\left(M_i\right), M_i = 0\right]\middle|M_i = 1\right\} \tag{6}$$

We estimated the propensity scores by applying a multinomial logit model from the first stage of our analysis. Hereby, we include the determinants of technological performance as well as their interaction terms. Using the computed propensity scores, we sub-classify the sample into five strata according to propensity score quintiles.[17] As we are interested in estimating the effects of mergers and because there are fewer merged firms than firms which did not engage in mergers, we create strata based on the estimated propensity scores for merger events, so that each stratum contains an equal number of mergers.[18] This ensures an adequate number of mergers in each stratum. To check for the adequacy of the propensity score model, some

[16] Descriptive data in Table 2 show that merged and non-merged firms in our sample suggest significant differences in the observed characteristics.

[17] Five sub-classes (quintiles) constructed from the propensity scores will often suffice to remove over 90% of the selection bias due to each of the covariates (Rosenbaum and Rubin, 1984).

[18] In general, strata boundaries can be based on the values of the propensity scores for both groups combined or for the treated or control group alone (D'Agostino, 1998).

analyses were conducted to assess the covariate balance in the five groups of matched merged and non-merged firms that will be discussed in the next section.

Using the propensity score subclassification, we can estimate the effects of a merger on innovation performance by taking the weighted average (by number of merged firms) of the within-strata average differences in performance outcomes between merged and non-merged firms. This is the average treatment effect on the treated referred to in the causal inference literature. The variance of the average treatment effect estimate is calculated by dividing the sample variances of the performance outcomes for merged and non-merged firms within each strata by the number of merged and non-merged firms within that strata, respectively; and then by averaging these across all five strata. Because the subclassification is based on propensity scores estimated from the data and therefore depends on the sample, the outcomes within each strata and between the strata are not independent (Du, 1998). However, one can make the "strong ignorability" assumption less restrictive by incorporating a wider array of firm characteristics in the analysis (Benjamin, 2003).

6 Empirical Results

6.1 Technological Determinants of a Merger

In this section, we examine the merger decision of the telecommunications equipment firms in a multivariate analysis. Given that both merging and non-merging firms are included in the sample, we can attempt to distinguish between the characteristics of merging firms in transaction events and the firms outside of the merger market. We estimate equation (2) using a multinomial logit model with four outcomes: to be an acquirer, to be acquired, to be a pooling merger, or to be not involved in a merger. There are substantial drawbacks associated with the use of the multinomial logit estimation because it assumes that the disturbances are independent across alternatives. This assumption suggests that if a firm was choosing between the four alternatives, then there is no relationship between a firm's disturbances for being an acquirer, a target, a partner in a pooling merger or does not involvement in a merger. The test of the maintained assumption of independence of irrelevant alternatives (IIA) will indicate whether the ratio of the choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives. In the context of this analysis, it is likely that merger behavior will not fulfill this requirement. In order to examine how the estimation results are affected by this property, four Hausman tests were conducted. The multinomial logit results are compared with those from a binomial logits between the non-merged firms sample and each of the samples of acquiring, acquired and pooling merged firms as well as between acquirer and pooling merger samples. The p-values associated with the resulting test statistics were .88, .93, .76, and .67, respectively. Therefore, the null hypotheses are not rejected each, which implies that the IIA assumption does not adversely affect the estimates. Furthermore, the results of the binomial logit regressions were almost identical to those of multinomial logit model. This also substantiates that the independence assumption is not a concern of our analysis, and we can utilize robust estimates of the variance of the estimated coefficients.[19]

Table 3 presents the marginal effects for the multinomial logit regressions. The statistics for the joint hypothesis and likelihood ratio tests are also reported. All estimated models are highly significant as indicated by the likelihood ratio tests of the null hypothesis that the slope coefficients are jointly zero, which are rejected at the 1 percent level using the chi-square test statistic.

INSERT TABLE 3 ABOUT HERE

Merging firms as a whole seem to have, on average, a significantly different innovative profile compared to that of non-merging firms. Larger firms, as measured by the book value of total assets, are more likely to engage in merger activity. This suggests that large firms are more willing to make use of their large and more stable internal funds to finance external R&D projects. A 100 percent increase in a firm's total assets is associated with a .0026 and .0005 percentage point increase in the likelihood of acquiring another firm and being involved in a pooling merger, respectively, which is a .37 and a 1.67 percent increase in each probability.

The significantly positive effect of the cash flow ratio on the likelihood to acquire another firm suggests that acquiring firms have considerable cash to run a larger firm and agency controls are imperfect. This is in accordance with the evidence that possessing the ability to finance a merger tends to precipitate acquisitions. Firms with a relatively low cash flow ratio tend not to engage in a merger due to their financial constraints. Thus, either imperfect agency concerns or availability of financing are significant constraints on acquisitions.

In the current sample, we do not find any statistically significant relationship between the variables confirming the growth opportunities of firms, which are growth in market value and Tobin's q, and the probability that a firm is engaged in a merger. This indicates that a transaction is likely to be financed with cash rather than with equity.

[19] Of course, alternative models could be equally well used. For instance, the nested logit model would partly overcome the limitation of the multinomial logit model (Greene, 2003). However, we do not specify a nested logit model as it requires that the choices at a given level are mutually exclusive. This is clearly not the case in our study, and we, therefore, remain with the multinomial logit model.

When we take the proposed determinants on innovative performance of the firms into account, then, at first, the merging firms are more likely to have a large accumulated citation-based patent stock. This evidence seems to be in accordance with the theoretical argument that a large stock of accumulated knowledge is essential if the acquirer (or one partner in a pooling merger) is to have the necessary absorptive capacity to identify the appropriate target (or another partner in a pooling merger). The fact that firms with a rather low accumulated knowledge stock are less likely to engage in a merger supports this evidence.

Next, firms with greater R&D and citation-based patent intensities have a greater propensity to undertake acquisitions. These results seem to mutually support the hypothesis that higher levels of relative absorptive capacity and the strengthening of its creation on the part of research-focused firms are necessary for those firms to incorporate and exploit new research into their R&D programs effectively. We also obtain a significantly negative coefficient of the dummy for acquiring firms when R&D expenditure is not reported, which are expected to have zero or low R&D intensity. Therefore, the acquirer are more likely to have non zero R&D input in the year before the merger. At the same time, the non-merging firms tend to have more frequent zero R&D intensity than merging firms.

After controlling for R&D and citation-based patent intensities, we find that the likelihood of becoming an acquirer is higher with the lower R&D productivity of firms. Although the acquiring firms experienced higher input and output in R&D, they seem to carry either a low number of patents and/or a relatively low-valued patents yield of R&D dollars before acquisitions. As mentioned above, large firms are often argued to have a lower R&D productivity than that of their somewhat smaller rivals because research conducted in most large laboratories is found to generate predominantly minor improvement inventions rather than new major inventions. This result suggests that an enhanced desire to acquire new technology and innovation-related assets driven by declining returns from the exploitation of the firms' existing knowledge base exists. At this step of the analysis, we are yet cautious about this indication, since the target probability regression provides insignificant results on marginal effects. The lack of preciseness in the target estimation may due to the fact that the probability of being acquired greatly varies among the small sample of target firms. We will come back to this point as some predications regarding the target firms' pre-merger performance can be derived from the next step of our analysis.

An interesting result is that firms with a poor accumulated citation-weighted patent stock and, at the same time, presenting higher R&D productivity tend to not engage in a merger. We ascribe these firms to be relatively young and with significantly new know-how. The negative effect of firm size on the propensity to stay outside of the merger activity also seems to point toward that direction. Moreover, the coefficient estimates of the multinomial logit model, which are not reported here, indicate that acquisition targets possessed a significantly large accumulated knowledge stock than the non-merged firms.

Finally, firms that experienced a low R&D output are more likely to be involved in a pooling, suggesting that the lack of innovation is an important driving force behind the merger activity. There is no significant relationship between R&D productivity and the propensity to go through a pooling merger that would further confirm this evidence.

6.2 Post-Merger Innovation Performance

The full impact of mergers on the innovation performance takes time and results may not be evident immediately. In order to capture the long-run post-merger performance, we examine the impact of a merger in year *t* on the change in outcomes from $t+1$ to $t+2$, $t+2$ to $t+3$ and $t+3$ to $t+4$.[20] The outcomes of the firms' innovation performance are defined as the annual growth rates of the innovation determinants, e.g., we analyze the post-merger annual percentage changes of innovation input and output, knowledge stock and research productivity. In order to derive the merger propensity score, we estimated the multinomial logit model of equation (2) with annual percentage changes of the innovation and financials variables used in our first step of analysis as well as their interaction terms. The sample shows a good overlap in the estimated propensities scores for merged and non-merged firms verifying that there are comparable firms that did and did not enter into a merger. As mentioned above, data in the region of propensity score overlap were subclassified into five blocks defined by the quintiles of the propensity scores for merged firms. We then used a two-way ANOVA to assess whether the propensity score balances each covariate between the merged and non-merged groups of firms. Each covariate is regressed on the merger and the propensity score stratum indicator and their interaction as factors. The insignificant effects of mergers and insignificant effects of the interaction between propensity score stratum and merger indicators determine that the distributions of the covariates within the sub-classes are the same for merged and non-merged firms.[21] The balance in covariates of merged and non-merged firms assures an unbiased estimate of the effect of a merger on the innovation performance (Dehejia and Wahba, 1990).

Table 4 reports our findings on the effects of mergers on innovation performance. The impact of mergers appears to be more concentrated in the first year following a merger. Herein, stronger results are obtained for our main variables which more strictly explain the firm's innovation performance.

[20] We cannot compare pre- and post-merger performance of merged firms with the matched sample of non-merging firms over the same time period because we lack pre-merger accounting data for one component of the merged entity for a significant fraction of our mergers.
[21] Before subclassification, we found using one-way ANOVA significant effects of mergers on more covariates.

First, the annual percentage change of R&D intensity displays a significantly positive sign in all three years following a merger. Hence, according to our previous result from the first stage of the analysis, this indicates that the strong R&D intensity of acquiring firms positively influences the assimilation of the external knowledge by supplementing in-house R&D effort. Moreover, it suggests that the firms engaged in the mergers did not depreciate their investments in R&D on behalf of financing the transaction.

INSERT TABLE 4 ABOUT HERE

Next, we find that mergers are followed by an improvement in the accumulated citation-based patent stock. In addition to the partners in a pooling merger, who possessed a large accumulated knowledge stock prior a merger, the targets also tend to be firms with highly valued patent stock. This result is in accordance with our prediction that accumulated knowledge stock confers an ability to recognize the new knowledge in environment and this ability seem to enhance the technological strengths even further.

The merged firms experience a significantly positive impact on the (citation-based) patent intensity compared to those outcomes that these firms would have reached if they had not been merged. Due to the fact that the acquiring firms had a higher citation-based patent intensity prior acquisitions, this effect suggests that an intensification of high-valued patents creation relative to the firm's assets base prior an acquisition generates a significantly high innovation output of the merged entity. Additionally, the pooling partners which faced some absence of innovation efficiency in terms of the innovation output seem to grow following a merger, potentially because the merger provided access to technological resources which the firms previously lacked.

Furthermore, the insignificant result on the post-merger research productivity suggests that the marginal returns from R&D investments do not change with respect to the innovation output. At the same time, merged and non-merged firms do not significantly differ in their financial characteristics such as cash flow ratio and Tobin's q, at least for the observation period.

Finally, we find a significant increase in the following variables reflecting the firms' economic performance. Firstly, there is a firm's size growth effect with respect to the annual percentage changes in the total assets as typically expected. Secondly, the positively significant increase in the annual growth of the market value on average confirms that, in the first year following the mergers, overall returns for shareholders are above those of the non-merged firms with similar characteristics.

7 Conclusions

This paper delivers insights into the desirability of M&A for the innovation performance of firms by analyzing the mergers that took place in the international telecommunications equipment industry from the late 1980s until the early 2000s. We provide evidence on strictly complementary as well as mutually supportive results. The overwhelming conclusion that arises from the analysis is that, on average, mergers realize significantly positive changes to the innovation performance of firms following a merger. The post-merger changes are in turn driven by both the success in R&D activity and the weakness in internal technological capabilities at acquiring firms prior to a merger.

The findings about the innovation-related characteristics of the merging firms have interesting implications for the propositions about the rationale of mergers set out in our theoretical section. According to the absorptive capacity theory (Cohen and Levinthal, 1989; Chesbrough, 2003), firms with greater R&D intensity and a larger stock of accumulated knowledge have a greater propensity to engage in the technological-related mergers, and these underlying higher levels of absorptive capacity convincingly indicate the necessity for the identification, the assimilation and the exploitation of the targets' technological knowledge. The analysis provides strong evidence, according to which firms with rapid R&D and firms that extensively apply the results of scientific advances to their own R&D results, e.g., inventions, acquire better quality patents.

We find support for the view that firms experiencing a decline in internal research productivity or which are more inefficient in inventive output are more likely to engage in an acquisition or a pooling merger, respectively, as an effort to boost their research pipelines. In effect, firms which face greater distress in the effectiveness of patenting activity appear to grow their invention intensity following a merger, which is probably because the merger provided a rectified access to the appropriate technological resources. However, increased patenting may not be directly resulting from higher R&D intensity and accordingly not related to protecting investment in R&D (Kortum and Lerner, 1998; Hall, 2001). Particularly, the increased (citation-weighted) patent intensity following a merger could stem from technological and managerial improvements. That is, the merged firms had redirected more of their R&D investments toward applied rather than basic research and/or improved their innovation management. As the mergers have, on average, an insignificant effect on the research productivity of the merged firms, this interpretation is quite tentative. Nevertheless, an increase in the inventive output intensity of the merged firms following a merger is suggestive of improvements to the underlying research portfolio as this is a direct measure of a firm's innovation performance.

Furthermore, we find that larger firms with strong internal funds to finance R&D are more likely to acquire and to engage in a pooling merger, whereas the firms which lack these characteristics are more likely to pursue technology internally. Solely relying on in-house R&D, non-merged firms are appear to be rather young and small market challengers, which are striving to rival the market establisher with a significantly new and/or advanced technology on their own. Contrary to these firms, the acquired firms seem to be experienced entrepreneurs that have succeeded in the past at generating larger and high-valued inventions.

The analysis reveals that mergers are, on average, a positive experience for shareholders, at least for a short-time span. Moreover, the finding that, in the long-run, mergers did not cut R&D spending suggests that post-merger R&D effort is not affected by financial resource constraints induced by the transaction and integration processes.

With respect to the average effects of mergers, the analysis has clearly shown that the merged firms faced different outcomes regarding the post-merger innovation performance. One potential explanation of the variability in the performances might be due to different financing of the mergers transactions. The decision on merger financing has important implications for merger capital structure, future profitability, subsequent financing choices and ownership structure. Therefore, it might have a significant impact on the R&D performance of a merger too. How and to what extend the merger financing choice affects firms' post-merger innovation deserves further investigations in future research.

References

Ahuja, G., and R. Katila (2001). Technological acquisitions and the innovation performance of acquiring firms: a longitudinal study, *Strategic Management Journal*, 22, pp. 197-220

Andrade, G., M. Mitchell, and E. Stafford (2001). New Evidence and Perspectives on Mergers, *Journal of Economics Perspectives*, 15, pp. 103-120

Benjamin, D.J. (2003). Does 401(k) eligibility increase saving? Evidence from propensity score subclassification. *Journal of Public Economics*, 87, pp. 1259-1290

Bresman, H., J. Birkinshaw and R. Nobel (1999). Knowledge transfer in international acquisitions, *Journal of International Business Studies*, 30, pp. 439-462

Brodt, T. and S. Knoll (2004). Innovation-Expansion In The ICT-Industries: R&D Acquisition and Integration Success From A Knowledge Perspective, mimeo

Brouwer, E. and A.H. Kleinknecht (1999). Innovative Output, and a Firm's Propensity to Patents, *Research Policy*, 28, pp. 439-462

Capron, L. and W. Mitchell (1997). Outcomes of International Telecommunications Acquisitions: Analysis of Four Cases with Implications for Acquisitions Theory, *European Management Journal*, 15 (3), pp. 237-251

Caves, R.E. (1982). Multinational Enterprise and Economic Analysis, *Cambridge Surveys of Economic Literature*

Chakrabarti, A, J. Hauschildt and C. Sueverkruep (1994). Does it pay to acquire technological firms?, *R&D Management*, 24, pp. 47-56

Chesbrough, H.W. (2003). Open Innovation: The New Imperative for Creating and Profiting from Technology, Boston, Harvard Business School Press

Christensen, C.M. (1997). The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail, McGraw-Hill Professional

Clark, K. and E. Ofek (1994). Mergers as a Means of Restructuring Distressed Firms: An Empirical Investigation, *The Journal of Financial and Quantitative Analysis*, 29 (4) pp. 541-565

Coase, R.H. (1937). "The nature of the firm", *Economica*

Cohen, W.M. and D.A. Levinthal (1989). Innovation and Learning: The Two Faces of R&D, *The Economic Journal*, 99, pp.569-596

Cohen, W.M. and D.A. Levinthal (1990). Absorptive capacity: a new perspective on learning and innovation, *Administrative Sciences Quarterly*, 35, pp. 569-596

D'Agostino, R.B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group, *Statistics in Medicine*, 17, pp. 2265-2281

Danzon, P.M., A. Epstein and S. Nicholson (2004). Mergers and Acquisitions in the Pharmaceutical and Biotech Industries, *NBER Working Paper*, 10536

Dehejia, R.H. and S. Wahba (2002). Propensity Score-Matching Methods For Nonexperimental Causal Studies, *The Review of Economics and Statistics*, 84(1), pp. 151-161

Di Minin, A. and C. Palmberg (2006). A case for non-globalization? The organization of R&D in the wireless telecommunications industry, *ETLA Discussion Paper*, 1047

Dosi, G. (1988). Sources, Procedures, and Microeconomic Effects of Innovation, *Journal of Economic Literature*, 26 (3), pp. 1120-1171

Du, J. (1998). Valid inferences after propensity score subclassification using maximum number of subclasses as building blocks, Ph.D. thesis, Harvard University

Duysters, G. and J. Hagedoorn (1998). Technological Convergence in the IT Industry: The Role of Strategic Technology Alliances and Technological Competencies, *Economic of Business*, 5(3), pp. 355-368

Duysters, G. and J. Hagedoorn (2000). Organizational Modes of Strategic Technology Partnering, *Journal of Scientific and Industrial Research*, 59, pp. 640-649.

Gerpott, T.J. (1995). Successful integration of R&D functions after acquisition: an exploratory empirical study, *R&D Management*, 25, pp. 161-178

Grandstrand, O., E. Bohlin, C. Oskarsson and N. Sjoberg (1992). External technology acquisition in large multi-technology corporations, *R&D Management*, 22, pp. 111-133

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature*, 28, pp. 1661-1697

Gugler, K., D.C. Mueller, B.B. Yurtoglu and C. Zulehner (2003). The Effects of Mergers: An International Comparison, *International Journal of Industrial Organization*, 21, pp. 625-653

Hall, B.H. (1990) The impact of corporate restructuring on industrial research and development, *Brookings Papers on Economic Activity*, 3, pp. 85-135

Hall, B. H. (1999) Mergers and R&D revisited, Paper presented at the Quasi-Experimental Methods Symposium, Econometrics Laboratory, UC Berkeley

Hall, B.H., M. Trajtenberg and A.B. Jaffe (2001). The NBER Patent Citation Data File: Lessons, Insight and Methodological Tools, *NBER Working Paper*, 8498

Hall, B.H. (2004) Exploring the Patent Explosion, *NBER Working Paper*, 10605

Hart, O. and J. Moore (1990). Property Rights and the Nature of the Firm, *Journal of Political Economy*, 98 (6), pp. 1119-58

Hart, P.D.S. and B. Ramanantsoa (1992). Strategic Technology Management, John Wiley&Sons

Hennart, J.-F. and S. Reddy (1997). The Choice between Mergers/Acquisitions and Joint Ventures: The Case of Japanese Investors in the United States, *Strategic Management Journal*, 18, pp. 1–12

Hirano, K., G.W. Imbens and G. Ridder (2002). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score, *Econometrica*, 71 (4), pp. 1161-1189

Hitt, M., R. Hoskisson and D. Ireland (1990). Mergers and Acquisitions and Managerial Commitment to Innovation in M-Form Firms, *Strategic Management Journal*, 11, pp. 29-47

Hitt, M.A., R.E. Hoskisson, R.A. Johnson and D.D. Moesel (1996). The Market for Corporate Control and Firm Innovation, *Academy of Management Journal*, 39 (5), pp. 1084-1119

Hundley, G., C.K. Jacobson and S.H. Par (1996). Effects of Profitability and Liquidity on R&D Intensity: Japanese and U.S. Companies Compared, *The Academy of Management Journal*, 39 (6), pp. 1659-1674

Jamison, M.A. (1998). Emerging Patterns in Global Telecommunications Alliances and Mergers, *Industrial and Corporate Change*, 7(4), pp. 695-713

Jovanovic, B. and P.L. Rousseau (2004). Mergers as Reallocation, *NBER Working Paper*, 9279

Kim, J.K. (2005). Telecommunications Merger Trends in the Context of the Convergence – Using the U.S. Merger Cases, *mimeo*

Kogut, B. and U. Zander (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology, *Organization Science*, 3 (3), pp. 383-397

Kortum, S. and Lerner, J. (1998). Stronger Protections on Technological Revolution: What is Behind the Recent Surge in Patenting?, *Research Policy*, 28, pp. 1-22

Kretschmer, K. (1997). Communication Standards and Patent Rights: Conflict or Coordination?, *Standards and Technology Annual Report*

Kusewitt, J.B. (1985). An Exploratory Study of Strategic Acquisition Factors Relating to Performance**,** *Strategic Management Journal*, 6 (2), pp. 151-169

Lane, P.J. and M. Lubatkin (1998). Relative absorptive capacity and interorganizational learning, *Strategic Management Journal*, 19, pp. 461-477

Larsson, R., L. Bengtsson, K. Henriksson and J. Sparks (1998). The interorganizational learning dilemma: Collective knowledge development in strategic alliances, *Organization Science*, 9, 285-305

Le Blanc, G. and H. Shelanski (2002). Merger Control and Remedies Policy in the E.U and U.S: the case of Telecommunications Mergers, *mimeo*

Li, F. and J. Whalley (2002). Deconstruction of the telecommunications industry: from value chains to value networks, *Telecommunications Policy*, 26(9/10), pp. 451-472.

Loderer, G. and H. Martin (1992). Postacquisition Performance of Acquiring Firms, *Financial Management*, 21(3)

Lubatkin, M. (1983). Mergers and the Performance of the Acquiring Firm, *The Academy of Management Review*, 8 (2), pp. 218-225

March, J.G. (1991). Exploration and exploitation in organizational learning, *Organization Science*, 2, pp. 71–87

Mork, R., A. Shleifer and R. Vishny (1990). Do managerial objective drive bad acquisitions? *Journal of Finance*, 45, pp. 31-48

Ornaghi, C. (2006). Mergers and Innovation: the Case of the Pharmaceutical Industry, *mimeo*

Powell, R.G. (1997). Modeling Takeover Likelihood, *Journal of Business Finance & Accounting*, 24 (7/8), pp. 1009-1030

Quah, D. (2000). Demand-driven knowledge clusters in a weightless economy, *mimeo*

Roeller, L.-H., J. Stenneck and F. Verboven (2001). Efficiency Gains from Mergers, *European Economy*, 5, pp. 31-128

Rosenbaum, P.R. and D.B. Rubin (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, pp. 41-55

Rosenbaum, P.R. and D.B. Rubin (1984). Reducing bias in observational studies using subclassification on the propensity score, *Journal of the American Statistical Association*, 79, pp. 516-524.

Rosenberg, E. (1997). Telecommunications Mergers and Acquisitions: Key Policy Issues and Options for State Regulators, The National Regulatory Research Institute

Rosenberg, N. (1990). Why Do Firms Do Basic Research (With They Own Money)?, *Research Policy*, 19, pp.165-174

Rubin, D.B. (1997). Estimating causal effects from large data sets using propensity scores, *Annals of Internal Medicine*, 127, pp. 757-763

Seth, A. (1990). Sources of Value Creation in Acquisitions: An Empirical Investigation**,** *Strategic Management Journal*, 11 (6), pp. 431-446

Shimizu, K., Hitt, M., Vaidyanath, D., and Pisano, V. (2004). Theoretical Foundations of Cross-Border Mergers and Acquisitions: A review of Current Research and Recommendations for the Future, *Journal of International Management*, 10, pp. 307–353

Sorenson, O. and J. B. Sorensen (2001). Finding the Right Mix: Franchising, Organizational Learning, and Chain Performance, *Strategic Management Journal*, 22, pp. 713 – 724.

Tadayoni, R. and K.E. Skouby (1999). Terrestrial digital broadcasting: convergence and its regulatory implications, *Telecommunications Policy*, 23 (2), pp. 175-199

Teece, D.J. (1996). Firm organization, industrial structure, and technological innovation, *Journal of Economic Behavior and Organization*, 31 (2), pp. 193-224

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations**,** *The RAND Journal of Economics*, 21 (1), pp. 172-187

Vermeulen, F. and H. Barkema (2001). Learning through Acquisitions, *Academy of Management Journal*, 44(3), pp. 457-476

Very, P., M. Lubatkin, R. Calori, and J. Veiga (1997). Relative Standing and the Performance of Recent European Mergers, *Strategic Management. Journal*, 18, pp. 593–615

Veugelers R. (2006). M&A and Innovation: a Literature Review, In Cassiman, B. and M.Colombo (eds), *Mergers and Acquisitions: The innovation impact*, Edward Elgar Cheltenham, pp. 37 - 62

Warf, B. (2003). Mergers and Acquisitions in the Telecommunications Industry, *Growth and Change*, 34 (3), pp. 321–344

Warner, A.G., J.F. Fairbank and H.K. Steensma (2006). Managing Uncertainty in a Formal Standards-Based Industry: A Real Options Perspective on Acquisition Timing, *Journal of Management*, 32 (2), pp. 279-298

Williamson, O.E. (1981). The Economics of Organization: The Transaction Cost Approach, *American Journal of Sociology*, 87 (3)

Wooldridge, J.M. (2002). Econometric Analysis of Cross Section and Panel Data, MIT Press

Table 1. Sample Statistics (n = 9,570 firm-years)

Variable Description	Mean	Standard Deviation
R&D Intensity	0.115	0.336
(Citation-weighted) Patent Intensity	0.019	0.097
(Citation-weighted) Patent Stock (Ln)	1.441	1.504
(Citation-weighted) Patent Productivity	0.237	1.194
Total Assets (Ln)	4.001	2.120
Annual Growth of Market Value (Ln)	1.519	3.236
Tobin's *Q*	2.091	3.259
Cash-Flow Ratio	-0.162	1.460
Indicator for Missing R&D Expenses	0.171	0.376
Indicator for Zero (Citation-weighted) Patent Intensity	0.512	0.500

Notes: The figures refer to the sample used for the estimation of the multinomial logit model (Table 3).

Table 2. Merging versus Non-Merging Firms before Matching

Variables	Mean (Standard Error) Merged Firms	Mean (Standard Error) Non-Merged Firms	t-statistic for difference in means
R&D Intensity	0.105 (0.005)	0.115 (0.005)	0.48
(Citation-weighted) Patent Intensity	0.014 (0.003)	0.02 (0.001)	0.85
(Citation-weighted) Patent Stock (Ln)	2.327 (0.152)	1.378 (0.029)	8.00***
(Citation-weighted) Patent Productivity	0.214 (0.067)	0.238 (0.024)	0.27
Total Assets (Ln)	5.344 (0.153)	3.914 (0.031)	10.89***
Annual Growth of Market Value (Ln)	1.410 (0.081)	1.611 (0.06)	0.52
Tobin's *Q*	2.476 (0.158)	2.037 (0.057)	2.01**
Cash-Flow Ratio	0.019 (0.016)	-0.174 (0.023)	2.10**
Indicator for Missing R&D Expenses	0.100 (0.018)	0.175 (0.005)	3.20***
Indicator for Zero (Citation-weighted) Patent Intensity	0.455 (0.030)	0.515 (0.007)	1.90*

Notes: Standard errors are given in parentheses. ***, ** and * difference in sample means is significantly different from zero at the 1%, 5% and 10% statistical level, respectively.

Table 3. Marginal Effects of the Propensity of Involvement in M&A Activity

	Acquirer	Target	Pooling Merger	No M&A
R&D Intensity	0.34e-02***	-0.62e-05	-0.20e-02	-0.13e-02
	(0.11e-02)	(0.34e-04)	(0.13e-02)	(0.17e-02)
(Citation-weighted) Patent Intensity	0.76e-05***	-0.45e-07	-0.39e-05**	-0.37e-05
	(0.28e-05)	(0.17e-06)	(0.21e-05)	(0.35e-05)
(Citation-weighted) Patent Stock (Ln)	0.34e-05**	0.20e-07	0.35e-05***	-0.70e-05***
	(0.16e-05)	(0.72e-07)	(0.11e-05)	(0.20e-05)
(Citation-weighted) Patent Productivity	-0.79e-05***	0.67e-07	0.15e-05	0.62e-05*
	(0.27e-05)	(0.23e-06)	(0.19e-05)	(0.34e-05)
Total Assets (Ln)	0.26e-02***	-0.54e-05	0.48e-03**	-0.31e-02***
	(0.41e-03)	(0.15e-04)	(0.19e-03)	(0.45e-03)
Annual Growth of Market Value (Ln)	-0.78e-06	-0.64e-08	0.88e-06	-0.89e-07
	(0.17e-05)	(0.28e-07)	(0.10e-05)	(0.20e-05)
Tobin's *Q*	0.36e-05	0.15e-06	0.82e-06	-0.46e-05
	(0.25e-05)	(0.58e-06)	(0.14e-05)	(0.30e-05)
Cash-Flow Ratio	0.17e-04***	0.53e-05	0.34e-05	-0.26e-04*
	(0.76e-05)	(0.14e-04)	(0.27e-05)	(0.16e-04)
Indicator for Missing R&D expenses	-0.44e-02***	0.16e-04	0.37e-03	0.40e-02**
	(0.17e-02)	(0.59e-04)	(0.11e-02)	(0.21e-02)
Indicator for Zero (Citation-weighted) Patent Intensity	-0.17e-02	-0.10e-04	0.12e-02	0.51e-03
	(0.14e-02)	(0.36e-04)	(0.84e-03)	(0.17e-02)
Mean of Dependent Variable (Percentage Points)	0.70	0.00	0.30	99.00
Observations	217	25	122	9,206
Log Likelihood	-1,350.60			
Restricted Log Likelihood	-1,590.54			
Prob > ChiSqd	0.00			

Notes: The marginal effects provide percentage point changes in the probability of an outcome. Standard errors are given in parentheses. Marginal effects are computed at means of explanatory variables. ***, ** and * indicate a significance level of 1%, 5% and 10%, respectively.

Table 4. Effects of M&A (Average Treatment Effects on the Treated)

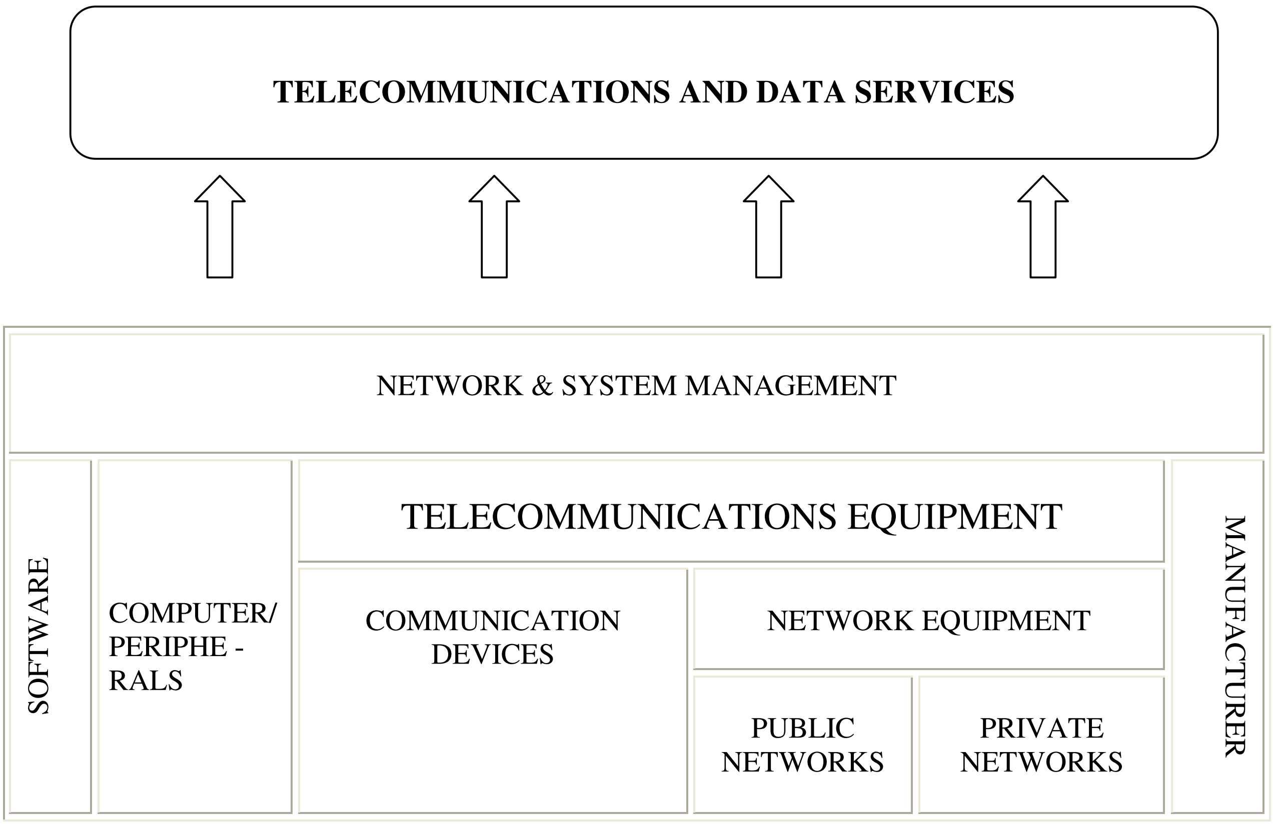
	First year ***(t+1 to t+2)***	***Second year*** ***(t+2 to t+3)***	***Third year*** ***(t+3 to t+4)***
R&D Intensity	0.139*** (0.045)	0.193*** (0.052)	0.228*** (0.039)
(Citation-weighted) Patent Intensity	0.083*** (0.004)	-0.113 (0.152)	-0.051 (0.436)
(Citation-weighted) Patent Stock (Ln)	0.0046*** (0.017)	0.004 (0.024)	0.018 (0.025)
(Citation-weighted) Patent Productivity	0.816 (0.626)	-0.006 (0.589)	0.238 (0.315)
Total Assets (Ln)	0.052** (0.026)	0.041 (0.026)	0.040 (0.028)
Annual Growth of Market Value (Ln)	0.338*** (0.103)	-0.027 (0.197)	0.124 (0.146)
Tobin's *Q*	2.500 (2.920)	2.076 (2.053)	1.694 (1.642)
Cash-Flow Ratio	-0.031 (0.874)	1.002 (3.016)	-1.052 (2.096)

Notes: Reported are means. Standard errors are given in parentheses. ***, ** and * indicate a significance level of 1%, 5% and 10%, respectively.

Appendix: Telecommunications Equipment Industry

The telecommunications equipment industry can be segmented into different sub-markets according to the nature of the user: communication devices, public and enterprise network equipment and system and network management. Communication devices encompass all equipment used by private and business customers to deploy voice and data telecommunications services. Public and enterprise network equipment comprises all equipment used by carriers and enterprises to provide voice and data network services. enterprise network equipment includes advanced PBXs (Private Branch exchanges) and key systems, LAN (Local Area Network) equipment such as hubs and interfaces and other equipment such as modems. Public network equipment comprises transmission equipment, packet and circuit switching equipment, routing equipment and cellular mobile radio infrastructure. System and network management encompasses all system infrastructure software. This software is used to manage the multitude of computing resources in a company, coordinate resources between servers and nodes on a network and operate the hardware platforms and communication networks.

Figure 4. Structure of the Telecommunications Equipment Industry



International Trade Policies and Agreements on Compatibility Standards for Industries with Network Externalities

Mikhail M. Klimenko

School of Economics
Georgia Institute of Technology
781 Marietta Street, NW, Atlanta, GA 30332-5750
mikhail.klimenko@econ.gatech.edu

August 1, 2006

Abstract

The paper considers a country (home) in which domestic and foreign firms produce partially incompatible products while heterogeneous consumers value both variety and a network externality. The presence of the network externality justifies home government intervention by means of a standard requiring the foreign firm to guarantee a minimum level of compatibility between its own product and the product of the domestic rival. The paper shows that the exporting (i.e., foreign) government also has incentives for conducting a policy affecting the degree of compatibility between the exported and import-competing products in the home market. Based on the analysis of the governments' incentives, we examine the equilibrium outcome of the non-cooperative game in which both government use policies toward compatibility. We show that the strategic use of such policies by the countries always results in an inefficient combination of trade volume and compatibility level. The paper then analyzes international agreements on policies toward compatibility and evaluates the existing provisions in the WTO legal system aimed at minimizing the trade-inhibiting impact of standards and regulations in the area of technical compatibility.

JEL classification numbers: L5; F13

Keywords: Trade Policy; Trade Agreements; Technical Standards; Network Externalities.

1. Introduction

A growing number of international trade disputes concerning domestic standards and regulations reflect declining importance of traditional trade barriers such as tariffs and quotas and the increasing reliance of governments on more subtle instruments of protection against foreign competition. The rising importance of standards in commercial policy matters led to active discussion of their trade-inhibiting implications in the analytical and empirical trade literature.[1] However, the formal economic analysis of standards as tools of trade policy has been focused primarily on quality, environmental, and labor standards. Relatively little attention has been given to the trade policy role of standards and regulations that ensure technical compatibility (or interoperability) among goods and services the consumption of which generates demand-side scale economies or network externalities. The trade policy implications of such technical compatibility standards and regulations are the main focus of this paper.

The international disciplines governing technical compatibility regulations and standards for goods and services are contained in the original General Agreement on Tariffs and Trade (GATT), as well as in the Agreement on Technical Barriers to Trade (TBT) and the General Agreement on Trade in Services (GATS), both of which were concluded during the Uruguay Round of multilateral trade negotiations. While all of these agreements prohibit the discriminatory use of standards, they do not deny any country the right to use standards and regulations to pursue legitimate domestic regulatory objectives. One such legitimate objective is to maintain seamlessness and integrity of national information and communication technology (ICT) networks and infrastructure.[2] The economic theory justification for this objective is associated with the notion of the positive network externality, which is maximized when all products constituting the network are perfectly compatible. Perfect compatibility can be achieved by imposing the standard, which mandates the adoption of a uniform technical design by all producers. However, for many types of networks the desired degree of integrity and seamlessness can be achieved by means of less stringent compatibility standards, which allow heterogeneity among technical designs embodied in the elements of the network and specify an acceptable range of performance characteristics that have to be guaranteed during interoperation of these elements.[3]

---

[1] See, for example, Bagwell and Staiger (2001), Fischer and Serra (2000), Gandal and Shy (2001), and Maskus et al. (2001).

[2] Information technology has been making an important contribution to globalization and economic growth. The share of ICT equipment in world merchandise exports has increased from 9% in 1990 to 13% in 1998. (See OECD, 2001.) Special characteristics of trade in ICT equipment and services are recognized in the WTO Information Technology Agreement (ITA) concluded in 1997. The primary goal of that agreement is the reduction of customs duties to zero on a limited range of information, communication, and broadcasting technologies. One of the goals of the Doha Round of the multilateral trade negotiations is to conclude a follow-up agreement (ITA II), which will expand product coverage and create clear international disciplines for trade-related ICT regulations and standards.

[3] The Organization for Economic Co-operation and Development (OECD) distinguishes four types of these standards for the ICT industry: interoperability, public network integrity, electromagnetic compatibility (EMC), and interference avoidance standards. (See OECD, 2000a, 2001.) While the interoperability and public network integrity standards are directly aimed at improving interconnection among devices constituting the network, the EMC and interference avoidance standards help to improve compatibility by reducing the congestion and signal interference in shared communication channels.

Motivating examples and evidence

There are a large number of examples of technical compatibility standards imposed for legitimate domestic public policy reasons that have had trade-inhibiting effects. Imports of wireless communication equipment for the licensed spectrum bands are often hindered by strict requirements for interoperability between the imported equipment and the networks already licensed and deployed in the domestic market.[4] Imports of wireless equipment for the unlicensed spectrum bands are sometimes inhibited by strict spectrum sharing, interference avoidance and data security standards, which can be complied with only through costly modifications of the imported equipment.[5]

Another example concerns suppliers of direct broadcasting satellite services and equipment intended for high-definition television (HDTV) in the E.U. and the U.S. These suppliers often complain that the main impediment to their entry into these markets are the excessive technical regulatory requirements for the signal conversion to ensure that the HDTV programming can be sent and received between transmitters and receivers based on the European and the U.S. systems. (See European Commission, 2001.)

Finally, with the introduction of Europe's own satellite navigation system, Galileo, the U.S. manufacturers of equipment based on the Global Positioning System (GPS)—and the U.S. airlines using this equipment—fear that the E.U. regulators may impose excessively strict requirements for interoperability between the GPS receivers installed in the aircraft and the Galileo-based transmitters in the European airports. Compliance with these requirements would involve substantial additional costs for the U.S. companies and might erode their competitive positions in Europe. (See O'Neil, 2001.)

