CERGE
Center for Economics Research and Graduate Education
Charles University Prague

Essays in Industrial Organization

Olena Senyuta

Dissertation

Prague, March 2015

Dissertation Committee

KREŠIMIR ŽIGIĆ (CERGE-EI, Charles University; chair)

PATRICK GAULÉ (CERGE-EI, Charles University)

AVNER SHAKED (CERGE-EI, Charles University)

Referees

STEFAN BEHRINGER (University Duisburg-Essen)

DIETMAR HARHOFF (Max Planck Institute for Innovation and Competition)

Table of Contents

Abstract

This dissertation deals with topics related to R&D investment in endogenous sunk costs markets. In particular, in Chapter 1 (which is co-authored by Krešimir Žigič), we build a theoretical model, where knowledge spillovers are introduced into Sutton's concept of endogenous sunk costs (investment in quality). We show that with spillovers increasing and the effectiveness of investment in raising quality decreasing, the lower bound on concentration for an industry decreases and ultimately collapses to zero when spillovers are large enough and/or effectiveness of investment is low enough. We also show that for an intermediate range of spillovers firms do invest in R&D although the market structure becomes fragmented as market size grows (there is no lower bound to concentration). In the second part, we allow firms to protect their investment against spillovers and focus on the symmetric equilibria, where all firms either protect their investment or do not protect it at all. We show that higher spillovers and/or lower effectiveness of investment may induce firms to protect themselves against spillovers, leading to higher investment in quality, and to more concentrated market structure. Thus, Sutton's result on the concentration bound is preserved.

We differentiate between ex post and ex ante knowledge spillovers. While the latter include only exogenous characteristics of the market environment, the former also account for possible protective actions undertaken by firms. In Chapter 2 we carry out empirical testing of the role of knowledge spillovers in an endogenous sunk costs model environment. First, we show that markets in which firms undertake intensive R&D expenditures are not likely to become fragmented, ceteris paribus, as the size of the market increases. On the contrary, for firms which do not undertake intensive R&D expenditures: the market is more likely to become fragmented as its size increases, and the effect of market size is even stronger if knowledge spillovers are high. We find that ex post knowledge spillovers have adverse impact on R&D incentives. On the other hand, firms with high ex ante knowledge spillovers are more likely to use private protective measures, possibly restoring incentives to invest in R&D.

Chapter 3 empirically investigates how the level of delegation of authority is related to the performance of an organization. Decentralized, horizontal organizational structure takes advantage of more efficient decision making, mainly due to more efficient use of "soft" information. The cost of such decentralization is the loss of control and the need to properly incentivise agents who are legitimately given the authority to make decisions. This is the trade-off organization faces when deciding on the level of authority delegation. The effect of authority delegation is studied using empirical data from the banking sector. Different specifications were used to estimate the effect of authority delegation on performance characteristics. Estimates demonstrate that more authority delegated has a positive effect on quantitative measures of bank performance; however, it decreases the quality of decisions taken. Results demonstrate that there is a trade-off between quantitative and qualitative performance characteristics. While a local bank branch is able to increase loan generation when more authority is delegated to it, there is also some evidence of loan quality deterioration.

Abstrakt

Tato dizertační práce pojednává o tématech souvisejících s investicemi do vědy a výzkumu (R&D) na trzích s endogenními utopenými náklady. Konkrétně v první kapitole (jejíž spoluautorem je Krešimir Žigič), konstruujeme teoretický model, ve kterém je do Suttonova konceptu endogenních utopených nákladů (investice do kvality) zaveden efekt přelévání znalostí. Ukazujeme, že při zvyšujících se efektech přelévání a snižující se efektivitou investic do zvyšování kvality, se Suttonova dolní mez koncentrace odvětví snižuje. A tato dolní mez klesá až k nule, pokud jsou efekty přelévání dostatečně významné a/nebo efektivita investování do zvyšování kvality dostatečně nízká. Také ukazujeme, že pro střední interval síly efektu přelévání firmy investují do R&D, přestože se tržní struktura fragmentuje s rostoucí velikostí trhu (není dolní hranice). Ve druhé části umožňujeme firmám své investice před efekty přelévání chránit. Zaměřujeme se na symetrické rovnováhy, ve kterých všechny firmy buď chrání svoje investice nebo je nechrání vůbec. Ukazujeme, že větší efekty přelévání a/nebo nižší efektivita investic do zvyšování kvality může přimět firmy, aby se před efekty přelévání chránily, což vede k vyšším investicím do kvality a k více koncentrované tržní struktuře. Suttonův výsledek ohledně meze koncentrace je tedy zachován.

Rozlišujeme mezi ex post a ex ante přeléváním znalostí. Ex ante zahrnuje pouze exogenní charakteristiky tržního prostředí, zatímco ex post bere v potaz i možné ochranné aktivity firem. Ve druhé kapitole empiricky testujeme roli přelévání znalostí v modelovém prostředí endogenních utopených nákladů. Zaprvé ukazujeme, že trhy, na kterých firmy vykazují intenzivní R&D náklady, se s rostoucí velikostí trhu, ceteris paribus, pravděpodobně nestanou fragmentované. Naproti tomu pro firmy, které nevykazují intenzivní R&D náklady, se trh pravděpodobně bude fragmentovat postupně tak, jak roste velikost trhu. A tento efekt velikosti trhu je silnější pokud existují velké efekty přelévání znalostí. Zjišťujeme, že ex post efekty přelévání mají negativní dopad na motivaci pro R&D investice. Na druhou stranu, firmy s vysokými ex ante efekty přelévání jsou více náchylné k použití soukromých ochranných opatření, které mohou případně obnovit motivaci pro investice do R&D.

Třetí kapitola empiricky zkoumá, jak delegování pravomocí souvisí s výkonem organizace. Decentralizovaná, horizontální organizační struktura využívá efektivnější rozhodování, zejména díky účinnějšímu používání "měkkých" informací. Cenou za takovouto decentralizaci je ztráta kontroly a potřeba nabídnout správné pobídky zástupcům, kteří dostávají legitimní oprávnění činit rozhodnutí. Toto je kompromis, který organizace musí dělat při rozhodování o úrovni delegovaných pravomocí. Dopad delegování pravomocí je studován na základě empirických dat z bankovního sektoru. Byly použity různé specifikace za účelem posouzení účinku delegování pravomocí na charakteristiku výkonu. Odhady ukazují, že více delegovaných pravomocí má pozitivní vliv na kvantitativní výkon banky, avšak snižuje kvalitu učiněných rozhodnutí. Výsledky ukazují, že dochází ke kompromisu mezi kvantitativními a kvalitativními charakteristikami výkonu. Zatímco místní pobočka banky je schopná zvýšit počet generovaných půjček, pokud k tomu bude mít více delegované pravomoci, existují rovněž určité důkazy o zhoršené kvalitě takovýchto půjček.

Acknowledgments

I would like to express my gratitude to my supervisor, Krešimir Žigič, for his continuous support and the overall guidance I needed to complete this work.

I also would like to thank Levent Čelik, Avner Shaked, John Sutton, Jan Kmenta, Francine Lafontaine, Jakub Steiner, Jeroen Hinloopen, Patrick Gaulé, Nikolas Mittag and Vilém Semerák for their valuable help with my work.

I am also thankful to participants of the INTERTIC Conference on Competition in High-Tech Markets, University of Milan, Bicocca (October 2010).

Financial support of the Grant Agency of the Czech Republic (grant no. P402/12/0961) and the Grant Agency of the Charles University (grant no. 56710) is gratefully acknowledged.

Last but not least, I am grateful to Martin Kuncl and my friends for their unconditional support, motivation and encouragement.

Prague, Czech Republic — Olena Senyuta
March 2015

Introduction

This dissertation studies two topics in the field of Industrial Organization: first, I theoretically and empirically study the problem of market structure with a focus on knowledge spillovers, and second, I empirically study the effect of control delegation inside an organization.

Specifically, the first two chapters of my dissertation focus on the theory of endogenous market structure. In this framework, the escalation of endogenous sunk cost expenditures prevents the market from fragmentation as the size of the market increases. Endogenous sunk costs are the result of investment in quality improvement in our model; we refer to these costs as R&D investment. We assume that customers consider every product a combination of useful features and qualities developed by the producer. R&D investment helps firms to develop and improve those features. We recognize the importance of knowledge spillovers in a setting with such quality improvements (consider hi-tech industries an example). Reverse-engineering and labor force flows are just a few examples of how knowledge spillovers might be realized in practice in such markets.

We start by assuming that knowledge spillovers are exogenous to the firms. In this setup, spillovers tend to decrease the lower bound of market concentration and when spillovers are strong enough, they completely eliminate it. Eliminating the lower bound does not, however, imply immediate elimination of a firm incurring endogenous sunk costs. We show that for an intermediate range of spillovers, firms invest in R&D although the market concentration becomes fragmented as market size grows. For very large spillovers, firms do not invest in quality improvement due to very strong disincentive effects of spillovers. Further, we assume that firms can protect their investment against spillovers. We focus on symmetric pure strategy Nash equilibria, where all firms either protect their investment or not. As a result, we show that unlike in the baseline case with exogenous spillovers, high knowledge spillovers may lead to a more concentrated market structure due to the possibility of firms' private protection from spillovers. The empirical analysis in Chapter 2 complements the theoretical results of Chapter 1. Regression analysis shows how incentives to invest in R&D and incentives to protect these investments respond to knowledge spillovers and market size.