A number of sectoral case studies and firm level surveys conducted by the OECD documented the empirical importance of technical compatibility regulations as instruments of commercial policy in network industries. For example, OECD (2000a) presented results of a survey of 20 manufacturers of terminal

---

[4] The second- and third-generation digital wireless communication networks in Europe are based on the Global System for Mobile Communications (GSM) and the Universal Mobile Telecommunications System (UMTS), respectively, both of which are sponsored by a group of European telecommunication companies. When a U.S. company, Qualcomm, developed an alternative wireless communications system called Code Division Multiple Access (CDMA), the European Telecommunications Standards Institute (ETSI) responsible for setting the European Union's telecommunications standards did not explicitly ban CDMA networks but imposed prohibitively strict roaming requirements on wireless network operators wishing to use the CDMA technology. These roaming requirements essentially imply that in order to sell their equipment in Europe the CDMA equipment makers will have to supply fully integrated dual-mode (CDMA/GSM) wireless handsets and cell stations that can ensure perfectly smooth roaming between the networks based on the different systems. The high cost of compliance with these requirements has made the exporting of CDMA equipment to Europe economically infeasible. For a discussion of the trade-inhibiting effects of ETSI's standardization practices see Clarke (1999) and Grindley, Salant, and Waverman (1999).

[5] For wideband Wireless Local Area Networks (WLAN) operating in the 2.4 GHz frequency band, otherwise known as Wi-Fi networks, China has developed its own standard *GB 15629.11-2003*, which is different from the U.S. backed standard specified by the *802.11* committee of the Institute of Electrical and Electronics Engineers (IEEE). The main difference between the two standards is in terms of encryption and quality-of-service specifications. In order to comply with the encryption compatibility requirement for the Chinese market, foreign Wi-Fi chip makers and equipment manufactures have to implement costly modifications of their products in accordance with China's WLAN Authentication and Privacy Infrastructure standard (WAPI). (See Mannion and Clendenin, 2003.) The U.S. government claimed that the WAPI encryption requirement for Wi-Fi equipment imports violates China's WTO commitments with regard to national treatment and market access and threatened to file a complaint against China in the WTO. (See Clendenin, 2004.) For WLAN operating in the 5 GHz frequency band, the E.U. telecom regulator ETSI has developed the High Performance European Radio Local Area Network 2 (HiperLAN2) standard, while most of the U.S. equipment makers support the IEEE 802.11a standard. Both the European and the U.S. equipment vendors acknowledge that it is technically feasible albeit costly to achieve interoperability between the two systems through dual-mode 802.11a/HiperLAN2 solutions based on flexible digital signal processing architectures. What prevents them from selling in each other's markets are the excessively stringent interference avoidance and power emission requirements set by the regulators in the U.S. and the E.U. (See Wong, 2001, and McLean, 2001).

telecommunications equipment (TTE) in the United States, Japan, the United Kingdom, and Germany. The majority of surveyed telecommunications companies indicated that because of the need to adapt products to meet technical specifications related to interoperability, interference avoidance and public network safety in export markets their production costs were greater than for an equivalent domestic manufacturer in the export market. The estimates of additional production costs ranged from 5 to 10 per cent.[6] 13 out of 20 surveyed telecommunications equipment firms indicated that they had been prevented from exporting products due to the magnitude of costs of compliance with technical compatibility standards in the export markets. (See OECD 2000a, p. 93.)

Another OECD case study focused on trade implications of a specific type of regulation of electrotechnical sector -- electromagnetic compatibility (EMC) standards. (See OECD, 2000b.) EMC standard includes a specification for a method to prevent damage to electricity supply networks from unwanted low-frequency harmonic current emissions produced by electrical products. The study indicated that the cost to industry of enforcing the EMC standard throughout the world could exceed US$50 billion per year.[7]

Main contribution of the paper and the relevant previous literature

Inspired by these examples and empirical evidence, this paper considers an analytical framework combining elements of well-known models of trade under imperfect competition with the approaches taken in the industrial organization literature in analyzing the role of compatibility-enhancing devices in industries with network externalities.[8] Specifically, we analyze a home-market "half" of the reciprocal-markets model in which a domestic firm and a foreign firm supply imperfectly compatible products while heterogeneous consumers value both variety and compatibility. This framework allows us to highlight three important trade-related aspects of compatibility issues, which were not considered in the earlier international trade literature.

First, opening the domestic market to international trade brings foreign competition, which directly benefits the consumers by expanding the variety of available products and reducing the prices. However, the adoption of the foreign product by some domestic consumers undermines the integrity of the domestic network because the foreign product is not perfectly compatible with the domestic product, and the country loses the positive network externality associated with the single technology nationwide network.

---

[6] Gandal (2001) provides qualitative evidence that cellular communications equipment firms are likely to dominate in their own domestic markets in which foreign equipment makers are subject to very strict compatibility standards. He notes that 63 per cent of all mobile phones sold in the U.S. are produced domestically while the corresponding shares of the domestic producers in the U.S. markets for audio and video equipment are, respectively, 19 and 26 per cent. One of the reasons for such a difference in the domestic marker shares of the U.S. firms in these industries undoubtedly has to do with greater compatibility compliance costs that foreign importers incur in order to gain access to the U.S. cellular communications equipment market compared to these costs in audio and video equipment markets.

[7] High costs caused by standards-related trade barriers in the area of telecommunications terminal equipment were also reported in the recent OECD (2001) case study of telecommunications sector. The study points out that differences over approaches to technical compatibility regulations across countries arise from different evaluation of types of harmful effects of incompatibility and interference for telecommunications network.

[8] The effects of compatibility-enhancing devices (e.g., converters, adapters, gateways) on technology adoption in a closed economy with network externalities were discussed informally by Braunstein and White (1985), David and Bunn (1988), David and Greenstein (1990), and David and Steinmueller (1994). Formal analyses of the effect of converters on the outcome of the strategic interaction between the rival firms and technology adoption by users were conducted by Katz and Shapiro (1985, 1986), Berg (1988), Economides (1988, 1991), Farrell and Saloner (1992), and Choi (1997).

Second, although the loss of the network externality can be reduced by increasing compatibility between the imported and the import-competing products, any compatibility enhancement (short of ensuring complete compatibility) is insufficient to overcome a distortion in the product adoption by the users. This distortion arises because the users adopt their most preferred product without taking into account the effect of their adoption decisions on the relative size of the rival networks. Under oligopoly, this inefficiency is aggravated by the distortion caused by the imperfectly competitive pricing behavior of the firms. Therefore, through its effect on the sorting of the users into the rival networks, the government compatibility policy should balance the network-related component of the social welfare and the "stand-alone" component, the latter being defined by the user utility derived from the product characteristics independently of the network effect.

The third aspect has to do with the incentives to achieve greater compatibility among the firms competing internationally in the imperfectly competitive environment. When the compatibility-enhancing technology (i.e., a converter or an adapter) creates symmetric benefits for consumers of the rival products each of the rival firms has insufficient incentives for enhancing compatibility because some of the enhanced network benefits accrue to consumers and to the rival firm. What makes this problem trade-policy relevant is that as long as some of the cost of attaining greater compatibility is borne by the foreign firm, the home government has the incentive to choose inefficiently stringent (from the global welfare maximization perspective) regulation or standard for compatibility between the rival goods. This creates the possibility of a *regulatory sham* in trade policy—i.e., the practice of using a legitimate regulatory objective to disguise a protectionist action.

The incentives affecting the optimal trade and industrial policies toward international oligopoly in the domestic market *without* network externalities are well understood in the trade literature (see, for example, Brander and Spencer (1984), Dixit (1984, 1988), Eaton and Grossman (1986, Section 6) and Cheng (1988)). This literature demonstrates that trade restrictions against foreign firms involve a trade-off between the negative effect on the domestic consumer surplus and the positive effect on the domestic firms' profits. Despite this trade-off, under fairly typical conditions the optimal trade policy is an import tariff, which allows the home government to shift rents from the foreign firms to the domestic firms and the domestic treasury. Moreover, if the import tariff is the only policy instrument available to the home government, then in the full optimum the government should use it not only for rent shifting but also for targeting the domestic consumption distortion arising from the non-competitive behavior of the domestic firms (see Dixit (1988)).[9]

Gandal and Shy (2001) use a three-country model to formally analyze government's incentives to recognize foreign standards when there are potentially both network effects and conversion costs (i.e., costs of attaining compatibility). They show that depending on the relative magnitude of conversion costs and network effects, the countries may agree to mutual standards recognition universally or in a restricted fashion through the formation of a standardization union, which excludes one country. Their approach and ours are similar in that each

[9] Brander and Spencer (1984) demonstrated that under homogenous good international Cournot duopoly in the domestic market the import tariff is the optimal policy if it does not reduce the domestic consumer surplus too much relative to the domestic firm's gain in market share. Eaton and Grossman (1986) generalized the Brander and Spencer result by pointing out that domestic firms often choose their actions on the basis of incorrect conjectures about the actions of foreign firms. If the government understands that the domestic firms' beliefs are incorrect and can precommit to tax/subsidy schemes, there is a role for government policy in correcting the distortions arising from the incorrect conjectures of the firms. Using a conjectural variations approach, Dixit (1988) showed that an import tax and a domestic production subsidy are optimal under various assumptions about the substitution between foreign and domestic goods in the domestic demand. Cheng (1988) used a similar approach and showed that for international duopoly in the domestic market an import tax and a domestic production subsidy are optimal under both Cournot and Bertrand conjectures.

highlights how government's incentives to use standards as strategic trade policy instruments depend on the inherent tension between costs of attaining greater compatibility and network effects. However, in their analytical framework, unlike ours, governments cannot use traditional instruments of trade policy such as tariffs and subsidies. Moreover, Gandal and Shy do not consider standards requiring only partial compatibility (i.e., in their framework the foreign technology, which is not recognized by the home country regulator, can only be made fully compliant with the domestic standard).

The international economics literature recognizes that countries have incentives to use domestic regulations and standards to affect the international competitiveness of their firms not only in imperfectly competitive markets but also in environments in which offer curves are well defined and countries are large enough to affect their terms of trade. For example, Bagwell and Staiger (2001) showed that the government of the importing country has the incentive to impose *unduly lax* standards on the domestic import-competing firms. The government of the exporting country can change the terms of trade to its advantage by imposing *excessively stringent* standards on exporters. If the home firms compete with the foreign firms in a third-country market, the unilaterally optimal home country standard will again be unduly lax in comparison with the home standard. If, in addition to the terms-of-trade driven externalities, there are other types of cross-border externalities, then the regulatory laxity toward exporters may create additional concerns about cross-boundary spillovers, "races to the bottom," and "regulatory chill," which have also been discussed by economists (see Bagwell et al. (2002)).

Overview of the structure of the paper and the main results

The basic analytical framework of the paper is introduced in Section 2, where we describe the equilibrium outcome of the price-setting game between the rival firms assuming that the degree of compatibility between their products is given. In Section 3, we examine how government motives for trade policies are affected by the presence of the distortions caused by the imperfect compatibility between the domestic and foreign products. After we clarify the trade policy incentives of the home government, we extend our model by assuming that in addition to the import tariff the home government can set a standard that establishes the minimum degree of compatibility between the foreign and domestic products. The standard reduces the loss of the social network benefit arising from the decision by some domestic users to adopt the foreign product. However, if the costs of compliance fall primarily on the foreign firm, then the home government has the incentive to impose an excessively stringent (from the global perspective) compatibility standard.[10] Section 3 analyzes the combination of the import tariff and the compatibility standard that would be optimal for the home government in this environment, as well as the optimal standard under a free trade agreement restraining the government's ability to use the import tariff.

In Section 4, we analyze the incentives of a foreign government to use a policy affecting the foreign firm's choice of the degree of compatibility between its own product and the home firm's product. The foreign

[10] Given the structure of our model, both rival firms prefer greater compatibility and can coordinate on its jointly efficient level for any allocation of the intellectual property rights over the interface. In practical terms, this can be achieved through a cross-licensing arrangement that ensures that costs and benefits of compatibility-enhancement are allocated according to the bargaining powers of the rivals in dividing the joint surplus from greater compatibility. Since we are interested in trade effects of the domestic compatibility standard, we adopted a simplifying assumption under which the entire cost of achieving compatibility is born by the foreign firm. This means that the burden of standard-compliance falls disproportionately on the foreign firm. Therefore, the compatibility standard we consider in this paper is inherently discriminatory. However, the difference in the standard-compliance costs that arises from the allocation of the intellectual property rights over the interface does not necessarily imply a violation of the national treatment principle. If the firm controlling the interface were domestic, then the burden of compliance with the standard would fall on the domestic firm.

government can use this policy alone or in combination with the export taxes or subsidies. The incentives of the foreign government to conduct a compatibility-suppressing (or compatibility-enhancing) policy are somewhat similar to the incentives to tax (or subsidize) the exporting firm's quality-enhancing investment or cost-reducing R&D, which have been addressed in the strategic trade policy literature (see, for example, Spencer and Brander (1983), Cheng (1988) and Zhou et al. (2002)). However, in our setting the optimum mix of trade and compatibility-enhancing policies of the foreign government is affected by the presence of the network externality. Section 4 shows that when compatibility affects the firm's fixed cost but not its marginal cost, the foreign government must combine the export tax and the subsidy for compatibility-enhancing investment. [11]

After we clarify the foreign and home governments' incentives for policies affecting the degree of compatibility between the imported and import-competing products, we analyze, in Section 5, the equilibrium outcome of the non-cooperative game in which the home government uses the compatibility standard and the foreign government uses a tax (or a subsidy) linked to the foreign firm's compatibility-enhancing effort. To simplify the analysis and to emphasize the trade-related implications of policies toward compatibility, we assume that a free trade agreement restrains the governments' ability to use trade taxes. In addition, in this section we depart from the assumption that compatibility does not affect the marginal cost of the foreign firm. In this setting, the governments' incentives for using trade taxes are deflected into their compatibility policies. We derive the governments' best response policy functions and analyze some of the properties of the Nash equilibrium of the compatibility policy game.

In Section 6, we examine international agreements on policies toward compatibility. First, we characterize the combination of compatibility policies that is jointly efficient for the two countries. Having identified the globally efficient policy combination, we analyze the inefficiencies that are present in the non-cooperative equilibrium. We show that the strategic use of compatibility policies by the countries always leads them to an inefficient combination of trade volume and compatibility level. Depending upon the strength of the network externality effect, there can be either an excessively high equilibrium level of compatibility (in combination with either too much or too little trade) or very low levels of both compatibility and trade.

In Section 7, we discuss our findings in the context of the existing provisions of the WTO legal system aimed at minimizing the trade-inhibiting impact of domestic technical standards and regulations. In particular we use the results of our formal analysis to evaluate whether the "least-restrictive means" principle of policing technical regulatory barriers to trade is sufficient to enable the WTO member countries to reach global efficiency in the presence of network externalities.

---

[11] Obviously, the conclusions regarding the sign of the optimal policies are sensitive to the assumption that the firms compete in prices. Since the goal of our analysis was to illustrate in principle how government incentives to pursue international trade policy goals using compatibility policy instruments are affected by the presence of the network externality and user preference heterogeneity, we decided to confine our analysis to the setting involving price-competing firms, which is common in the industrial organization literature on compatibility. An analysis of trade policy in an environment in which the firms' choice variables are strategic substitutes can be developed using, for example, the Cournot framework with rational expectations employed by Katz and Shapiro (1985).

2. The basic model: consumer demand and costs

The analysis is conducted within a two-country framework in which a domestic firm A and a foreign firm B supply two products for the home country market.[12] Although domestic users have heterogeneous preferences regarding the products, they also value compatibility between the products they adopt and those adopted by other users because compatibility allows them to experience positive network externality. While users can achieve perfect compatibility if they adopt the product of the same firm, users of different firms' products can enjoy only partial compatibility. The extent of compatibility enjoyed by any two users of the different products is determined by the parameter $\gamma\in[0, 1]$, representing the fraction of the full compatibility benefit that the two users could realize if they adopted the product of the same make.[13]

In modeling the consumer preferences, we follow Farrell and Saloner (1992). Specifically, we consider a unit mass of domestic users, each of which has an inelastic demand for one unit of the products. The users differ in terms of their taste index $s\in[0,1]$, which determines the "stand-alone" value of the product to them (i.e., users' willingness to pay for the product regardless of the network externality). Assuming that users' relative preferences for product A over product B increase in s, the stand-alone utility accruing to a user with index s from adopting the products is given by:

$$\begin{cases} a+s \text{ if she adopts product } A; \\ a+(1-s) \text{ if she adopts product } B, \end{cases} \tag{1}$$

where constant a represents the part of the utility that is independent of the underlying product characteristics.[14]

All users share linear preferences for the network benefit. Specifically, if the share of users who adopt product A is x then the network-related component of the user's utility is $nx + n\gamma(1 - x)$ if she adopts product A and $n(1-x) + n\gamma x$ if she adopts product B where the parameter $n > 0$ measures the strength of the network externality. [15] Therefore, the consumer surplus of the type s user is:

---

[12] By ignoring the domestic market in the foreign country we focus on one of the two national markets of a reciprocal markets framework. A usual caveat about the conditions of market segmentation applies. The results of our analysis can be readily extended to a full reciprocal markets model incorporating the domestic markets of both countries. However, such an extension would involve substantially more notation without generating additional policy-relevant insights.

[13] It is helpful to think about the users as ICT operators providing the services to the unmodeled end consumers. While the end consumers of the ICT services may not appreciate the difference between the alternative technical designs of the substitute technologies underlying the services they purchase, the providers of those services are much more savvy and have clear preferences with regard to both the stand-alone quality of the technology they adopt and compatibility with the rival providers. In that interpretation, the duopolists are suppliers of equipment to the service operators.

[14] Parameter a captures the reservation price of consumers. We adopt a standard approach employed in applications based on the so-called address models of horizontal product differentiation by assuming that a is sufficiently large to ensure that the reservation price constraint is not binding in the equilibrium with two firms supplying the market. (See Economides (1984) and Neven et. al. (1991)). We also assume that a is large enough to guarantee that no user abstains from purchase even under autarky when the domestic market is supplied by the domestic monopolist producing a single product. Although these assumptions suppress the market-size and market-structure effects, they allow us to highlight the issues related to compatibility in the context of international trade.

[15] More generally, the network-related component of the user's utility is $N(x + \gamma(1 - x))$ if she adopts product H and $N((1-x) + \gamma x)$ if she adopts product F where $N(\cdot)$ denotes the network benefit function. $N(\cdot)$ is increasing and concave

$$\begin{cases} a + s + nx + n\gamma(1-x) - P_A & \text{if she adopts product } A; \\ a + (1-s) + n(1-x) + n\gamma x - P_B & \text{if she adopts product } B, \end{cases} \tag{2}$$

where P_A and P_B are the prices for products A and B.[16]

We assume that the products are incompatible ex ante (i.e., by design) but partial compatibility between them can be achieved by means of an ex post compatibility-enhancing modification, which can be incorporated into the main technical design of one or both rival products. Typically, the amount of control the rival producers have over the degree of compatibility between their products depends upon the allocation of the intellectual property (IP) rights over the technical interfaces. To simplify the analysis, we assume that the degree of compatibility γ is controlled by the foreign firm *B*, which can unilaterally undertake a compatibility-enhancing modification of its product.[17] We follow the existing literature by focusing on the symmetric two-way compatibility-enhancing modification, which requires altering only one of the two rival products but confers the same compatibility benefit on the users of both products (see, for example, the discussion of the two-way converters in Choi (1997)). Although these simplifying assumptions are rather severe in terms of the representation of ownership and control of the interface, they allow this paper to focus on the trade effects of the domestic standard that regulates the foreign supplier's choice of compatibility with the domestic product.[18]

For simplicity, the production costs of the firms under complete incompatibility are assumed to be zero. However, the foreign firm's compatibility-enhancing decision affects its production cost. We consider two alternative cost structures. In this section and the next one, we assume that to achieve the degree of compatibility γ

(convex) if the network externality is characterized by decreasing (increasing) returns to the average level of compatibility attained by the users. Auriol and Benaim (2000) identified the degree of concavity of the network benefit function with the user's tolerance to incompatibility between rival technologies; the more concave the network benefit function the more tolerant is the user to incompatibility. According to Auriol and Benaim's classification, communication networks are neutral to incompatibility, i.e., characterized by the linear network benefit function of the kind considered in this paper. Although we assumed that the network benefit function is linear, the conclusions of our paper remain valid for a nonlinear form of the function, as long as it is not too concave, i.e., the users are not too tolerant to incompatibility.

[16] Given this specification of preferences, if the user s prefers product *B*, then so does every other user in the interval [0, s]. Similarly, if the user s prefers product *A*, then so does every other user in the interval [s, 1].

[17] A possible interpretation of this assumption is that the domestic firm's interface-related IP rights have already expired while those of the foreign firm have not because the foreign technology is newer than the domestic. In regard to the asymmetry in the interface control between the producers of the two complementary products, see also MacKie Mason and Netz (2002). They discuss technical design strategies, which allow a firm controlling the IP rights over one of the two complementary technical systems to extend the boundary of its control to include the IP rights over the entire interface through which the two systems can interoperate. According to MacKie Mason and Netz, such strategies are quite common in the information technology industry.

[18] Farrell and Saloner (1992) discuss an alternative structure in which two incompatible products are supplied by a duopoly and the converters are supplied by the independent perfectly competitive firms. The consumers are free to choose whether to buy the products with or without converters. In that framework, there are three pure-strategy equilibria in the product adoption game among the users. The first equilibrium ("full standardization") is the adoption of the same product by all users. There is no need to buy converters in this case because users achieve perfect compatibility by virtue of using the same product. The second equilibrium is "perfect incompatibility," which occurs when the users split into two equal groups according to the product they adopt but no one adopts the converter. The third equilibrium ("conversion") implies imperfect compatibility because the users in the larger group buy one of the products without the converter and the users in the smaller group buy the rival product and the converter. In the third equilibrium, it is either only the users of product A or only the users of product B who buy the converters; there cannot be an equilibrium with both groups of users buying the converters. The identity of the minority group of users (i.e., A-users or B-users) who pay for the converters is indeterminate. In the context of our model, the fact that the foreign firm controls the interface and bears the entire cost of compatibility enhancement implies that the users of the foreign product implicitly pay for the converter embedded in the foreign product.

the foreign firm has to incur a fixed cost $F(\gamma)$.[19] Moreover, after the compatibility-enhancing modification, the unit cost of the foreign product becomes $C > 0$ regardless of the specific level of γ chosen by the foreign firm. The investment cost of compatibility, $F(\gamma)$, and the marginal investment cost of compatibility, $F'(\gamma)$, are assumed to be strictly increasing for all $\gamma \in [0, 1]$. In addition, in order to ensure the existence of the interior equilibrium with positive sales of the foreign product in the domestic market we assume that $F(0) = F'(0) = 0$ and $\lim_{\gamma \to 1} F'(\gamma) = \infty$.[20] In Section 5, we will consider an alternative cost structure in which compatibility enhancement affects the unit cost of the foreign firm but does not involve any fixed cost.[21]

The firms compete on prices after firm B has already made an investment ensuring the degree of compatibility γ. Assuming that both firms have positive sales in the domestic market, the condition identifying the marginal consumer s_B, who is indifferent with respect to the domestic product A and the foreign product B, is given by

$$a + (1 - s_B) + ns_B + n\gamma(1 - s_B) - P_B = a + s_B + n(1 - s_B) + n\gamma s_B - P_A. \tag{3}$$

Rearranging equation (3) gives the demand functions for products A and B:

$$s_A = 1 - s_B = \frac{1}{2} - \frac{(P_A - P_B)}{2(1 - n(1-\gamma))} \text{ and } s_B = \frac{1}{2} - \frac{(P_B - P_A)}{2(1 - n(1-\gamma))}. \tag{4}$$

The profits of firms A and B, respectively, are: $\Pi_A = P_A s_A$ and $\Pi_B = (P_B - C)s_B$. The Nash equilibrium prices are determined by solving the problem of profit maximization simultaneously for the two firms:[22]

$$\begin{cases} P_A = 1 - n(1-\gamma) + \dfrac{C}{3} \\ P_B = 1 - n(1-\gamma) + \dfrac{2C}{3}. \end{cases} \tag{5}$$

Therefore, the equilibrium sales and profits of the two firms, respectively, are:

[19] $(1-\gamma)$ can be interpreted as the loss of the network benefit, which is attributed to product performance degradation due to the imperfections of the compatibility-enhancing technology. For example, in wireless telephony, users of multi-mode wireless phones usually experience a greater number of dropped calls and shorter battery life when these phones are used for roaming in wireless networks based on communication protocols or radio frequencies that are different from the users' "native" networks.

[20] An example of the function satisfying these assumptions is $F(\gamma) = \beta\gamma(1 - (1 - \gamma)^{\alpha})$, where $1 > \alpha > 0$, $\beta > 0$.

[21] The assumption that the cost of achieving a degree of horizontal or vertical product differentiation is sunk prior to the determination of prices and output is well established in the literature. See, for example, Gabszewicz and Thisse (1979) and Zhou et al. (2002). In our model, the fixed cost can be interpreted of as a cost of the interface-related R&D that must be undertaken in order to increase the compatibility between the two products.

[22] It is easy to verify that the second-order conditions are satisfied. However, d'Aspremont et al. (1979) showed that in the Hotelling-type address models a pure strategy Nash equilibrium in prices may fail to exist for some parameter values because the profit functions are not quasiconcave. To ensure the existence of the equilibrium, the firms' price strategies must be price undercutting-proof. The condition for proofness against price undercutting in our model is $|P_A - P_B| < (1 - n(1-\gamma))$. Because it follows from equations (5) that $|P_A - P_B| = C/3$, the price-undercutting proofness condition is equivalent to $C/3 < 1 - n(1 - \gamma)$. The latter inequality follows from the assumption that the compatibility-enhancing technology is not too inefficient (see the discussion in footnote 23 below): $C < 1 - 2n(1 - \gamma) \Rightarrow C < 1 - n(1 - \gamma) \Rightarrow C/3 < 1 - n(1 - \gamma)$. As can be seen from equations (6), the condition $C/3 < 1 - n(1 - \gamma)$ also ensures that in equilibrium both firms have positive sales.

$$s_A = \frac{1}{2} + \frac{C}{6(1 - n(1-\gamma))}; \quad s_B = \frac{1}{2} - \frac{C}{6(1 - n(1-\gamma))} \tag{6}$$

and

$$\Pi_A = \frac{\left(1 - n(1-\gamma) + \frac{C}{3}\right)^2}{2(1 - n(1-\gamma))}; \quad \Pi_B = \frac{\left(1 - n(1-\gamma) - \frac{C}{3}\right)^2}{2(1 - n(1-\gamma))}. \tag{7}$$

Since enhanced compatibility provides symmetric benefits for users of both products but raises the cost of only the foreign product, the market share of the domestic firm must exceed that of the foreign firm. This can be seen from equations (6). Therefore, we are going to refer to the domestic product as the dominant product and to the foreign product as the minority product.

When the foreign product is not perfectly compatible with the domestic product (i.e., $\gamma < 1$), the introduction of the former into the domestic market undermines the integrity of the domestic network and may have a negative effect on the home country welfare. On the one hand, the availability of product *B* increases the welfare of those users who prefer it to product *A*. On the other hand, since the network of product *B* users is only partially compatible with the dominant network of product *A* users, each user who adopts product *B* would have conferred greater total network externality on the society by adopting product *A*. Since *B* users do not take into account the negative effect of their product adoption decisions on the society, in equilibrium there are more *B* users than is socially optimal. Farrell and Saloner (1992) demonstrated that the problem of overadoption of the minority product occurs regardless of whether the rival technologies are supplied by perfectly competitive firms or by the oligopoly. However, under oligopoly the problem is aggravated by the fact that the producer of the minority product can attract even more users by undercutting the above-cost price of the dominant firm.

Opening the domestic market to trade benefits those domestic consumers who prefer the foreign product. However, the home country welfare is negatively affected by the shift of some of the pure rents abroad and the loss of the network benefits due to the disintegration of the homogenous domestic network into two partially compatible smaller networks. When the negative welfare effects outweigh the positive one, the home country is better off under autarky than with trade. Since the focus of this paper is on trade policy, we make an assumption to the effect that with a rent-capturing import tariff the home country prefers trade to autarky. Specifically, we assume that the compatibility-enhancing technology is not too inefficient: $C < 1 - 2n$.[23]

[23] In a closed economy, if the compatibility-enhancing technology is not efficient enough to satisfy the inequality $C < 1 - 2n(1-\gamma)$, then the homogenous network based on a single technology is better than the two partially compatible networks even when the market share of the minority technology is optimized. Therefore, in the open economy context, the violation of this inequality implies that even if the import tax captured all rents of the foreign supplier the homogenous network based on the domestic technology under autarky would still be better than the network based on the partially compatible domestic and foreign technologies. In order to make the trade policy analysis in our model meaningful, we assume that the conversion technology is sufficiently efficient that the heterogeneous foreign/domestic network in combination with the rent-capturing import tariff is better than the autarky-induced homogenous network based on the domestic technology. The assumption about the conversion technology's efficiency also guarantees that the model is price undercutting-proof (see the discussion in footnote 22). To ensure that the inequality $C < 1-2n(1-\gamma)$ is valid for all $\gamma \in [0, 1]$, we assume that $C < 1-2n$ and $n < 1/2$.

3. Optimal trade and standardization policies of the home country.

The home government faces a number of distortions, which it can target by means of an import tariff and a compatibility standard. First, there is the strategic distortion, which creates the rent-shifting incentive for policy intervention (see Brander and Spencer (1984) and Dixit (1988)). Second, there is the consumption distortion.[24] However, because the present model assumes that consumers have inelastic unit demand and that the market is fully covered, the consumption distortion is evident not in the presence of consumers who abstain from purchasing but in the inefficient relative size of the dominant and the minority networks. In other words, the dominant network is too small because too many users buy the minority product. Third, the firms (or only the foreign firm in our case) have insufficient incentives for making their products compatible.

After we clarify the home government's incentives that determine its choice of the import tariff for a given level of compatibility between the foreign and domestic products, we extend our model by assuming that in addition to the import tariff the home government can set a compatibility standard. The standard can be chosen either simultaneously with the import tariff or prior to it.[25] In either case, the standard reduces the welfare loss due to the disintegration of the domestic network and shifts the cost of achieving greater compatibility onto the foreign firm, which controls the interface between the two products.

Optimal tariff

With a specific tariff τ, the profit function of the foreign firm becomes: $\Pi_B = (P_B - C - \tau)s_B$. The equilibrium prices are thus

$$P_A = 1 - n(1-\gamma) + \frac{(C+\tau)}{3} \text{ and } P_B = 1 - n(1-\gamma) + \frac{2(C+\tau)}{3}. \tag{8}$$

Therefore, the equilibrium sales of the firms in the domestic market are:

$$s_A = \frac{1}{2} + \frac{C+\tau}{6\big(1-n(1-\gamma)\big)} \text{ and } s_B = \frac{1}{2} - \frac{C+\tau}{6\big(1-n(1-\gamma)\big)}. \tag{9}$$

The foreign firm will have positive sales on the domestic market if $C + \tau < 3(1 - n(1 - \gamma))$, which can be interpreted as a condition requiring that the loss of the network benefit due to imperfect compatibility is not too large. If this condition is satisfied, then both firms have positive sales, and the equilibrium profits are then given by

[24] In the trade policy literature, the consumption distortion that arises from the supplier's market power is commonly targeted with a production subsidy (see Cheng, 1988). However, since it is not very realistic to assume that anti-competitive behavior is rewarded with a subsidy, we assume that the home government does not use the production subsidy.

[25] As we are going to demonstrate below, the standard will always be binding when it is used in combination with the tariff. Therefore, we can ignore the effect of the tariff on the foreign firm's choice of compatibility.

$$\Pi_A = \frac{\left(1-n(1-\gamma)+\frac{C+\tau}{3}\right)^2}{2\left(1-n(1-\gamma)\right)} \text{ and } \Pi_B = \frac{\left(1-n(1-\gamma)-\frac{C+\tau}{3}\right)^2}{2\left(1-n(1-\gamma)\right)}. \tag{10}$$

The choice of the tariff by the home government is affected by a number of incentives. As equations (10) show, the tariff shifts some of the foreign firm's profit to the domestic firm. When no other policy instrument is available to the home government, the level of the optimum tariff also reflects its second-best role in targeting the distortionary overadoption of the imported product.

The welfare of the home country is simply the sum of the consumer surplus, the domestic firm's profit, and the tariff revenue:

$$\begin{aligned} W_H(\tau) &= \int_0^{s_B}(a+1-x)dx + \int_{s_B}^{1}(a+x)dx + n\left(1-2s_B(1-s_B)(1-\gamma)\right) - P_B s_B + \tau s_B \\ &= a + 1/2 + n + \left(1-2n(1-\gamma)\right)s_B(1-s_B) - P_B s_B + \tau s_B \end{aligned} \tag{11}$$

The first two terms on the right-hand side of the first equality in (11) represent the stand-alone benefits of users who adopt either product B (users with the taste parameters $s \in [0, s_B]$) or product A (users with tastes $s \in [s_B, 1]$). The third term is the maximum network benefit attainable when all users adopt the same product minus the loss due to imperfect compatibility. The last two terms are the cost of the imported products and the home tariff revenue.

To obtain an expression for the optimal tariff, we differentiate (11) with respect to τ:

$$\frac{\partial W_H}{\partial \tau} = \frac{6-3\tau+n(1-\gamma)\left(C+4(\tau-3)+6n(1-\gamma)\right)}{18(1-n(1-\gamma))^2}. \tag{12}$$

If the loss of the network benefit due to incomplete compatibility, $n(1-\gamma)$, is not too large, the welfare function $W_H(\tau)$ is concave in τ.[26] Setting $\partial W_H(\tau)/\partial\tau$ equal to zero and rearranging the terms, we find that the optimum tariff is positive and is given by the expression:

$$\tau^* = \frac{6+n(1-\gamma)\left(C+6n(1-\gamma)-12\right)}{3-4n(1-\gamma)}. \tag{13}$$

Equation (13) implicitly assumes that the optimal tariff is not prohibitive—i.e., that the foreign firm has positive sales when the tariff is τ^*. However, this need not be the case. The import tariff becomes prohibitive if it reaches the level that sets s_B in equation (9) to zero: $\tau^P = 3(1 - n(1-\gamma)) - C$. It is straightforward to verify that when the condition for sufficient efficiency of the compatibility-increasing technology is satisfied (i.e., $C < 1 - 2n$) the optimal tariff is less than the prohibitive tariff: $\tau^* < \tau^P$. Therefore, although the optimal tariff reduces trade, it does not shut it out completely.

[26]The exact condition for the concavity of the welfare function in the tariff is $n(1-\gamma) < ¾$. Note that this inequality is valid under the condition that the compatibility-enhancing technology is efficient (i.e., $C < 1 - 2n$), which we assume throughout this paper.

The effect of compatibility on the optimal tariff

It is often suggested that greater compatibility reduces product differentiation, which, in turn, sharpens competition and leads to lower price.[27] In this paper, however, greater compatibility with the rival's product reduces the dependency of the firm's profit on the size of its own network and, therefore, makes the profit less sensitive to the firm's market share. This effect blunts competition for market share and leads to higher prices. Thus, as products become more compatible, the firms exert weaker competitive pressure on each other and their profits increase. This implies that greater compatibility between the rival firms strengthens the home government's incentives for conducting the rent-capturing trade policy. This is confirmed by the sign of the derivative of the optimal tariff with respect to the compatibility parameter γ:

$$\frac{\partial \tau^*}{\partial \gamma} = \frac{3n\left(4(1-n(1-\gamma)) - C\right)\left(1-2n(1-\gamma)\right)}{\left(3-4n(1-\gamma)\right)^2} > 0 \,. \tag{14}$$

It is instructive to compare the import tariff with the production tax the government would impose on the minority firm in a closed economy if both firms were domestic. Assuming that the fixed cost of compatibility has already been sunk, the production tax targets only the overadoption of the minority product that is creating the inefficiency in the relative size of the two networks.[28] The optimal production tax t^D can be derived by solving the first-order condition for maximization of welfare of the country with two domestic rivals producing partially compatible products:

$$\max_{t^D}\left(a + 1/2 + n + \left(1-2n(1-\gamma)\right)s_B(1-s_B) - s_B C\right), \tag{15}$$

where firm B's equilibrium market share is now a function of the production tax: $s_B = \frac{1}{2} - \frac{C + t^D}{6\left(1-n(1-\gamma)\right)}$. After verifying the second-order condition, simple rearrangement of the first-order condition leads to the following expression for the optimum production tax on the domestic minority duopolist:

$$t^D = \frac{C\left(2-n(1-\gamma)\right)}{1-2n(1-\gamma)}. \tag{16}$$

Because $\partial t^D/\partial\gamma < 0$ while $\partial\tau^*/\partial\gamma > 0$, the difference between the optimal tax on the domestic minority firm in a closed economy and the optimal tariff against the foreign minority firm in an open economy increases as γ becomes larger. The reason for this result is that the larger the value of γ the less severe is the problem of overadoption and the higher is the profit of the foreign firm. Since the production tax on the domestic minority firm

[27] Greater compatibility between the rival products can make them better complements and substitutes at the same time. By making them better complements, greater compatibility increases consumer willingness to pay for each of the two products. By making the networks based on the two rival technologies better substitutes, greater compatibility can lead to more intense competition between the firms. See Berg (1988) and Economides (1991) on the effects of compatibility on competition between firms in a closed economy with network externalities.

[28] The extent of the overadoption distortion can be measured by the wedge between the socially optimal share of the domestic minority firm $\hat{s}_B = \frac{1}{2} - \frac{C}{2\left(1-2n(1-\gamma)\right)}$ and its equilibrium share $s_B = \frac{1}{2} - \frac{C}{6\left(1-n(1-\gamma)\right)}$. The optimal production tax defined by equation (16) pushes the minority firm's equilibrium share to the socially optimal level.

only targets the overadoption distortion, while the import tariff targets the overadoption distortion and recaptures the foreign profit, the domestic tax and the import tariff differ the most when the common incentive for their use (i.e., the problem of overadoption) is small.[29]

Optimal compatibility standard

We now turn to the characterization of the optimal compatibility standard chosen by the home government. Under the optimum tariff against the foreign firm, the home country welfare is given by:

$$W_H(\tau^*(\gamma)) = a + 1/2 + n + \frac{\left(C - 1 - 2n(1-\gamma)\right)^2}{4(3 - 4n(1-\gamma))}. \tag{17}$$

Because greater compatibility between the rival technologies benefits all domestic consumers and reduces the problem of overadoption of the minority product, the home country welfare under the optimal tariff increases with increases in the degree of compatibility:

$$\frac{\partial W_H(\tau^*(\gamma))}{\partial \gamma} = \frac{n\left(1-2n(1-\gamma)-C\right)\left(2-2n(1-\gamma)+C\right)}{(3-4n(1-\gamma))^2} > 0. \tag{18}$$

Does this mean that the home government should require complete compatibility by setting the standard γ^H = 1? Any standard that exceeds the level of compatibility at which the foreign firm breaks even is exclusionary (i.e., prohibitively stringent) because it prevents the foreign firm's production for the home country market.[30] If the complete compatibility standard is exclusionary, then setting $\gamma^H = 1$ can induce autarky. Since under international oligopoly it is not uncommon that autarky is preferred to trade (e.g., see Fung (1988)), we compare the home country welfare under trade subject to the optimal tariff with the home country welfare under autarky. Under autarky, all consumers belong to the homogenous network based on the product supplied by the domestic firm. Therefore the autarkic welfare is given by

$$W_H^{Au} = \int_0^1 (a + x + n)dx = a + 1/2 + n. \tag{19}$$

By comparing (17) and (19), it is easy to see that with the optimal tariff levied on the foreign firm the home country welfare is higher with trade than under autarky if and only if $n(1 - \gamma) < ¾$. This condition is true under the assumption $C < 1 - 2n$, which has been adopted in this paper. The above results can be summarized as:

Proposition 1: If the compatibility-enhancing technology is sufficiently efficient to satisfy the condition $C < 1 - 2n$, then the optimal policy of the home country government is to impose the import tariff given by equation (13) and the most stringent non-exclusionary compatibility standard γ^H.

---

[29] It is worth noting that the optimal tax against the domestic minority firm does not become zero even when the two rival products are perfectly compatible (i.e., $\gamma = 1$). This is because the problem of overadoption of the minority (i.e., the more costly) product is present even when the rival products are perfectly compatible. The main cause of the overadoption problem is not incomplete compatibility, but rather the non-competitive pricing behavior of the firms. When the products are perfectly compatible and supplied by perfectly competitive firms, the problem of overadoption disappears; i.e., under perfect competition and perfect compatibility the equilibrium market shares of the rival technologies are socially optimal.

[30] Given the tax τ, the exclusionary standard exists if $\Pi_B(1, \tau) < F(1)$, where $\Pi_B(\gamma, \tau)$ is defined by (10).

If the complete compatibility standard is non-exclusionary (i.e., $\Pi_B(1, \tau^*) > F(1)$), the home government will set $\gamma^H = 1$. If the complete compatibility standard is exclusionary (i.e., $\Pi_B(1, \tau^*) < F(1)$), the home government will set the standard just below the lowest exclusionary level of γ. This is the highest standard under which the foreign firm would enter the home market. If the cost function $F(\gamma)$ satisfies the assumptions introduced in Section 2, the foreign firm's net profit function $\Psi(\gamma) = \Pi_B(\gamma, \tau^*) - F(\gamma)$ is quasi-concave and satisfies the conditions: $\Psi(0) > 0$ and $\partial\Psi(\gamma)/\partial\gamma\big|_{\gamma=0} > 0$. Therefore, the solution to $\Psi(\gamma) = 0$ is well defined and we can determine γ^H by finding the level of compatibility under which the foreign firms breaks even:

$$\Psi(\gamma) = \frac{(C-1+2q)^2(1-q)}{2(3-4q)^2} - F(\gamma) = 0\,, \tag{20}$$

where $q = n(1-\gamma)$.

Is the home country compatibility standard binding? The foreign firm chooses the level of compatibility to maximize its profit net of the fixed cost of achieving compatibility:

$$\gamma^B = \arg\max_{\gamma} \Psi(\gamma)\,. \tag{21}$$

Again, given the properties of the net profit function $\Psi(\gamma)$ (i.e., quasi-concavity, $\Psi(0) > 0$, and $\partial\Psi(\gamma)/\partial\gamma\big|_{\gamma=0} > 0$), it follows that $0 < \gamma^B < \gamma^H$. Therefore, if the government can impose the optimal tariff, it will combine the tariff with the binding but non-exclusionary compatibility standard implicitly defined by (20). Such a combination of policies achieves several goals: (1) it gives the domestic users access to the foreign product, (2) it minimizes the loss of the network externality caused by that access, and (3) it allows the government to recapture some of the profit earned by the foreign product supplier.

When the home government's ability to tax imports is curtailed by a trade agreement, the homogenous network based only on the domestic product is preferred to the heterogeneous network based on the partially compatible foreign and domestic products even if an efficient compatibility-enhancing technology is available. In other words, autarky is preferred to free trade.

Proposition 2: Under a free trade agreement, the home government will set the compatibility standard at a prohibitively high level if the exclusionary compatibility standard exists (i.e., if ($\Pi_B(1,0) < F(1,0)$)). If there is no exclusionary standard under the free trade agreement (i.e., if ($\Pi_B(1,0) > F(1,0)$)), then the home country compatibility standard will be ineffectual: $\gamma^H < \gamma^B$.

The possibility that autarky may be preferred to free trade under international oligopoly is a well-known fact in the international trade literature.[31] In the present framework, the home country's welfare is higher under autarky because the profit lost by the domestic firm after opening the home market to free trade exceeds consumer surplus gains. If the home government cannot extract foreign profits through a sufficiently high tariff, then autarky

[31] Fung (1988) compared free trade under international duopoly in the home country market with the domestic monopoly under autarky. He discussed the conditions under which autarky may be better than free trade when the differentiated product duopolists compete in quantities (i.e., in the Cournot-Nash setting). Fischer and Serra (2000) considered a framework in which autarky is induced by a prohibitively high quality standard that applies to the domestic as well as foreign firms. They investigated the conditions under which the standard-induced autarky is welfare superior to free trade without a standard when the firms are homogenous-good quantity competitors.

ensures greater total welfare for the home country than trade. Therefore, if a trade agreement restrains the use of tariffs but not standards, the home government will use the standardization policy to induce autarky. This can be achieved by choosing any standard above the level of compatibility that sets the foreign firm's net profit to zero:

$$\frac{\left(1-n(1-\gamma)-\frac{C}{3}\right)^2}{2(1-n(1-\gamma))}-F(\gamma)=0\,. \tag{22}$$

Without trade taxes, setting a binding but non-exclusionary standard will only shift a greater share of the total surplus to the foreign firm. The fact that the positive effect of γ on the foreign profit dominates its positive effect on the domestic profit and consumer surplus is evident from the sign of the derivative of the home country welfare with respect to γ when the tariff is set to zero: $\frac{\partial W_H(0)}{\partial \gamma}=-\frac{C^2 n\left(2-n(1-\gamma)\right)}{18\left(1-n(1-\gamma)\right)^3}<0$.

Therefore, if neither complete exclusion of the foreign firm nor recapturing of the foreign profit through an import tariff is feasible for the home government, it will set a non-binding standard, which will have no effect on the foreign firm's choice of compatibility level.

4. Foreign policies when compatibility does not affect the marginal production cost.

Having discussed the policies of the home country government, we now turn to the policies of the foreign government toward the export market. Since the consumer surplus of the foreign citizens is not involved, the foreign government's only motivation for policy intervention is rent shifting. However, unlike the earlier literature on strategic trade policies, this paper considers two distinct commitment mechanisms through which the equilibrium outcome of the game between the firms in the export market can be changed to the advantage of the exporting country. First, the government can use the export tax to commit the exporting firm to a less-aggressive price-setting behavior, helping it to achieve the Stackelberg outcome of the game. Second, the foreign firm's compatibility-enhancing investment acts as a mechanism for self-imposed commitment to less-aggressive price setting.

Although both mechanisms—the export tax and the compatibility-enhancing investment—help to shift rents from the home firm to the foreign firm and the foreign country treasury, the former mechanism interferes with the latter. The export tax reduces the foreign firm's incentives to invest in greater compatibility with the home firm's technology. To overcome the negative effect of the export tax on the firm's incentive to invest in compatibility, the foreign government has to subsidize the firm's investment. Therefore, to achieve the full optimum the foreign government must combine the export tax with the subsidy for compatibility-enhancing investment. The foreign government's policies in this setting can be compared with other models of government policies toward the imperfectly competitive export market. Spencer and Brander (1983) showed that if the oligopolistic competition is Cournot and the exporting firm can make a marginal cost-reducing investment, the optimal policy is a combination of the export subsidy and the investment tax. If the firms compete in Bertrand fashion, then the optimal policy mix is the export tax and the investment subsidy (see Eaton and Grossman (1986) and Cheng (1988)).[32] While the export tax ensures the firm's commitment to less-aggressive pricing, the investment subsidy induces the firm to choose the investment that maximizes the direct impact of reducing the marginal cost on the profit. In contrast to these models, the imperfectly competitive exporter considered in this section has no own strategic incentives to underinvest because the compatibility-enhancing investment does not affect the firm's marginal cost.[33] It is the government's use of the export tax for relaxing price competition in the export market that runs into conflict with the exporter's incentive to invest in compatibility. Therefore, the export tax needs to be complemented by a subsidy for compatibility-enhancing investment.

This result holds regardless of whether the export tax and the compatibility investment subsidy are applied simultaneously before the investment is in place or in a sequential manner (i.e., the investment subsidy before the investment and the export tax afterward, but before the production stage). However, to fix the ideas, we assume that the foreign government chooses the investment subsidy before the firm undertakes the compatibility-enhancing investment and only after that chooses the export tax.[34]

With a specific export tax σ, the profit function of the foreign firm, which has already undertaken the compatibility-enhancing investment, is given by: $\Pi_B = (P_B - C - \sigma)s_B$. The profit of the home firm is defined as in Section 2. After solving for the Nash equilibrium in prices, we obtain the equilibrium profit and sales of the foreign firm in the market of the home country:

$$\Pi_B(\sigma) = \frac{\left(1 - n(1-\gamma) - \frac{C+\sigma}{3}\right)^2}{2\left(1 - n(1-\gamma)\right)} \text{ and } s_B = \frac{1}{2} - \frac{C+\sigma}{6\left(1 - n(1-\gamma)\right)}. \tag{23}$$

To determine the optimal export tax, we maximize the sum of the foreign firm's profit and the export tax revenue with respect to σ: $\max_{\sigma}\left(\Pi_B(\sigma) + s_B\sigma\right)$. Rearranging the first-order condition for this maximization problem[35] yields the expression for the optimal export tax:

$$\sigma^* = \frac{3(1-n(1-\gamma)) - C}{4}. \tag{24}$$

Under the optimum export tax, the market share of the foreign firm is given by

$$s_B = \frac{3}{8} - \frac{C}{8(1-n(1-\gamma))} = \frac{3(1-n(1-\gamma)) - C}{8(1-n(1-\gamma))}. \tag{25}$$

---

[32] Under price competition, the exporting firm under-invests in marginal cost reduction because, in addition to the direct positive effect of such an investment on the firm's profit, there is an opposing indirect effect related to the strategic complementarity of price-setting actions. Lower marginal cost makes the exporter more aggressive in lowering the price, which, in turn, triggers a price reduction by the competitors and subsequently lowers the profits of both firms. Because of this negative indirect effect, the exporting firm moderates its investment in marginal cost reduction. To overcome this distortion, the government imposes the policy mix of the export tax and the investment subsidy.

[33] In fact, both the direct and the indirect effects of higher compatibility on the exporting firm's profit are positive. Higher compatibility directly increases the firm's marginal revenue because it raises the network-related quality of the product. Higher compatibility also indirectly helps the foreign firm to earn higher profit because it reduces the intensity of competition between the firms.

[34] The sequence of policy decisions might be justified by the fact that the firm's response to the investment subsidy involves changing its investment behavior, which generally takes longer than the production stage response to trade taxes.

[35] Note that the second-order condition is satisfied, as well: $\frac{d^2W_F}{d\sigma^2} = \frac{-2}{9(1-n(1-\gamma))} < 0$.

With the optimum export tax, the foreign firm's profit net of the subsidized investment cost of the converter is given by:

$$\pi_B = \frac{\left(1-n(1-\gamma)-\frac{C+\sigma^*}{3}\right)^2}{2\left(1-n(1-\gamma)\right)} - (1-\phi)F(\gamma) = \frac{\left(C-3\left(1-n(1-\gamma)\right)\right)^2}{32\left(1-n(1-\gamma)\right)} - (1-\phi)F(\gamma)\,, \tag{26}$$

where ϕ represents the proportion of the cost of the compatibility-enhancing investment covered by a subsidy to the foreign firm. If ϕ takes a negative value, we interpret it as an investment tax.

The foreign government chooses the subsidy to maximize the country's welfare, which equals the foreign firm's net profit plus the export tax revenue minus the cost of the investment subsidy:

$$W_F = \pi_B + s_B \cdot \sigma^* - \phi F(\gamma) = 2\left[\frac{\left(C-3\left(1-n(1-\gamma)\right)\right)^2}{32\left(1-n(1-\gamma)\right)} - \frac{1}{2}F(\gamma)\right], \tag{27}$$

where second equality is obtained by rearrangement using the definitions of the optimum export tax (24) and the foreign firm's market share (25).

A comparison of equations (26) and (27) suggests that to induce the foreign firm to choose the welfare-maximizing level of compatibility of its technology with the rival technology, γ, the foreign government has to offer the firm a subsidy that covers up to half the cost of investment in compatibility.

Proposition 3: The optimum policy of the foreign government is a combination of the export tax defined in (24) and a subsidy for compatibility-enhancing investment.

So far, we have considered the trade and product compatibility policies of the two governments separately. Now, we must ask: What happens if both countries heed their strategic incentives and commit themselves to using these policies? Specifically, what happens if the policy interaction between the governments involves the import tax and the compatibility standard on the side of the home government and the subsidy for compatibility-enhancing investment and the export tax on the side of the foreign government? To answer this question, we must consider a multi-stage game in which the policies toward compatibility on both sides are applied either prior to the application of trade taxes or simultaneously with them. The analysis of such a model is technically complicated and is not presented here. However, it turns out that regardless of the timing of applying the compatibility-enhancing policies (i.e., prior to or simultaneously with the trade taxes) the *nature* of the desirable policies would be the same as described above, and only the *magnitude* of the chosen instruments would change. The home government would still prefer an import tax in combination with the most stringent non-exclusionary compatibility standard, and the foreign government would prefer an export tax in combination with an investment subsidy.

5. Non-cooperative compatibility policies when compatibility affects the marginal production cost.

Thus far, we have examined the case in which achieving compatibility with the domestic product has no effect on the marginal production cost of the exporting firm but requires it to incur a fixed cost. In this section, we consider a framework in which greater compatibility with the rival product affects the foreign exporter's marginal cost. This can happen when enhancing compatibility requires costly modification of each individual unit of the foreign product. For example, it may be that in order to achieve compatibility the foreign firm has to embed a compatibility-enhancing device (i.e., a converter) in each unit of its product. For simplicity, we assume that attaining compatibility does not involve any fixed cost.

Because the level of compatibility between the products affects the foreign firm's marginal cost, the foreign government now has two strategic reasons for trying to influence the firm's compatibility choice. First, as we showed in the previous section, a higher level of compatibility increases the foreign firm's marginal revenue by raising consumer willingness to pay for the products and by relaxing price competition between the firms. Second, more compatibility means higher marginal cost, which makes the firm less aggressive in price setting and also, as a consequence, dampens competition.

As in the previous sections, the home government continues to use the minimum compatibility standard. However, to concentrate on the trade-related implications of policies toward compatibility, we assume in this section that a trade agreement between the two countries prevents them from using traditional trade taxes.[36] Instead, the foreign government now uses a subsidy (or a tax) aimed at the level of compatibility chosen by the foreign firm. Although the compatibility subsidy/tax affects the foreign firm's unit cost, it is different from the export subsidy/tax because it is directly linked to the firm's choice of compatibility with the domestic product.

In telecommunications and information technology, there are a number of examples of government policies targeted at the compatibility of exports with rival products in the export market. Broadly speaking, any policy measure, including but not limited to explicit taxes (or subsidies), that makes it harder (or easier) for an exporter to achieve greater compatibility between its own product and the installed base of the rival firms' products in the export market can be interpreted as a compatibility-suppressing tax (or compatibility-enhancing subsidy). Examples of such compatibility policy measures are taxes (or subsidies) on the interface-related R&D affecting interoperability between the rival technologies and taxes (or subsidies) on the R&D that increases differentiation between the designs of the products.[37] In the latter case, an R&D subsidy leading to greater ex ante differentiation

[36] To make a meaningful analysis of implications of the trade agreement restraining the use of trade taxes but not policies aimed at compatibility, we assume that the available compatibility-enhancing technology is sufficiently efficient that even the complete compatibility standard does not exclude the foreign firm from the home market (i.e., even after choosing $\gamma = 1$ the foreign firm still has a positive market share in the home market). This assumption allows us to ignore the possibility that autarky may be preferred to untaxed trade even in the setting where a compatibility standard affects the marginal cost of the foreign firm.

[37] Such R&D may affect the exporter's compatibility-enhancing decision either by requiring it to incur a sunk cost, which we considered in the previous section, or by affecting its unit cost, as we discuss in this section. Therefore, while we considered in the previous section the compatibility-enhancing subsidy (or compatibility-suppressing tax), which takes the form of an investment subsidy (or tax), in this section we analyze the effect of a unit cost subsidy (or tax).

between the rival products (i.e., differentiation by design) will have an effect similar to the compatibility-suppressing tax because it increases the cost of making the products compatible ex post.[38]

5.1 Unconstrained optimal policy of the foreign government

Let $C(\gamma)$ be the unit cost of achieving the level of compatibility γ, where $C(\gamma) \geq 0$, $C'(\gamma) \geq 0$, and $C''(\gamma) \geq 0$. To ensure the existence and uniqueness of equilibrium, we also assume that $C'(\gamma)$ becomes infinite in the limit as γ approaches one. In order to make it profitable for the foreign firm to enter, we assume that the total and marginal costs of the "first unit" of compatibility equal zero (i.e., $C(0) = C'(0) = 0$). In summary, we assume:

$$C(0) = C'(0) = 0;\ C'(\gamma) > 0 \text{ and } C''(\gamma) > 0 \text{ for } \gamma > 0;\ \lim_{\gamma \to 1} C'(\gamma) = \infty.$$ [39]

Denote by θ the foreign tax ($\theta > 0$) or subsidy ($\theta < 0$) per unit compatibility cost. Assuming that $\theta > -1$, the foreign firm faces a strictly positive cost of compatibility, and its profit is given by: $\Pi_B(\gamma, \theta) = \left(P_B - (1+\theta)C(\gamma)\right)s_B$. Since the tax (or subsidy) θ and the compatibility level γ are chosen prior to the stage in which the firms compete in prices, the equilibrium prices and the foreign firm's profit and sales are determined in a manner similar to the way they were determined in Section 2:

$$\begin{cases} P_A = 1 - n(1-\gamma) + \dfrac{C(\gamma)(1+\theta)}{3} \\ P_B = 1 - n(1-\gamma) + \dfrac{2C(\gamma)(1+\theta)}{3}, \end{cases} \tag{28}$$

and

$$\Pi_B = \frac{\left(1 - n(1-\gamma) - \frac{C(\gamma)(1+\theta)}{3}\right)^2}{2\left(1 - n(1-\gamma)\right)} \text{ and } s_B = \frac{1}{2} - \frac{C(\gamma)(1+\theta)}{6\left(1 - n(1-\gamma)\right)}. \tag{29}$$

The level of compatibility chosen by the foreign firm for a given tax θ satisfies the following first-order condition for maximization of Π_B with respect to γ:

[38] Examples in wireless telecommunications include the subsidies for the R&D related to spectrum-sharing and interference-avoidance technologies embedded in multimode wireless handsets and base stations. Such subsidies are used by the governments of Finland and South Korea. An example of policies stimulating greater design differentiation and, as a result, leading to less compatibility is the Chinese government-funded efforts in developing distinct homegrown technical designs on a wide range of technology from audio/video compression and optical-disk technology to operating systems and mobile-phone network protocols. Such policy recently helped a consortium of Chinese firms to develop a third-generation wireless communications system called Time Division Synchronous Code Division Multiple Access (TD-SCDMA). If the state-run China Telecomm adopts TD-SCDMA as a technology for the next generation of wireless networks, China's wireless telephone networks will be compatible neither with the European wideband-CDMA (WCDMA) equipment nor with the U.S.-backed CDMA2000 equipment.

[39] These assumptions greatly simplify the analysis by ensuring the existence of the interior equilibrium with partial compatibility and positive sales of the foreign product in the domestic market. An example of a function satisfying these assumptions is $C(\gamma) = \gamma(1 - (1-\gamma)^{\alpha})$ where $0 < \alpha < 1$.

$$\frac{\partial \Pi_B}{\partial \gamma} = \frac{\left(3(1-q)-(1+\theta)C(\gamma)\right)\left(n(1+\theta)C(\gamma)+(1-q)\left(3n-2(1+\theta)C'(\gamma)\right)\right)}{18(1-q)^2} = 0\,, \tag{30}$$

where $q = n(1-\gamma)$. Given the assumption about the efficiency of the compatibility-enhancing technology, the first-order condition (30) is equivalent to[40]

$$n(1+\theta)C(\gamma)+(1-q)\left(3n-2(1+\theta)C'(\gamma)\right)=0\,. \tag{31}$$

Totally differentiating (31) and solving for the foreign firm's choice of the compatibility level shows: [41]

$$\gamma = f(\theta, n)\,, \tag{32}$$

where the function $f(\theta, n)$ is characterized by $f_1 < 0$, $f_{11} > 0$, $f_2 > 0$, $f_{12} > 0$.

As expected, the higher the compatibility tax the lower the level of compatibility chosen by the firm. The negative effect of the tax on compatibility suggests that the foreign government experiences conflicting motives in its choice of policy toward compatibility. On one hand, the government's inability to use trade taxes deflects the strategic rent-shifting motive into the compatibility policy. Under price competition, this motive dictates that the government should increase the firm's marginal cost by imposing a compatibility tax. The tax commits the firm to a less-aggressive price-setting strategy and thereby moves the outcome of price competition between the firms to the Stackelberg equilibrium of the game in which the foreign firm is a leader. On the other hand, the tax also induces the firm to choose a lower level of γ, which leads to a reduction in the firm's profit. Therefore, the government's strategic trade motive for using the compatibility tax is moderated by the positive effect of compatibility on the firm's profit.[42]

The foreign government chooses the compatibility tax/subsidy to maximize the country's welfare, which is equal to the sum of the foreign firm's profit and tax revenues:

$$W_F(\theta) = \Pi_B(f(\theta,n),\theta)) + \theta C(f(\theta,n))s_B(f(\theta,n)) = \left(P_B(f(\theta,n)) - C(f(\theta,n))\right)s_B(f(\theta,n))\,.$$

The solution to this maximization problem defines the optimal compatibility tax/subsidy:

$$\theta^* = \arg\max_{\theta}\left[\left(P_B(f(\theta,n)) - C(f(\theta,n))\right)s_B(f(\theta,n))\right]. \tag{33}$$

[40] The condition ensuring that the compatibility-enhancing technology is sufficiently efficient that a prohibitive standard does not exist (for a given θ) is $C(1)(1+\theta) < 3(1-n(1-\gamma))$, where $C(1)$ is the unit cost of achieving full compatibility. More precisely, for a given γ the no-autarky condition is $s_B > 0$, which is equivalent to $C(\gamma)(1+\theta) < 3(1-n(1-\gamma))$. Because of the convexity of $C(\gamma)$, this inequality is harder to satisfy when γ is higher. To ensure that the condition is satisfied for any $\gamma \in [0, 1]$, we assume that $C(1)(1+\theta) < 3(1-n(1-\gamma))$. Note that this assumption implies that we again limit our analysis to the relatively efficient compatibility-enhancing technologies.

[41] The second-order condition for profit maximization is satisfied because, given our assumptions about the cost function $C(\gamma)$, the foreign firm's profit is concave in γ.

[42] Since $f_{1,2} > 0$, a greater network externality effect mitigates the foreign firm's negative response to the compatibility tax. Therefore, the foreign government will be more careful in heeding its rent-shifting incentive by means of the compatibility tax when the network externality effect is strong.

Proposition 4: The unconstrained optimal policy of the foreign government is the compatibility tax: $\theta^* > 0$. The weaker the network externality effect and the more efficient the conversion technology, the larger is the tax.

Proof: See Appendix.

Changes in the strength of the network externality effect (i.e., the parameter n) and in the efficiency of the compatibility technology lead to changes in the magnitude of the unconstrained optimal policy but not in its sign, because the optimal policy is always a compatibility tax rather than a subsidy. The government uses the compatibility-suppressing tax for rent-shifting purposes more aggressively if this policy is less costly in terms of the network externality benefit forgone due to imperfect compatibility, $n(1 - \gamma)$. This loss is smaller when the network externality effect is weak (i.e., n is small) and/or the compatibility technology is efficient enough to allow the firm to choose a high γ despite the tax.

5.2 Foreign best response compatibility policy

We are now ready to examine the best response compatibility policy of the foreign country. The function $f(\theta)$, which represents the level of compatibility chosen by the foreign firm in response to θ, is depicted by the thin curve in Figure 1. The foreign unconstrained optimal compatibility tax θ^* and the corresponding choice of compatibility level by the foreign firm, $\gamma^* = f(\theta^*)$, is represented by point F_2. A home compatibility standard more lax than γ^* (i.e., $\gamma < \gamma^*$) will not be binding on the foreign firm. Denoting the best response compatibility tax of the foreign country by $\theta^F(\gamma)$, we have $\theta^F(\gamma) = \theta^*$ for all $\gamma < \gamma^*$. In other words, when the home standard is not binding, the foreign government is free to choose its unconstrained optimal compatibility tax. This tax optimally balances the effects it has on price competition in the export market through its imposition of higher production cost on the foreign firm and through its influence on the firm's choice of compatibility between the products. The corresponding part of the foreign best response function is represented by the segment F_1F_2 of the thick dark curve in Figure 1.

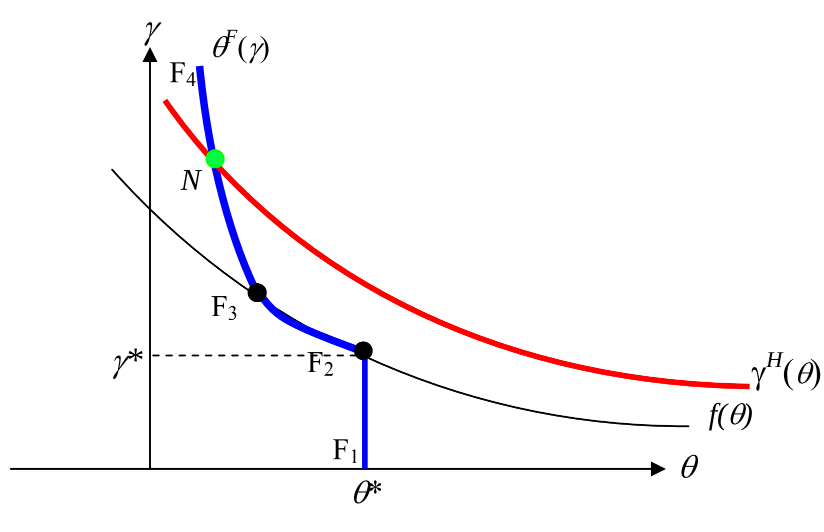


Figure 1.

For a home compatibility standard tighter than γ^* (i.e., $\gamma > \gamma^*$), the foreign government has two options. It can either impose a tax that will induce the foreign firm to choose the level of compatibility exactly equal to the compatibility standard or it can choose a tax that will not have any effect upon the foreign firm's choice of compatibility. If it does the former, the best response tax is given by the inverse function of $f(\theta)$—i.e., $\theta^F(\gamma) = f^{-1}(\gamma)$.

When the home country compatibility standard is binding and sufficiently tight, the foreign government does not need to be concerned that its compatibility policy reduces the foreign firm's incentives for making its technology compatible with the home firm's technology. Therefore, for a sufficiently tight compatibility standard in the export market the foreign government does not have to moderate its tax on compatibility. Instead, the government uses the tax θ only as an instrument for increasing the foreign firm's production cost and, thus, committing it to a more relaxed price-setting behavior in the home market. Thus, the foreign government chooses a compatibility tax different from $f^{-1}(\gamma)$. In Figure 1, the foreign best response curve $\theta^F(\gamma)$ deviates from the curve defined by $\theta = f^{-1}(\gamma)$. Specifically, the government chooses θ in such a way that the unit tax $\theta \cdot C(\gamma)$ is equal to the optimal export tax given by equation (24) in Section 4:

$$\theta \cdot C(\gamma) = \frac{3(1-n(1-\gamma)) - C(\gamma)}{4}. \tag{34}$$

The result is that when the standard imposed by the home government is sufficiently tight the foreign government's optimum rent-capturing compatibility tax $\theta^F(\gamma)$ is given by:

$$\theta = \frac{1}{4}\left(\frac{3(1-n(1-\gamma))}{C(\gamma)} - 1\right). \tag{35}$$

When the compatibility standard is not too tight (but tighter than γ^*) the foreign government will choose to follow $\theta^F(\gamma) = f^{-1}(\gamma)$ over (35). As γ approaches γ^* from above, $f^{-1}(\gamma)$ becomes arbitrarily close to the unconstrained optimal compatibility tax θ^*; whereas the right-hand side of (35) does not. Because the foreign welfare function is continuous, the convergence of $f^{-1}(\gamma)$ to θ^* as γ approaches γ^* implies that for some (perhaps small) range of compatibility standards the foreign government will set the tax inducing the firm to choose γ, which is exactly equal to the minimum level of compatibility required by the home country standard. The corresponding part of the foreign best response function is represented by the segment F_2F_3 in Figure 1.

The location of the intersection of the curves defined by (35) and $\theta = f^{-1}(\gamma)$ depends upon the strength of the network externality and the degree of convexity of the cost function $C(\gamma)$. In Figure 1, we depict a case in which the intersection of the curves is at point F_3, which is to the left of the point representing the unconstrained optimal tax θ^*. However, the curves may intersect to the right of the point θ^*. In that case, the graph will look like the one shown in Figure 2. In the figure, the discontinuity in the foreign country's best response function $\theta^F(\gamma)$ occurs at a value of γ at which the foreign country's welfare function has two local maxima yielding the same level of welfare. In both Figures 1 and 2, the section of the foreign best response function corresponding to (35) is depicted by the segment F_3F_4.

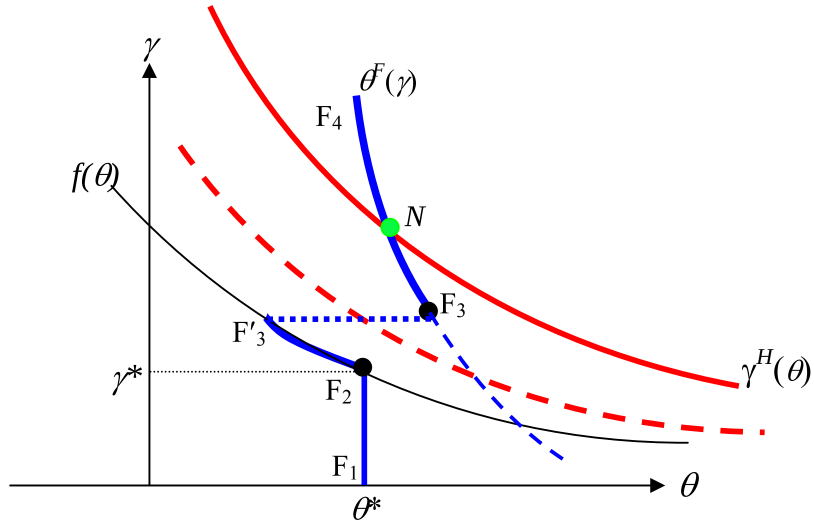


Figure 2.

5.3 The home country best response standard and the Nash equilibrium of the compatibility policy game

As we demonstrated in Section 3, if the home government's ability to tax imports is curtailed by a trade agreement it would prefer autarky to untaxed trade. When γ does not affect the foreign firm's marginal cost and there is no prohibitive (i.e., autarky-inducing) level of γ, then the minimum compatibility standard is ineffective as a policy instrument because the home government prefers less compatibility than the firm controlling the interface. However, this is not the case when compatibility affects the marginal cost of the foreign firm. When the cost function $C(\gamma)$ is increasing and convex, the compatibility standard can help the domestic firm to capture a greater share of pure profits in the imperfectly competitive market. Therefore, in this section, unlike the situation considered in Section 3, the compatibility standard allows the home government to shift rents in the "right" direction—i.e., to the home firm. In this setting, it is appropriate to say that the free trade agreement deflects the government's rent-shifting motivation into the standardization policy.

The best response function for the home country represents the level of compatibility standard, which maximizes the home country's welfare given the foreign country's compatibility-suppressing policy variable θ. The home country welfare is given by $W_H = a + 1/2 + n + (1 - 2\,n(1 - \gamma))s_B\,(1 - s_B) - s_B P_B$, where s_B and P_B are given, respectively, by (28) and (29).

The home government's best response compatibility standard will be binding for any compatibility policy chosen by the foreign government. To see this, note that given θ, a non-binding standard $\gamma^H < f(\theta)$ has no effect on the foreign firm's compatibility decision. By raising the standard above the level $f(\theta)$, the home government can achieve two goals. First, it can increase the value of the network based on the partially compatible technologies for the domestic users, and, second, it can reduce the pure rents earned by the foreign firm. Therefore the home government's best response standard will be above the level of compatibility preferred by the foreign firm: $\gamma^H \geq f(\theta)$. This can be verified by checking that $\partial W_H/\partial\gamma > 0$, when evaluated at $\gamma = f(\theta)$.

However, because compliance with the standard increases the marginal cost of the minority product relative to the dominant product, tightening the standard exacerbates the overadoption distortion. For sufficiently high γ, the negative welfare effect from tightening the standard dominates the positive welfare effect. This follows from the fact that $\partial W_H/\partial\gamma < 0$ when evaluated at γ sufficiently close to one. Therefore, the home government will not mandate full compatibility by setting the standard $\gamma^H = 1$. Instead, it will set the standard γ^H in such a way that $f(\theta) < \gamma^H < 1$.

For $\gamma^H > f(\theta)$, the best response function for the home country can be derived from the first-order condition for the home country welfare maximization: $\partial W_H/\partial\gamma = 0$. The home country's marginal welfare benefit from tightening the standard reflects an increase in the consumer surplus due to greater compatibility between the domestic and foreign products and an increase in the profit of the home firm due to less intense competition with the foreign firm. The marginal social cost of tightening the compatibility standard is reflected in the negative effect on the consumer surplus exerted by higher prices. Because the price difference between the rival products increases as the standard becomes tighter, some consumers switch to the less-preferred but cheaper domestic product, and those who continue to buy the foreign product pay a higher price. The welfare-maximizing standard sets the marginal social benefit of greater compatibility equal to the marginal social cost. The first-order condition for home welfare maximization is:[43]

$$\gamma = 1 - \frac{2}{3n} + \frac{(1+\theta)CC'(3-2q)(1-q) - 3C'(1-q)^3}{3n^2(1+\theta)C^2}, \tag{36}$$

where, as before, $q = n(1-\gamma)$. Equation (36) implicitly defines the home country's best response compatibility standard $\gamma^H(\theta)$. Equation (36) implicitly defines the home country's best response compatibility standard $\gamma^H(\theta)$. By totally differentiating (36), we verify that $\gamma^H(\theta)$ is decreasing and convex: $d\gamma^H/d\theta < 0$, $d^2\gamma^H/d\theta^2 > 0$. In Figures 1 and 2, $\gamma^H(\theta)$ is represented by the lighter solid curve.

When the pure-strategy Nash equilibrium in the compatibility policy game exists, the best response curves $\gamma^H(\theta)$ and $\theta^F(\gamma)$ intersect. Such equilibrium is stable if at the intersection point the absolute value of the slope of the curve corresponding to $\theta^F(\gamma)$ exceeds the absolute value of the slope of $\gamma^H(\theta)$.[44] However, it is possible that the curves $\gamma^H(\theta)$ and $\theta^F(\gamma)$ do not intersect. For example, referring to Figure 2, if the home country's best response curve $\gamma^H(\theta)$ is located very close to the curve $f(\theta)$, then $\gamma^H(\theta)$ might pass through the discontinuity in the foreign country's best response curve. This possibility is illustrated by the dashed curve $\gamma^H(\theta)$ drawn close to $f(\theta)$. In that case, there are no pure-strategy Nash equilibria in the compatibility policy game.

[43] Equation (36) is the reduced form of the first-order condition: $\partial W_H/\partial\gamma = 0$. Note that the second-order condition for home welfare maximization is satisfied because, given our assumptions about the cost function $C(\gamma)$, W_H is concave in γ.

[44] Whether the stability condition is satisfied depends upon the degree of the curvature of the cost function and the strength of the network externality. We are more likely to observe a stable Nash equilibrium when n is larger and the convexity of $C(\gamma)$ is greater.

6. International coordination of policies toward compatibility.

This section examines international agreements on policies toward compatibility. We begin by characterizing the combination of the compatibility policies that is jointly efficient for the two countries. Having identified the globally efficient policy combination, we analyze the inefficiencies that are present in the non-cooperative equilibrium. By comparing the Nash equilibrium outcome with the globally efficient policy combination, we can discern the changes in the countries' policies toward compatibility that should arise as a result of an international treaty.

6.1 Jointly efficient compatibility policies

The jointly efficient policy combination maximizes the aggregate welfare of the two countries:

$$W_W = a + 1/2 + n + (1 - 2\,n(1 - \gamma))s_B\,(1 - s_B) - s_B C.$$

When foreign consumers are not involved, the problem of the two-country joint welfare maximization under international duopoly is equivalent to the welfare maximization problem for a closed economy with a domestic duopoly.

With the increasing unit cost of compatibility, raising γ leads to a greater marginal cost difference between the firms. As a result, enhancing compatibility has a mixed effect on the aggregate welfare. While the increase of γ raises the network-related component of the consumer utility, the increase in the unit cost of compatibility, $C(\gamma)$, worsens the inefficiency due to overadoption of the minority technology.

If the compatibility tax (or subsidy) were the only policy instrument available to the governments for joint welfare maximization, its optimal level would reflect the opposing welfare effects of compatibility. On one hand, the problem of overadoption of the minority technology by users suggests that the policy should be a tax ($\theta \geq 0$). On the other hand, the minority firm's socially insufficient incentive for making its product compatible with the rival's product calls for a compatibility-enhancing subsidy ($\theta < 0$). A high enough tax θ can completely eliminate the problem of overadoption but will also suppress the minority firm's incentive to invest in compatibility. Therefore, if the countries use both policy instruments (i.e., the compatibility tax and the standard), it would be optimal for them to target the overadoption distortion with a *positive* tax and induce the firm to select the socially optimal level of compatibility by means of a binding standard.[45]

For a given value of γ, the jointly optimal *unit* compatibility tax, $\theta^W \cdot C(\gamma)$, is equal to the closed economy optimal production tax on the minority duopolist defined by equation (16) in Section 3. Therefore, the jointly optimal compatibility tax is given by:

$$\theta^W(\gamma) = \frac{2 - n(1-\gamma)}{1 - 2n(1-\gamma)}. \tag{37}$$

[45] In a closed economy context, the fully optimal social outcome could be achieved if the government could impose the consumption tax on the minority technology users and give the compatibility-enhancing subsidy to the minority technology producer. Although we abstract from such policies in our open economy model, if they were feasible they would allow the countries to achieve a higher level of welfare.

The jointly optimal compatibility standard, $\gamma^W(\theta)$, can be derived from the first-order condition for the maximization of the aggregate world welfare with respect to γ. The condition equates the marginal social benefit of tightening the standard to the marginal social cost:

$$\frac{\partial\big((1-2n(1-\gamma))s(1-s)\big)}{\partial\gamma} = \frac{\partial(sC)}{\partial\gamma}.$$

Rearranging the first-order condition, we find that $\gamma^W(\theta)$ is given implicitly by

$$\gamma = 1 - \frac{1}{n} + \frac{9(1-q)^3(C'-n) + nC^2(1+\theta)\big(3-q(\theta-2)\big)}{nCC'(1+\theta)\big(5-2q(\theta-2)-\theta\big)}, \qquad (38)$$

where, as before, $q = n(1-\gamma)$.[46]

Totally differentiating (38) shows that the function $\gamma^W(\theta)$ is characterized by $d\gamma^W/d\theta < 0$ and $d^2\gamma^W/d\theta^2 > 0$. Comparing the levels of compatibility jointly preferred by the governments with the level preferred by the foreign firm, we confirm that for any tax $\theta > 0$ the jointly optimal standard is binding on the firm: $f(\theta) < \gamma^W(\theta)$. Moreover, since the rent-shifting motive is not present under joint welfare maximization, the jointly optimal compatibility standard will be lower than the best response standard of the home government: $\gamma^W(\theta) < \gamma^H(\theta)$. Therefore, we find that for $\theta > 0$: $f(\theta) < \gamma^W(\theta) < \gamma^H(\theta)$.

In Figures 3 (a, b, c) the jointly optimal standard $\gamma^W(\theta)$ and tax $\theta^W(\gamma)$ are depicted by thin curves. The intersection of the curves corresponding to $\theta^W(\gamma)$ and $\gamma^W(\theta)$ gives us the efficient combination of the tax and the standard: (θ^W, γ^W). Depending upon the parameter values, this combination can be represented by the points *W, W**, and *W*** in the figures.

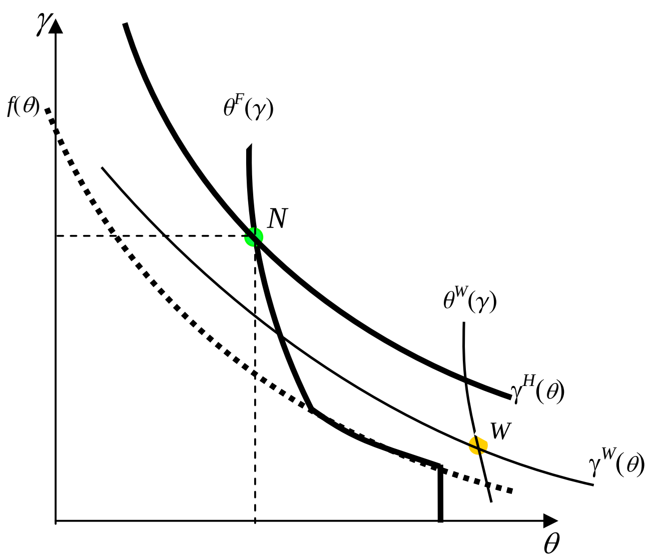


Figure 3a.

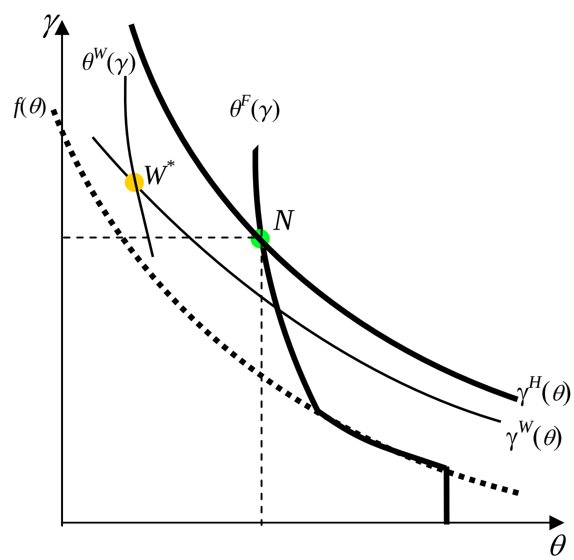


Figure 3b.

[46] The second-order conditions for the maximization of W_W with respect to γ and θ require $W_{\gamma\gamma}, W_{\theta\theta} < 0$, and $W_{\gamma\gamma}W_{\theta\theta} - W_{\gamma\theta}^2 > 0$. These conditions are satisfied if the cost function $C(\gamma)$ is characterized by a sufficient degree of curvature (or convexity). We assume that this is the case in our model.

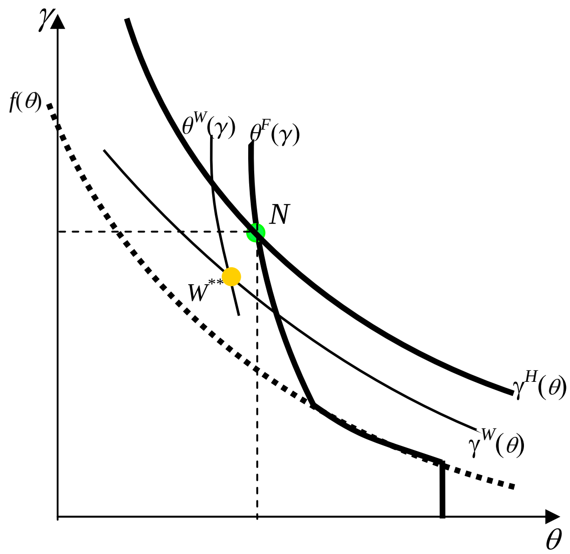


Figure 3c.