The third chapter contributes to the debate about costs and benefits of horizontal organizational structures as compared to vertical organizations. Choice of ideal organizational structure includes an important trade-off. More horizontal, less vertically integrated organizations prove to be more effective in using "soft" information (non-verifiable and non-transferable), compared to more hierarchical organization. The use of this information can increase the effectiveness of decisions taken. But with more authority delegated to the lower levels of the hierarchy, the costs and the span of monitoring become higher. Using loan performance data from local bank offices, I study the effects of more control delegation to lower levels of hierarchy. Regression estimates show that more authority delegated has a positive effect on quantitative measures of local bank office performance. However, I also find evidence that more authority delegated might also decrease the quality of the decisions made.

Chapter 1

Managing Spillovers: an Endogenous Sunk Cost Approach [1]

Co-authored by Krešimir Žigič

1.1 Introduction

In his influential book, John Sutton (1991) provides a theory that explains why some markets remain highly concentrated. His theory predicts that in the presence of a certain type of sunk costs there is lower bound on the level of concentration in an industry. More precisely, the number of firms in a free entry equilibrium would reach some finite number, even if the size of the market approaches infinity. The reason is that the sunk costs "escalate" as market size grows. The special type of sunk costs that lead to such an outcome is coined "endogenous sunk costs". Sutton (1991) focuses on advertising outlays as the premier type of endogenous sunk costs, but any R&D expenditures, such as cost-reducing investment, and investment into quality, can be considered an endogenous sunk cost. Finally, note that in Sutton's approach, both endogenous sunk costs and market concentration are endogenously determined in industry equilibrium by such parameters as market size and efficiency of the sunk costs in affecting the market outcome (say, preferences of consumers).

[1] **An earlier version of this Chapter was published in CERGE-EI working paper series: Senyuta, Olena and Žigić, Krešimir. 2012. Managing spillovers: an endogenous sunk cost approach. Economics Institute, Academy of Sciences of the Czech Republic, WP472**

Much like Sutton (1991) and Sutton (2007), we focus on the markets in which incurring endogenous sunk costs is an essential feature of competition, but these sunk costs stem from an investment in product quality improvement rather than advertisement. Moreover, we introduce the knowledge or R&D spillovers stemming from firms' investment in product quality[2]. A firm's effective quality of the good is thus influenced by both the firm's own investment in quality, and investment in quality by other firms. In other words, a firm's product quality is a sum of its own quality innovations, and some portion of quality innovations developed independently by other firms. Thus, spillovers are assumed to be mutual; each firm benefits from spillovers coming from the other firms ("receiving spillovers"), but at the same time each firm involuntarily provides spillovers to all other firms in an industry ("giving away spillovers"). These features are consistent with the fact that innovations and imitations may be complements and reinforce each other (see Shenkar, 2010).

As for the empirical relevance of such a setup, one of the stylized facts about R&D investment (endogenous sunk costs in this case) is knowledge diffusion and imperfect appropriability of innovations. Reverse-engineering[3], labor force flows and strategic alliances among firms, among others, may serve as examples of such mutual knowledge spillovers and the mode by which they can be practically realized in an industry (see Shenkar, 2010, for many examples of these kind of knowledge leakages); see also Senyuta and Žigić (2012) for more detailed description of the modes of knowledge diffusion and for related literature on it.

In the basic version of our model, we treat R&D spillovers as exogenous to firms (captured by a single parameter) in the sense that firms cannot affect the intensity of those spillovers, while in the second part of the paper, we allow for the possibility for firms to manage spillovers (protect themselves from "giving away spillovers"). By that we mean deliberate actions of the firms to constrain giving away spillovers and to prevent a leakage of relevant knowledge to its competitors. In this case, we distinguish between ex-ante spillovers (which are exogenously given from the point of view of the firm), and ex-post spillovers, which are spillovers (if any!) that remain after the firms' protective actions. In other words, in the basic version of the model we consider only ex-ante spillovers, while in the extended model we allow firms to use protective measures and so the notion of ex-

[2]Note, however, that our analysis would be basically the same for the type of advertising known as "informative advertising" that spills over to the competitors and beneficially affects them.