6.2 Pareto-improving agreements on compatibility

By comparing the efficient policy combination with the policies emerging in a non-cooperative equilibrium, we can determine the likely features of a negotiated agreement on compatibility. As Figures 3 (a, b, c) illustrate, the exact magnitude of the jointly efficient policies relative to the Nash equilibrium policies can vary depending upon the parameters of the model. However, regardless of the "disagreement" levels of welfare, which countries can guarantee themselves without policy cooperation, an agreement based on an efficient policy combination can always be reached in an international negotiation, assuming that the countries can make welfare transfers to one another as part of the agreement. The assumption that countries are able to make such welfare transfers can be motivated by the fact that trade negotiations frequently involve cross-country linkages among a large number of issues. Therefore, it is appropriate to assume that countries use these linkages to effectively make transfers.[47]

Comparing equations (35) and (37), we observe that for any γ the size of the jointly optimal compatibility tax relative to the non-cooperative compatibility tax of the foreign country depends upon the strength of the network externality, n.[48] Two qualitatively distinct examples are shown in Figures 3 (a, b). When the network externality is strong (i.e., n is large), the overadoption distortion is more severe. Therefore, the jointly optimal compatibility tax targeting this distortion tends to be larger than the best response tax of the foreign country. This possibility is illustrated in Figures 3 (a) in which the point W corresponding to the jointly optimal policy

[47] Hoekman (1993), for example, pointed out that negotiating countries exchange concessions both within and across issues. Cross-issue linkages may allow agreement even if within-issue exchange of concessions proves insufficient to generate an improvement on the status quo for all concerned.

[48] By rearranging the inequality $\theta^W(\gamma) > \theta^F(\gamma)$, we obtain: $C > (1-q)(1-2q)/(3-2q)$. When the cost function is sufficiently convex and satisfies this paper's assumptions regarding $C(\gamma)$, this inequality holds for large values of n but is reversed when n is small.

combination is located to the right and below the point *N,* which represents the Nash equilibrium policy outcome. By contrast, when the network externality is weak (i.e., n is small) the overadoption distortion is less severe and the jointly optimal tax tends to be smaller than the foreign best response tax. This situation is depicted in Figure 3 (b), in which the curve corresponding to $\theta^W(\gamma)$ is drawn closer to the vertical axis of the graph than the curve representing $\theta^F(\gamma)$. As a result, the jointly optimal policy combination point W^* is located to the left and above the Nash equilibrium point *N*.

Figures 3 (a, b) illustrate two principal ways in which the governments can use the compatibility factor to raise the component of their joint welfare associated with a network externality effect. First, they can increase compatibility *between* the two rival networks by tightening the standard. Second, they can take the advantage of perfect compatibility *within* the dominant network by expanding its size. However, as we discussed above, under the increasing unit cost of compatibility there is a conflict between the two ways of increasing the network benefits. When γ affects the marginal cost difference between the rival firms, increasing compatibility between the networks worsens the overadoption distortion, which by definition means worsening the inefficiency of the dominant network's size. The optimal size of the dominant network can be restored by means of the tax aimed at the overadoption of the minority technology. But the drawback of the tax is that it increases the cost of achieving greater compatibility between the rival networks.

When the network externality is strong, it is more efficient to emphasize the benefits of perfect compatibility within the dominant network than to achieve a higher (but still imperfect) compatibility between the rival networks. Therefore, when n is large we may be able to observe the situation depicted in Figure 3 (a). The figure illustrates the example in which the lack of policy cooperation in the area of compatibility leads to excessive compatibility between the rival networks at the cost of overadoption of the minority technology. The latter results in the sub-optimal size of the dominant network within which the users enjoy perfect compatibility. Therefore, without policy cooperation, too much of the imperfect inter-network compatibility is substituted for the perfect compatibility within the dominant network.

When the network externality is weak (small n), raising compatibility between the rival networks has a greater positive impact on the countries' joint welfare than the expansion of the dominant network. This is the case illustrated in Figure 3 (b), in which the jointly optimal policy combination has a tighter standard and a smaller tax than the non-cooperative policy combination. To understand why this is so, note that the weakness of the network externality implies that the overadoption distortion is small even for high values of γ. Therefore, a smaller tax is needed to target this distortion. As a result, the governments have a stronger joint incentive to increase compatibility between the minority network and the dominant networks.

One notable feature of the examples illustrated in Figures 3 (b, c) is the fact that the jointly optimal tax, which targets only the overadoption distortion, is less than the unilateral rent-capturing tax of the foreign country. As a result, the standard adopted by the countries under the agreement maximizing their joint welfare may be tighter than even the protectionist unilateral standard of the home country (as shown in Figure 3 (b)).

7. Discussion and Conclusion

In the Introduction, we mentioned the disciplines for policing technical regulatory barriers that exist within the legal system of the WTO. Similar disciplines exist within the legal systems of the European Union and NAFTA. Reduced to bare essentials, all of these disciplines are based on the requirement that regulatory policy objectives be achieved in the manner that minimizes impediments to commerce and open markets. This requirement in the international law is referred to as the least restrictive means principle.[49]

The exact language used to formulate this principle in the international trade agreements varies. The original GATT agreement (Article XX) states that domestic regulatory measures should not constitute a means of "arbitrary or unjustifiable discrimination between countries" or "a disguised restriction on international trade." Article 2.1 of the Standards Code adopted during the Tokyo Round says that governments should not adopt standards and regulations "with a view of creating obstacles to international trade" or with " the effect of creating unnecessary obstacles to international trade". The Uruguay Round Agreement on Technical Barriers to Trade (TBT) proposes a balancing test, which requires measuring negative trade effects against the putative benefits of regulation. As stated in Article 2.2 of the TBT Agreement, "technical regulations shall not be more trade-restrictive than necessary to fulfill a legitimate objective, taking account of the risks non-fulfillment would create." Article VI (4) (b) of the General Agreement on Trade in Services (GATS) introduced the *necessity test*, which invalidates domestic standards and regulations that are "more burdensome than necessary to ensure the quality of the service." In the context of standards and regulations governing the supply of services characterized by network externalities, quality-of-service standards include standards aimed at ensuring the integrity of the networks.

Legal scholars identify the rationale behind the least trade restrictive means principle with the goal of maximizing "the sum of net gains from trade and net gains from regulation" or "the net benefits of domestic consumers and producers as well as foreign producers" (Trachtman, 1998). Most economists agree that the appropriate economics interpretation of the least restrictive measure is one that identifies it with the global welfare-maximizing measure, or the measure that a welfare-maximizing government would choose if all producers supplying the domestic market were domestic by origin (see, for example, Baldwin (1970), Engel (2000), and Fischer and Serra (2000)).

Our model suggests that the unilaterally optimal compatibility standard of the importing country is "more burdensome than necessary," in that it does not maximize the joint welfare of the importing and the exporting countries. However, as should be apparent from the preceding formal analysis, an exchange of commitments between countries to use the least restrictive measures in their own domestic markets is necessary but not sufficient to ensure the globally efficient levels of trade and compatibility, regardless of whether international welfare transfers are feasible or not. Although such commitments guarantee access to the markets of the importing countries, they do not restrain the compatibility policies of the exporting countries. The latter are important because global efficiency requires an agreement specifying not only the relative size of the networks based on the imported and the import-competing products (which can be pinned down by the reciprocal market access agreement alone)

[49] Other important principles of the internal law designed to prevent regulatory protectionism are non-discrimination, obligation to give reasons and advance notice, transparency, precedence of performance regulations over design regulations, mutual recognition, reliance on international standards, and the 'sham' principle. Alan Sykes (1995) provided a thorough overview of these principles and suggested that all of them may be considered as corollaries of the least restrictive means principle.

but also the level of compatibility between the rival products within each domestic market. Only such an agreement can guarantee that exporting as well as importing countries internalize the costs that their compatibility policies impose on trade partners.

This point is related to the conclusion reached by Bagwell and Staiger (2001), who suggest that if the only type of the negative externality the countries can exert on each other is associated with the terms-of-trade effect, then the global efficiency can be attained through a reciprocal market access agreement setting the world prices at the efficient level.[50] However if, in addition to the terms-of-trade driven externalities, countries exert other types of cross-border externalities on their trade partners, then the reciprocal exchange of commitments regarding market access alone will be insufficient to achieve global efficiency.

In our model, the compatibility policy in the exporting country exerts negative externality on the importing country not only through the price charged by the exporting firm but also through the exporting firm's choice of the degree of compatibility with the import-competing product.[51] The latter type of cross-border externality arises because incomplete compatibility of the foreign product with the domestic product undermines the integrity of the domestic network. When the exporter's choice of compatibility level affects its unit cost, the exporting government's compatibility policy has implications for both the relative size of the rival networks in the export markets and the level of compatibility between them. In this situation, an agreement on market access can ensure the optimal relative size of the networks but will not ensure the optimal level of inter-network compatibility. Therefore, in the context of our model, the globally efficient trade agreement should incorporate commitments not only with regard to market access but also with regard to the level of compatibility between the products.

Specifically, if the governments can use trade taxes and subsidies as well as national policies toward compatibility of the kind we have described in this paper, the optimal trade agreement between them must include both reciprocal market access commitments that pin down the relative size of the rival networks within each national market and commitments regarding the level of compatibility between the rival products. After that, the countries can choose their preferred trade taxes and compatibility regulations only to the extent that their choices do not alter their market-access commitments and their commitments on inter-network compatibility.

Appendix

Proof of Proposition 4:

Omitting the subscript denoting the foreign firm's variables, the first-order condition for welfare maximization is given by

$$P_\gamma f_\theta s + s_\gamma f_\theta P - C' f_\theta s - s_\gamma f_\theta C = (P_\gamma - C')\, f_\theta s + (P - C) s_\gamma f_\theta = 0,$$

which is equivalent to

$$(P_\gamma - C')s + (P - C)s_\gamma = 0. \tag{A1}$$

[50] Bagwell et al. (2002) also argue that the need to supplement market access commitments with a cross-country linkage of specific regulatory and trade policies arises only if the countries do not have access to the full spectrum of sufficiently flexible trade policy instruments for offsetting the terms-of-trade effects of domestic regulations.

[51] The effect of the rent-shifting policies on the prices set by the imperfectly competitive exporters is similar to the effect on world prices of the trade policies aimed at improving the exporting country's terms of trade. Both types of policies lead to higher export prices.

Given our assumptions about the cost function $C(\gamma)$, the welfare function W_F is concave in θ, and the second-order condition is satisfied.

Differentiating the foreign firm's market share (29) with respect to γ gives:

$$s_\gamma = \frac{(1+\theta)\big(Cn - C'(1-n(1-\gamma))\big)}{6(1-n(1-\gamma))^2}. \tag{A2}$$

Because we are concerned with $\theta > -1$, the sign of s_γ is the same as the sign of $Cn - C'(1-n(1-\gamma))$. Notice that, given the assumptions about the cost function (i.e., $C'>0$, $C''>0$, $\lim_{\gamma\to 1} C'(\gamma) = \infty$ and $C(0)= C'(0) = 0$), we have $C(\gamma) < \gamma C'(\gamma)$, $\forall\gamma \in [0, 1]$. Therefore, $C(\gamma) < C'(\gamma)$, $\forall\gamma \in [0, 1]$. Because our assumption about the efficiency of the converter (see footnote 21) implies that $n \in (0, ½)$, it follows that $Cn < C'(1-n(1-\gamma))$. Therefore, for $\theta > -1$, we have $s_\gamma < 0$.

Under imperfect competition, the mark-up of the foreign firm is positive. Hence, we have $(P - C)s_\gamma < 0$. Therefore, any θ^* (greater than minus one) that solves equation (A1) and under which the foreign firm has a positive market share ($s(\theta^*) > 0$) must satisfy $P_\gamma(\theta^*) - C'(\theta^*) > 0$, $\forall\gamma \in [0, 1]$. This condition is equivalent to $C'(2\theta^* -1) > -3n$, $\forall\gamma \in [0, 1]$. It follows that $\theta^*>1/2$. ■

References

1. Auriol, E. and M. Benaim, (2000), "Standardization in Decentralized Economies," *American Economic Review* 90, 550-570.
2. D'Aspremont, C, J. J. Gabszewics and J.-F. Thisse (1979), "On Hotelling's "Stability in Competition," *Econometrica* 47, 1145-1150.
3. Bagwell, Kyle, Mavroidis, Petros C., and Robert W. Staiger (2002) "It's a Question of Market Access," *The American Journal of International Law* 96, 56-76.
4. Bagwell, Kyle, and Robert W. Staiger, (2001) "Domestic Policies, National Sovereignty and International Economic Institutions," *Quarterly Journal of Economics* 116, 519-562.
5. Baldwin, Richard E. (1970), *Nontariff Distortions of International Trade*, Washington, D.C.: Brookings Institution.
6. Berg, S. V., (1988), "Duopoly Compatibility Standards with Partial Cooperation and Standards Leadership," *Information Economics and Policy* 3, 35-53.
7. Brander, J. A., and B. J. Spencer (1984), "Tariff Protection and Imperfect Competition," in H. Kierzkowski (ed.), *Monopolistic Competition and International Trade*. Clarendon Press, Oxford. pp. 194-206.
8. Braunstein, Y., and L. White, (1985), "Setting Technical Compatibility Standards: An Economic Analysis," *The Antitrust Bulletin* 30, 337-355.
9. Cheng, Leonard K. (1988), "Assisting Domestic Industries under International Oligopoly: The Relevance of the Nature of Competition to Optimal Policies," *American Economic Review* 78, 746-58
10. Choi, Jay Pil (1997), "The Provision of (Two-Way) Converters in the Transition Process to a New Incompatible Technology," *Journal of Industrial Economics* 45, 139-53.

11. Clarke, P., (1999), “U.S. warns Europe to open 3G standards,” *Electronic Engineering Times,* Manhasset, July 19, 1.
12. Clendenin, M., (2004) “Intel balks at China's Wi-Fi security spec -- Government pressured to reconsider WAPI scheme,” *Electronic Engineering Times*, March 15, 1.
13. Dixit, A. K. (1984), “International Trade Policy for Oligopolistic Industries,” *Economic Journal* 94, 1-16.
14. Dixit, A. K. (1988), “Anti-Dumping and Countervailing Duties under Oligopoly,” *European Economic Review* 32, 55-68.
15. David, P. A., and J. A. Bunn, (1988), “The Economics of Gateway Technologies and Network Evolution: Lessons from Electricity Supply History,” *Information Economics and Policy* 3, 165-202.
16. David, P. A. and S. Greenstein, (1990), “The economics of compatibility standards: An introduction to recent research,” *Economics of Innovation and New Technology* 1, 3-41.
17. David, P. A and W. E. Steinmueller, (1994), "Economics of Compatibility Standards and Competition in Telecommunication Networks," *Information Economics and Policy* 6, 217-242.
18. Eaton, J. and Grossman, G. (1986): “Optimal Trade and Industrial Policy under Oligopoly,” *Quarterly Journal of Economics* 101, 383-406.
19. Economides, N. (1984), “The Principle of Minimum Differentiation Revisited,” *European Economic Review* 24, pp. 345-368.
20. Economides, N., (1988), "Variable Compatibility without Network Externalities," Discussion Paper No. 145, Studies in Industry Economics, Stanford University.
21. Economides, N., (1991), "Compatibility and the Creation of Shared Networks," in *Electronic Services Networks: A Business and Public Policy Challeng*e, edited by Margaret Guerrin-Calvert and Steven Wildman, Praeger Publishing Inc., New York: 1991.
22. Engel, E., (2000), “Poisoned Grapes, Mad Cows and Protectionism,” *Journal of Policy Reform* 4, 91–111.
23. European Commission, (2001), *Report on United States Barriers to Trade and Investment.* Brussels, July 2001.
24. Farrell, J., and G. Saloner, (1992), "Converters, Compatibility, and the Interfaces," *Journal of Industrial Economics* 40, 9-36.
25. Fischer, R. and Serra P., (2000), “Standards and Protection,” *Journal of International Economics* 52, 377-400.
26. Fung, K. C. (1988), “Strategic Trade Policies, Differentiated Duopoly and Intra-industry Trade,” *International Economic Journal* 2, 19-34.
27. Gabszewicz, J.J. and Thisse, J.F., (1979), “Price competition, quality and income disparities,” *Journal of Economic Theory* 2, 340–359.
28. Gandal, Neil, (2001), “Quantifying the Trade Impact of Compatibility Standards and Barriers: An Industrial Organization Perspective,” in Keith E. Maskus and John S. Wilson, editors, *Quantifying the Impact of Technical Barriers to Trade: Can It Be Done?* (Ann Arbor: University of Michigan Press), pp. 137-153.
29. Gandal, N., and O. Shy, (2001), “Standardization Policy and International Trade," *Journal of International Economic*s 53, 363-383.
30. Grindley, P., Salant, D. and L. Waverman (1999), “Standards Wars: The Use of Standard Setting as a Means of Facilitating Cartels – Third Generation Wireless Telecommunications Standard Setting,” *International Journal of Communications Law and Policy* (IJCLP), Issue 3, Summer, [IJCLP Web-Doc 2-3-1999; http://www.ijclp.org],

31. Hoekman, Bernard (1993), "Multilateral Trade Negotiations and Coordination of Commercial Policies," in Stern, R. (ed.), *The Multilateral Trading System: Analysis and Options for Change.* Ann Arbor: University of Michigan Press.
32. Katz, M. and C. Shapiro, (1985), "Network Externalities, Competition and Compatibility," *American Economic Review* 75, 424-440.
33. Katz, M. and C. Shapiro, (1986), "Product Compatibility Choice in a Market with Technological Progress," *Oxford Economic Papers* 38, 146-165.
34. MacKie Mason, Jeffrey and Janet S. Netz, (2002), "Manipulating Interface Standards as an Anti-Competitive Strategy," Mimeo, presented at 2002 Telecommunications Policy Research Conference.
35. Mannion, P. and M. Clendenin, (2003) "China throws Wi-Fi a curve -- Mandatory security scheme incompatible with 802.11," *Electronic Engineering Times*, December 22, 1.
36. Maskus, Keith, Otsuki, Tsunehiro and John S. Wilson, (2001), "An Empirical Framework for Analyzing Technical Regulations and Trade," in Keith E. Maskus and John S. Wilson, editors, *Quantifying the Impact of Technical Barriers to Trade: Can It Be Done?* (Ann Arbor: University of Michigan Press), pp. 29-58.
37. McLean, Heather, (2001), "Intel WLAN can forget about Europe," accessed at www.silicon.com/a47267 on September 10, 2001.
38. Neven, D., Norman, G. and J.-F., Thisse, (1991) "Attitudes towards foreign products and international price competition," *Canadian Journal of Economics* 24, 1-11.
39. Organization for Economic Development and Cooperation (2000a), *An Assessment of the Costs for International Trade in Meeting Regulatory Requirement*s. TD/TC/WP(99)8/FINAL, Paris: OECD.
40. Organization for Economic Development and Cooperation (2000b), *Standardization and Regulatory Reform: Selected Cases.* TD/TC/WP(99)47/FINAL, Paris: OECD.
41. Organization for Economic Development and Cooperation (2001), *Standards-Related Barriers and Trade Liberalization: Telecommunications Sector Study – Interim Report*. TD/TC/WP(2001)11/REV1, Paris: OECD.
42. O'Neil, Kim (2001), "Galileo - European Satellite Navigation System," Advanced Aviation Technology Ltd. accessed at http://www.aatl.net/publications/galileo.htm on September 26, 2001.
43. Spencer B. and J. Brander, (1983), "International R & D Rivalry and Industrial Strategy", *Review of Economic Studies* 50, 707-722.
44. Sykes, A.O. (1995), *Product Standards for Internationally Integrated Goods Markets.* Brookings, Washington.
45. Trachtman, J., (1998) "Trade and . . .. Problems, Cost-Benefit Analysis and Subsidiarity," *European Journal of International Law* 9 (1), 32-85.
46. Wong, W., (2001), "Wireless networking wars hit speed bump," CNET News.com at http://news.cnet.com/news/0-1004-200-4952376.html accessed on February 26, 2001.
47. Zhou D., Spencer B., and I. Vertinsky, (2002), "Strategic trade policy with endogenous choice of quality and asymmetric costs," *Journal of International Economics* 56, 205–232.

Network Structure and the Long Tail of Electronic Commerce[1]

Gal Oestreicher-Singer and Arun Sundararajan

Leonard N. Stern School of Business, New York University
44 West 4th Street, New York, NY 10012
goestrei@stern.nyu.edu, asundara@stern.nyu.edu

(Draft: please do not circulate)

Summary: We report on a research project that studies how network structures affect demand in electronic commerce, using daily data about the graph structure of Amazon.com's co-purchase network for over 250,000 products. We describe how the presence of such network structures alters demand patterns by changing the distribution of traffic between ecommerce web pages. When this traffic distribution generated by the presence of the network is less skewed than the intrinsic or "real world" traffic distribution, such network structures will even out demand across products, leading to a demand distribution with a longer tail. We estimate an econometric model to validate this theory, and report on preliminary confirmation by contrasting the demand distributions of products within over 200 distinct categories on Amazon.com. We measure the overall extent to which a product influences the network by adapting Google's PageRank algorithm, applying it to a weighted composite of graphs over four distinct 7-day periods, and we characterize the demand distribution of each category using its Gini coefficient. Our results establish that categories whose products are influenced more by the network structure have significantly flatter demand distributions, which provides an additional explanation for the widely documented phenomenon of the long tail of ecommerce demand.

1. Introduction

There are numerous networks associated with electronic business. Some of these can describe the relationships between consumers who communicate product information and influence each others purchasing, others can describe how the demand for different products are related based on shared purchasing patterns, and yet others may describe the patterns of trade between firms.

[1] We thank Vasant Dhar, Nicholas Economides, William Greene, Panos Ipeirotis, Roy Radner and seminar participants at New York University, the Second Annual Statistical Challenges in Electronic Commerce Research Symposium and the 34th Annual Telecommunications Policy Research Conference for their feedback.

In this paper, we report on a project that aims to model and measure the effect that such network structures have on outcomes in electronic commerce. A good example of such a structure is the network of product pages on an ecommerce site. Each product on an ecommerce site has a network position, which is determined by the products it links to, and those that link to it. If one imagines the process of browsing an ecommerce site as being analogous to walking the aisles of a physical store, then the ecommerce aisle structure is the this graph of interconnected products, and the network position of a product in this graph is analogous to its virtual shelf placement. A product that is linked to by an intrinsically popular one is likely to enjoy an increase in sales on account of this aspect of its network position. A product linked to by hundreds of others is likely to get more "network traffic" more than one linked to by just a few. Thus, both the structure of the networks and the nodes that comprise them seem to matter.

We measure the extent to which the position of a product in such a network structure will affect its demand, based on the idea that the network structure redirects the flow of consumer attention, which results in a redistribution of traffic and demand. The manner in which attention is redistributed can be measured using certain properties of the network structure, and these properties can be associated with observed variations in both individual and aggregate product demand. One specific prediction of our theory is that network structures with common degree distributions will even out traffic between products (nodes), thereby reducing demand inequity between products.

Our preliminary evidence is base on econometric estimates of how the intensity of the network structure affects the demand distributions of over 250,000 products within over 200 distinct categories on Amazon.com. Briefly, we compute an adapted version of the PageRank coefficient for each node of four composites of seven daily instances of Amazon.com's co-purchase network. We characterize the demand distribution of each category by constructing its Lorenz curve and measuring its Gini coefficient. We show that when network structure has a greater influence (when the average Weighted PageRank is higher) on a category, its demand distribution displays significantly lower inequity (its Gini coefficient is significantly lower). In other words, the presence of the network structure flattens ecommerce demand distribution. This provides a new explanation for the widely documented long tail of ecommerce demand.

A recent literature on "network games" (for example, Bramoulle and Kranton, 2005, Galeotti et. al., 2006, Sundararajan, 2006) has begun studying theoretically how the properties of the equilibria of specific classes of games played on a graph depend on network structure. A more detailed survey is available in Sundararajan (2006). In empirical work, Aral, Brynjolfsson and Van Alstyne (2006) study the influence of network structures on IT worker productivity, relating technology use and social network characteristics to economic measures, and providing evidence that the structure and size of workers' communication networks, including such social network metrics as betweenness and structural holes are highly correlated with performance. Features based on network structure have been shown to improve the predictions of data mining models used for targeted marketing (Hill, Provost and Volinsky, 2006). There has been some prior research in marketing which aimed to assess the structure of preferences for products based on purchase data. These studies have used scanner panel data, and are based on a similar notion: that such data contain important information based on revealed preference about the structure of brand preferences both within and across product categories. For example, the time series of purchasing data has been used to compute segment-product distances (Ramaswamy and DeSarbo, 1991), segment consumers with respect to brand preferences (Russell and Kamakura, 1997, Matthias, Bauer and Hammerschmidt, 2002) and build probabilistic models that provide spatial representations of product structure (Erdem, Imai, and Kean,1999). These are based on co-purchase bundles, and none of these papers has used a co-purchase network, or exploited any structural properties of these networks in making inferences about the nature of demand.

Network structures have received a significant amount of attention from sociologists studying

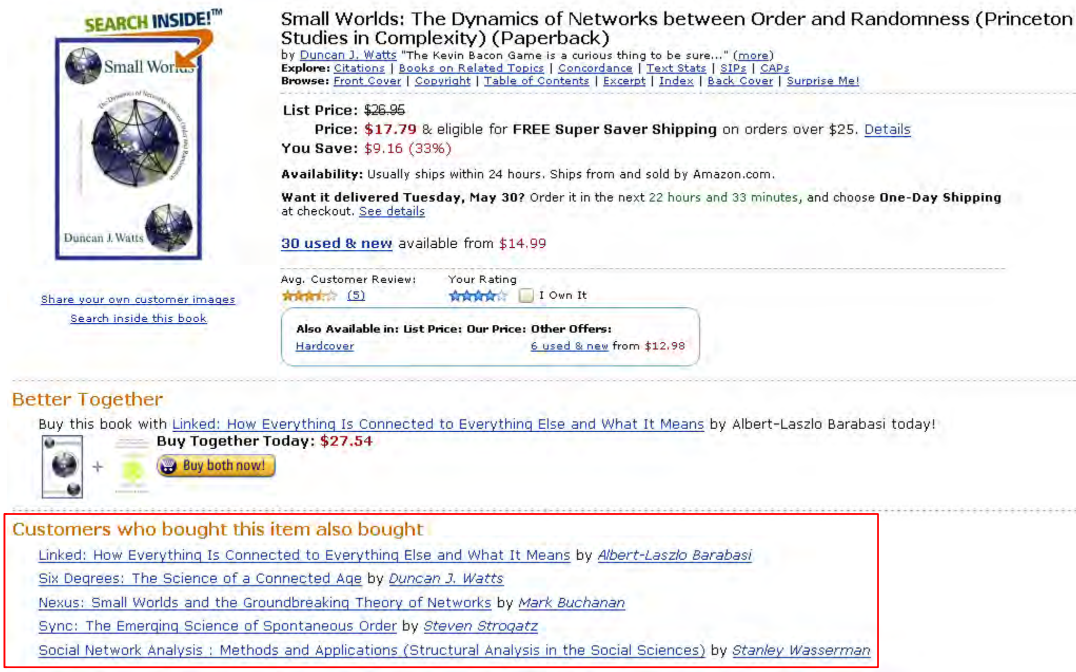


Figure 2.1: Illustrates sample copurchase links on Amazon.com

relationships between people, from physicists and computer scientists studying the Web. Our work is also related to a growing literature on using network structures to create sophisticated ranking algorithms, one well-known contribution being the PageRank algorithm of Brin and Page, 1998 (for further information, see Langville and Meyer, 2005).

2. Data

We collect daily product, pricing, demand and “network” information for over 250,000 books sold on Amazon.com. Each product on Amazon.com has an associated webpage. Those pages each have a set of “co-purchase links”, which are hyperlinks to the set of products that were co-purchased most frequently with this product on Amazon.com. This set is listed under the title "Customers who bought this also bought" and is limited to 5 items (see Figure 3.1).

Conceptually, the co-purchase network is a directed graph in which nodes correspond to

Figure 2.2: Illustrates a subset of paths in the graph

products, and edges to directed co-purchase links. We collect data about this graph using a Java-based crawler, which starts from a popular book and follows the co-purchase links using a depth-first algorithm. At each page, the crawler gathers and records information for the book whose webpage it is on, as well as the co-purchase links on that page, and terminates when the entire connected component of the graph is collected. This process is repeated daily.

We have chosen to focus on books because they are in the product category with by far the largest number of individual titles, whose product set is relatively stable (compared to electronics, for instance), and it seems to be a class of products for which the network we study would actually matter.

The data collection began in August 2005 and is currently ongoing. The graph is traversed every day, and we thus have over 300 co-purchase graphs collected so far. Apart from the co-purchases, each book's ISBN, list price, sale price, category affiliation, secondary market activity, author, publisher, publication date, and consumer ratings are gathered. A sample part of the graph is illustrated in Figure 3.2

The following data that we gather are available for each book on the copurchase graph, for each day.

ASIN: a unique serial number given to each book by Amazon.com. Different editions and different versions have different ASIN numbers.

List Price: The publisher's suggested price.

Sale Price: The price on the Amazon.com website that day.

Copurchases**:** ASINs of the books that appear as its copurchases.

SalesRank: The sales rank is a number associated with each product on Amazon.com, which measures its demand of relative to other products. The lower the number is, the higher the sales of that particular product.

Category Affiliation: Amazon.com uses a hierarchy of categories to classify its books. Thus, each book is associated with one or more hierarchical lists of categories, starting with the most general category affiliation, and ending with the most specific one. For example:

Subjects > *Business & Investing* > *Biographies & Primers* >*Company Profiles*

(for "The Search" by John Batelle).

Using the second level of the hierarchy, there are 1472 such categories across all books sold, of which between 203 and 225 have 100 or more nodes represented in our copurchase network, the minimum category size we analyze.

Author: The name of the author or authors of the book.

Publisher: The name of the publisher of the book.

Publication date: The date of publication of the book (by that publisher).

3. Characterizing ecommerce demand and its distribution

In order to relate the network position of a product to variation in its demand, we do the following:

1. Infer demand levels from the SalesRank data reported by Amazon.com, thereby associating a periodic demand level with each product.

2. Characterize the extent to which the network structure influences a product based on its network position.

3. Associate variation in (2) with variation in (1) at both a product-specific level of analysis and at a group-specific level of analysis.

3.1. Estimating demand from Amazon.com salesranks

To estimate the actual level of demand, $Demand(j)$, of a book from its sales rank, $SalesRank(j)$, we use a conversion model suggested by Goolsbee and Chevalier (2003) and by Brynjolfsson, Hu and Smith (2003).

$$Log[Demand(j)] = a + bLog[SalesRank(j)] \tag{3.1}$$

This formula to convert SalesRank information into demand information was first introduced by Chevalier and Goolsbee (2003). Their goal was to estimate demand elasticity. Their approach was based on making an assumption about the probability distribution of book sales, and then fitting some demand data to this distribution. They choose the standard distributional assumption for this type of rank data, which is the Pareto distribution (i.e., a power law).

In a later study, Brynjolfsson, Hu and Smith (2003) use data provided by a publisher selling on Amazon.com to conduct a more robust estimation of the parameters of the formula. They estimate the parameters $a = 10.526, b = -0.871$.

We have used the latter estimate in our study. In future work, we propose to conduct an independent purchasing and demand estimation experiment in order to update these estimates. Note, however, that since our results are all based on comparisons between categories, the fact that these parameters are dated are unlikely to affect our results directionally.

3.2. Quantifying the distribution of demand: the Gini coefficient

Next, we compute the Gini coefficient of each category of books. The Gini coefficient is a measure of distributional inequality, a number between 0 and 1, where 0 corresponds with perfect equality (in our case: where all the books in that category have the same demand) and 1 corresponds with perfect inequality (where one book has all the demand, and all other books have zero demand).

The Gini coefficient is based on the Lorenz curve, a widely used depiction of distributional equality, most commonly used to compare income distributions across regions and time. In our analysis, the Lorenz curve of a category's demand ranks the products in increasing order of sales, then plots the cumulative fraction $L(\rho)$ of sales associated with each ascending rank percentile ρ, where $0 < \rho \leq 1$. More precisely, define $N = \{1, 2, 3, ..., n\}$ as the set of all books in a category of size n, and recall that $q(i)$ is the demand for book i. To compute the Lorenz curve, we define, for each book i, $R(i)$ as the size of the set $\{x : x \in N, q(x) \leq q(i)\}$, which is the set of all products with demand less than or equal to that of i. $R(i)$ is thus simply the (inverse) rank of the product within its category, with the product with the lowest demand having the lowest rank. Next, define

$$S(r) = \{y \in N, R(y) \leq r\}, \tag{3.2}$$

which is the set of product indices whose rank is less than or equal to r. Then, for each percentile

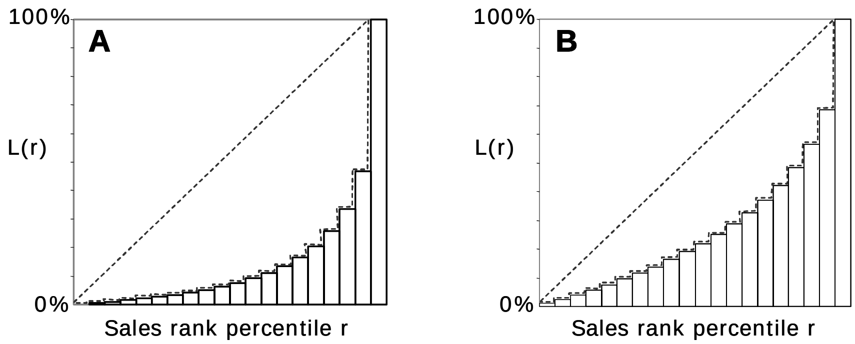


Figure 3.1: Illustrates the Lorenz curves and the Gini coefficients for two categories in our data set: Computers and Internet: Web Development (A) and Science: Chemistry (B) respectively. Their Gini coefficients are about 0.75 and 0.5 respectively.

ρ (which corresponds to the books ranked ρn or lower), the Lorenz curve is defined by:

$$L(\rho) = \frac{\sum\limits_{y \in S(n\rho)} q(y)}{\sum\limits_{y \in N} q(y)}. \tag{3.3}$$

Notice that the Lorenz curve is increasing and convex.

The Gini coefficient is computed as twice the area between the Lorenz curve $L(\rho)$ and the 45-degree line between the origin and $(1, 1)$. We calculate it for each category by first computing the entire area above the Lorenz curve, the Lorenz upper area:

$$LU = \sum_{y=1}^{n} \left[1 - L(y/n)\right], \tag{3.4}$$

and then using the identity

$$Gini = 2(LU) - 1. \tag{3.5}$$

Figure 4.1 illustrates this computation for two categories in our data set.

The Gini coefficient is especially suitable for this study for a variety of reasons. Most importantly, it measures inequality in the demand distribution, regardless of the category's average demand (or popularity), which facilitates comparing different categories despite their intrinsic differences and independent of their scale.

3.3. Two measures of network influence: Immediate Influence and Weighted PageRank

We have developed two different measures of network influence:

ImmediateInfluence: This is a measure of the traffic which flows into a product's webpage from its neighbors in the network. It is based on the assumption that the influence exerted by each product is proportionate to its total incoming traffic, is divided equally and flows to those products it has direct co-purchase links to (note, that this model does not allow back clicking). It therefore captures the influence of a product's immediate neighbors. Therefore, the influence that the co-purchase network has on demand depends on two factors: the local structure of the network and the amount of traffic associated with each link in the network. To combine the demand information with the structure of the network, we construct the ImmediateInfluence variable in the following way:

$$ImmediateInfluence(i) = \sum_{j \in G(i)} \frac{Demand(i)}{OutDegree(i)}, \tag{3.6}$$

where $G(i)$ is the set of books that link to book i.

WeightedPageRank: This is based on Google's PageRank algorithm, and iteratively computes the influence of the entire network on each product over time, although ignoring variations in intrinsic traffic across pages. It operates on an "average graph", constructed as a weighted composite of a time series of co-purchase networks. The original PageRank algorithm provides a ranking of the "importance" of web pages based on the link structure of the "web" created by the hyperlinks between the pages. This ranking forms the basis for Google's search engine.

The PageRank algorithm is based on a simple model of behavior – a random surfer. This surfer follows any one of the links on a page with equal probability or jumps to a random page with probability $(1 - \alpha)$ (this probability is also referred to as the "dumping factor", and is what differentiates PageRank from a commonly used notion of "centrality" in social network theory). The algorithm divides a page's PageRank evenly among its successors in the network. The ranking of a page ends up being the long run steady-stage probability that a random surfer who starts at a random page will visit the specific page. Thus, a page can gain a high ranking by either having many pages pointing to it or having few highly ranked pages pointing to it. The PageRank of all pages in the network is computed iteratively, until some convergence estimator is met.

We adapt the PageRank algorithm to account for the fact that we wish to measure the average influence the network has on a product over four successive one-week periods. In our adapted model:

$$WeightedPageRank(i) = \frac{(1-\alpha)}{n} + \alpha \sum_{j \in G(i)} Weight(j,i) \frac{WeightedPageRank(j)}{\sum\limits_{k \in F(j)} Weight(j,k)}, \tag{3.7}$$

where $Weight(j,i)$ is the fraction of the 7 days that the link was present on the copurchase graph. The contrast between SalesRank and PageRank is illustrated in Figure 4.4.

It is important to note that while this measure is widely used in ranking algorithms (such as Google's), we use the fact that fundamentally, Weighted PageRank measures the probability that a "random surfer" will arrive at a hyperlinked page if he were to traverse just the hyperlinks of the network. In other words, a product with a higher Weighted PageRank is more likely to get traffic from the network than one with a lower Weighted PageRank, and this therefore measures the *extent* to which the network structure we are interested in – the co-purchase network – *influences* the product in question.

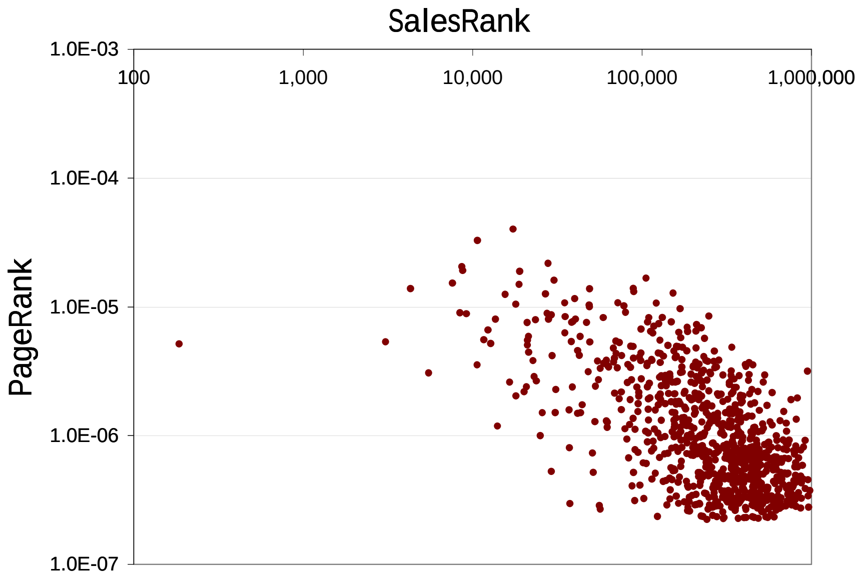


Figure 3.2: Plots SalesRank versus PageRank for a sample of the data. Illustrates the fact that while they are weakly (negatively) correlated, there are factors beyond network position that affect a product's demand.

The Weighted PageRank and the Immediate Influence measures described above are two different measures, which differ in the following key ways:

1. Weighted PageRank does not take the demand or intrinsic traffic variation across books into account. It is based only on the structure of the network. In contrast, Immediate Influence is based on both the structure of the network and the demand associated with each page.

2. The two measures use information about the structure of the network differently. Immediate Influence only includes the information about the immediate neighbors of the page, while Weighted PageRank measures the influence of the entire network.

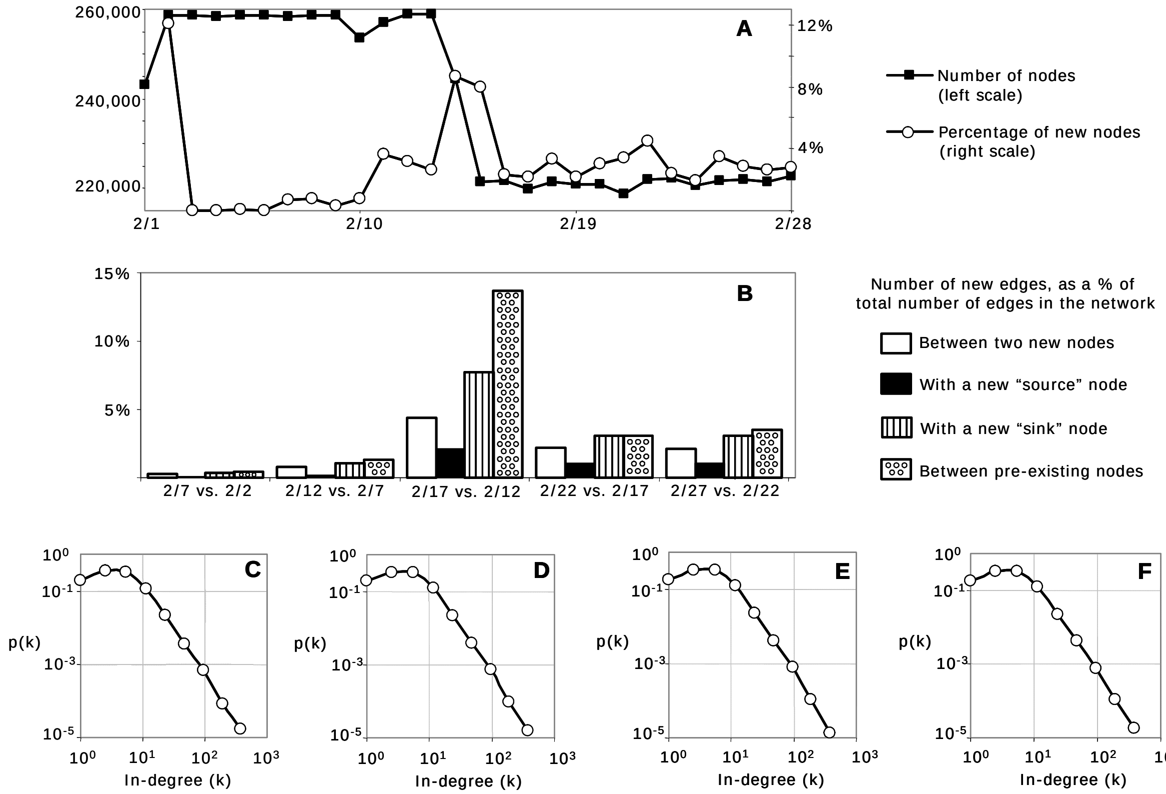


Figure 4.1: Graph statistics on the daily copurchase networks.

4. How network structure influences demand and its distribution

The results we discuss in what follows were obtained using data for four distinct one-week periods between February 1st and February 28th . There is a seasonal demand pattern associated with sales around Valentine's Day in our data set, and we observed substantial changes in the edges of the co-purchase graph during this period (close to 20% of the edges changed). As will be seen, despite such changes in the identities of the books linking to each other, our results remained relatively stable. Some summary statistics of the daily graphs are in Figure 5.1.

We first study the variation between demand for each individual product, and the corresponding ImmediateInfluence of that product. Our results indicate that the immediate influence of a product explains a significant amount of the variation in the demand for the product. Our

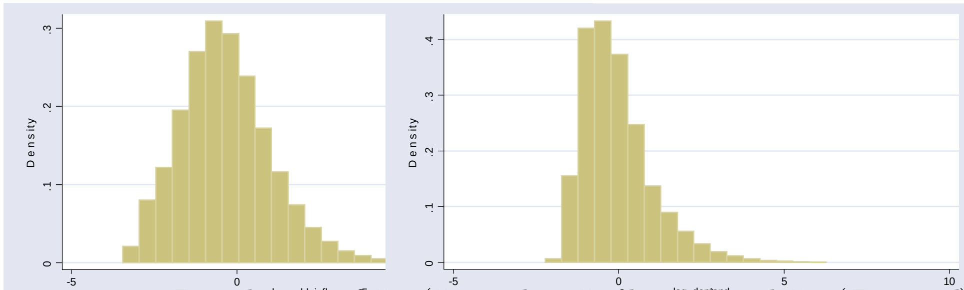


Figure 4.2: Distribution of $Log(ImmediateInfluence)$ and $Log(Demand)$

final section reports on refinements we are working on that account for endogeneity in these estimates.

What is more pertinent to our main results is the contrast between the distribution of demand and Immediate Influence, which is illustrated in Figure 5.2. A casual examination of the distribution of demand and influence suggests the effect that the network structure might have on the demand distribution. As required by the definition of ImmediateInfluence, both have the same mean (influence is simply demand being redistributed), but the range and standard deviation of influence are larger than that of demand. This leads one to suspect that suspect that the network redistributes demand in a more “equal’ manner, a hypothesis we report on in the next section.

We now “zoom out” to an aggregated level (the category affiliation) and study how the network structure affects the market and the demand distribution. To study the effect of the network structure on the demand distribution, we group the books according to category affiliation. We test the hypothesis that a higher average weighted PageRank for a category will be associated with a lower Gini coefficient for that category. Following our interpretation of PageRank, a category whose products collectively have a higher average weighted PageRank, is, all else being equal, one whose products are influenced more by the network. The summary statistics for average weighted PageRank (and other variables we use as controls) are presented

Variable	Range	Mean	StdDev
GINI	$0.39 - 0.96$	0.55	0.12
AVGDEMAND	$0.92 - 22.02$	3.04	3.07
AVGPAGERANK	$1.93 \times 10^{-6} - 6.03 \times 10^{-6}$	3.36×10^{-6}	6.32×10^{-7}
PAGERANKVAR	$3.97 \times 10^{-12} - 4.23 \times 10^{-10}$	5.74×10^{-11}	6.73×10^{-11}
SIZE	$100 - 11,179$	$1,087$	$1,722$
MIXING	$0.01 - 0.80$	0.32	0.18

Figure 4.3: Sample summary statistics

in Figure 5.3.

Using this data, we estimate the following reduced-form econometric model:

$$\begin{aligned} Log[GINI] &= a + b_1 Log[AVGDEMAND] + b_2 Log[AVGPAGERANK] \\ &\quad + b_3 Log[PAGERANKVAR] + b_4 Log[SIZE] + b_5 Log[MIXING] \end{aligned}$$

We chose a logarithmic specification because it facilitates ease of interpretation of the coefficients (in percentage terms), and because the empirical distributions of the transformed variables were more suitable for OLS regression.

The results of this estimation, presented in Figure 5.4, are striking. Based on a comparative analysis across over 200 categories of products, they establish that an increase in the extent to which the network structure is influential leads to a flattening of demand, or an increase in the relative demand for niche (rather than blockbuster) products. That is, we find that the average Weighted PageRank of the books in the category (the AVGPAGERANK variable) is negatively associated with the Gini coefficient of the category. This confirms that an increase in the extent to which network structure influences demand flattens the distribution of demand, or leads to a longer tail for demand, a phenomenon widely observed in electronic commerce (Anderson, 2004). Notice that this coefficient is not just statistically significant, but is economically significant as

Variable	Coefficient	Estimated Values (Standard Error)			
		2/1-2/7	2/8-2/14	2/15-2/21	2/22-2/28
constant	*a*	-1.97 (0.39) ***	-1.93 (0.37) ***	-2.19 (0.45) ***	-2.05 (0.43) ***
log[AVGDEMAND]	*b1*	0.26 (0.00) ***	0.25 (0.00) ***	0.23 (0.00) ***	0.24 (0.00) ***
log[AVGPAGERANK]	*b2*	-0.15 (0.04) ***	-0.15 (0.04) ***	-0.18 (0.04) ***	-0.17 (0.04) ***
log[PAGERANKVAR]	*b3*	0.03 (0.00) ***	0.03 (0.00) ***	0.03 (0.00) ***	0.03 (0.00) ***
log[SIZE]	*b4*	0.03 (0.00) ***	0.03 (0.00) ***	0.03 (0.00) ***	0.04 (0.00) ***
log[MIXING]	*b5*	-0.01 (0.00)	-0.01 (0.00)	-0.02 (0.00) *	-0.02 (0.01) *
	Adj. R-squared	86.85%	82.72%	81.04%	82.36%
indicates significance at the * 5%, ** 1%, and ***0.1% levels					

Figure 4.4: How network structure affects the distribution of ecommerce demand.

well. The highest average PageRank is generally about 3 times the lowest average PageRank. A doubling of average PageRank decreases the Gini coefficient by between 15% and 18% (since this is a log-log regression), which is pretty close to one standard deviation of Gini relative to its mean. Surprisingly, these results persist across four different weeks, one of which had substantial seasonal variation.

Moreover, the variance of the Weighted PageRank of different books within a category is positively correlated with the category's Gini coefficient. That is, after controlling for differences in average Weighted PageRank, a higher variance in the ranking (measured by PAGERANKVAR) is associated with increased inequality. To understand this result, consider two categories, both with the same average Weighted PageRank. Category A, where all books has the same Weighted PageRank and Category B, where there are a few books with a much higher than average Weighted PageRank, and correspondingly a number of books with a lower than average Weighted PageRank. It seems reasonable to expect that the flattening effect will be stronger for category A than for category B. After all, most of the traffic that goes into category B goes to the same few books and is likely to enhance the inequality in demand, thus increasing the Gini coefficient. In contrast, all books in category A get the same additional traffic from the network, so the relative differences in demand decrease, thus flattening the demand distribution.

The number of books in a category has a positive effect on the Gini coefficient. The categories in our data had between 100 and over 10,000 books in them. It is natural to assume that when all else is equal, a category with over 10,000 books is more likely to have higher variance in the demand for its books than a category with about 100 books.

Further, the average demand of the category has a positive effect on the Gini coefficient of the category. A straight forward interpretation of these results is that as the intrinsic demand increases, the added demand due to network traffic has a lower relative effect on the distribution of demand. To understand this result, consider two categories, both with the same average Weighted PageRank. Category A, with low average demand and Category B, with high average demand. Since both categories have the same average Weighted PageRank, they receive the same traffic from the co-purchase network (same number of consumers "flowing in"). This means they sell the same number of books to consumers who arrived at the books' pages via the co-purchase network. The network traffic has a flattening effect in both cases. In other words, the fraction of demand, which can be attributed to the best selling books, is lower. However, the impact that same number of additional copies sold will have on the fraction of demand that come from the best selling books will be lower for category A. Thus, since the traffic from the network accounts for a smaller fraction of category A's sales, the flattening effect will be smaller in magnitude.

5. Conclusions and ongoing work

We have briefly outlined a new economic theory of how network structures in electronic commerce might affect demand and cause ecommerce demand to be different from what is observed in traditional bricks-and-mortar commerce. We have gathered a new and unique data set comprising hundreds of observations of a giant component of the co-purchase network of Amazon.com, along with the relevant economic variables for each of its constituent products. We have provided the first evidence that the presence of these network structures can cause changes in demand patterns that are consistent with the observed "long tail" of ecommerce demand. We do so by adapting the PageRank algorithm to measure the influence that the network structure has on each product, and then contrasting variations in the average such measure across categories that have different demand distributions. To the best of our knowledge, ours is the first study of its kind.

Our current work aims to extend these results in the following salient ways:

- Rather than being the "random" surfers used by the PageRank model, ecommerce consumers are strategic. They tend to visit more popular products more often, and their purchasing is affected by other economic variables like price, customer reviews and product age. Our first extension aims to develop a model of a "strategic surfer" that is grounded in more familiar economic theory, but with retains sufficient structure to allow the iterative estimation of the "importance" of the network. We have made substantial progress on this front, solving a first model. This gives us a basis for a structural model.

- The methods we have used so far do not explicitly separate demand that is caused by the presence of the hyperlinks associated with the network structure with the demand variation that complementary products might naturally experience together. We have developed and tested a preliminary model of network-induced peer effects that separates demand using a two-stage spatial autoregressive model towards identifying network-induced endogenous effects and separating them from demand covariation that is on account of shared characteristics or product complementarity. We hope to report further on these results in June 2007.

6. References

1. Anderson, C. The Long Tail. http://www.wired.com/wired/archive/12.10/tail_pr.html

2. Aral, S. Brynjolfsson, E., and Van Alstyne, M., 2006. Information, Technology and Infor-

mation Worker Productivity: Task Level Evidence. Mimeo, MIT.

3. Brin, S., and Page, L., 1998.The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 33:107–117.

4. Brin, S., Page L., Motwani, R., and Winograd, T., 1999.The PageRank citation ranking: bringing order to the Web. Technical Report 1999-0120, CS Department, Stanford University.

5. Brynjolfsson, E., Smith, M. D., and Hu, Y. J., 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. *Management Science,* 49.

6. Brynjolfsson, E., Smith, M. D., and Hu, Y. J., 2006. From Niches to Riches: Anatomy of the Long Tail. Sloan Management Review 47(4), 67-71.

7. Butler, B., 2001. Membership Size, Communication Activity, and Sustainability: The Internal Dynamics of Networked Social Structures. *Information Systems Research* 12, 346–362.

8. Erdem, T., Imai, S., Kean, M. P., 1999. Econometric Modeling of Competition: A Multi-Category Choice-Based Mapping Approach,. *Journal of Econometrics* 89.

9. Ghose, A., M. Smith, and R. Telang. 2006. Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact. *Information Systems Research* 17, 3-19.

10. Goolsbee, A., and Chevalier, J. A., 2003. Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics* (1) 203-222.

11. Hill, S. F. Provost and C. Volinsky, 2006. Network-based Marketing: Identifying Likely Adopters via Consumer Networks. *Statistical Science* 21, 256-276.

12. Kane, J. and Alavi., M., 2006. Casting the Net: Towards a Theory of Multimodal Knowledge Networks. Mimeo, Emory University.

13. Langville, A., and Meyer, C., 2005. Deeper Inside PageRank. *Internet Mathematics*, 1(3):335–380.

14. Matthias, S., Bauer, H. H., and Hammerschmidt, M., 2002. Structuring Product Markets: An Approach Based on Customer Value. *Journal of the American Marketing Association.*

15. Newman, M. E. J., 2003, The structure and function of complex networks. *SIAM Review* 45

16. Ramaswamy, V., DeSarbo, W. S., 1990, SCULPTRE: A New Methodology for Deriving and Analyzing Hierarchical Product-Market Structures from Panel Data. *Journal of Marketing Research* 27(4)

17. Russell, G. J., Kamakura, W. A. 1997, Modeling Multiple Category Brand with Household Basket Data. *Journal of Retailing* 73(4)

18. Schoberth, T. Preece, J. and Heinzl, A., 2003. Online communities: a longitudinal analysis of communication activities. *Proceedings of the 36th Hawaii International Conference on Systems Sciences.*

19. Sundararajan, A., 2006. Local Network Effects and Complex Network Structure. *Contributions to Theoretical Economics* 6 (1), in press.

20. Watts, D. J. and Strogatz, S. H., 1998, Collective dynamics of 'small-world' networks. *Nature* 393

Durable Goods, Innovation and Network Externalities

Daniel CERQUERA*

Work in Progress!

March 15, 2007

Abstract

We develop a model of R&D competition between an incumbent and a potential entrant with network externalities and durable goods. We show that the threat of entry eliminates the commitment problem that an incumbent may face in its R&D decision due to the goods durability. Moreover, a potential entrant over-invests in R&D and an established incumbent might exhibit higher, equal or lower R&D investments in comparison with the social optimum. In our model, the incumbent's commitment problem and the efficiency of its R&D level is determined by the extent of the network externalities.

Keywords: Network externalities, Durable Goods, Innovation, Imperfect Competition.

JEL Classification: D21, D85, L13, O31.

*cerquera@zew.de. Centre for European Economic Research (ZEW), Mannheim, Germany. Research Group Information and Communication Technologies. I would like to thank Jay P. Choi, Toker Doganoglu, Stefan Mittnik, Monika Schnitzer and Sven Rady for helpful comments and suggestions, as well as seminar participants at the University of Munich and the 2006 European Economic Association Annual Congress in Vienna and the 2006 Latin American and Caribbean Economic Association Annual Meeting in Mexico City. Part of this research was completed at the Center for Information and Network Economics at the University of Munich. Generous financial support from Volkswagen Stiftung and the Munich Graduate School of Economics made this research possible. All errors are mine.

1 Introduction

An industry exhibits network externalities when the benefit that consumers enjoy from purchasing one or several of its goods depends on the number of other consumers that use the same and/or compatible products. For the firms in those sectors (e.g. software, telecommunications, consumer electronics, etc.), the presence of network externalities implies that the attractiveness of their products is a function of their quality-adjusted prices and the potential benefit attached to their expected network sizes (i.e. installed bases).[1]

Those products (i.e. network goods) tend to be characterized by two features closely related. Durability and rapid technological progress.[2] Durability implies that network goods tend to "wear out" not as a result of physical deterioration, but as a consequence of technical obsolescence; a feature due to technological progress. For example, a given software (or mobile phone, or video game, etc.) can be functional for a long time. However, the utility derived by its use tend to be dissipated due to new (and actually very frequent) developments that are more closely related to consumers needs and tastes.

This paper considers a stylized network industry where these two features, durability and technological progress, are analyzed together. In particular, we propose a model of R&D competition between an incumbent and a potential entrant and consider the implications of the durability of network goods. Our main objective is to isolate the role of network externalities and analyze the social efficiency of the R&D incentives of the firms in this industry.

We depart from the current literature by considering, simultaneously, an oligopolistic setup, endogenous R&D processes and durable goods. Therefore, this paper is not only closely related to the literature on durable goods and to the literature on technological

[1] See Katz and Shapiro (1985, 1986) and Farrel and Saloner (1985, 1986) for seminal treatments, and Katz and Shapiro (1994) and Economides (1996) for surveys on network markets.

[2] See Katz and Shapiro (1999) for an informal analysis of antitrust in software markets, where these two characteristics are explicitly considered.

progress in network industries, but represents a first step in bridge them together.

The economic literature has highlighted the role that durability plays in the evolution of a market dominated by a monopolist. In particular, the conventional problem for the monopolist is that, having sold a durable good, there is an incentive to reduce price later to bring into the market those consumers that would not pay the initial high price. However, consumers realize that the monopolist has such an incentive to reduce price once they have purchased and those that value the good less highly will withhold their purchase until price falls. For this reason the monopolist is unable to extract as much money from the market as would be possible with a pre-commitment of "no future price reductions". The fact that in the absence of commitment the monopolist may act against his own profitability implies a "time-inconsistency" problem (i.e. choices that maximize current profitability might not maximize overall profitability).

This notion was first discussed by Coase (1972) and has been labelled as the "Coase Conjecture".[3] Since its formulation, the Coase Conjecture has been theoretically developed in several papers that consider the robustness of the basic observation.[4]

The essential problem is that the monopolist's actions in the future provide competition for the company in the present market.[5] If the monopolist is able to lease the good, distort technology or implement buy back procedures then more profit can be extracted from the market since these strategies restrict the aftermarket.[6] Failing this the monopolist has an incentive to reduce durability or make the good obsolete after a period of time.[7] The existing analysis of durability in the presence of network externalities has intended,

[3]Strictly, the Coase Conjecture refers to a limiting case. It states that in the absence of commitment and if the monopolist may adjust its prices frequently enough, the successive price reductions lead to marginal cost pricing and the subsequent loss of market power.

[4]See, for example, Bagnoli, Salant and Swierzbinski (1989), Bulow (1982), Gul, Sonnenschein and Wilson (1986), and Stokey (1981).

[5]The price of a durable good attempts to extract current and future surplus, however, future surplus depends on future actions that are not realized when the price is set.

[6]See Fudenberg and Tirole (1998), Kahn (1986) and Waldman (1997).

[7]See Bulow (1986), Hendel and Lizzeri (1999), Rust (1986) and Waldman (1993), Grout and Park (2005).

as the main literature on durability, to verify the validity of the Coase Conjecture.[8]

However, the implications of durability are much broader than the pricing commitment problem considered in the analysis of the Coase Conjecture. In particular, the result that a monopolist in the absence of commitment may affect its own overall profitability applies in several contexts. In fact, as pointed out by Waldman (2003), any present and future action that affects the future (relative) value of the monopolist's used goods might be subject to the "time-inconsistency" described above. One leading case of such actions is a firm's R&D expenditures which, by definition, affect the (relative) value of used (or previously sold) goods.[9]

In the presence of network externalities, the similar analysis of introduction of new durable goods has been analyzed.[10] However, this literature is focused on a monopolistic setup and considers the production of new technologies as exogenous. Hence, and to the best of our knowledge, there is no analysis that consider explicitly the process of endogenous R&D processes in the presence of network externalities and durable goods.

The paper presented by Ellison and Fudenberg (2000) is the closest to ours and is actually our departure point. In that paper, the authors consider a monopoly that operates in a two-period framework and produces durable network goods. In the first period the monopoly produces a good with a given low quality and, subsequently, has the choice of introducing an improved version in the second period. Network externalities play a role because the improvement of the old good implies backward compatibility. That is, consumers of the new good enjoy network benefits from the entire population, while consumers of the old good only enjoy network benefits from consumers of the same good.[11]

In their model, there is an inflow of new consumers in the second period and, with consumer homogeneity, the paper shows that the monopolist has the incentive to introduce

[8]See Bensaid and Lesne (1996), Cabral et al. (1999) and Mason (2000).
[9]See Waldman (1996), Fishman and Rob (2000) and Nahm (2004).
[10]See Choi (1994) and Ellison and Fudenberg (2000).
[11]A case of this situation was evidenced by the launch of Microsoft Word 97. Consumers of Word97 were fully compatible with consumers of Word95 but the opposite did not hold.

the improved good, even though the monopolist's overall profits (and social surplus) is reduced. That is, in the absence of commitment the monopolist's choice that maximizes current (second period) profits does not maximize overall profitability.

We present a model that extends that of Ellison and Fudenberg (2000) by introducing and endogenous R&D process in the production of the new technology, and consider the role of a potential entrant. We show results not present in the Ellison and Fudenberg (2000) analysis. In particular, we consider a two-period framework with an incumbent, a potential entrant and an inflow of new consumers. Consumers are homogeneous and participate in a market with durable network goods.

In the first period, there is a first group of consumers that buy a network good from the established incumbent. Before the second period starts, a potential entrant appears in the market and, jointly with the incumbent, decides on an investment level that will allow him to compete in the second period. This R&D process is stochastic. By investing a certain amount, both firm determine the probability that in the second period they are able to produce a new product that is quality-improved relative to the existing good produced by the incumbent. Conditional on the success or failure of the innovation process, both firms compete in price in the second period when a new group of consumers arrive.

We analyze the incentives to innovate for both firms, we compare it to the social optimum and investigate the role of the network externalities. With our simplified approach, we are able to isolate the impact of network externalities and reach three main results.[12]

First, the threat of entry reverses the commitment problem that a monopolist (without such threat) may face in its R&D decision given good durability. This result is not present in the current literature and follows from the role that R&D incentives play in deterring entry. In our case, the monopolist's commitment problem arises only due to the presence of network externalities.

Second, the levels of R&D determined by market outcome might differ from the so-

[12]The robustness of these results with respect to the assumed functional forms is the objective of current work.

cially optimal levels. In particular, a potential entrant always over-invests (as an entry strategy) and an established incumbent might exhibit higher, lower or equal R&D levels in comparison with the social optimum. This result suggests that successful entry takes place too often in comparison with the social optimum.

And third, the extent of network externalities is the crucial parameter in the efficiency of the incumbent R&D level. In fact, it is only the presence of network externalities that permits, potentially, to the established incumbent to provide an efficient level of innovation. Without network externalities (or very low network effects), it is shown that the incumbent firm always under-invests in R&D efforts. This result sheds some light on the debate whether a dominant incumbent in a network industry provides sufficient innovation to the society.

The paper is organized as follows. The next section presents the model. Section 3 presents the analysis of its equilibrium. Section 4 computes the social optimum and compares it with the results of the market outcome. Finally, section 5 concludes and discusses some areas of further research.

2 The Model

We consider a model of a network industry with durable goods based on that of Ellison and Fudenberg (2000).[13] There are two periods denoted by $t = 1$ and $t = 2$ with a group of homogeneous consumers arriving in each period. In period 1 there is a monopolist incumbent that is challenged in period 2 by a potential entrant. In period 2, firms compete in prices with quality differentiated products. Quality is determined through endogenous and stochastic R&D processes carried out in period 1.

[13] We construct our model to make Ellison and Fudenberg (2000) a particular case of the one presented here.

2.1 Supply Side and R&D Process

In period 1, an incumbent monopolist, I, produces a durable network good with quality level q_1 (i.e. stand-alone value). The good lasts two periods after which it vanishes. We consider the case of product innovations where, subject to R&D expenditures, the incumbent might be able to produce a network good of better quality to be introduced in period 2. In our model, this process of innovation is carried out at the end of period 1. In addition, we assume that the outcome of the R&D process is stochastic with two possible outcomes, success or failure. This outcome is realized at the beginning of period 2.

In particular, we consider an R&D process where the incumbent firm determines the probability s_I that the innovation process is successful. Higher investments (i.e. higher probability of success) entail higher costs. These costs are summarized by means of a function $C(s_I)$ that is increasing in the probability of success s_I. For simplicity, we assume that $C(s_I) = \frac{as_I^2}{2}$, where a is a cost parameter.

In the case of success, the innovation is achieved and allows the incumbent firm to produce a "new" network good with quality q_2 in period 2, where $q_2 = q_1 + q_\Delta$ and q_Δ is the extent of the innovation. q_Δ is assumed to be constant and greater than zero. If the innovation process is unsuccessful, the incumbent produces in period 2 the same "old" good with low quality q_1. It is assumed that the achievement of the innovation do not preclude the incumbent to produce the "old" good in period 2.

In addition, we introduce a potential entrant, E, that intends to compete with the incumbent in period 2. In order to be able to enter the market, the potential entrant must invest in R&D to develop a network good. The entrant's innovation process takes place simultaneously with that of the incumbent firm. It is assumed, that the innovation process for the potential entrant is identical to the one of the incumbent firm. Therefore, the potential entrant must determine the probability s_E, that its innovation process succeeds. If so, the entrant is able to produce the "new" good with quality q_2 in period 2. It is assumed that in the case of unsuccessful innovation, the entrant firm stays out of the market (i.e. it cannot produce the old quality network good).

As in Ellison and Fudenberg (2000), we assume that the network goods are backward compatible. That is, consumers of the new good enjoy network benefits from all users (i.e. users of new and old goods), while consumers of the old good only enjoy network benefits from consumers of the same good (e.g. Word97 vs. Word95).[14]

It is further assumed that the firms cannot change the quality of the goods once they are already produced. Marginal costs of production are independent of quality and set equal to zero. For simplicity the discount factor is equal among firms and normalized to 1.

2.2 Demand Side and Expectation Formation Process

The demand side represents the core of the model. In each period there is a group of N_t homogeneous consumers arriving in the market and, for convenience, we normalize $N_1 + N_2 = 1$. Consumers exhibit a per-period unitary demand for a network good and buy as soon as they reach the market. This implies that the N_1 consumers make a purchase decisions in period 1 and in period 2. Given durability, this is not a trivial implication.

To see this, note that the price charged to the N_1 consumers in period 1 tries to extract period 1 and 2 surpluses (i.e. the good is durable). However, period 2 surplus is affected by the outcome of the R&D processes, the prices of the two firms in period 2 and the N_1 and N_2 consumers' choices. Therefore, the willingness to pay of the N_1 consumers in period 1 depends on their beliefs on how the firms are going to behave in period 2. This gives rise to the commitment problem discussed in the introduction.

Consider first period 1. The first group of consumers, with size N_1, arrives at the beginning of period 1, finds only the incumbent's good and observes its price (to be

[14]Note that the assumption of backward compatibility implies that, conditional on successful innovation, the surplus offered by the new good is independent of the identity of the firm that produces it.

derived below). We model utility by assuming that each consumer in N_1 derives a first-period benefit (gross of price) from buying from the incumbent firm given by $q_1 + \alpha x - c$. In this expression, q_1 is the quality of the good, α is a parameter that measures the extent of the network benefits, x is the number of users of compatible goods[15] and c is a cost of learning to use the network good. We introduce the following assumptions.

Assumption 1. $2q_1 > 0$. *N_1 always consume the old good in period 1.*

By introducing assumption 1, the model implies that even in the case where network benefits are equal to zero, first period consumers always consume. This assumption will allow us to analyze the model with very small (or non-existent) network benefits and compare the results with the case where network externalities are important without introducing discontinuities in the consumers' behavior.

Assumption 2. $q_1 + \alpha N_1 - c_u > 0$. *It is optimal for N_1 to consume in both periods.*

The previous assumption 2 is introduced to avoid the possibility of N_1 consumers waiting to period 2 to consume.[16] This assumption reduces the number of cases to be analyzed, and allows us to focus on the results we are interested in.[17]

Of course, the overall benefit enjoyed by consumers in N_1 also depends on period 2 choices to be explained below. Note that at the beginning of period 2, the outcome of the innovation process is realized depending on the investment decisions. Hence, there are four possible cases in period 2; no firm innovates; only the incumbent or only the entrant innovates; and both firms achieve the innovation.

Now consider period 2. When the N_1 consumers reach the beginning of period 2, they observe the outcome of the innovation process. If the innovation is achieved, the N_1 consumers evaluate the incremental utility from purchasing (i.e. upgrading to) the

[15] Note that given the homogeneity of the consumers $x = N_1$ in period 1.

[16] In order to maintain the order of the exposition, the parameter c_u (i.e. the cost of upgrading) is introduced below.

[17] See, for example, Choi and Thum (1998) for the analysis when consumers can wait to adopt a network good.

new generation of the good and decide accordingly.[18] Therefore, they compare the benefit (gross of price) from the new good $q_2 + \alpha(N_2 + x) - c_u$ with the second-period benefit of staying with the old good $q_1 + \alpha x$. c_u is the cost of learning to use the new generation (i.e. cost of upgrading). It is assumed that $c_u < c$. As common in models with network externalities, the equilibrium value of x depends on the way consumers form expectations about other consumers behavior.

We assume that consumers are able to coordinate on the outcome that maximize their surplus (i.e. Pareto-Optimal coordination equilibrium).[19] In other words, consumers are able to coordinate on the choice that maximize joint surplus. Thus, they compare $q_2 + \alpha - c_u$ with $q_1 + \alpha N_1$ and, in consequence, the incremental utility from upgrading is given by $q_\Delta + \alpha N_2 - c_u$. Hence, whenever $q_\Delta + \alpha N_2 - cu > 0$ upgrade by the N_1 consumers takes place, otherwise the N_1 consumers do not buy the new good and stay with the old one. We denote this (candidate) price of upgrading as p_u.

In period 2, a second group of consumers with size N_2 arrives in the market. This group of consumers observes the outcome of the innovation process, observes prices (to be derived below) and makes purchase decisions. In particular, it is assumed that whenever the innovation is successful (either by the incumbent, the entrant or both) the N_2 consumers do not exhibit any preference for the old good produced by the incumbent. That is, the willingness to pay of N_2 consumers for the new generation of the good is equal to $q_2 + \alpha - c$.[20] We denote this (candidate) price as p_n. Note that given the assumption of backward compatibility, consumers of the new good enjoy the full network benefits (i.e αx with $x = 1$).

In the case that the innovation does not take place (i.e. no firm innovates), the N_2

[18] Recall that for the N_1 consumers the identity of the firms that produces the new good in period 2 is irrelevant (footnote 14).

[19] See Katz and Shapiro (1986) and Farrell and Katz (2005).

[20] This assumption allows the incumbent monopolist to extract the full consumers surplus in the case without entry. Therefore, it permits us to conclude that any reduction in the monopolist's profit implies a reduction in social welfare.

consumers decides for the old good with a willingness to pay equal to $q_1+\alpha-c$. We denote this (candidate) price as p_o. Therefore, analogous to Ellison and Fudenberg (2000), it is the choice of the N_1 consumers in period 2 that represents the most important part of the analysis.

In the next section we present the main results of the market outcome.

3 Market Outcome

In this section we consider the optimal pricing decision and the private incentives to innovate of the two firms. As a benchmark, we consider first the monopoly case. This analysis will allow us to compare the present paper with the current literature, to analyze the impact of network externalities and highlight the main results we obtain in comparison with Ellison and Fudenberg (2000). Once the monopoly case is considered, we analyze the model where the incumbent monopolist faces the threat of entry. In both cases, we consider the commitment as well as the no commitment case given its role in the durability literature discussed in the introduction.

As has been widely highlighted in the literature, the no commitment case is equivalent to focus on the Subgame-Perfect Nash-Equilibrium (SPNE), and the commitment case corresponds to the Nash-Equilibrium (NE) of the global multi-stage game.

3.1 A Monopoly Model

In order to solve the monopoly model, we first solve for the period 2 demands, profits and price equilibria. Then, we turn to the investment decision at the end of period 1 and derive the commitment and the no commitment case.

3.1.1 Second Period - Pricing Decision

Before deriving the equilibrium prices conditional on the outcome of the R&D process, it is important to note that the value of p_u is critical to the analysis because it describes the situation where upgrade takes place.

Assumption 3. $p_u > 0$. *Whenever the new good is produced, it is optimal for N_1 to upgrade.*

We focus on the analysis, unless otherwise noticed, for cases when assumption 3 holds. (i.e. upgrade is possible and optimal) and later on we present a brief discussion considering the case when assumption 3 does not hold.

Note that price competition depends on the outcome of the innovation process, therefore, there are two cases to consider according to the success or failure of the monopolist's innovation process.

Monopolist does not innovate. In this case, the monopolist still produces the old good with quality q_1 in period 2. As explained before, the N_1 consumers do not make any purchase decision (they already have the only existing good) and the N_2 consumers buy the old good if the price is less or equal to the maximum surplus offered by the good (i.e. $p \leq p_o$). Therefore, given the homogeneity of consumers, the incumbent charges p_o to the N_2 consumers that are his only revenue source and extract their full surplus.

Monopolist does innovate. In this case, the new generation of the good with quality q_2 is produced by the monopolist. Under assumption 3 and the coordination assumption, it is optimal for the N_1 consumers to upgrade if the price charged is less or equal to the incremental surplus offered by the new good (i.e. $p \leq p_u$). Again, given consumer homogeneity, the monopolist charges p_u to the N_1 consumers. Using similar arguments, it can be shown that the monopolists charges p_n to the N_2 consumers. Note that innovation increase the source of revenues for the incumbent.

Table (1) summarizes the pricing decision by the monopolist in period 2 conditional

	Monopoly does not innovate	**Monopoly does innovate**
Monopolist's Prices	0 $q_1 + \alpha - c$	$q_\Delta + \alpha N_2 - c_u$ $q_2 + \alpha - c$

Table 1: Period 2 - Pricing Decision - Monopoly case

on the outcome of the R&D process. Each cell in the table shows the price charged to the N_1 and the N_2 consumers, respectively.

3.1.2 First Period - Investment Decision

Suppose that to obtain the improved quality in period 2, the monopoly has to invest and succeed according to the R&D process described above. That is, the monopolist must decide the probability s that in period 2 the innovation is achieved and the new generation of the good with quality q_2 is produced.[21] The cost of choosing the probability s is given by $C(s) = \frac{as^2}{2}$, where a represents a cost parameter. Assume that consumers coordinate on the Pareto-optimal equilibrium. Then, if the innovation is successful, for $q_\Delta + \alpha N_2 - c_u > 0$ (i.e. assumption 3 holds) in period 2 the N_1 consumers upgrade and pays a price p_u and the N_2 consumers adopt the new good. If $q_\Delta + \alpha N_2 - c_u < 0$ (i.e. given that the innovation is achieved and assumption 3 does not hold) the N_1 consumers do not upgrade and the N_2 consumers adopt the new technology. If the innovation is not achieved, the N_1 consumers do not make any decision and the N_2 consumers adopt the old good. Consider the case where assumption 3 holds, then, the investment problem of the monopolist and the end of period 1 is given by,

$$\max_s \Pi_M = N_1 p_1 + s(N_1 p_u + N_2 p_n) + (1-s)(N_2 p_o) - \frac{s^2}{2} \tag{1}$$

with,

$$p_1 = q_1 + \alpha N_1 - c + s(s_n - p_u) + (1-s)(s_o)$$

[21] Note that if the innovation can be achieved with certainty and at no cost, the analysis is the one presented in Ellison and Fudenberg (2000)

In this expressions, we have simplified considering $a = 1$ $p_u = q_2 - q_1 + \alpha N_2 - c_u$, $p_n = q_2 + \alpha - c$, $p_o = q_1 + \alpha - c$, $s_n = q_2 + \alpha - c_u$ and $s_o = q_1 + \alpha$.

In this expression, $N_1 p_1$ corresponds to the period 1 revenues, $s(N_1 p_u + N_2 p_n) + (1 - s)(N_2 p_o)$ are the period 2 revenues and $\frac{s^2}{2}$ is the cost attached to the innovation process.

Consider the revenues obtained in period 1. As can be seen, p_1 extracts the full surplus enjoyed by the N_1. In particular, $q_1 + \alpha N_1 - c$ represents the period 1 surplus and $s(s_n - p_u) + (1 - s)(s_o)$ is the expected period 2 surplus that is conditional on the outcome of the innovation. That is, with probability s the innovation is achieved and, given assumption 3 holds, it is optimal for the N_1 consumers to upgrade in period 2 with a net surplus of $s_n - p_u$. On the other hand, if the innovation is not achieved, the period 2 net surplus of the N_1 consumers is equal to s_o.

Importantly, note that the price charged in period 1, p_1, depends on the level of investment because the surplus that the N_1 consumers enjoy in period 2 is uncertain at the beginning of period 1. Moreover, observe that $\frac{\partial p_1}{\partial s} = -\alpha N_2 < 0$. This observation implies that through investment, the monopolist reduces the future value of its good sold in period 1. Therefore, a higher R&D investment reduces the willingness to pay from the N_1 consumers in period 1 as the durability literature suggests. At the same time, a higher investment level increases the probability of introducing a new generation of the network good in period 2, and in consequence, expected period 2 revenues are increased. As we will see, it is the interaction (i.e. trade-off) between these two effects that represents the main impact of durability in the R&D incentives by the monopolist and highlights the role of commitment.

The revenues obtained in period 2 presented by the second and third term of equation (1) have an straightforward interpretation. In the following, we solve for the optimal investment decision given the problem stated in equation (1). We first present the no commitment case and then the commitment case.

3.1.3 No Commitment Case

Under no commitment the analysis of the SPNE rule out any non-credible threats by the monopolist. Therefore, consumers in period 1 determine their willingness to pay considering the case of what would the monopolist do after the N_1 consumers have made their period 1 purchasing decision. In other words, solving backwards and considering the R&D level that maximizes second period profits for the monopolist, we obtain the following first-order condition,

$$0 = N_1 p_u + N_2 p_n - N_2 p_o - s^{nc}$$

It can be seen that the second-order condition for an interior solution also holds. Thus, the corresponding optimal level of investment in the absence of commitment by the monopolist is given by,

$$s^{nc} = q_\Delta - N_1 c_u + \alpha N_1 N_2 \tag{2}$$

Before analyzing this result, we solve first for the commitment case.

3.1.4 Commitment Case

In this case, the monopolist is able to internalize the negative impact that his investment decision has on the first period prices (i.e. recall $\frac{\partial p_1}{\partial s} < 0$). Therefore, by considering the NE of the global multi-stage game, we obtain the following first order condition,

$$0 = -N_1 N_2 \alpha + N_1 p_u + N_2 p_n - N_2 p_o - s^c$$

Analogously, the second-order condition for an interior solution holds and the optimal level of investment provided that the incumbent is able to commit is given by,

$$s^c = q_\Delta - N_1 c_u \tag{3}$$

As can be readily seen from the preceding analysis, $s^{nc} > s^c$ holds for any parameter configurations. This results is not surprising and is in line with the traditional literature. It says that without commitment, the monopolist has the incentive to invest more than in the presence of commitment because it does not internalize the negative impact of its investment level on the price charged in period 1. Moreover, it is evident that the difference between the two investment levels is equal to $\alpha N_1 N_2$ which vanishes when the network externalities are not present (i.e. $\alpha = 0$). This implies that the effect of commitment is completely isolated and will allow us to conclude that any inefficiency, if present, will be solely due to the presence of network externalities.[22]

This result is stated in the following proposition.

Proposition 1. *Without the threat of entry, the monopolist invests more in the absence of commitment than it would be the case if commitment is possible. This difference is only due to the presence of network externalities.*

In addition, comparing the two profit levels (solving for the corresponding optimal investment levels in equation (1)) it can be shown that $\Pi_M^c - \Pi_M^{nc} = \frac{(N_1)^2(N_2)^2\alpha^2}{2}$ which is unambiguously positive. Again, this result highlights the main commitment problem on the R&D incentives of a monopolist that arises in the presence of durable goods (see Waldman (1996)). That is, once a monopolist does not have the possibility to commit to future R&D investments, its optimal decision affects negatively its overall profitability. Importantly, note that the previous result vanishes if $\alpha = 0$.

In addition, given that consumers are homogeneous, the monopolist is able to extract all the surplus from the consumers and, therefore, the absence of commitment reduces social surplus.

[22]This result also holds in the Ellison and Fudenberg (2000) paper.

Proposition 2. *For the monopoly case, the absence of commitment in the R&D investment implies a lower social surplus compared to the case when commitment is possible. This result is only due to the presence of network externalities*

The analysis of the monopoly model presented two main results. First, the presence of network externalities implies a commitment problem in the investment decision by the monopolists. This commitment problem is represented by an over-investment in comparison with the case where commitment is possible. And second, due to the presence of network externalities, the commitment problem implies a lower overall profit and an associated lower social welfare. These results are in line with the current literature and represent the benchmark for comparison for our analysis of entry.

3.2 A Model with Entry

In this subsection we extend the monopoly analysis presented above and consider the case of a potential entrant. Keeping the same framework, we model the case of an incumbent monopolist that serves the entire market in period 1 and must compete with a potential entrant in period 2. As explained before, entry is conditional on innovation and, therefore, both firms invest in developing a new technology at the end of period 1. At the beginning of period 2 the outcome of the innovation process is realized and price competition takes place.

As in the analysis of the monopoly case, the investment decision depends on the equilibrium concept adopted, namely, SPNE or NE, which characterizes the no commitment and commitment case, respectively. In order to proceed, we first solve for the period 2 demands, profits and price equilibria that follow from Bertrand competition. Then, we turn to the strategic investment decision at the end of period 1 and derive the commitment and the no commitment case.

3.2.1 Second Period - Price Competition

As in the monopoly analysis and in order to simplify exposition, we assume in what follows that assumption 3 holds. Note that price competition depends on the outcome of the innovation process, therefore, there are four cases to consider according to the success or failure of a given firm's innovation process, and the identity of that firm.

No firm innovates. In this case, no firm achieves the innovation. In consequence, the incumbent firm still produces the old good with quality q_1 in period 2 and the entrant firm has no production. As explained before, the N_1 consumers do not make any purchase decision (they already have the only existing good) and the N_2 consumers buy the old good if the price is less or equal to the total surplus they get from it. Therefore, the incumbent is able to charge p_o to the N_2 consumers that are his only revenue source in period 2. Note that this case, ex-post, is identical to the monopoly case without innovation.

Only Incumbent innovates. In this case, the new generation of the good is produced by the incumbent and the entrant does not enter the market. Therefore, given the assumption that the consumers are able to coordinate on the Pareto-Optimal equilibrium, the incumbent charges p_u to the N_1 consumers and p_n to the N_2 consumers. Note that innovation increase the source of revenues for the incumbent. Given that entry does not take place, this case is, ex-post, identical to the monopoly case with successful innovation.