[3]Reverse-engineering is disassembling of the product to learn how it was built and how it works.

post spillovers appear. These protective measures, in addition to patents and copyrights, also include also costly private protection that firms undertake to reduce or eliminate spillovers if they find it optimal. In some cases, spillovers might be characterized as information leakage or imitations that are on the border of intellectual property rights (IPR) violations and cannot be effectively suppressed by public IPR protection (patents or copyrights). In this case, private or technical protection (see Střelický and Žigić, 2011; Scotchmer, 2006, chapter 7) is an example of managing giving away spillovers.

Note that this extended setup (in which firms manage spillovers) can be also viewed as a situation in which both public (patents, copyrights, etc.) and private (secrecy, increasing product complexity, masquing, etc.) IPR protections are present. More specifically, the ex ante spillovers can be considered the information leakages that exist despite public protection like copyright or even patents (and are, as we argue above, borderline or actual IPR violations). Ex post spillovers, on the other hand, can be considered the information leakages that remain after the firms add their private protection on top of already existing public IPR protection.

Our analysis has several aims: Firstly, we investigate the robustness of the lower bound on concentration in the above setup in which knowledge spillovers are exogenous, and study the impact of spillovers on the equilibrium values such as endogenous sunk costs or market concentration. More specifically, we aim to study the incentives of a firm to invest in quality enhancement in the presence of knowledge spillovers and to analyze how an interplay between spillovers, market size, the effectiveness of R&D investment in quality improvement (henceforth shortened to "the effectiveness of investment") and free entry affects endogenous sunk costs (that is, R&D outlays) and, consequently, market concentration. In this respect, we decompose the change of endogenous sunk costs induced by change in market size into i) entry and ii) escalation effects and then study how the size of spillovers and the size of markets impacts these two effects and, consequently, the total change in endogenous sunk costs. Secondly, we study how levels of spillovers and results of investments in quality improvement will impact a firm's decision to allow giving away spillovers. We thereby investigate the interaction between public and private protection which are simultaneously allowed in our extended setup, such as the impact of relaxation of public protection. Thirdly, we analyze how the possibility to restrain "giving away spillovers" affects the lower bound of concentration and the level of endogenous sunk costs. Finally, we also investigate how the level of effectiveness of investment affects the endogenous sunk costs and, consequently, the market concentration in the situation when firms manage spillovers.

The effect of spillovers on the lower bound of market concentration is not only an interesting theoretical exercise, but also provides important insight to antitrust authorities, given the empirical relevance of spillovers. Competition offices would surely benefit from knowledge of how the actual market concentration deviates from the corresponding lower bound in order to assess the possible barriers to entry and, consequently, the degree of competitiveness in an industry.

To best of our knowledge, our analysis is novel and, as we will discuss in the next section, rather different in its focus compared to related literature that builds on Sutton's (1991) seminal work. Moreover, it yields several interesting testable hypothesis. For instance, when spillovers are exogenous, market concentration and its lower bound will be lower, and may even disappear when spillovers exceed a particular threshold and when the effectiveness of investment falls beyond certain critical value. Another testable hypothesis arises when firms manage spillovers. Then, large enough ex ante spillovers may induce a more concentrated market structure due to the possibility of firms' private protection from spillovers.

The rest of the paper is organized as follows. In Subsection 1.2 we briefly review the related literature, while in Subsection 1.3 we present the basic model in which spillovers are assumed exogenous to the firms and study i) the effects of spillovers and the effectiveness of R&D investment on the lower bound of concentration, ii) R&D and profit disincentives due to spillovers and iii) the effect of market size on endogenous sunk costs under different levels of spillovers and its decomposition into entry and escalation effects. In Subsection 1.4, we allow the firms to eliminate giving away spillovers by means of private protection if they find it optimal and study how this added feature affects the relationship between the market size and concentrations for different levels of initial or ex ante spillovers. Moreover, we also study how the effectiveness of investment and the firm's cost of protection affects firms' decisions whether or not to manage spillovers. Finally, in Subsection 1.5 we conclude.

1.2 Survey of the Literature

Despite the indisputable importance of Sutton's (1991) and (2001) works, intriguingly enough, subsequent theoretical and empirical research in this area is relatively scant.

As for theoretical work that builds on Sutton's setup, there are various directions and themes on which the subsequent theoretical literature progressed and focused. Bresnahan (1992) was one of the first to review Sutton's (1991) concept of endogenous sunk costs in relation to the existing market structure literature. In particular, Bresnahan (1992) concludes that it would be necessary to use the strategic approach put forward by Sutton for further research of industry structure and concentration (that is, to account for the possible existence of endogenous sunk costs in an industry and its economic impact).