Only entrant innovates. In this case, the entrant innovates and is able to produce the new generation of the good in period 2. Therefore, the entrant firm is able to capture the N_2 consumers and charges p_n to them. In addition, and assuming that he can identify the N_1 consumers (i.e. the entrant can offer a cross-subsidy), the price charged to them is p_u subject to the coordination assumption discussed above.[23]

Both firms innovate. In this case, both firms achieve the innovation and compete with homogeneous products in a homogeneous market. Thus, Bertrand competition drives

[23]Note that if the entrant cannot offer a cross-subsidy, the price charged to the N_1 is in any case equal to the incremental benefit that those consumer enjoy by purchasing the new good from the entrant firm.

Firm	Both Firms Innovate	Incumbent Innovates	Entrant Innovates	No Firm Innovates
Incumbent's Prices	0 0	$q_\Delta + \alpha N_2 - c_u$ $q_2 + \alpha - c$	0 0	0 $q_1 + \alpha - c$
Entrants's Prices	0 0	0 0	$q_\Delta + \alpha N_2 - c_u$ $q_2 + \alpha - c$	0 0

Table 2: Period 2 - Price Competition - Entry case

prices and period 2 profits to zero.

Table (2) summarizes the pricing decision in period 2 conditional on the outcome of the R&D process. Each cell in the table shows the price charged to the N_1 and the N_2 consumers, respectively.

3.2.2 First Period - Investment Decisions

After deriving the equilibrium prices from the competition in period 2 between the incumbent and the potential entrant, we are able to analyze the optimal investment decisions by the two firms. Note that in the case of the threat of entry, the investment decisions are derived strategically.

As explained before, the investment decisions correspond for the firms to choose the probability, s_k for $k \in I, E$, that the innovation is achieved in period 2. In addition, there is a cost $C(s_k) = \frac{as_k^2}{2}$ associated with a given probability s, where a correspond to a cost parameter.

The overall problem of the incumbent firm is given by,

$$\max_{s_I} \Pi_I = N_1 p_1 + s_I(1 - s_E)(N_1 p_u + N_2 p_n) + (1 - s_I)(1 - s_E)(N_2 p_o) - \frac{s_I^2}{2} \tag{4}$$

with,

$$p_1 = q_1 + \alpha N_1 - c + s_I(1 - s_E)(s_n - p_u) + (1 - s_I)(1 - s_E)(s_o)$$

In this expressions, we have simplified considering $a = 1$ $p_u = q_2 - q_1 + \alpha N_2 - c_u$, $p_n = q_2 + \alpha - c$, $p_o = q_1 + \alpha - c$, $s_n = q_2 + \alpha - c_u$ and $s_o = q_1 + \alpha$.

In this expression, $N_1 p_1$ corresponds to the period 1 revenues, $s_I(1-s_E)(N_1 p_u + N_2 p_n)$ are the period 2 revenues that can be obtained if the incumbent firm is the only innovator, $(1-s_I)(1-s_E)(N_2 p_o)$ are the period 2 revenues for the case where no firm innovates, and $\frac{s_I^2}{2}$ is the cost attached to the innovation process. Recall that if the two firms innovate, profits are dissipated due to the price competition and that there is no revenues for the incumbent if the potential entrant is the unique innovator.

Consider the revenues obtained in period 1. As can be seen, p_1 extracts the full surplus enjoyed by the N_1 by charging the total surplus enjoyed in period 1 (i.e. $q_1 + \alpha N_1 - c$) and the expected surplus enjoyed in period 2 (i.e. $s_I(1 - s_E)(s_n - p_u) + (1 - s_I)(1 - s_E)(s_o)$). Moreover, as in the monopoly case, the period 1 price charged by the incumbent decreases with its own investment level. In particular, $\frac{\partial p_1}{\partial s_I} = -\alpha N_2(1 - s_E) < 0$. This observation implies that through a higher level of investment, the incumbent firm reduces the willingness to pay of the N_1 consumers in period 1. At the same time, and similar to the monopoly case, higher investments boost period 2 revenues. However, investments in the context analyzed in this subsection play an additional role: deter entry. Therefore, we analyze not only the trade-off between more revenues in period 1 or 2, but also consider the preemptive role of investments.

Analogously, the problem of the entrant firm is given by,

$$\max_{s_E} \Pi_E = s_E(1 - s_I)(N_1 p_u + N_2 p_n) - \frac{s_E^2}{2} \tag{5}$$

Again, we have simplified using $a = 1$ $p_u = q_2 - q_1 + \alpha N_2 - c_u$, $p_n = q_2 + \alpha - c$, $p_o = q_1 + \alpha - c$, $s_n = q_2 + \alpha - c_u$ and $s_o = q_1 + \alpha$. Note that the entrant can only have positive revenues if it is the unique innovator. In addition, it is important to highlight that the fact that the potential entrant has no period 1 revenues, it will not face any commitment problem. However, given that the investment levels are obtained strategically, the behavior of the incumbent has an important impact on the behavior of the potential entrant.

3.2.3 No Commitment Case

As in the monopolist problem, this case is obtained by focusing on the SPNE. Accordingly, the first-order condition for the incumbent firm taking into account only second period profits is given by,

$$0 = (1 - s_E)(N_1 p_u + N_2 p_n) - (1 - s_E)(N_2 p_o) - s_I^{nc} \tag{6}$$

Considering equation (5), the SPNE concept provides the first-order condition for the entrant firm given by,

$$0 = (1 - s_I)(N_1 p_u + N_2 p_n) - s_E^{nc} \tag{7}$$

It can be seen that the second-order conditions for an interior solution are satisfied. Thus, solving equations (6) and (7) provides the equilibrium R&D levels for the incumbent and the entrant firm in the absence of commitment by the incumbent firm. Again, note that given that the entrant firm only competes in period 2, it has no choice concerning a committed action. Before analyzing the results, we calculate first the commitment case.

3.2.4 Commitment Case

As should be clear by now, the NE of the global game represents the commitment solution and provides the following first-order condition for the investment level by the incumbent. That is,

$$\begin{aligned} 0 = {} & N_1((1 - s_E)(s_n - p_u) - (1 - s_E)(s_o)) \\ & + (1 - s_E)(N_1 p_u + N_2 p_n) - (1 - s_E)(N_2 p_o) - s_I^c \end{aligned} \tag{8}$$

Analogously, the first-order condition for the entrant firm is,

$$0 = (1 - s_I)(N_1 p_u + N_2 p_n) - s_E^c \tag{9}$$

As in the case of no commitment, solving equations (8) and (9) provides the equilibrium investment levels for both firm in the presence of commitment of the incumbent firm. In order to simplify the analysis (given the large number of parameters), we consider the behavior of the best response functions described by the first order conditions. Given the specifications on the R&D processes, from observations of equations (6) and (7) for the no commitment case, and equations (8) and (9) for the commitment case, the best response functions are linear and therefore provide a unique equilibrium. Moreover, they are downward sloping implying strategic substitutability in the investment levels. We require and additional assumption to guarantee the existence of an economically plausible equilibrium.

Assumption 4. $q_2 < 1 + c_u - \alpha$. *The best response functions that describe the incentives to innovate are stable.*

As can be seen, assumption 4 restricts the size of the innovation. This assumption guarantees, in addition to provide stability to the best response functions, that for any parameter configurations, the probabilities of success lie on the interval $(0, 1)$. Figure 1 shows the behavior of the best response functions and suffices to provide the main results.

As can be seen from the figure, $R_E(s_I)$ represents the best respond function for the entrant as a function of the investment level of the incumbent firm. This function is obtained from solving equation (7) for s_E^{nc}.[24] Equivalently, the best respond functions for the incumbent firm, $R_{Inc}(s_E)$ and $R_{Ic}(s_E)$, are obtained from solving equations (6) and (8) for s_I^{nc} and s_I^c, respectively. It can be shown that under assumption 4 the best response functions lie always on the positive quadrant and below 1.

[24]Note that the form of $R_E(s_I)$ does not depend on the presence of commitment because the entrant only competes in period 2. Therefore, $R_E(s_I)$ can also be obtained from solving equation (9) for s_E^c.

In particular, the analysis of the market outcome is summarized in Figure 1. Figure 1a shows the case where network externalities are important and Figure 1b shows the case without network externalities. Figure 1a shows two main results. First, independent of the presence of commitment, the potential entrant always invest more that the incumbent firm. That is, in any case the equilibrium lies below the 45 degree line. And Second, as explained above, in the absence of commitment, the incumbent firm does not internalize the negative effect that its own investment has on his first period price and, therefore, invest more than it would be the case if commitment is possible. As a consequence, once commitment is considered the incumbent corrects its R&D expenditures negatively. This correction implies a stronger incentive for the entrant to innovate and, hence, increases the entrant's level of investment. In Figure 1 this is represented through the fact that the commitment equilibrium lies below and to the right of the no commitment equilibrium. This result holds for any parameter configuration satisfying the assumptions of the model.

Proposition 3. *Independent of the possibility of commitment by the incumbent, the potential entrant always invests in R&D more than the incumbent firm. Moreover, this difference is increased if commitment is possible.*

In addition, from equations (6) and (8) it can be shown that the difference between the commitment and no commitment case is only due to the presence of network externalities. This is represented in Figure 1 by the fact that the difference between the best response function of the incumbent without commitment lies above the best response function in the presence of commitment. In particular, the difference between the points at which both lines intersect the vertical axis is always positive and equal to $\alpha N_1 N_2$. Therefore, the strategic impact of entry is completely isolated. Figure (1b) shows a particular case with $\alpha = 0$.

Proposition 4. *The difference in the optimal investment levels with or without commitment is only due to the presence of network externalities.*

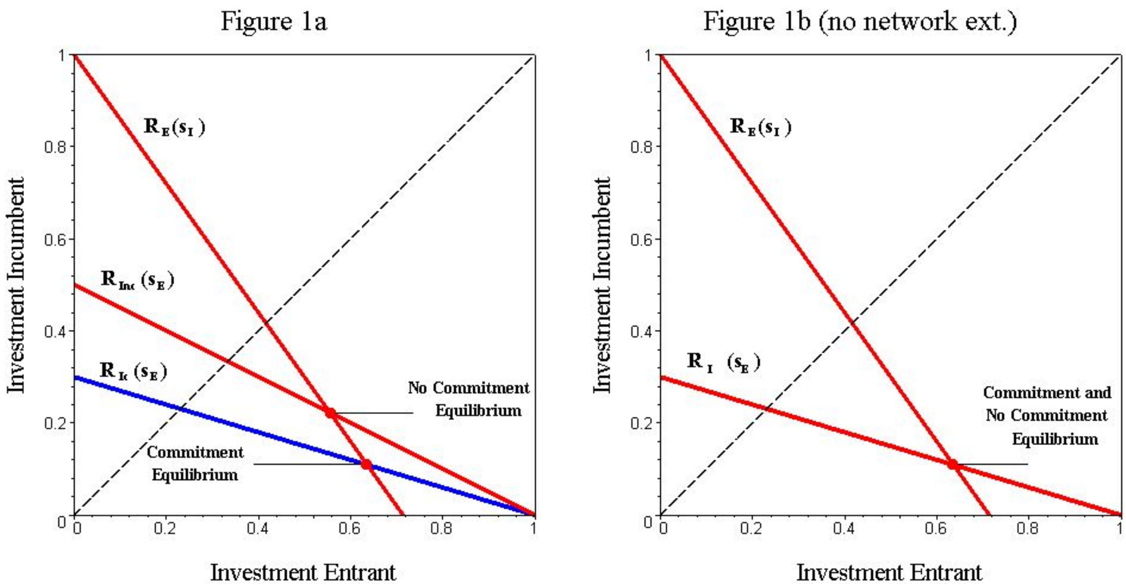


Figure 1: Best Response Functions - Market Outcome

Importantly, several numerical analyses suggest that, for some parameter configurations, the profit of the incumbent is higher in the absence of commitment than it would be the case if commitment is possible. That is, the threat of entry implies that in some special cases it is strategically optimal for the incumbent to increase its R&D investment as a mechanism to response to the potential entrant. This result is in clear contrast with the monopoly analysis presented before and, therefore, extends the analysis of Ellison and Fudenberg (2000).

The result that the threat of entry may eliminate the commitment problem of a monopolist in durable goods market has been analyzed by Bucovetsky and Chilton (1986), Ausubel and Deckenere (1986) and Vettas (2001). However, to the best of our knowledge, there is no analysis that considers the role of R&D incentives in this situation and, therefore, our result differs from the current literature.

Proposition 5. *With the threat of entry, the incumbent firm may achieve a higher profit by strategically not committing its investment level. This is in contrast to the case without the threat of entry.*

One of the main objectives of this paper is to analyze the social efficiency of the incentives to innovate in the presence of network externalities and durable goods. This is the purpose of the next section.

4 Social Optimum

In the previous section we obtained the incentives to innovate in an industry that exhibits network externalities and durable goods. In particular, we considered the monopoly case and concluded that, in line with the current literature, in the absence of commitment the monopolist has incentive to invest in R&D in excess of what it would maximize its overall profits. Moreover, we showed that the negative impact of this over-investment was reflected in lower social welfare and it was a consequence of the presence of network externalities.

Subsequently, we analyzed the case where the monopolist is faced by a potential entrant. Interestingly, we were able to conclude that due to the threat of entry, the commitment problem exhibited in the monopoly case by the incumbent firm was not present anymore. Even thought the absence of commitment was reflected in higher investments because the incumbent is not able to internalize the negative impact on his period 1 pricing, the threat of entry, and the induced higher level of investment, more than compensated the lower period 1 revenues by increasing the expected period 2 profits.

However, it is important to analyze the social efficiency of the results obtained in the previous section. Therefore, and as a major objective of this paper, the present section consider the problem faced by a social planner that maximizes social surplus. In particular, we obtain the socially optimal R&D incentives and compare our results with the ones obtained before for the case of the market outcome. Moreover, we investigate the role of network externalities in the potential social inefficiencies that may arise.

Assuming that the social planner is able to produce the two goods, set prices equal to zero, induce adoption and invest in R&D, its problem can be written as,

$$\max_{s_I,s_E} W = N_1 p_1^s + s_I s_E (N_1 s_n + N_2 p_n) s_I (1 - s_E)(N_1 s_n + N_2 p_n)$$
$$+ s_I (1 - s_E)(N_1 s_n + N_2 p_n) + (1 - s_I)(1 - s_E)(N_1 s_o + N_2 p_o) \qquad (10)$$
$$- \frac{s_I^2}{2} - \frac{s_E^2}{2}$$

with,

$$p_1^s = q_1 + \alpha N_1 - c$$

As before, we have simplified taking into account $a = 1$ $p_u = q_2 - q_1 + \alpha N_2 - c_u$, $p_n = q_2 + \alpha - c$, $p_o = q_1 + \alpha - c$, $s_n = q_2 + \alpha - c_u$ and $s_o = q_1 + \alpha$.

Equation (10) is obtained by calculating, for each period, the maximum social surplus that can be enjoyed by the entire population given that the social planner can induce adoption. In addition, the assumption that the social planner invests in the two technologies simply reflects a risk diversification strategy. That is, ex-ante, it is impossible for the social planner to realize which technology will be successful in period 2. Also, note that investing in both technologies is an efficient strategy given the quadratic form of the costs associated with the innovation process.

Note that for the social planner problem the SPNE and the NE coincide. Therefore, we can calculate the first-order condition that provide the socially optimal investment level. This expressions are,

$$0 = (1 - s_E)(N_1 s_n + N_2 p_n) - (1 - s_E)(N_1 s_o + N_2 p_o) - s_I^w \qquad (11)$$

$$0 = (1 - s_I)(N_1 s_n + N_2 p_n) - (1 - s_I)(N_1 s_o + N_2 p_o) - s_E^w \qquad (12)$$

As can be seen from equations (11) and (12), the social planner invests equally in both technologies. This is due to the fact that the social planner internalizes the costs of the projects. Moreover, straightforward manipulations of equations (11) and (8) show that the best response function of the social planner is identical to the one exhibit by the incumbent firm in the presence of commitment. This implies that in order to compare the social optimum with the results from the market outcome we should consider the results presented in Figure 1 with the level of investment produced by the incumbent's best response function in the presence of commitment. Given that the social planner invest equally in both technology, the social optimal level of investments is reached in the intersection of the incumbent's best response function with commitment and the 45 degree line. This is presented in Figure 2.

Figure 2 provides two interesting results. First, it shows that the entrant firm, unambiguously, always over-invests in R&D in relation to the socially optimal amount. That is, independent of the presence of commitment by the incumbent, the market equilibrium always lie to the right of the social optimum. This result is due to the fact that a successful innovation represents the only possibility for the potential entrant to make positive profits.

Proposition 6. *The potential entrant unambiguously exhibit an over-investment in comparison with the social optimum. This result is independent of the possibility of commitment by the incumbent firm.*

In addition, it can be observed in Figure 2 that in the absence of network externalities or for sufficiently low values of α the incumbent firm always under-invests in R&D. However, depending on the extent of the network externalities (i.e the value of α) the incumbent firm may exhibit a lower (Figure 2a), equal (Figure 2b) or higher (Figure 2c) level of investment compared with the social optimum. This result follows from numerical simulations.

Proposition 7. *Depending on the extent of the network externalities, the incumbent* *firm may exhibit a lower, equal or higher investment level in comparison with the social optimum.*

This result sheds some light on the controversy around the efficiency of the observed market structure in network industries. As has been pointed in the literature (and observed in reality), network industries are characterized by the presence of few successful incumbents. This observed structure has led regulation authorities to consider whether the high level of concentration is detrimental for the socially optimal level of innovation undertaken in these industries. Our analysis shows that there is no clear answer to that questions and that the measurement of the extent of network externalities may be crucial for policy purposes. Hence, any conclusion must be based on a formal analysis and this paper is a small step in that direction.

5 Conclusions

We presented a model of R&D competition between an incumbent and a potential entrant in market with durable goods and network externalities. In particular, we analyzed the market outcome and the social efficiency of the incentive to innovate in the presence of uncertain innovation processes. The robustness of the presented results with respect to the assumed functional forms is the objective of current work.

We found three main results. First, the threat of entry reverses the commitment problem that a monopolist (without such threat) may face in its R&D decision given the durability of the network goods. This result is not present in the current literature on R&D and follows from the role that R&D incentives play in deterring entry. In our case, the monopolist's commitment problem arises only due to the presence of network externalities.

Second, the levels of R&D determined by market outcome might differ from the socially optimal levels. In particular, a potential entrant always over-invests (as an entry

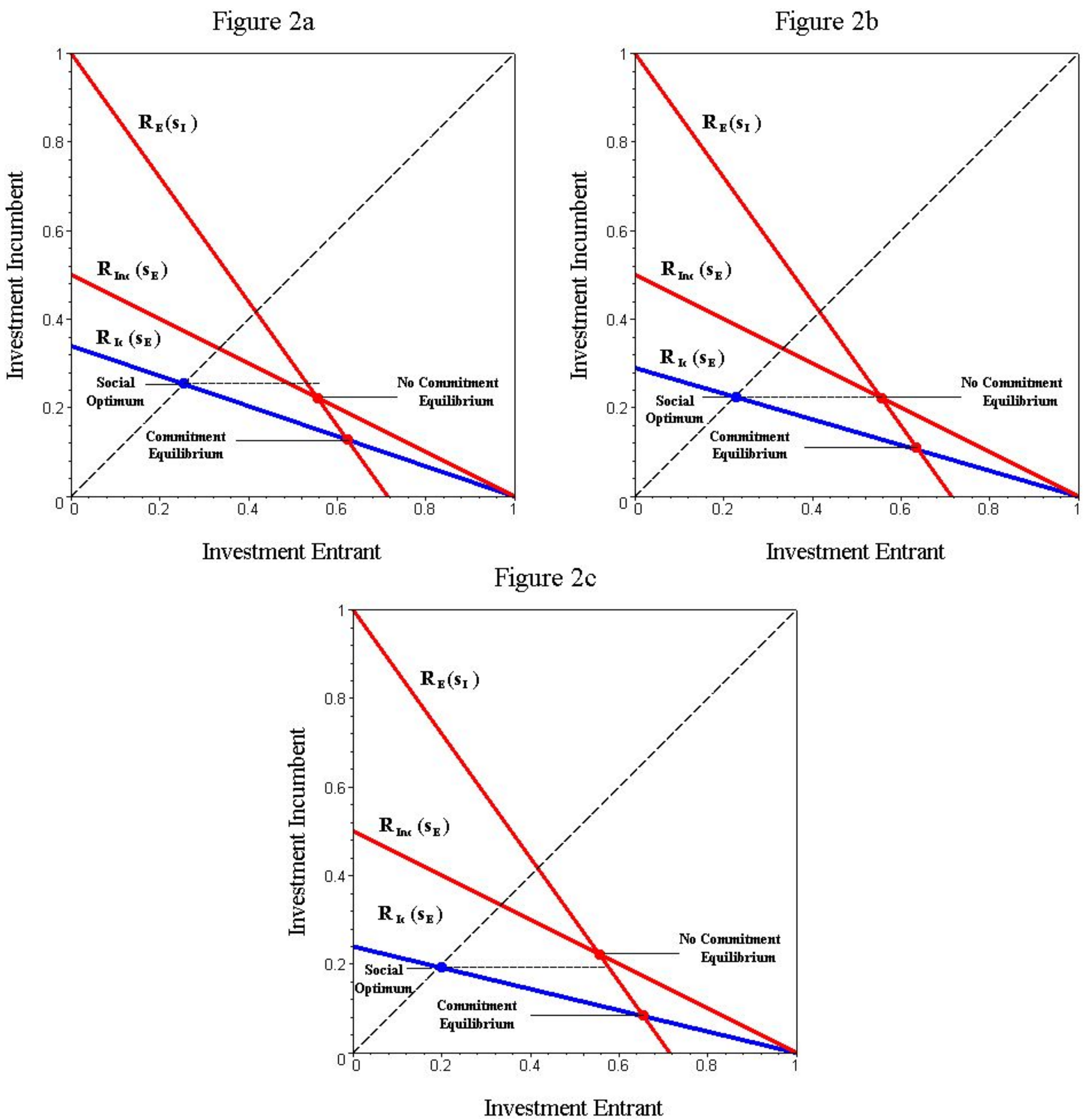


Figure 2: Best Response Fncs. - Social Optimum

strategy) and an established incumbent might exhibit higher, lower or equal R&D levels in comparison with the social optimum. This result suggests that successful entry takes place too often in comparison with the social optimum.

And third, the extent of network externalities is the crucial parameter in the efficiency of the incumbent R&D level. In fact, it is only the presence of network externalities that permits, potentially, to the established incumbent to provide an efficient level of innovation. Without network externalities (or very low network effects), it is shown that the incumbent firm always under-invests in R&D efforts. This result sheds some light on the debate whether a dominant incumbent in a network industry provides sufficient innovation to the society.

We recognize several areas of further research in the area of R&D incentives in the presence of network externalities and durable goods. To reduce the dependence on initial conditions and parameter assumptions, a fully dynamic model may shed light on some more realistic characteristics of industry evolution inside the framework analyzed in current paper. In addition, the analysis of compatibility decisions must also be considered given its obvious relevance in these industries but for the time being beyond the scope of the present paper. Finally, a more detailed (or alternative) description of the consumers' coordination assumptions may enrich the results.

References

Ausubel, and Deckenere, (1987), "One is Almost Enough for Competition", *RAND Journal of Economics*, 18, 255-274.

Bagnoli, M., S. Salant, and J., Swierzbinski (1989), "Durable-goods Monopoly with Discrete Demand", *Journal of Political Economy*, 97, 1459-1478.

Bensaid, B., and Lesne, J.P., (1996), "Dynamic Monopoly Pricing with Network Externalities", *International Journal of Industrial Organization*, 14, 837-855.

Besen, S. and J. Farrell, (1994), "Choosing How to Compete: Strategies and Tactics in Standardization", *Journal of Economic Perspectives*, vol. 8, pp. 117-131.

De Bijil, P.W.J. and S. Goyal, (1995), "Technological Change in Markets with Network Externalities", *International of Industrial Organization*, 13, 307-325.

Bucovetsky, S. and J. Chilton, (1986), "Concurring Renting and Selling in a Durable Goods Monopoly under Threat of Entry", *RAND Journal of Economics*, 17, 261-275.

Bulow, J., (1982), "Durable Goods Monopolists", *Journal of Political Economy*, 15, 314-332.

Bulow, J., (1986), "An Economic Theory of Planned Obsolescence", *Quarterly Journal of Economics*, 101, 729-749.

Cabral, L.M.B., D. Salant, and G. Woroch, (1999), "Monopoly Pricing with Network Externalities", *International Journal of Industrial Organization*, 17, 199-214.

Coase, R., (1972), "Durability and Monopoly", *Journal of Law and Economics*, 15, 143-143.

Choi, J.P., (1994), "Network Externalities, Compatibility Choice, and Planned Obsolescence", *Journal of Industrial Economics*, 42, 167-182.

Ellison, G., and D. Fudenberg, (2000), "The Neo-Luddite's Lament: Excessive Upgrades in the Software Industry", *RAND Journal of Economics*, 31, 253-272.

Farrell, J. and M. Katz, (2005), "Competition or Predation? Consumer Coordination, Strategic Pricing and Price Floors in Network Markets", *Journal of Industrial Economics*, LIII, 203-231.

Farrell, J. and G. Saloner, (1985), "Standardization, Compatibility and Innovation", *Rand Journal of Economics*, vol. 16, 70-83.

Farrell, J. and G. Saloner, (1986), "Installed Base and Compatibility: Innovation, Product Preannouncement, and Predation", *American Economic Review*, 76, 940-955.

Fudenberg, D. and J. Tirole, (1998), "Upgrades, Tradeins, and Buybacks", *RAND Journal of Economics*, 29, 235-258.

Fudenberg, D. and J. Tirole, (2000), "Pricing a Network Good to Deter Entry", *Journal of Industrial Economics*, XLVIII, 373-390.

Grout, P.A., and I. Park, (2005), "Competitive Planned Obsolescence", *RAND Journal of Economics*, 36, 696-612.

Gul, F., H. Sonnenschein, and R. Wilson, (1986), "Foundations of Dynamic Monopoly and the Coase Conjecture", *Journal of Economic Theory*, 39, 155-190.

Hendel, I., and A. Lizzeri, (1999), "Adverse Selection in Durable Goods Markets", *American Economic Review*, 89, 1097-1115.

Kahn, C. M., (1986), "The Durable Goods Monopolist and Consistency with Increasing Costs", *Econometrica*, 54, 275-294.

Katz, M. and C. Shapiro, (1985), "Network Externalities, Competition and Compatibility", *American Economic Review*, vol. 75, 424-440.

Katz, M. and C. Shapiro, (1986), "Technology Adoption in the Presence of Network Externalities", *Journal of Political Economy*, 94, 822-84.

Katz, M. and C. Shapiro, (1992), "Product Introduction with Network Externalities", *Journal of Industrial Economics*, 40, 55-84.

Katz, M. and C. Shapiro, (1994), "Systems Competition and Network Effects", *Journal of Economic Perspectives*, vol. 8, pp. 93-115.

Kristiansen, E.G., (1996), "R&D in Markets with Network Externalities, *International Journal of Industrial Organization*, Vol. 14, pp. 769-784.

Mason, R., (2000), "Network Externalities and the Coase Conjecture", *European Economic Review*, 44, 1981-1992.

Reinganum, J., (1989), "The Timing of Innovation: Research, development and diffusion". In R. Schmalensee and R. Willig, (eds.) (1989). *Handbook of Industrial Organization*, North-Holland. 849-908.

Shapiro, C. and H. Varian, (1999). Information Rules: A strategic guide to the network economy. *Harvard Business School Press.*

Stokey, N., (1981), "Rational Expectations and Durable Goods Pricing", *Bell Journal of Economics*, 12, 112-128.

Waldman, M., (1993), "A New Perspective on Planned Obsolescence", *Quarterly Journal of Economics*, 108, 273-283.

Waldman, M., (1996), "Planned Obsolescence and the R&D Decision", *RAND Journal of Economics*, 27, 583-595.

Waldman, M., (1997), "Eliminating the Market for Secondhand Goods: An Alternative Explanation for Leasing", Journal of Law and Economics, 40, 61-92.

Waldman, M., (2003), "Durable Goods Theory for Real World Markets", *Journal of Economic Perspectives*, 17, 131-154.

Does IT-Outsourcing Increase Firm Success? A Firm-level Investigation for Germany

Jörg Ohnemus

Centre for European Economic Research (ZEW), Mannheim

March 15, 2007

Abstract

Using German firm-level data, an endogenous switching regression model is employed for labor productivity and firms' decision to source out basic IT-services. This approach takes possible complementarities between IT-outsourcing and production input factors into account and allows IT-outsourcing to affect any factor of the production function. Estimation results show that IT-outsourcing firms produce more efficiently than non-IT-outsourcing firms. Furthermore, they have a significantly larger output elasticity of ICT-capital. ICT-capital and IT-outsourcing can for that reason be interpreted as complementary factors positively affecting firms' labor productivity. An additional analysis indicates that IT-outsourcing, in the medium term, has no negative effect on firms' growth rate.

Keywords: IT-Outsourcing, Productivity, Endogenous Switching Regression, Labor Force Growth

JEL-Classification: C21, D24, J21, J24

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–(0)621–1235–354
Fax: +49–(0)621–1235–4354 or +49–(0)621–1235–333
E-Mail: ohnemus@zew.de

1 Introduction

Outsourcing, especially information technology (IT) outsourcing received a lot of attention in public debate during the last couple of years. Big (IT) outsourcing deals with Asian or East European companies lately raised fears in industrialized nations about its negative consequences for the local labor market. But so far, most of the IT-outsourcing activities in Germany, for example, are still done locally. A recent study by the Centre for European Economic Research (Zentrum für Europäische Wirtschaftsforschung; 2005) shows that 94 percent of German firms, which are outsourcing their IT-services partly or fully, exclusively cooperate with local IT-vendors. Only 6 percent of firms, which are mainly big manufacturing companies, are active in so called IT-offshoring. It seems that IT-offshoring is not playing such a mayor role compared to the whole IT-outsourcing market. Nevertheless, even the outsourcing of IT-services to nearby local external vendors has consequently a negative effect on jobs in the outsourcing companies, because employees which were assigned to take care of those services are not needed anymore. But subsequently, those jobs have to be created at least to some extend in the IT-service providing sector.

A natural question to ask in relation with IT-outsourcing is the following: Can firm performance be improved through the outsourcing of information technology? In most companies IT does not belong to the core competencies of the firm. Consequently, it would be beneficial for them to source out if they can find an appropriate provider, which does the same work for at least the same costs and quality. This would help the firms' management to concentrate on their core competencies again. As a result, one could observe a better future performance of firms that source out their IT-activities in terms of productivity increases and in the potential to hire employees working for the core business of the firm.

After establishing a positive impact of ICT-investment on firms' productivity in numerous empirical analysis of the last decade, there is still little evidence about the effects of IT-outsourcing to external vendors on the outsourcing firms' economic performance. The aim of this paper is to close this gap using German firm-level data collected in 2000 and 2004 by the Centre for European Economic Research. The research focuses thereby on the effects of IT-outsourcing on labor productivity and the growth rate of firms' labor force (as a measure of firm success). To analyze labor productivity I use cross-sectional data for the year 2004 and employ an endogenous switching regression model (Bertschek and Kaiser; 2004; Bertschek et al.; 2006). This specifications allows to take two aspects into account: First, their might be potential simultaneity between labor productivity and IT-outsourcing. Causality can go in either direction, IT-outsourcing might affect labor productivity or the other way around since firms might outsource their IT-task in order to increase productivity. Second, firms are allowed to produce according to different production function regimes depending on their decision to source out basic IT-services to external providers. With this flexible representation, the presence of complementarities between IT-outsourcing and the factor inputs can be accounted for. Firms' labor force growth rate between 1999 and 2003 is examined by an instrumental variable approach, accounting for the potential endogeneity of IT-outsourcing on the firms growth rate.

The empirical results show, that ICT-capital has a significantly higher contribution on labor productivity for firms involved in IT-outsourcing, suggesting that those firms exploit their ICT-capital more efficiently. Furthermore, also the constant terms in the two regimes, representing multifactor productivity, are significantly different and higher for IT-outsourcer. Therefore, firms' which use external vendors to take care of their IT-services produce more efficiently over all, compared to their non-IT-outsourcing counterparts. A supplementary analysis of the firms' labor force growth rate indicates that IT-outsourcing has no negative effects on the medium term growth of firms'.

The rest of this paper is structured as follows: Section 2 gives a short overview of the related literature, before in section 3 the model and the data for the labor productivity analyzes is described. Section 4 presents the results of the endogenous switching regression estimation. Finally, in section 5 the effect of IT-outsourcing on the labor force growth rate is examined and section 6 concludes.

2 Literature Review

There is a large body of literature which is dealing with outsourcing or offshoring (outsourcing across country boarders) of material and service inputs, both on industry and on firm-level. Service inputs thereby mostly include computer services, albeit on a very aggregated level. Comprehensive overviews of this literature are given in Olsen (2006) and Heshmati (2003). One branch of the literature is dealing with the differences in productivity growth in manufacturing and service industries, arguing that the outsourcing of services helped to promote productivity growth in the manufacturing sector (i.e. Siegel and Griliches, (1992); Fixler and Siegel, (1999); ten Raa and Wolff, (2001)). Amiti and Wei (2006) investigate the effects of offshoring services and material inputs on total factor productivity (TFP) and labor productivity using industry level data for the United States. They find that service offshoring has a positive significant effect on TFP and that it accounts for 11–13 percent of labor productivity growth over the period 1992 to 2000.[1] One of the earliest papers dealing with the impact of service outsourcing on productivity on the firm-level is presented by Görzig and Stephan (2002). They find that firms tend to overestimate the benefits accruing from outsourcing of business-services, as opposed to material outsourcing. In contrast, Girma and Görg (2004) find that the outsourcing intensity of industrial services in some industries is positively related with labor productivity and total factor productivity growth. For the impacts of offshore outsourcing on productivity, see for example Görg and Hanley (2003; 2004; 2005) and Görg et al. (2005).

Most of the literature dealing with IT-outsourcing, is concerned with identifying factors that have an impact on firms' decision to source out information technology. Thereby,

[1]Material offshoring only accounts for 3 to 6 percent of labor productivity growth in this time span.

they mainly draw upon economic theories such as transaction cost or production theory. For example Loh and Venkatraman (1992) find that production cost advantages are the main reason for IT-outsourcing. But until now there is little empirical evidence about whether IT-outsourcing has a significant economic contribution for the outsourcing firm at all. In contrast, the positive impacts of information technology investment on productivity is by now a well established fact which is documented in several empirical firm-level studies (i.e. Brynjolfsson and Hitt (1996; 2000; 2003), Greenan and Mairesse (2000) or Hempell (2005), to mention a few). To amend this line of literature, an examination whether IT-outsourcing contributes to the performance of the outsourcing firm would be interesting. This is exactly what is done in this paper.

3 Estimating Productivity Effects

In order to investigate the effect of IT-outsourcing on firm-level labor productivity, I apply an endogenous switching regression model within a production function framework.[2] The endogenous switching model assumes that some unobserved factors affect the IT-outsourcing decision and labor productivity simultaneously. In contrast to traditional instrument variable methods, the endogenous switching regression allows IT-outsourcing to change the entire set of partial productivity elasticities instead of a priori restricting IT-outsourcing to act only as a shift parameter in the productivity equation (Bertschek and Kaiser; 2004, p. 395). The estimation results indicate that labor productivity and the outsourcing of IT-services are simultaneously determined. In the next section the used model is described in more detail.

[2]See for example Bertschek and Kaiser (2004) and Bertschek et al. (2006). For further details about the endogenous switching regression approach see Maddala (1983)

3.1 The Model

The model assumes that firm i produces according to a Cobb-Douglas production technology. The output Y_i of firm i is a function of capital K_i, ICT-capital ICT_i, labor L_i and a vector $\mathbf{X}'_i$ with further variables determining the production process of firm i:

$$Y_i = A_i\, K_i^{\alpha}\, ICT_i^{\beta}\, L_i^{\gamma}\, \mathbf{X}_i'^{\theta}. \tag{1}$$

The term A_i represents differences in production efficiency not related to the input factors and reflects multifactor productivity. The exponents α, β and γ denote the output elasticities with respect to capital, ICT-capital, and labor. After taking logs on both sides and adding an i.i.d. error term denoted by u_i, Equation (1) can be rewritten as:

$$ln(Y_i) = ln(A_i) + \alpha ln(K_i) + \beta ln(ICT_i) + \gamma ln(L_i) + \mathbf{X}'_i\theta + u_i. \tag{2}$$

After subtracting $ln(L_i)$ from both sides, labor productivity in log output per employee $ln(\frac{Y_i}{L_i})$ is then given by:

$$ln\left(\frac{Y_i}{L_i}\right) = ln(A_i) + \alpha ln(K_i) + \beta ln(ICT_i) + (\gamma - 1) ln(L_i) + \mathsf{x}'_i\theta + u_i. \tag{3}$$

The following terms state the two different productivity regimes:

$$\begin{aligned} ln\left(\frac{Y_i}{L_i}\right)_{OUT} &= ln(A_{i,OUT}) + \alpha_{OUT} ln(K_i) + \beta_{OUT} ln(ICT_i) + \cdots \\ &\quad (\gamma_{OUT} - 1) ln(L_i) + \mathsf{x}'_i\theta_{OUT} + u_{i,OUT} \end{aligned} \tag{4}$$

$$\begin{aligned} ln\left(\frac{Y_i}{L_i}\right)_{nOUT} &= ln(A_{i,nOUT}) + \alpha_{nOUT} ln(K_i) + \beta_{nOUT} ln(ICT_i) + \cdots \\ &\quad (\gamma_{nOUT} - 1) ln(L_i) + \mathsf{x}'_i\theta_{nOUT} + u_{i,nOUT}. \end{aligned} \tag{5}$$

If firm i sources out IT-activities to an external provider, its labor productivity is given by equation (4) and if no outsourcing of IT-activities takes place, labor productivity can be stated as in equation (5). The subscripts OUT and $nOUT$ thereby denote the two productivity regimes *"do source out IT-services"* and *"do not source out IT-services"*.

The endogenous switching regression approach takes into account that firms with and without IT-outsourcing differ in observable and unobservable characteristics. If unobservable factors, which influence the decision to source out IT-services, also influence the firms' productivity, the expected values of the error terms in equations (4) and (5) are different to zero ($E(u_{i,OUT}) \neq 0$ and $E(u_{i,nOUT}) \neq 0$) and OLS estimation would lead to inconsistent results. To take account of this unobserved heterogeneity, one includes $E(u_{i,nOUT})$ and $E(u_{i,OUT})$ as additional explanatory variables in equation (4) and (5).

Therefore, it is necessary to analyze which firms are involved in outsourcing activities. The IT-outsourcing decision of the firm is positive if the expected gains from outsourcing are bigger than the associated costs. Thus, firm i is deputing an external vendor with taking care of its IT-services if the cost per employee associated with outsourcing C_i are smaller than the corresponding productivity profits. The latent variable

$$I_i^* = a\left(ln\left(\frac{Y_i}{L_i}\right)_{OUT} - ln\left(\frac{Y_i}{L_i}\right)_{nOUT}\right) - C_i + \epsilon_i \tag{6}$$

represents the difference between the productivity gains (weighted by the term a which represents the effect of the productivity gains from IT-outsourcing on the decision to outsource) and the costs arising from IT-outsourcing. The outsourcing decision is unaffected by the productivity differences if $a = 0$. The selection mechanism for observing IT-outsourcing then is

$$OUT_i = \begin{cases} 1 & \text{if} \quad I_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The selection equation is estimated in reduced form, $I_i^* = \mathbf{Z}_i\mathbf{\Pi} + \epsilon_i$, were the parameter vector $\mathbf{Z}_i$ includes variables that explain labor productivity and the variables that influence the cost of IT-services outsourcing.[3]

An assumption of the model is, that the error terms of the production functions $u_{i,OUT}$ and $u_{i,nOUT}$ as well as the error term of the selection equation ϵ_i are jointly normally distributed. The expected values of $u_{i,OUT}$ and $u_{i,nOUT}$ conditional on ϵ_i can be stated as follows:

$$E(u_{i,OUT}|\epsilon_i \leq \mathbf{Z}_i\mathbf{\Pi}) = \rho_{u_{OUT}\epsilon}\left(-\frac{\phi(\mathbf{Z}_i\mathbf{\Pi})}{\Phi(\mathbf{Z}_i\mathbf{\Pi})}\right), \tag{8}$$

$$E(u_{i,nOUT}|\epsilon_i \leq \mathbf{Z}_i\mathbf{\Pi}) = \rho_{u_{nOUT}\epsilon}\left(\frac{\phi(\mathbf{Z}_i\mathbf{\Pi})}{1-\Phi(\mathbf{Z}_i\mathbf{\Pi})}\right). \tag{9}$$

The parameters $\rho_{u_{nOUT}\epsilon}$ and $\rho_{u_{OUT}\epsilon}$ measure the covariance of the error terms of the production function for IT-outsourcers and non-IT-outsourcers with the selection equation. To obtain consistent estimates for both regimes, one has to use these conditional expected values as additional explanatory variables in the production function estimation. The resulting model is estimated as *full information maximum likelihood*(FIML).

In most cases it is difficult or even impossible to observe the direct costs C_i related to the IT-outsourcing process. An alternative way to solve this problem is to detect factors that are likely to influence the cost of IT-outsourcing and thus might have an effect on a

[3]At least one instrument variable, that explains IT-outsourcing but has no impact on labor productivity has to be added.

firm's decision to contract out IT-tasks to external providers. In this paper the following factors are assumed to influence this process:

- **works council:** A works council is regarded as having a negative effect on the IT-outsourcing decision, since it is presumable to expect that a works council is always strongly against the outsourcing (no matter what kind of activities) of formerly in-house production to an external provider. If the firm is giving away substantial parts of their in-house production, a works council would lose influence within the firm. Also, long negotiations with the works council to achieve an agreement about outsourcing will increase the costs of the outsourcing process.

- **export share:** Exporting firms are more exposed to market pressure compared to firms only active at home because they face worldwide competition. They are used to adjust more quickly to changes in the market environment and therefore, the costs of IT-outsourcing are assumed to be lower for those firms.

- **foreign subsidiary:** A similar argument holds for firms with a foreign subsidiary/location. Again, these firms are more confronted with international competition, resulting in smaller adjustment cost for the implantation of IT-outsourcing. Furthermore, multinational firms are usually presumed to employ superior technologies to domestic firms (Markusen; 1995) making IT-outsourcing even more favorable. On the other hand, firms' with a foreign subsidiary may have the possibility or are even forced to use the IT-department of the group (Bertschek and Müller; 2006). The sign of the effect then depends upon which argument dominates.

- **IT-applications:** The number of software and internet related applications adopted by the firm is supposed to have a negative effect on outsourcing. It is reasonable to assume that firms using more IT-applications have a more complex IT-infrastructure, which is being better looked after by internal IT-specialists than by an external provider. Furthermore, in those firms one can regard the IT-infrastructure as belonging to their core competencies. They rely heavily on a perfect functioning of their IT-applications in the production process. A decision to outsource core competencies is rather unlikely.

- **firm age:** For older firms the cost of implementing IT and reorganizing the production process is probably more expensive than for younger firms. According to Christensen and Rosenbloom (1995) new/younger firms are more flexible. Thus, they are more likely to adopt a new technology. Following this argument, younger firms might be more inclined to adopt a new business model, in our case this is IT-outsourcing to an external provider.

- **Y2K-consulting:** Firms which made use of consulting services for the *Year 2000 Problem* are considered to tend more to outsource IT-tasks, because they are already experienced in working together with an external IT-service provider.

3.2 The Data

The data set used for the estimation of the endogenous switching regression model results from a CATI-survey (computer-aided telephone interview) conducted in 2004 by the Centre for European Economic Research (ZEW). The survey has a special focus on the diffusion and the use of information and communication Technologies (ICT) in German companies. The data set originally contains detailed information for more than 4,000 firms with five and more employees, stratified by industry affiliation[4], size class and location (West/East Germany). Besides detailed information on ICT aspects concerning the firms, the data set contains additional information about total sales, the number of employees, the skill structure of the work force, investments, export share and various other variables.

[4]For a detailed description of the sectors included in the survey, see Table 13 in the Appendix.

Tables 1 and 2 compare industry and size structure of the employed data set with the values for the whole German population of firms in the respective industries and size classes.[5] Due to item-nonresponse the estimation sample reduced to 2,385 observations.[6] Regarding the share of firms in each industry, there are quite big differences to observe. Especially *retail trade* and *other business-related services* are substantially under-represented in the sample. But taking a look at total sales, the industry structure is represented fairly well. The majority of firms in the estimation sample are small and medium-sized firms (with respect to the number of employees). More than 88 percent of the firms have less than 50 employees. Taking a look at the generated sales in each size-class, the sample and the population expound a fairly equal distribution.

To enrich the data set and to construct value added as an indicator of output for the performed estimation, two other data sources are used to amend the final data set for estimation. The following paragraph describes the information additionally implemented and the source of this information.

From survey data, only total sales are available as a measure of firms' output. Since there is no further information for intermediate input, using sales in a production function framework might induce an omitted variable bias, since industries that operate at the end of the value chain (i.e. *wholesale* or *retail trade*) resort more strongly to intermediate goods than industries operating at an earlier stage of the value chain (Hempell; 2005; Schreyer and Pilat; 2001). To control for those differences, I calculated the shares of real value added at the NACE two-digit industry level using data from the German

[5] The sector *electronic processing and telecommunication* is excluded from the sample, due to the fact that firms providing IT-services to other companies are mainly located in this sector.

[6] A check of systematic differences in the anatomy of firms (with respect to firm size, sector affiliation, regional affiliation, capital and IT-capital) that have to be left out due to item-nonresponse indicates that these firms are missing at random.

statistical office.[7] The firm-specific values for total sales are then multiplied by those industry-specific shares. A further data source utilized, is the data base of the *Verband der Vereine Creditreform* (VVC), Germany's largest credit rating agency which was also used to draw the original survey sample. This database provides additional information about the founding date of the companies.

The survey questionnaire covered the whole range of IT-activities companies potentially need in running their business, asking further if the firm has outsourced each specific activity to an external service provider. The range of the covered activities lasts from basic IT-services, like *hard- and software installation* to more sophisticated services such as *software programming* and *IT-security*. To construct an outsourcing variable, I decided to choose only basic IT-activities[8], which are required in almost every firm. The constructed dummy variable takes the value one if a firm outsources at least one of those three basic IT-services completely and zero otherwise.

An overview of the IT-outsourcing intensity by industry affiliation is given in Table 3. A total of more than 39 percent of firms are involved in IT-outsourcing. The intensity lies in most of the industries slightly above the mean value, only the share of outsourcing firms in the *electrical engineering* industry and the *technical services* industry deviates quite substantially from the mean, with a share value of 23 percent and 29 percent, respectivly from the mean. Figure 1 shows the relative frequencies of IT-outsourcing in relation to firm size measured by $ln(employees)$. Since the outsourcing variable is binary, the relative frequencies are obtained by grouping the variable $ln(employees)$ into equispaced intervals. The size of the individual dots reflects the number of firms in each group. After a slight increase in the outsourcing intensity, the frequency drops sharply

[7]The online data access of the German Statistical Office (GENESIS) is providing this data, based on National Accounts (Volkswirtschaftliche Gesamtrechnung des Bundes).

[8]In detail, those are the (i) *installation of hard- and software*, (ii) *computer system maintenance* and (iii) *user assistance and support*.

and continues to fall for the firm-group with $4.75 \leq ln(L) < 5.25$.[9] There is no clear structure in the data for fairly large firms (with more than 1400 employees). At the same time the data set contains only a small proportion of large firms, as indicated by the size of the dots.

Since there is no data available for the physical capital stock of the firms, I used, as in Bertschek and Kaiser (2004), gross investment as an empirical proxy for the capital stock. This approach is a potential drawback of this study, but without panel data at hand, the calculation of firms' capital stocks through the perpetual inventory method (see for example Hempell; 2005) is not possible. Unfortunately, a couple of firms in the original data set have a missing value for investments or report that they have zero investments.[10] For the firms reporting zero investments, it seems reasonable to assume, that investment is positive but low and is rounded by the interviewee to zero. To tackle this problem, the value of investment for firms that report zero investment is set to the 10 percent quantile of their respective industry and size class. The investment value of firms having a missing value is replaced by the median value of their respective industry and size class.[11]

In the survey, there is also no information for ICT-capital available. Using ICT-investment would have been a solution to this problem because ICT-capital depreciates rather quickly. Unfortunately, data for ICT-investment isn't available, either. Therefore, I employed the percentage share of employees working mainly at a computerized workplace as a general proxy for ICT-capital in the organization in question.

[9]This range is approximately equivalent to 70–115 employees.

[10]With an econometric specification of the production function in log values, these firms would have been excluded from the sample.

[11]A total of 438 replacements were carried out, 112 for zero values and 326 for missing values.

Table 4 shows the descriptive statistics of the variables used in the estimation of the endogenous switching regression model. The quantitative variables $labor\ productivity$, $capital$, $employees$ and $export\ share$ refer to the year 2003. The other variables refer to the year 2004. The mean labor productivity (in value added terms) is €109,766 with the median value being substantially lower at around €55.000. Mean investment is €3,083100 and the average firm size amounts to 294 employees. More than 22 percent of the sample firms are located in East Germany. A works council exists in 40 percent of the firms and only a small sub-sample of 14 percent has a foreign subsidiary or foreign location. The average export share amounts to 17 percent. *University degree* and *vocational education* reflect the skill structure of the work force. The average amount of employees with the highest degree of education being university degree is 18.7 percent. More than 59 percent of the work force completed a vocational education. As indicated by the firm age dummy variable, the largest part of the firms in the sample is eight years and older. 54 percent of the firms apply a medium amount of IT-applications (software or internet related applications).

4 Results of the the Endogenous Switching Regression Model

4.1 Productivity Estimation

The estimation results of the two regime equations, with and without IT-outsourcing, from the endogenous switching regression model are presented in Table 5. The dependent variable *labor productivity*, as well as the variables *capital* and *labor* are transformed into their logarithmic values for estimation (see also section 3.1). After the estimation, Wald tests for identical coefficients in the two regimes are carried out, the results thereof are shown in Table 7. Results for the selection equation estimation are presented in section 4.2.

As one can see in Table 5, there is a highly significant and positive effect of capital, ICT-capital and labor on labor productivity in both regimes, with the partial elasticities for capital and labor being slightly higher in the regime without outsourcing activities.[12] But identity of the partial productivity estimates of capital and labor cannot be rejected at the usual significance levels (see Table 7). For ICT-capital, the partial elasticity is higher in the IT-outsourcing regime, with a significant difference to the regime without outsourcing. Additionally, a Wald test for identity of the factor inputs (capital, ICT-capital and labor) as well as for the identity of the entire set of variables included in the productivity equation reject identity. This result implies, that firms that make use of outsourcing their IT-activities seem to exploit their ICT-capital more efficiently than firms not involved in IT-outsourcing.

The sum of the three input elasticities amounts to 1.0578 in the regime without IT-outsourcing and to 1.0006 in the regime with IT-outsourcing. In the first case, the null hypothesis of constant returns to scale can be rejected (Wald: $\chi 2 = 6.92$; *p*-value = 0.0085). For IT-outsourcers, on the other hand, the constant returns to scale hypothesis cannot be rejected (Wald: $\chi 2 = 0.00$; *p*-value = 0.9855). However, a test for identical returns to scale for the two regimes is only weakly rejected (Wald: $\chi 2 = 2.8718$; *p*-value = 0.0901).

The indicators for the qualification structure of the work force, *university degree* and *vocational education*, as well as the presence of a *works council* and the *export share* show significantly positive contributions to labor productivity in both regimes. A *foreign subsidiary* of the firm, as well as a high amount of *IT-applications* only have a significant positive effect on productivity in the regime without IT-outsourcing. *Firm age* is insignificant in both regimes. For none of the before mentioned partial elasticities, identity of

[12]The estimated coefficients for the labor input correspond to $(\gamma - 1)$. Adding one to the estimated coefficient yields the partial output elasticity of labor.

the parameters cannot be rejected at the usual significance levels.

The dummy variables indicating whether a firm is located in East Germany has a significant and negative coefficient, reflecting lower labor productivity in East Germany. Interestingly, the difference between the two regimes is highly significant, leading to the result that East German IT-outsourcing firms are less productive than their non-outsourcing counterparts. Some, but not all of the industry dummies are significant (the base category is *metal and machine construction*). In this setting the coefficients of the sector dummies have no specific economic interpretation. They rather control for different measurement of labor productivity and other factors across industries.

An important result is the difference between the constant terms which reflect multi-factor productivity, in both regimes. The parameter is significantly larger in the IT-outsourcing regime. This implies the interesting result that firms being active in IT-outsourcing produce more efficiently than firms which do not outsource.

Table 5 also shows the correlation coefficients between the error term of the labor productivity equations and the error term of the selection equation, $\rho_{u_{OUT}\epsilon}$ and $\rho_{u_{nOUT}\epsilon}$. The two correlation coefficients are both positive and individually significant. Further, they are also jointly significant (see Table 8, last row). This implies that treating IT-outsourcing as truly exogenous for labor productivity is not appropriate. Because of the negative signs of the correlation coefficients, an unanticipated productivity shock would lead to a decrease in firms' propensity to source out IT-services.

4.2 Selection Equation

All the variables included in the estimation of the productivity regimes are included in the selection equation. Additionally, one instrument variable that explains the IT-outsourcing decision, but has no impact on labor productivity, is included.

The instrumental variable chosen for this purpose is *Y2K-consulting*. This dummy variable indicates whether a firm drew on external consultancy in conjunction with the year 2000 problem (also known as the Y2K problem, the millennium bug, and the Y2K Bug). The year 2000 problem was the result of a practice in early computer program design that caused some date-related processing to operate incorrectly for dates and times on and after January 1, 2000. Since computer technology is widely used in companies, virtually all firms were equally confronted with the thread of the year 2000 problem. The final decision to use consulting services depended upon the managements valuation of how seriously the Y2K problem would affect the firms normal course of business. This valuation of the year 2000 problems seems not related to productivity. On the other hand, firms that were already engaged in the "outsourcing" of the Y2K problem are more experienced in the usage of external help in solving IT-problems and therefore are more disposed to outsource IT-activities.

Table 6 contains the estimation results of the selection equation. For the model to be applicable, the explanatory restrictions have to be significant.[13] As one can see in the first line of Table 6, the Wald-statistic for joint significance of the entire set of explanatory variables is highly significant ($\chi^2 = 59.4594$; p-value $= 0.0000$). This suggests that the chosen exclusion restrictions are valid and hence the entire model is valid, too. A closer look on the individual coefficients shows that being located in a foreign country has a significantly negative effect on IT-outsourcing. This could be explained by the better availability of IT-resources in a group of companies. Furthermore, the coefficient for *Y2K-consulting* is positive and highly significant, inducing a strong effect of year 2000 consulting on the subsequent IT-outsourcing decision. There are no significant effects observable for the other identifying restrictions. For the factor inputs *ICT-capital* and *labor* the partial coefficients are negative and highly significant. This makes sense, since larger firms and firms with a higher IT-intensity have their own IT-departments which

[13]The chosen explanatory restrictions are *Y2K-consulting*, *works council*, *IT-applications*, *export share*, *foreign subsidiary* and *firm age*. For further details, see also section 3.1

take care of the IT-services needed in-house. Consequently, a test for joint significance of all three factor inputs is highly significant (see the second line in Table 8), indicating that the decision to source out IT-services is influenced by productivity differences.

5 IT-Outsourcing and the Growth of the Firm

After studying the effects of IT-outsourcing on labor productivity, I now want to check if there is a medium to long run impact of IT-outsourcing on the growth rate of the firms' labor force. Usually, in the short run, we would expect that the labor force of outsourcing firms is reduced, due to the fact that previous in-house IT-services are now done by external service providers. The jobs associated with this in-house production are now redundant, consequently leading to a downsize of the workforce. However, the long run effect of IT-outsourcing on the labor force growth rate is not clear.

An instrumental variable approach, taking a potential endogeneity of IT-outsourcing into account, is employed to investigate this question:

$$ln(employees^{2003})_i - ln(employees^{1999})_i = \alpha + \beta ITout_i + \mathbf{X}_i'\gamma + u_i, \qquad (10)$$

where $ln(employees)_i^{2003} - ln(employees)_i^{1999}$ is the log growth rate of firm i's workforce, $ITout_i^{2000}$ is a dummy variable indicating if the firm is outsourcing IT-services in 2000 and the vector $\mathbf{X}_i'$ contains all the other explanatory variables of labor force growth. As in the previous section, *Y2K-consulting* is used as an instrument for IT-outsourcing.

Based on equation (10) two specifications are estimated: The first specification is rather parsimoniously specified with the number of employees, the skill structure of the labor force, firm age and sector specific dummies as well as a location dummy as explanatory variables. In the second specification additional information is added about the IT-intensity of the firms (number of IT-applications, the share of employees working at a computerized workplace, the rate of change in computer workers, the rate of change of IT-specialists) and variables about the degree of internationalization of the firm (export share and wether the firm has a foreign location/subsidiary).

5.1 The Data

The data used for this analysis is basically drawn from the first wave of the ICT-survey conducted in the year 2000. The survey in 2000 has a focus on the scarcity of IT-specialists in German firms with more than 5 employees, but also includes questions about the diffusion and the use of ICT in German companies. Again, the data set originally contains detailed information for more than 4,000 firms stratified by industry affiliation, size class and location (West/East Germany).

Matching the data for 2000 and 2004 allows to construct the growth rate of the firms' labor force for the period 1999 to 2003, as well as the change in employees working at a computerized workplace and the change in the number of IT-specialists employed by firm i for the period 2000 to 2004.[14] Since only less than 1,300 firms took part in both waves of the ICT-survey, this leads to a serious reduction in the number of observation. After dropping observations due to item-nonresponse, the data set reduced to 907 observations for Specification I and 646 observations for Specification II, respectively.

The explanatory variable for IT-outsourcing is generated from data provided by the firms in 2000. As in the labor productivity estimation, only the outsourcing of basic IT-services is considered in the analyzes. Again, the constructed dummy variable indicating whether a firm is involved in IT-outsourcing takes the value one if at least one basic IT-task

[14] Computer workers are measured as percentage share of total employees and $(((\textit{share of computer employees})^{2004}/(\textit{share of computer employees})^{2000}) - 1) * 100$ is the growth rate of employees working at a computerized workplace. Employees specialized in information technology are reported in absolute numbers. Due to the fact that a lot of firms report zero IT-specialist in the base year 2000, I calculated a log growth rate adding 1 to each observation: $ln(\textit{IT-specialist}^{2004} + 1) - ln(\textit{IT-specialist}^{2000} + 1)$.

is sourced out to an outside provider and zero otherwise. Table 9 shows the number of firms and the percentage share of IT-outsourcing firms for each industry and both models under consideration.[15] *Wholesale trade* reveals the highest share of outsourcing firms with 63 and 65 percent. The least active firms in Specification I are located in the *electrical engineering* industry (28 percent) and in the *technical services* industry for Specification II.

The descriptive statistics for all variables used in the estimation is presented in Table 10. The mean value for the labor force growth is negative, being slightly lower in Specification I. Basically, the rest of the comparable variables are almost of identical magnitude in both models. The $\triangle$*standard wages* (99–03), which measures the increase in standard wages during the period 1999 to 2003 deserves a closer look. This information is provided by the German Statistical Office on NACE two-digit industry level. Besides the growth variables, all other variables refer to the year 2000.

5.2 Results

For examination of the labor growth rate, a 2SLS estimation procedure is employed. The first stage results for the instrumental variable estimation show a highly significant positive coefficient for *Y2K-consulting* in both model specifications, supporting its use as an instrument for IT-outsourcing. Year 2000 consulting activity is considered as exogenous to the growth rate of the labor force. The estimation results of the first stage regression are displayed in Table 11.

Table 12 reports the final estimation results. Turning to the parsimonious model (Specification I) first, it can be seen that the coefficient for IT-outsourcing is positive but

[15] The sector *electronic processing and telecommunication* is again excluded from the sample. For further details see footnote 5.

insignificant. To control for the initial size of the firm, *ln(employees)*, its squared and cubic term are included into the regression. There is a negative effect of the firm size on the growth rate of the labor force which diminishes as indicated by the positive coefficient of *ln(employees)*^{2} and increases later on again. The share of highly skilled employees has a positive and significant effect on firm growth. Furthermore, firms' aged between 4 and 7 years tend to grow more than their counterparts in the base group (firms older than seven years), a result that seems plausible since younger firms are generally smaller and thus have more potential to grow. If we add variables for the IT-intensity and internationalization of the firms to the regression equation (Specification II), the result regarding the effect of IT-outsourcing chances slightly. Now the coefficient is weakly significant, indicating a positive effect of IT-outsourcing on the subsequent growth rate of the firm. Whereas export share and a foreign subsidiary have no effect, a low number of IT-applications has a negative impact on growth. Interestingly, an increase in the share of employees with a computerized workplace leads to a reduction of total employment. A possible explanation therefore might be an efficiency increase through computer workplaces making other jobs superfluous. For the last variable reflecting firms' IT-intensity, the change in the number of IT-specialists employed, the coefficient is positive and highly significant.

6 Conclusions

The aim of this paper is to analyze the effects of IT-outsourcing on different measures of firm performance. In a fist step, the relationship between IT-outsourcing and firms' labor productivity is investigated using an endogenous switching regression model which separates firms into two regimes, IT-outsourcers and non-IT-outsourcers and this takes into account that IT-outsourcing might affect the productivity elasticities of the input factors. In a further step, medium to long run effects of IT-outsourcing on firms' labor force growth rate are examined, thereby using an instrumental variable approach to account for the possible endogeneity of IT-outsourcing on the growth rate of the labor force. For both analysis, firm-level data from a comprehensive survey conducted by the Centre for European Economic Research (ZEW) in the years 2000 and 2004 was available.

Using an endogenous switching regression model to account for the simultaneity between IT-outsourcing and labor productivity is justified by the jointly and individually significant correlation parameters of the productivity equations and the selection equation. Therefore, the right model for productivity analyzes is chosen. One important result shows, that firms which are active in sourcing out basic IT-services to external vendors, are more productive overall. This is expressed by the significantly higher coefficient for the constant term, reflecting multifactor productivity, in the regime with IT-outsourcing. Furthermore, firms that outsource IT, seem to exploit their IT-capital more efficiently than their counterparts without outsourcing. ICT-capital and IT-outsourcing can for that reason be interpreted as complementary factors positively affecting firms' labor productivity.

After determining the impact of IT-outsourcing on productivity it is interesting to see if there is also an effect of outsourcing on the firms' labor force growth rate observable. Two specifications are analyzed. The first, rather parsimonious specification, shows no significant impact on the growth rate, although the coefficient is positive. The second specification, which includes variables to account for the IT-intensity and the internationalization of the firms, gives evidence for a positive relationship of IT-outsourcing on firms' subsequent medium term growth rate. Since the effect is only weakly significant, this result has to be treated with caution. But it is reasonably safe to say, that IT-outsourcing is in the medium run not to blame for job displacement in firms active in sourcing out their IT-services.

Appendix

Table 1: Switching Regression – Comparison of Sample and Population by Industry

	Number of firms			Sales	
	Sample		Population[1]	Sample	Population[1]
Industries	#	%	%	%	%
consumer goods	239	10.02	5.45	6.31	7.60
chemical industry	144	6.04	0.58	16.00	6.11
other raw materials	206	8.64	2.09	6.07	4.11
metal and machine construction	293	12.29	4.71	6.25	6.57
electrical engineering	164	6.88	1.20	4.75	3.76
precision instruments	221	9.27	0.76	5.21	0.88
automobile	163	6.83	0.39	10.96	7.89
wholesale trade	129	5.41	9.52	5.92	14.56
retail trade	174	7.30	35.71	12.91	14.33
transportation and postal services	177	7.42	9.47	5.08	4.40
banks and insurances	117	4.91	1.06	17.22	20.88
technical services	196	8.22	5.01	1.22	0.81
other business-related services	162	6.79	21.37	2.11	4.93
Total	2,385	100.00	100.00	100.00	100.00

Note: [1]Share of German firms with five and more employees in the respective industry in 2004. [2]Share of sales of German firms with five and more employees in the respective industry in 2004. **Source:** German Statistical Office, ZEW and own calculations.

Table 2: Switching Regression – Comparison of Sample and Population by Size-Class

	Number of firms			Sales	
Size-Class	Sample		Population[1]	Sample	Population[2]
(# employees)	#	%	%	%	%
5-9	307	12.87	52.91	0.22	5.78
10-19	333	13.96	20.24	0.43	4.23
20-49	481	20.17	15.09	1.62	6.84
50-99	398	16.69	5.75	3.06	6.61
100-249	375	15.72	3.56	7.41	9.72
250-499	210	8.81	1.32	9.34	10.00
500 and more	281	11.78	1.12	77.91	56.82
Total	2,385	100.00	100.00	100.00	100.00

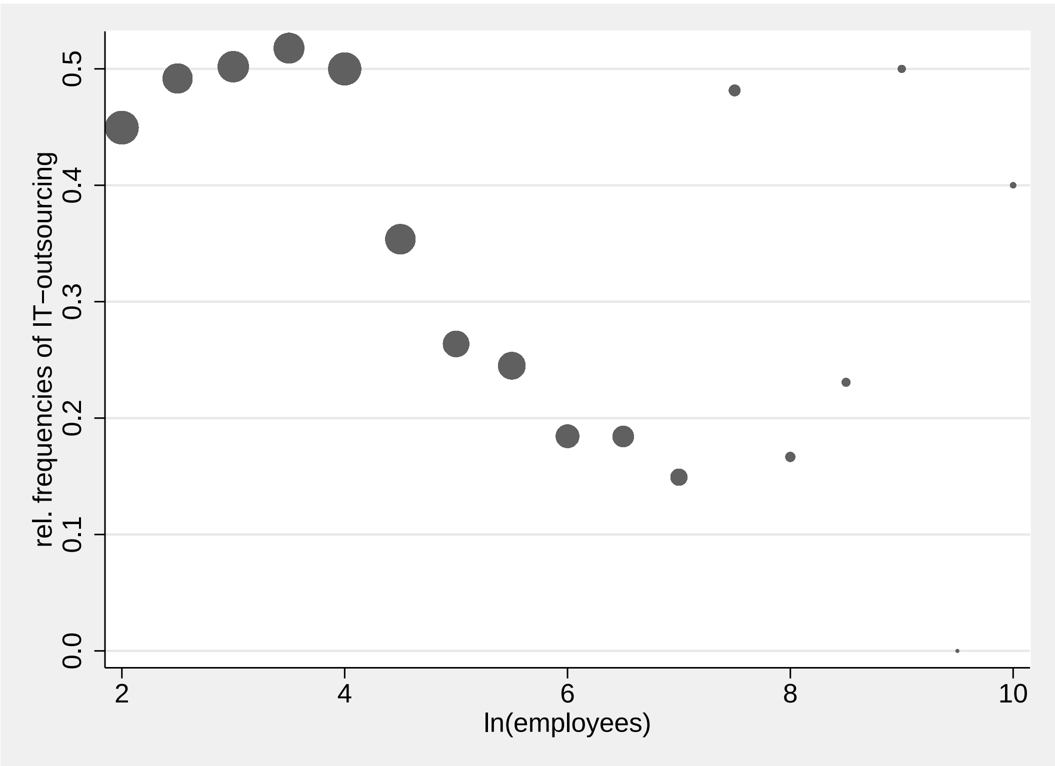
Note: [1]Share of German firms in the respective industry and size-class in 2004. [2]Share of sales of German firms in the respective industry and size-class 2004. **Source:** German Statistical Office, ZEW and own calculations.

Table 3: Switching Regression – Number and Share of Firms Involved in IT-Outsourcing by Industry

	IT-Outsourcing	
Industry	**# of firms**	**%**
consumer goods	108	45.19
chemical industry	57	39.58
other raw materials	84	40.78
metal and machine construction	115	39.25
electrical engineering	38	23.17
precision instruments	84	38.01
automobile	64	39.26
wholesale trade	59	45.74
retail trade	74	42.53
transportation and postal services	73	41.24
banks and insurances	47	40.17
technical services	56	28.57
other business-related services	76	46.91
Total	935	39.20

Note: Number and share of firms involved in basic IT-outsourcing in 2004.

Figure 1: Switching Regression – IT-Outsourcing vs. Firm-Size



Note: $ln(employees)$ grouped into equi-spaced intervals versus the relative frequencies of outsourcing. The size of the dots indicates the number of firms in the considered interval.

Table 4: Switching Regression – Summary Statistics

Variable	Mean	Standard Dev.	10% Quant.	50% Quant.	90% Quant.	Dummy Variable
labor productivity[§]	109.7662	236.6749	23	55	197	
capital[†]	3,083.10	23,121.58	15	220	5,000	
ICT-capital[‡]	42.1992	30.7758	9	30	95	
employees	293.6319	1,212.44	8	55	550	
IT-outsourcing	0.3920	0.4883	0	0	1	yes
Y2K-consulting	0.5367	0.4988	0	1	1	yes
East Germany	0.2247	0.4175	0	0	1	yes
university degree	18.7079	22.7126	0	10	55	
vocational education	59.2151	24.9582	20	63	89	
works council	0.4038	0.4908	0	0	1	yes
foreign subsidiary	0.1421	0.3493	0	0	1	yes
export share	16.9644	24.7887	0	3	60	
firm age ($\leq$ 3 years)	0.0528	0.2237	0	0	0	yes
firm age (4 $\leq$ years $\leq$ 7)	0.1711	0.3766	0	0	1	yes
firm age ($\geq$ 8 years)	0.7761	0.4169	0	1	1	yes
IT-application (0–3 appl.)	0.2264	0.4186	0	0	1	yes
IT-application (4–7 appl.)	0.5480	0.4978	0	1	1	yes
IT-application (8–10 appl.)	0.2256	0.4180	0	0	1	yes
consumer goods	0.1002	0.3003	0	0	1	yes
chemical industry	0.0604	0.2382	0	0	0	yes
other raw materials	0.0864	0.2810	0	0	0	yes
metal and machine construction	0.1229	0.3283	0	0	1	yes
electrical engineering	0.0688	0.2531	0	0	0	yes
precision instruments	0.0927	0.2900	0	0	0	yes
automobile	0.0683	0.2524	0	0	0	yes
wholesale trade	0.0541	0.2262	0	0	0	yes
retail trade	0.0730	0.2601	0	0	0	yes
transportation and postal services	0.0742	0.2622	0	0	0	yes
banks and insurances	0.0491	0.2160	0	0	0	yes
technical services	0.0822	0.2747	0	0	0	yes
other business-related services	0.0679	0.2517	0	0	0	yes
Number of observations			2385			

Note: [§]Value added per employee (in 2003) in €1,000. [†]Capital is proxied by gross investment. [‡]ICT-capital is proxied by the share of employees working mainly at a computerized workplace. **Source:** ZEW ICT-survey 2004 and own calculations.

Table 5: Switching Regression – Estimation Results of the Regime Equations

Estimation Results for ...	**... regime without IT-Outsourcing**		**... regime with IT-Outsourcing**	
	Coef.	**S.E.**	**Coef.**	**S.E.**
constant	2.5610***	0.1671	3.2957***	0.1468
capital	0.1160***	0.0202	0.1073***	0.0200
ICT-capital	0.0052***	0.0011	0.0097***	0.0013
labor	-0.0634**	0.0296	-0.1164***	0.0328
East Germany	-0.2575***	0.0553	-0.3990***	0.0596
university degree	0.0053***	0.0017	0.0059***	0.0023
vocational education	0.0024**	0.0011	0.0036***	0.0012
works council	0.1566***	0.0567	0.1350*	0.0703
export share	0.0030***	0.0009	0.0050***	0.0012
foreign subsidiary	0.1826***	0.0645	0.1248	0.1099
firm age ($\leq$ 3 years)	-0.1480	0.0935	-0.0890	0.1175
firm age (4 $\leq$ years $\leq$ 7)	-0.0395	0.0573	-0.0346	0.0697
IT-application (4–7 appl.)	0.0777	0.0622	0.0124	0.0594
IT-application (8–10 appl.)	0.1637**	0.0764	0.0870	0.0830
consumer goods	-0.0533	0.0755	-0.1811**	0.0806
chemical industry	0.1411	0.0978	0.2270**	0.1052
other raw materials	-0.0120	0.0755	0.1075	0.0931
electrical engineering	-0.0581	0.0733	-0.2990***	0.1141
precision instruments	0.0617	0.0769	0.0364	0.0875
automobile	-0.1907**	0.0938	-0.1014	0.0972
wholesale trade	1.2506***	0.1332	1.1792***	0.1273
retail trade	0.7946***	0.0981	0.8815***	0.1141
transportation and postal services	0.2274**	0.0953	0.1702*	0.0915
banks and insurances	0.6955***	0.1795	0.3651*	0.2169
technical services	-0.0387	0.1057	-0.4325***	0.1237
other business-related services	0.1774	0.1193	0.2434*	0.1275
σ_{nOUT}	0.8322***	0.0440	–	–
σ_{OUT}	–	–	0.7485***	0.0270
$\rho_{u_{nOUT}\epsilon}$	-0.7120***	0.0853	–	–
$\rho_{u_{OUT}\epsilon}$	–	–	-0.1859*	0.1053
	Wald Tests for Joint Significance			
	χ^2	***p*-value**	χ^2	***p*-value**
factor inputs	63.38	0.0000	86.74	0.0000
sector dummies	198.33	0.0000	214.31	0.0000
entire specification	12,500.44	0.0000	4,756.06	0.0000

Note: *,** and *** indicate significance at the 10%, 5% and 1% level respectively. Robust standard errors are shown in the S.E. columns.

Table 6: Switching Regression – Estimation Results of the Selection Equation

	Coefficient	**Standard Error**
constant	0.3938**	0.1674
capital	-0.0168	0.0229
ICT-capital	-0.0035***	0.0012
labor	-0.1294***	0.0362
East Germany	0.1242*	0.0664
university degree	-0.0048**	0.0020
vocational education	0.0007	0.0014
works council	-0.0827	0.0742
export share	-0.0011	0.0013
foreign subsidiary	-0.1843*	0.0953
firm age ($\leq$ 3 years)	-0.0408	0.1183
firm age (4 $\leq$ years $\leq$ 7)	0.0823	0.0733
IT-application (4–7 appl.)	-0.0275	0.0721
IT-application (8–10 appl.)	-0.0547	0.0920
Y2K-consulting	0.4921***	0.0676
consumer goods	0.0438	0.1127
chemical industry	-0.0229	0.1328
other raw materials	-0.0367	0.1149
electrical engineering	-0.3346***	0.1267
precision instruments	-0.0825	0.1131
automobile	-0.0326	0.1278
wholesale trade	-0.0682	0.1461
retail trade	-0.2424*	0.1276
transportation and postal services	-0.1213	0.1249
banks and insurances	-0.1248	0.1828
technical services	-0.1514	0.1392
other business-related services	0.1074	0.1364

Note: *,** and *** indicate significance at the 10%, 5% and 1% level respectively. Robust standard errors are shown in the Standard Error column.

Table 7: Switching Regression – Wald Test for Identity of the Coefficients

	χ^2	***p*-value**
constant	11.3291	0.0008
capital	0.0959	0.7567
ICT-capital	6.9225	0.0085
labor	1.5060	0.2197
East Germany	3.1137	0.0776
university degree	0.0438	0.8342
vocational education	0.5454	0.4602
works council	0.0595	0.8072
export share	1.8704	0.1714
foreign subsidiary	0.2146	0.6432
firm age ($\leq$ 3 years)	0.1621	0.6872
firm age (4 $\leq$ years $\leq$ 7)	0.0031	0.9553
IT-application (4–7 appl.)	0.5975	0.4395
IT-application (8–10 appl.)	0.4801	0.4884
consumer goods	1.3948	0.2376
chemical industry	0.3635	0.5466
other raw materials	1.0602	0.3032
electrical engineering	3.3398	0.0676
precision instruments	0.0502	0.8228
automobile	0.4714	0.4923
wholesale trade	0.1548	0.6940
retail trade	0.3446	0.5572
transportation and postal services	0.1999	0.6548
banks and insurances	1.4059	0.2357
technical services	6.0777	0.0137
other business-related services	0.1492	0.6993
returns to scale	2.8718	0.0901
set of input factors	11.1995	0.0107
sector dummies	16.4509	0.1714
entire specification	86.0863	0.0000

Table 8: Switching Regression – Wald Test for Joint Significance of the Selection Equation Coefficients

	χ^2	***p*-value**
entire set of explanatory variables	59.4594	0.0000
factor inputs	35.7940	0.0000
sector dummies	16.6933	0.1615
entire productivity equation	174.6572	0.0000
entire selection equation	263.9536	0.0000
Wald Test for Joint Significance of the Entire Switching Regression Estimation		
entire switching regression	28,823.55	0.0000
correlation coefficients	31.0229	0.0000

Table 9: Labor Force Growth – IT-Outsourcing in 2000 by Industry

	IT-Outsourcing			
	Spec. I		**Spec. II**	
Industry	**# of firms**	**%**	**# of firms**	**%**
consumer goods	36	52.17	27	57.45
chemical industry	34	53.13	25	53.19
other raw materials	47	55.29	35	52.24
metal and machine construction	44	47.83	33	55.00
electrical engineering	21	28.00	13	25.49
precision instruments	36	45.57	27	50.00
automobile	30	39.47	22	43.14
wholesale trade	29	63.04	21	65.63
retail trade	30	58.82	16	50.00
transportation and postal services	31	46.27	24	47.06
banks and insurances	31	50.82	24	50.00
technical services	21	29.17	11	22.92
other business-related services	27	38.57	21	36.21
Total	417	45.98	299	46.28

Note: Number and share of firms involved in basic IT-outsourcing in 2000.

Table 10: Labor Force Growth – Summary Statistics

	Spec. I		Spec. II		Dummy
Variable	**Mean**	**S.D.**	**Mean**	**S.D.**	**Variable**
labor force growth (99–03)	-0.0863	0.7978	-0.0580	0.7736	
IT-outsourcing	0.4598	0.4987	0.4628	0.4990	yes
Y2K-consulting	0.5369	0.4989	0.5464	0.4982	yes
employees	337.8258	1,686.22	329.5588	1,890.28	
university degree	20.5882	22.7284	20.6133	23.1428	
vocational education	58.4846	25.1827	58.6749	25.6653	
East Germany	0.2249	0.4178	0.2415	0.4283	yes
firm age ($\leq$ 3 years)	0.1047	0.3064	0.1053	0.3071	yes
firm age (4 $\leq$ years $\leq$ 7)	0.1731	0.3785	0.1625	0.3692	yes
firm age ($\geq$ 8 years)	0.7222	0.4482	0.7322	0.4432	yes
IT-application (0–8 appl.)			0.2926	0.4553	yes
IT-application (9–12 appl.)			0.4659	0.4992	yes
IT-application (13–15 appl.)			0.2415	0.4283	yes
$\triangle$ standard wages (99–03)	10.2693	0.8488	10.2330	0.9000	
computer workers			46.3715	31.1368	
$\triangle$ computer workers (00–04)			53.8846	260.7942	
$\triangle$ IT–specialists (00–04)			-0.0515	0.8828	
foreign subsidiary			0.1889	0.3917	yes
export share			15.2647	22.8790	
consumer goods	0.0761	0.2653	0.0728	0.2599	yes
chemical industry	0.0706	0.2562	0.0728	0.2599	yes
other raw materials	0.0937	0.2916	0.1037	0.3051	yes
metal and machine construction	0.1014	0.3021	0.0929	0.2905	yes
electrical engineering	0.0827	0.2756	0.0789	0.2699	yes
precision instruments	0.0871	0.2821	0.0836	0.2770	yes
automobile	0.0838	0.2772	0.0789	0.2699	yes
wholesale trade	0.0507	0.2195	0.0495	0.2172	yes
retail trade	0.0562	0.2305	0.0495	0.2172	yes
transportation and postal services	0.0739	0.2617	0.0789	0.2699	yes
banks and insurances	0.0673	0.2506	0.0743	0.2625	yes
technical services	0.0794	0.2705	0.0743	0.2625	yes
other business-related services	0.0772	0.2670	0.0898	0.2861	yes
Number of observations	907		646		

Source: ZEW ICT-survey 2000, 2004 and own calculations.

Table 11: Labor Force Growth – IV-Estimation - First Stage Results

Dependent Variable:	**Spec. I**		**Spec. II**	
IT-Outsourcing (in 2000)	**Coef.**	**S.E.**	**Coef.**	**S.E.**
ln(employees)	0.4532***	0.1304	0.4336***	0.1523
ln(employees)^2	-0.1036***	0.0261	-0.1019***	0.0304
ln(employees)^3	0.0064***	0.0016	0.0065***	0.0019
university degree	-0.0008	0.0010	-0.0007	0.0013
vocational education	0.0008	0.0008	0.0000	0.0010
East Germany	0.0140	0.0410	0.0066	0.0479
firm age ($\leq$ 3 years)	0.0114	0.0521	0.0486	0.0606
firm age (4 $\leq$ years $\leq$ 7)	0.0104	0.0434	0.0459	0.0519
IT-application (0–8 appl.)			0.0864*	0.0488
IT-application (13–15 appl.)			-0.1064**	0.0478
$\triangle$ standard wages (99–03)	0.0230	0.0299	0.0253	0.0326
computer workers			-0.0006	0.0008
$\triangle$ computer workers (00–04)			0.0000	0.0001
$\triangle$ IT–specialists (00–04)			0.0595***	0.0210
foreign subsidiary			0.1076*	0.0551
export share			-0.0002	0.0009
consumer goods	0.0649	0.0763	0.0269	0.0915
chemical industry	0.0168	0.0816	-0.0554	0.0985
other raw materials	0.0885	0.0763	-0.0204	0.0908
electrical engineering	-0.1641**	0.0729	-0.2295**	0.0904
precision instruments	-0.0463	0.0756	-0.1017	0.0927
automobile	-0.0640	0.0762	-0.1096	0.0974
wholesale trade	0.0903	0.0846	0.0139	0.1033
retail trade	0.0360	0.0883	-0.1339	0.1122
transportation and postal services	0.0143	0.1045	-0.0730	0.1196
banks and insurances	0.0212	0.0844	-0.0326	0.1059
technical services	-0.1341*	0.0809	-0.2541**	0.1023
other business-related services	-0.0797	0.0789	-0.1734*	0.0965
Y2K-consulting	0.1400***	0.0331	0.1460***	0.0394
Constant	-0.3752	0.3799	-0.2431	0.4328
Centered R^2	0.1051		0.1538	
Number of observations	907		646	

Note: *,** and *** indicate significance at the 10%, 5% and 1% level respectively. Robust standard errors are shown in the S.E. columns.

Table 12: Labor Force Growth – IV-Estimation - Second Stage Results

Dependent Variable:	**Spec. I**		**Spec. II**	
labor force growth (99–03)	**Coef.**	**S.E.**	**Coef.**	**S.E.**
IT-outsourcing	0.4825	0.3753	0.7743*	0.4168
ln(employees)	-0.6818**	0.3363	-0.4887	0.3357
ln(employees)^2	0.1899**	0.0768	0.1299*	0.0744
ln(employees)^3	-0.0162***	0.0053	-0.0109**	0.0050
university degree	0.0057***	0.0021	0.0047**	0.0026
vocational education	0.0009	0.0014	0.0023	0.0017
East Germany	-0.0699	0.0555	-0.0293	0.0624
firm age ($\leq$ 3 years)	-0.0952	0.0888	-0.0747	0.1015
firm age (4 $\leq$ years $\leq$ 7)	0.1244*	0.0741	0.0694	0.0821
IT-application (0–8 appl.)			-0.1849**	0.0755
IT-application (13–15 appl.)			0.0684	0.0896
$\triangle$ standard wages (99–03)	-0.0674	0.0412	-0.0730*	0.0409
computer workers			0.0017	0.0013
$\triangle$ computer workers (00–04)			-0.0000	0.0002
$\triangle$ IT–specialists (00–04)			0.2991***	0.0644
foreign subsidiary			0.0237	0.1088
export share			-0.0020	0.0014
consumer goods	0.0523	0.1221	-0.0711	0.1375
chemical industry	0.1421	0.1091	0.1480	0.1227
other raw materials	0.0302	0.1119	0.1148	0.1330
electrical engineering	0.1984	0.1317	0.3105*	0.1700
precision instruments	0.1696	0.1122	0.1557	0.1423
automobile	0.2239*	0.1179	0.1433	0.1500
wholesale trade	0.0772	0.1183	-0.0660	0.1417
retail trade	-0.0931	0.1129	-0.0673	0.1772
transportation and postal services	0.0642	0.1750	-0.0098	0.1796
banks and insurances	0.0414	0.1471	-0.0338	0.1881
technical services	-0.1840	0.1592	-0.0788	0.1946
other business-related services	-0.0741	0.1615	-0.0490	0.1970
Constant	0.9497	0.5914	0.6639	0.5944
Number of observations	907		646	

Note: All variables pertain to the year 1999 or 2000, except those variables inducing a rate of change (i.e. $\triangle$ standard wages, $\triangle$ computer workers, $\triangle$ IT-specialists). *,** and *** indicate significance at the 10%, 5% and 1% level respectively. Robust standard errors are shown in the S.E. columns.

Table 13: Industry Classification

Industry	Explanation	NACE
consumer goods		
	manufacture of food products, beverages and tobacco	15-16
	manufacture of textiles and textile products	17-18
	manufacturing of leather and leather products	19
	manufacture of wood and wood products	20
	manufacturing of pulp, paper and paper products; publishing and printing	21-22
	manufacturing n.e.c.	36-37
chemical industry		
	manufacture of coke, refined petroleum products and nuclear fuel	23
	manufacture of chemicals, chemical products and man-made fibres	24
other raw materials		
	manufacture of rubber and plastic products	25
	manufacture of non−metallic mineral products	26
	manufacture of basic metal	27
metal and machine construction		
	manufacture of fabricated metal products (except machinery and equipment)	28
	manufacture of machinery and equipment n.e.c.	29
electrical engineering		
	manufacture of office machinery and computers	30
	manufacture of electrical machinery and apparatus n.e.c.	31
	manufacture of radio, television and communication equipment and apparatus	32
precision instruments		
	manufacture of medical, precision and optical instruments, watches and clocks	33
automobile		
	manufacturing of transport equipment	34-35
wholesale trade		
	wholesale trade and commission trade (except of motor vehicles and motorcycles)	51
retail trade		
	sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	50
	retail trade (except of motor vehicles and motorcycles), repair of personal and household goods	52
transportation and postal services		
	land transport, transport via pipeline	60
	water transport	61
	air transport	62
	supporting and auxiliary transport activities; activities of travel agencies	63
	post and courier activities	64.1
banks and insurances		
	financial intermediation	65-67
electronic processing and telecommunication		
	computer and related activities	72
	telecommunications	64.2
technical services		
	research and development	73
	architectural and engineering activities and related technical consultancy	74.2
	technical testing and analysis	74.3
other business-related services		
	real estate activities	70
	renting of machinery without operator and of personal and household goods	71
	legal, accounting, book keeping and auditing activities; tax consultancy; market research and public opinion pools; business and management consultancy; holdings	74.1
	advertising	74.4
	labor recruitment and provision of personnel	74.5
	investigation and security services	74.6
	industrial cleaning	74.7
	miscellaneous business activities n.e.c.	74.8
	sewage and refuse disposal, sanitation and similar activities	90

References

Amiti, M. and Wei, S. (2006). Service Offshoring and Productivity: Evidence from the United States, *Working Paper 11926*, NBER, Cambridge.

Bertschek, I., Fryges, H. and Kaiser, U. (2006). B2B or Not to Be: Does B2B E-Commerce Increase Labour Productivity?, *International Journal of the Economics of Business* **13**(3): 387–405.

Bertschek, I. and Kaiser, U. (2004). Productivity Effects of Organizational Change: Microeconometric Evidence, *Management Science* **50**(3): 394–404.

Bertschek, I. and Müller, M. (2006). Productivity Effects of IT-Outsourcing: Semiparametric Evidence for German Companies, *in* S. Sperrlich, W. Härdle and G. Adinli (eds), *The Art of Semiparametrics*, Physica-Verlag, Heidelberg, pp. 130–154.

Brynjolfsson, E. and Hitt, L. (1996). Paradox Lost? Firm-level Evidence on the Returns to Information Systems Spending, *Management Science* **42**(4): 541–558.

Brynjolfsson, E. and Hitt, L. (2000). Beyond Computation: Information Technology, Organizational Transformation and Business Performance, *Journal of Economic Perspectives* **14**(4): 23–48.

Brynjolfsson, E. and Hitt, L. (2003). Computing Productivity: Firm-Level Evidence, *The Review of Economics and Statistics* **85**(4): 793–808.

Christensen, C. and Rosenbloom, R. (1995). Explaining the Attacker's Advantage: Technological Paradigms, Organizational Dynamics, and the Value Network, *Research Policy* **24**(2): 233–257.

Fixler, D. and Siegel, D. (1999). Outsourcing and Productivity Growth in Services, *Structural Change and Economic Dynamics* **10**: 177–194.

Girma, S. and Görg, H. (2004). Outsourcing, Foreign Ownership, and Productivity: Evidence from UK Establishment-level Data, *Review of International Economics* **12**(5): 817–832.

Greenan, N. and Mairesse, J. (2000). Computers and Productivity in France: Some Evidence, *Economics of Innovation & New Technology* **9**(3): 275–315.

Görg, H. and Hanley, A. (2003). International Outsourcing and Productivity: Evidence from Plant Level Data, *Research Paper 2003–20*, The University of Nottingham.

Görg, H. and Hanley, A. (2004). Does Outsourcing Increase Profitability?, *Discussion Paper 1372*, IZA, Bonn.

Görg, H. and Hanley, A. (2005). International Outsourcing and Productivity: Evidence from the Irish Electronics Industry, *North American Journal of Economics and Finance* **16**(2): 255–269.

Görg, H., Hanley, A. and Strobl, E. (2005). Productivity Effects of International Outsourcing: Evidence from Plant Level Data, *Working paper*, The University of Nottingham.

Görzig, B. and Stephan, A. (2002). Outsourcing and Firm-level Performance, *Discussion Paper 309*, DIW, Berlin.

Hempell, T. (2005). What's Spurious, What's Real? Measuring the Productivity Impacts of ICT at the Firm-Level, *Empirical Economics* **30**: 427–464.

Heshmati, A. (2003). Productivity Growth, Efficiency and Outsourcing in Manufacturing and Service Industries, *Journal of Economic Surveys* **17**(1).

Loh, L. and Venkatraman, N. (1992). Determinants of Information Technology Outsourcing: A Cross-Sectional Analysis, *Journal of Management Information Systems* **9**(1): 7–24.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs, Cambridge University Press, Cambridge.

Markusen, J. (1995). The Boundaries of Multinational Enterprises and the Theory of International Trade, *Journal of Economic Perspective* **9**(2): 169–189.

Olsen, K. (2006). Productivity Impacts of Offshoring and Outsourcing: A Review, *OECD Science, Technology and Industry Working Paper 2006/1*, OECD.

Schreyer, P. and Pilat, D. (2001). Measuring Productivity, *OECD Economic Studies* **33**(2001/2): 127–170.

Siegel, D. and Griliches, Z. (1992). Purchased Services, Outsourcing, Computers, and Productivity in Manufacturing, *in* Z. Griliches (ed.), *Output Measurement in the Service Sectors*, Univiversity of Chicago Press, Chicago, IL, pp. 429–460.

ten Raa, T. and Wolff, E. (2001). Outsourcing of Services and the Productivity Recovery in U.S. Manufacturing in the 1980s and 1990s, *Journal of Productivity Analysis* **16**: 149–165.

Zentrum für Europäische Wirtschaftsforschung (2005). Internet bestimmt den Geschäftsalltag, *ICT–Report*, ZEW, Mannheim.

Conference on The Economics of Information and Communication Technologies
ENST Paris, June 21-22, 2007
Call for papers

The willingness to pay for online music in the presence of copying: An empirical investigation

First draft (December 2006)

Fabrice Le Guel* & Fabrice Rochelandet**

* Maître de conférences/Assistant professor
University of Nice Sophia Antipolis
GREDEG (www.gredeg.cnrs.fr) UMR CNRS 6227
250 rue Albert Einstein – Batiment 2 – 06560 Valbonne
& ADIS (www.adislab.net), University of Paris 11
fabrice.le-guel@gredeg.cnrs.fr

** Maître de conférences/Assistant professor
ADIS (www.adislab.net), University of Paris 11
Bureau D103, 54, bd Desgranges, F-92331 Sceaux cedex, France
fabrice.rochelandet@u-psud.fr

Keywords

Digital goods, music business models, willingness to pay, ordered Logit model

Abstract

This paper examines the willingness to pay (WTP) for an original –a legitimate copy of a work– when perfect digital copies are freely available. In so doing, we highlight some conditions for the viability of new business models facing the fast development of "everything is for free" environments. We use micro-data about WTP obtained from contingent valuation technique. We test them using an ordered Logit model. We suggest three results: (1) There exists a positive WTP for online music in the presence of copying and it is lesser than prevailing market price; (2) This WTP doesn't depend so much on demographics but rather on ethical concern, usability, the choice of titles, and the structure of the social embeddedness of individuals; (3) The importance of the social embeddedness suggests that alternative models to BtoC paying downloads can be viable.

1. Introduction: "From free to fee"?

Nowadays online music mainly consists in two kinds of access: on the one hand, online sales – with the representative model of iTunes Music Store– and on the other hand, free access through illegal file sharing networks. The latter one prevails as a major way to access music on the Internet and more recently, among mobile phones. This success is not so much explained by the so-called "free mentality" but rather by the willingness of copiers to access a wider diversity of titles and the range of possible uses unimpeded by digital devices (Rochelandet, Le Guel, 2005).

There has been much debate about the impact of P2P file-sharing networks on content markets. P2P file-sharing is suspected to decrease the sales of records while impeding the emergence of new profitable online business. Accordingly, in OECD countries, copyright legislations have been reinforced to fight against P2P sharing. But in fact, such a legal evolution presupposes only one kind of business model according to which individuals have to pay an average price (0.9 euros for each music tracks) for enjoying a use-restricted content. From our point of view, such a 'magic' price is not so the result of some natural matching between demand and supply but much more the result of an inefficient replication of traditional business practices on the Internet. New business models such as those of In2TV, YouTube and Jamendo suggest that alternative arrangements can be implemented face to a purely BtoC commerce. They appear, in fact, to be better fitted to the main feature of the Internet, namely to be a decentralized networks grounded on sharing behavior.

This paper examines the WTP for an original –a legitimate copy of a work– when perfect digital copies are freely available. In particular, the study deals with the willingness to pay (WTP) for legal online music despite the possibility of file-sharing. Among the various factors economic literature has already highlighted to explain the impact of copying on the overall sales of music figure the free nature of digital copies, the ethical concern about the impact of file-sharing and the perception of legal risk. But until now, no study has tested them simultaneously on a large sample. We try to fill this gap by using the contingent valuation technique to obtain micro-data about WTP and by testing them through an ordered Logit model. In so doing we expect to build further some criteria in order to evaluate alternative online services in the field of music and movie business. We suggests three results: (1) There exists a positive WTP for online music in the presence of copying and it is lesser than prevailing market price; (2) This WTP doesn't depend so much on individuals' demographics but rather on ethical concern, usability, the choice of titles, and the structure of their social embeddedness; (3) The importance of social embeddedness suggests that models alternative to BtoC paying downloads can be viable.

2. Literature and Theory

2.1. P2P literature: welfare considerations and alternative models

P2P literature is made of two approaches. The first one envisages the impact of P2P sharing on the market for originals: Does P2P file-sharing represent a social cost or a social benefit? How to design law in order to eliminate or to promote it? The second approach analyses the efficient running of P2P systems as methods of delivering contents: Why do people contribute or not for resources on those networks? How to eliminate free-riding? (*inter alia* Krishnan *et al.*, 2003, Vishnumurthy *et al.*, 2003). Crossing these two approaches consists in evaluating new business models grounded on P2P sharing networks as an alternative to online sales per unit (the "ITMS model").

The first approach is quite well documented by Liebowitz (2006), which aims at demonstrating that file-sharing should have a negative impact on the content industries. We demonstrate elsewhere that his demonstration can be contested by econometric analysis with strong methodological choice and when conducted on a large heterogeneous sample (Rochelandet, Le Guel, 2005). Nevertheless, according to Liebowitz, substitution effect, sampling, (local and global) network effects and indirect appropriability constitutes the main factors used to examine whether or not file-sharing can actually cause damages to the industry. By questioning those arguments generally used to suggest the positive impact of file-sharing on record sales, he suggests that "the substitution effect is quite simple to analyze. The copy is treated as a substitute for the original. If the copy is identical or close in quality to the original, and if the cost of making the copy is low, the copy for a price of zero dominates the original at its positive price." More generally, his analysis challenges the idea according to which file-sharing can represent an innovation by generating viable and more profitable business models than those prevailing nowadays.

Conversely, some studies put forward the new opportunities offered by P2P sharing. Therefore, it appears more relevant to consider file-sharing and new behavior of consumers as the conditions for the emergence of new models, which can mix paying and free access rather than one unique model relying on the payment for use-restricted contents. Nevertheless, studies on file-sharing have not yet addressed the question how much money people are willing to pay for accessing legally these resources and for getting legitimate copies when free copies are available. Moreover, none considers the very mechanisms through which such a WTP can be increased. Our purpose is to fill this gap by envisaging new business models in online music. In particular, we suppose that the exploitation of the very features of sharing networks (network effects and direct social interactions) can increase the WTP for both digital contents and new online services enabling viable business models based on direct exchanges among individuals.

2.2. Contingent valuation and cultural goods

Contingent valuation refers to the various survey-based technique to valuate the non-market goods and services by asking a sample of individuals how much they would be prepared to pay for a specified change in the supply of a given public good. Economic issue arises from the fact that individuals or organizations derive utility from those resources, while it proves difficult to evaluate such benefits. Contingent valuation constitutes one of the major techniques to measure them, widely used in environmental studies. In the field of cultural economics, it mainly concerns heritage and historic sites (Schuster, 2003).

Why and how evaluate the WTP of individuals in the field of digitized cultural goods? Music contents considered both as market goods and as non-traded goods when they are illegally shared. It could be interesting here to determine the value individuals put on music titles in order to evaluate the degree of substitution between originals and digital copies.

Our approach consists in evaluating the willingness to pay for an original when a perfect digital copy is freely available. Such an assessment is crucial to study the viability of new business models based on sharing behavior. On the one hand, the WTP could be zero (or most of the individuals are not prepared to pay for) ; in this case, paying models cannot coexist in the presence of free/unimpeded models, because individuals will always preferred the free copy to the original.

On the other hand, if the WTP is positive, then heterogeneous business models can develop all together. In this case, the issue is to evaluate the WTP according to individuals and to understand why it is positive. Some explanations have been suggested such as the income, the taste for cultural goods (music or movies), the consumption habits created by the past purchases of original CDs, the ethical concerns of individuals (who can considered that copying is a bad thing) and the risks associated with copying such as lawsuit, the risk of being attacked by virus or spyware…(*inter alia* Holm, 2003, Buxmann *et al.*, 2004, Liebowitz, 2006). Rational individuals thus are supposed to make a balance between the utility a specified content confers to them and the various costs they incur by getting a copy rather an original exemplar.

In addition, some recent papers in management literature explore the willingness to pay for online services (Ye et al., 2004, Chyi, 2004, Shih, 2003, Gefen, 2003). They suggest that relying exclusively on advertising revenues is not always a sustainable business model for the online delivery of contents. Models based on free access to online resources (online newspapers, databases…) are not considered as viable models because advertising revenues are uncertain over Internet and not enough to finance the production of those information goods. Moreover, free (lawful) access is supposed to challenge and substitute for traditional "offline" goods. Thus, it proves crucial to change or to diversify the sources of revenues. Two classical business methods consist in: (1) online unit sales of digital contents directly to consumers (for example, iTunes) and (2) subscription models (for example, MusicMe in France).

Those empirical studies thus are useful to determine the WTP for online services that can be currently freely accessed. They suggest that the WTP for online services is influenced by the perceived convenience these services provide for individuals and their online experience and habit. They envisage also other determinants such as fairness – one must pay for a service because it is a duty or because it might disappear through lack of revenues –, better quality expectations – one expects that paying services will be of better quality than the same services delivered freely –, "free mentality" – one has been accustomed to getting free services and one find illegitimate to pay for them –, and so on.

Besides, charging fees for services is often supposed to be incompatible with illegal file-sharing success. Many commentators suggest expanding strong enclosure strategies using DRM technologies in order to strengthen the expansion of paying models. Two scenarios then are to be considered according to which illegal files swapping disappear or do not. In the second case, new generations of technology appear and could ground upon direct exchanges amongst individuals (Rochelandet & Le Guel, 2005). So the issue is not so much to fight illegal sharing behavior but rather to understand precisely how to extract some value from them.

3. Willingness to pay and its determinants

3.1. Variables and Hypotheses

The explained variable is denoted by *WTP*. It represents the sum that an individual would accept to pay for an original exemplar of a musical track when perfect digital copies are freely available from her or his neighbors or through a P2P network[i].

We can envisage two cases according to the level of *WTP*.

First case: The WTP is equal to zero

At least a high significant proportion of individuals (more than 60% for instance) are not prepared to pay for originals when they can get a copy for free. In this case, we can infer that paying and "free" models could not coexist, because copiers do never accept to buy original content, even at a more personalized price. So there are two incompatible solutions: either producers and retailers have to design models either entirely grounded on digital enclosure (DRM), or they produce some models of distribution totally grounded on the free availability of contents, for example on sales of complementary goods, advertising or exploitation of private data.

Second case: The WTP is positive

At least a high proportion of people are prepared to pay a certain sum for original whereas they can get copies for free. In this case, there can be a coexistence of paying and free models. The question then is: How to articulate the two kinds of models in order to extract enough value from consumer to recoup fixed costs and to make profit? According to us, a first step to tackle this difficult question is to understand why this value is positive. In other words, what are the different determinants to be tested?

In addition to demographics, different variables can be suggested:

1. The **purchase of CDs** (denoted by *CULTSPEND*)**:** The more an individual buys original CDs, the more she/he is willing to pay for online music. This idea is grounded on the habits individuals take when they buy CDs. However, such a proposition could be questioned, because the utility derived from the use of CDs is not the same as in the case of online music. Dematerialized music is not the same product of CDs. For instance, it is not possible to lend or resale legal online music. We suppose nevertheless that this variable may be significant.

Proposition 1: *The level of CD purchase impacts significantly and positively WTP.*

2. The **ethical concerns** (denoted by *ETHIC*) of the individual regarding the copying of copyrighted works: Individuals can consider copying as unfair regarding the intellectual effort of artists or to endanger the existence of the record industry. *ETHIC* indicates psychological costs the individuals

bear when they feel ethically wrong the fact of copying contents. Here, we suppose that individuals make their calculus integrating other variables than the sole preference for music and relative prices of originals and copies. In a previous study, we show that this variable reduces the intensity of copying over P2P networks. Similarly, *ETHIC* can increase the *WTP* for originals of individuals, which are more ethically concerned.

The index *ETHIC* was built by requesting respondents to scale –between "do not agree", "quite not agree", "agree" and "fully agree"– their ethical concerns about copying behavior through four questions: "According to you, copying (1) endangers the movie and record markets; (2) affects the income of authors and artists; (3) does not respect the work of authors and artists; (4) is blamable in itself." We confer the value 1, 2, 3, 4 for each scaled variable and then add up them.

Proposition 2: *The ethically awareness is strongly and positively related to WTP.*

3. The **legal risk** (denoted by *LEGRISK*), namely the perceived likelihood to be caught and sanctioned for illegal activities. This variable represents another cost that individuals integrate in their rational calculus and perception. The impact of those heterogeneous perceptions of risks among individuals regarding their WTP for originals is supposed to be positive. It can be assimilated as an 'avoidance' cost: The more risk-adverse an individual is, the more she/he will be prepared to pay in order to save on the perceived cost to be caught and sanctioned for illegal file-sharing.

The variable *LEGRISK* was build by asking respondents to choose between four perceived levels of risk: no risk, low risk, medium risk and high risk. One key fact to be noted is that a wide campaign against copying was led shortly before we began our survey. So we suppose that respondents are quite aware of the risks associated with such practice. Thus, we consider the WTP for originals by ranging well-informed copiers from risk-adverse ones to risk-lovers.

Proposition 3: *The more risk-adverse is an individual facing the risk of being sanctioned by law, the greater she/he will be willing to pay for legitimate copies.*

d. The **social embeddedness** (assessing by *Herding* and *Herdingbis*): Individuals are supposed to be lesser willing to pay for originals the fewer is the proportion of copiers in their social neighborhood.

Individuals are supposed to be influenced both in their preferences and choices by the preferences and choices of their social neighboring (friends, family, colleagues). The more there are copiers in their relatives and friends, the more individuals tend to be copiers (local interaction effect, Rochelandet, Le Guel, 2005), and therefore the more they preferred copies rather than originals.

So we make the following proposition:

Proposition 4: *The proportion of copiers in the social neighboring of individuals impacts negatively* WTP.

e. The richness of the supply: the **diversity of music titles** (denoted by *Music diversity*) and the **set of possible uses** (denoted by *Usability*)

Online music features are strongly linked to the way it is delivered to individuals. Concerning illegal file-sharing, a greater diversity of titles can be found over P2P networks than over commercial platforms. Similarly, shared contents through P2P networks are not protected by DRM devices and are available at MP3 format that allows to play music tracks whatever the portable players and to share them among friends. By contrast, paying online music suffers from these features by delivering a limited range of use-restricted contents. In other words, lack of diversity and usability impacts negatively the WTP for originals. According to Sundararajan (2004), 'restricting the rights of usage that contribute to customer value [reduces] this value'. We suppose therefore that the greater the set of uses is, the more the individuals derive utility and are willing to pay for the content.

Therefore, we make the two following propositions:

Proposition 5: *The more the diversity of contents associated with paying online music is judged as insufficient, the lower is WTP.*

Proposition 6: *Individuals' desire for unrestricted usability impacts negatively WTP.*

The other independent variables that may explain WTP may be grouped as **demographics** (education, socio-professional group/occupation, household structure and income). The increase in age of the respondent is expected to reduce the intensity of copying. Younger people will be more open in their use of newly introduced ICTs. In fact, age usually reflects many other variables favorable to the intensity of a copying activity such as technical skills and income. Furthermore we hypothesize a positive impact of income level on the intensity of copying. The influence of the other demographics can be positive, negative or neutral.

The table 1 summaries the independent variables used in our econometric test.

Table 1: Variables

VARIABLE	DESCRIPTION	EXPECTED SIGN
Demographics - Gender - Age - Income	 - - -	 (indeterminate) (+) (+)
Cultural spending	CD & DVD purchases	(+)
Herding	The proportion of copiers in the social neighboring (relationship density)	(-)
Ethic	Ethical concern towards copying	(+)
Legal risk	Perception of legal risks (to be caught)	(+)
Usability	The set of uses possible from originals in comparison to copies (online music)	(-)
Music Diversity	Perception of cultural diversity associated with originals (compared to file-sharing)	(-)
Herdingbis (friends)	The proportion of copiers in friendship	(-)

3.2. Data and methodology

This article measures and explains the willingness to pay for originals when digital copies are freely available and whatever the method or technology used to get them. We apply survey methodology to measure the value individuals are prepared to pay for an online music track.

We base our analysis on primary data gathered in January and February 2005: 2,828 individuals were surveyed using a paper survey directly addressed to individuals and a Web-based survey. Note that non-response bias leads to a reduction in the size of the sample depending on the considered variables in the model. Sample reduction represents no more than 26% (maximum, i.e. for the general model). To simplify missing data correction, we choose listwise deletion approach (Allison, 2001). Further paper will use multiple imputation procedure for incomplete mixed data (Schafer, 1997). The sample bias due to Web-based survey had been corrected using a post-stratification method implemented with SAS macro (CALMAR method, INSEE[ii]). However, there is no significant bias and therefore we choose to present our model without post-stratification.

The description of the sample figures in the table 2.

Table 2: Description of the sample

Variables	Mean	Min	Max
Dependent variable (WTP) **0 euros**	26,04 %	0	1
0.10 to 0.30 euros	31,58 %	0	1
0.40 to 0.50	28,38 %	0	1
0.60 to 0.90	14 %	0	1
Gender (Ref : man)	78 %	0	1
Age	22,58 %	0	1

<25			
25-30	19,05 %	0	1
31-40	23,16 %	0	1
41-50	15,45 %	0	1
>50	19,74 %	0	1
Income (index)	4.23 (sd=2,29)	1	9
Cultural spending (index)	2.09 (sd=1,27)	1	7
Ethic (index)	6.39 (sd=2,60)	4	16
Legal risk (index)	2.64 (sd=0,91)	1	4
Herding (index)	2.81 (sd=1,11)	1	4
Usability (index)	2.81 (sd=1,14)	1	4
Music diversity (index)	3.33 (sd=0 ,94)	1	4
Herdingbis (friends)	53 %	0	1

Note that about 74% of respondents state that they are willing to pay for original music track. Their WTP ranges from 0.1 to 0.9 euros and the mean is far from being null (0.30 €). However, the standard deviation of 0.27 € suggests that the WTP is not superior to 0,6 €: there is a significant request for lower prices of originals.

Because the WTP for originals is an ordered qualitative variable, we cannot use Tobit model. Instead, we use an ordinal regression model (Train, 2003) to forecast the determinants of the willingness to pay for an original when copies are freely available. y^* is a latent variable divided into J ordinal categories:

$$y_i = m \quad if \quad \tau_{m-1} \le y_i^* \le \tau_m \quad for\ m = 1\ to\ J$$

Specifically:

$$y_i = \begin{cases} 1 \Rightarrow 0\ Euros\ if\ \ \tau_0 = -\infty \le y_i^* < \tau_1 \\ 2 \Rightarrow 0.10\ to\ 0.30\ Euros\ if\ \ \tau_1 \le y_i^* < \tau_2 \\ 3 \Rightarrow 0.40\ to\ 0.50\ Euros\ if\ \ \tau_2 \le y_i^* < \tau_3 \\ 4 \Rightarrow 0.60\ to\ 0.90\ Euros\ if\ \ \tau_3 \le y_i^* < \tau_4 = +\infty \end{cases}$$

We define the structural model for latent variable as $y_i^* = x_i\beta + \varepsilon_i$ where i is the observation and ε a random error vector.
Cutpoints τ_1 through τ_{J-1} are estimated and we suppose that $\tau_0 = -\infty$ and $\tau_J = +\infty$.
We can define the maximum likelihood as the product of four components:

$$\begin{cases} \Pr(y = 1|x) = F(\tau_1 - x\beta) \\ \Pr(y = 2|x) = F(\tau_2 - x\beta) - F(\tau_1 - x\beta) \\ \Pr(y = 3|x) = F(\tau_3 - x\beta) - F(\tau_2 - x\beta) \\ \Pr(y = 4|x) = 1 - F(\tau_3 - x\beta) \end{cases}$$

$$L = \prod_{i=1}^{N} \Pr(y = 1|x)^{y_{i1}} . \Pr(y = 2|x)^{y_{i2}} . \Pr(y = 3|x)^{y_{i3}} . \Pr(y = 4|x)^{y_{i4}}$$

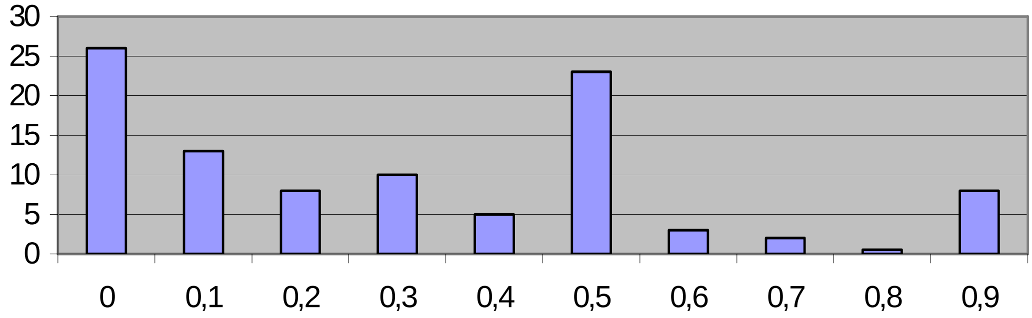
with N=2828 respondents and $y_{im} = 1\ if\ i\ chooses\ m,\ 0\ otherwise$.
We use maximum likelihood estimation to determine β vector of parameters.

4. Empirical results

4.1 Willingness to pay estimates

Figure 1 presents a histogram of respondents' stated willingness to pay.

Figure 1: Respondents' stated willingness to pay (€ and %)



Only 26% of respondents exhibited a zero WTP for originals, whereas 8% were prepared to pay the full price of an original. 66% are not willing to pay the full (charged) price of an original music track but they were willing to pay a positive sum ranged from 0.1 euro to 0.8 euro with a majority between 0.1 and 0.5.

4.2 Ordered Logit model results

This section discusses the results that were derived from estimating the ordered Logit model of the willingness to pay for an original sound track when a perfect digital copy is freely available. These results are displayed in the table 3 (see appendix 1).

The first column displays the variables used to test the *WTP* variable. The results of the general model figure on the second column. Subsequent columns partition the general model into eight specific ones to test the propositions suggested in the previous section. Note that some variables incorporated in the general model are not significant due to the correlations among some of them. When we estimate independently those determinants, they become significant at a level less than 5% or 1%.

As expected, we validate our various propositions. Main variables are generally significant and of the predicted sign. Note that some variables are quite robust (*Age, Cultural spending, Ethic, Herding, Herdingbis, Music diversity*). Other variables prove less robust (*Legal risk, Income, Usability*). As usual in economic studies on information technology and usage, gender is not significant.

All the results (significance and sign) are summarized in the table 4.

Table 4: Summary of the results

VARIABLE	EFFECTS ON WTP
Demographics - Gender - Age - Income	 (indeterminate) (+)*** (+) (its significance depends on the co-tested variables)
Cultural spending	(+)**/***
Herding	(-)**/***
Ethic	(+)***
Legal risk	(+)**/not significant in general model
Usability	(-)***/not significant in general model
Music Diversity	(-)***
Herdingbis (friends)	(-)***

5. Implications and conclusion

A first result is that individuals do not take only the free feature of digital copies into consideration when they value originals. So the "free mentality" hypothesis is not fully tenable to justify and foster the legal and technical enclosure against illegal file-sharing.

In fact, our study suggests that there is not a simple opposition between those who are willing to pay for legal online music and those who are not. Individuals are heterogeneous in their valuation of originals so that it should be profitable to exploit these differences in WTP. Perhaps there is a potential market to exploit and likely to generate higher levels of profit than traditional business models. In fact, individuals integrate other variables into their calculus such as ethical concerns, cultural diversity and usability of contents. Note that cultural behavior and valuation are strongly embedded into the social neighboring of individuals. Innovative distribution models should fuel and exploit social interactions in order to be viable and source of higher remuneration for artists and producers. Online music retailers have to modify their business models if they want to impose them over the Internet.

5.1. Old and new pricing models

Because the WTP is positive, pricing strategies could be a solution to allow content producers and retailers to appropriate sufficient remuneration and therefore to permit a mutually beneficial coexistence of file-sharing and paying models. According to Shapiro & Varian (1998), due to the zero cost of additional copy, a low price can be sufficient to recoup high fixed costs of production of information goods. So pricing models different from the one prevailing currently (namely ITMS model) can be suggested.

(1) A price reduction could be sufficient to lead low-value consumers to buy online music and so to generate sales enough to compensate the fixed costs. This is the idea suggested by Chen & Png (2003) according to which price reduction is more socially desirable than legal fighting against illegal copying. However, this low price strategy has two drawbacks: The first one is that individuals tend to assess and declare a WTP higher than it would be if they actually have to pay. The second one is that both retailers and copyright holders are not prepared to accept a reduction in their margins and royalties. Thus, such a cut in prices may encounter resistance from copyright holders. All in all, our study shows that this price reduction has to be significant because the majority of positive valuations are ranged below half of the full price of an original music track.

(2) Price discrimination constitutes a pricing model that could enhance both consumers' surplus and producers' benefits. It could be implemented according to the individual preferences. Although it would be unrealistic because of the inherent costs of detection and revelation of preferences, this strategy is conceivable through the exploitation of personal data over P2P networks (in this way, file-sharing could serve to facilitate the implementation of price discrimination over paying models by revealing preferences of copiers!). Price discrimination could also be made according to the product features. Buxmann et al. (2004) suggest that music industry should price-discriminate contents according their nature: current hits, older titles, rarities and works of new artists. But once again, individuals could overstate their WTP and revenues might be not enough to cover fixed costs.

5.2. Innovation

By contrast to those classical pricing models, our study suggests that music industry should not try to replicate traditional models but rather innovate in order to increase the WTP of individuals (see also Rochelandet, 2005). In so doing, WTP is not considered a fixed variable. According to our econometric results, two variables allow to increase WTP and support innovation in the field of online music: Allowing more uses by implementing less DRM protection –at least by ensuring interoperability between standards– and offering a much wider set of choices in terms of music titles may enhance the WTP for originals. Of course, there could be a substitution effect (Liebowitz, 2006) associated with a weaker technical enclosure. But two facts counterbalance such an opinion. First, WTP increase can be enough to generate enough revenue. Secondly, our analysis has also suggested that most of copiers are ethically concerned people, which could be sensitive to significant innovation from record industry and new comers.

So we predict that a radical innovation in business model could enhance social welfare by increasing authors' remuneration, consumers' surplus and middlemen's margins. In particular, alternative models might be more efficient than the iTunes model if they respect two conditions:
(1) A stronger matching with actual cultural behavior (need for cultural diversity, wide usability, recommendations among friends, and so on);

(2) A greater integration of the new conditions of demand formation on the Internet (new forms of social interactions and information dissemination, new possibilities of use…).

Further investigation will consist in examining the WTP for the right to download and share freely copyrighted contents. Our next step will consider P2P sharing as a non-market service that could be translated into a paying model. Online sharing of contents gives people some utility through the contents they enjoy afterwards and the information they get about the shared contents. In counterpart, they could be willing to pay to be entitled to access contents without restriction and whatever the sharing technology they use. Such approach does not assess the existing services, but an entitlement. It helps to evaluate the models based on an unimpeded access to cultural works: Who would be prepared to pay? How to increase their willingness to pay? How to induce to pay those who wouldn't be prepared to?

6. References

BUXMANN, P., POHL, G., JOHNSCHER, P., STRUBE, J., GROFFMANN, H.D. (2004) "Strategies for digital music markets: Pricing and the effectiveness of measures against pirate copies: Results of an empirical study", working paper.

CHEN, Y. and PNG, I. (2003) "Information goods pricing and copyright enforcement: Welfare analysis", *Information Systems Research*, 14(1), 107-123

CHYI, Hsiang Iris (2005) "Willingness to Pay for Online News: An Empirical Study on the Viability of the Subscription Model", *Journal of Media Economics*, 18(2), 131-142.

DOU, Wenyu (2004) "Will internet users pay for online content?", *Journal of Advertising Research*, December, 349-359.

HOLM, H.J. (2003), ‘Can economic theory explain piracy behavior?’, *Topics in Economic Analysis & Policy*, 3(1)

KRISHNAN R., SMITH, M.D., TELANG R. (2003), « The economics of peer-to-peer networks », *Journal of Information Technology Theory and Applications*, 5(3)

LIEBOWITZ (2006) "File Sharing: Creative Destruction or Just Plain Destruction?", *Journal of Law & Economics*, 49(1), 1-28

McFADDEN, D., TRAIN K.E. (1996) “Consumers' Evaluation of New Products: Learning from Self and Others”, *Journal of Political Economy*, 104(4), 683-703

OBERHOLZER, F., STRUMPF, K. (2004) “The Effect of File Sharing on Record Sales An Empirical Analysis”, working paper

PEITZ M., WAELBROECK P. (2003) “Piracy of Digital Products; A Critical Review of the Economics Literature”, CESIFO working paper n°1071

ROB, R., WALDFOGEL J. (2004), 'Piracy on the High C’s: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students', *Journal of Law & Economics*, 49(1), 29-62

ROCHELANDET, F., LE GUEL, F. (2005) "P2P music-sharing networks: Why legal fight against copiers may be inefficient?, avec Fabrice Le Guel, *Review of Economic Research on Copyright Issues*, 2(2), 69-82.

ROCHELANDET, F. (2005), "Unauthorised sharing through P2P networks: A digital pollution?", *Journal of Network Industries*, 5(1), 25-45

SHAPIRO C., VARIAN, H. (1998), *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press

SUNDARARAJAN, A. (2004) "Managing digital piracy: Pricing and protection", *Information Systems Research*, 15(3), 287-308.

TAKEYAMA, L.N. (1994), "The Welfare Implications of Unauthorized Reproductions of Intellectual Property in the Presence of Demand Network Externalities", *Journal of Industrial Economics*, 42(2), 155-166

TRAIN, K.E. (2003), *Discrete choice methods with simulation*, Cambridge University Press

VISHNUMURTHY, V., SANGEETH C., EMIN G.S. (2003), “KARMA: A Secure Economic Framework for Peer-to-Peer Resource Sharing”, Working Paper, Cornell University

ZENTNER, A. (2006), “Measuring the Effect of Music Downloads on Music Purchases”, *Journal of Law & Economics*, 49(1), 63-90

Appendix 1: Econometric results

Table 3: Ordered Logit Model (Dependent variable: WTP)

Variables	General Model	Demog	Cultural spending	Ethic	Legal risk	Herding	Usability	Music diversity	Herdingbis (friends)
Gender (Ref : man)	0.066 (0.62)	-0.143 (-1.58)	-0.104 (-1.22)	-0.025 (-0.28)	-0.131 (-1.45)	-0.092 (-1.02)	0.049 (0.47)	0.064 (0.62)	-0.084 (-0.84)
Age (Ref :<25) **25-30**	0.454*** (3.27)	0.376*** (2.88)	0.293*** (2.76)	0.487*** (3.68)	0.372*** (2.84)	0.344*** (2.63)	0.402*** (2.96)	0.447*** (3.28)	0.344** (2.53)
31-40	0.558*** (3.96)	0.534*** (4.16)	0.516*** (5.12)	0.663*** (5.11)	0.550*** (4.28)	0.406*** (3.11)	0.505*** (3.75)	0.560*** (4.15)	0.465*** (3.46)
41-50	0.673*** (4.24)	0.682*** (4.86)	0.679*** (5.94)	0.797*** (5.61)	0.712*** (5.05)	0.502*** (3.47)	0.653*** (4.35)	0.691*** (4.62)	0.631*** (4.23)
>50	0.448*** (2.72)	0.735*** (5.42)	0.766*** (7.18)	0.760*** (5.53)	0.770*** (5.64)	0.455*** (3.11)	0.594*** (3.96)	0.561*** (3.74)	0.506*** (3.37)
Income	0.019 (0.95)	0.037** (2.13)		0.024 (1.37)	0.036** (2.04)	0.045** (2.57)	0.031 (1.62)	0.026 (1.36)	0.038** (2.05)
Cultural spending	0.088*** (2.77)		0.066** (2.44)						
Ethic	0.173*** (9.21)			0.195*** (12.80)					
Legal risk	0.063 (1.37)				0.094** (2.29)				
Herding	-0.104** (-2.42)					- 0.191*** (-5.12)			
Usability	-0.052 (-1.37)						-0.134*** (-3.82)		
Music diversity	-0.109** (-2.27)							-0.240*** (-5.51)	
Herdingbis (friends)	-0.257*** (-3.07)								-0.358*** (-4.46)
LL	-2671	-3308	-3770	-3205	-3292	-3293	-2775	-2793	-2866
τ_1	-0.026 (-0.09)	-0.576 (-4.38)	-0.576 (-5.32)	0.716 (4.32)	-0.406 (-2.66)	-0.980 (-6.38)	-0.817 (-4.70)	-1.236 (-6.13)	-0.759 -4.99
τ_2	1.481 (5.15)	0.813 (6.15)	0.812 (7.43)	2.173 (12.73)	0.985 (6.39)	0.420 (2.75)	0.616 (3.54)	0.200 (1.00)	0.661 4.35
τ_3	3.171 (10.76)	2.364 (16.91)	2.344 (19.99)	3.827 (20.83)	2.536 (15.75)	1.979 (12.53)	2.202 (12.21)	1.789 (8.76)	2.233 13.99

*: Significant at 10%, **: Significant at 5%, ***: Significant at 1%. (...) Student coefficient.

[i] 'Perfect' means that copies are very closed to originals from a technical viewpoint.
[ii] http://www.insee.fr/en/home/home_page.asp. See 'Classification, Definitions – Methods' page, 'Statistical Tools' page and download CALMAR Macro.

Dépôt légal : 2007 – 4ème trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 1242-5125 ENST S (Paris) (France 1983-9999)