The importance of the endogenous sunk costs concept was reaffirmed in other papers which studied market structure. Carlton (2005), for instance, reconsiders the concept of entry barriers, and essentially shows that they should be modeled as dynamic phenomena as Sutton's approach suggested. Matraves and Rondi (2005) compare horizontally and vertically differentiated markets using Sutton's concept of endogenous sunk costs and show that in markets with horizontal product differentiation "escalation effect" is not present, so markets become fragmented as market size increases. Vasconcelos (2006) builds a model of endogenous coalition creation (merger of firms) in markets with both exogenous and endogenous sunk costs and shows that in a market characterized by exogenous sunk costs a monopoly coalition of firms is unsustainable: as the size of the market increases, more firms prefer to enter the market, and the upper bound on market concentration decreases. In a market characterized by endogenous sunk costs, however, an upper bound to concentration does exist. Vasconcelos (2006), thus extends Sutton's model to allow for endogenous merger decisions.

Behringer (2014) studies the viability of the positive lower bound in a market with intra- and inter-firm network effects. The author assumes that "perceived" quality of the product depends on the quantity produced by the firm itself (intra-firm network effects) and by other firms in the market (inter-firm network effects), and uses the example of the video game console market, where information about the product (at early stages) was spread by "word of mouth". Thus, the more firms sell today, the higher amount they will sell tomorrow. Unlike this work, Behringer (2014), however, does not model investment in R&D (endogenous sunk costs) explicitly, and spillover effects in his model stem from the network effects rather than from knowledge leakages. Despite that, Behringer (2014), shows that when inter-firm network effects are sufficiency high, the lower bound on concentration decreases to zero as market size grows to infinity.

Another somewhat related analysis to Sutton (1991) and Sutton (2007) can be found in Vives (2008), who generalizes models of free entry with cost-reducing R&D investments, in the stage prior to the product market competition. Vives (2008) shows that increasing market size typically leads to an increase in the number of firms but less than proportionately, and thus increases the innovation incentives of individual firms. So the potential negative effect of an increase in the number of firms on the incentive to innovate is mitigated by the size of the market effect[4].

Concerning the related empirical literature, most of it focuses on testing for the presence of endogenous sunk costs in an industry and its consequences for market concentration and competition and so this literature is more relevant for our forthcoming analysis than the theoretical literature reviewed. There is an important distinction between exogenous and endogenous sunk costs given their respective roles and impact on the nature of competition in given markets (Type I and Type II markets respectively, following Schmalensee (1992)).

In Type I markets, market size does not affect sunk costs. Therefore, as the market grows, more firms enter, and the market is likely to become more fragmented. On the other hand, an increase in market size leads to an increase in R&D expenditures (or other endogenous sunk costs) by the incumbents on Type II markets who, in turn, create the barriers to entry for other firms, and the market is more likely to remain concentrated. Thus, the predictions about the market size and market concentration are clear in both cases: i) an increase in the market size in exogenous sunk costs markets does not influence R&D investment, but leads to new firm entry and decreased market concentration; ii) in endogenous sunk costs markets an increase in the market size leads to an increase in R&D investment, and consequently limits or prevents the entry of new firms.

One of the examples of empirical tests of the endogenous sunk costs theory is Dick (2007) paper, which focuses on the banking industry. The author conjectures that the banking industry is an endogenous sunk costs market. Using geographical definitions of the bank markets in USA (defined by metropolitan statistical areas), the author finds that the correlation between market size and concentration is close to zero, and provides evidence that banks invest extensively in quality to raise the barriers to entry, which shows that banking is likely to be a Type II industry. Furthermore, the paper investigates quality provision in the banking industry. All quality measures are positively associated with market size (advertising and branch intensity are among the measures of quality), and so dominant banks in the markets provide higher quality than fringe banks.

[4]This Vives (2008) prediction was tested by Coscollá-Girona et al. (2011) using the panel data of Spanish manufacturing firms (1990-2006). Empirical evidence shows that the market size variables (which measure the competitive pressure) positively and significantly influence the incentives to conduct product and process innovations.

Another empirical paper which tests endogenous sunk costs theory is Robinson and Chiang (1996). The authors use a heterogeneous sample of consumer and industrial product markets and show that in exogenous sunk costs markets, minimum values of concentration decline towards zero as market size increases, and in the endogenous sunk costs markets concentration remains bounded away from zero. Similarly, Matraves (1999) finds that endogenous sunk costs play a crucial role in the pharmaceutical industry and affect market structure. A paper by Bronnenberg et al. (2005) studies consumer package goods markets in geographical dimension and finds that a fixed number of advertised brands exists across markets of varying size, and that concentration is bounded from below in advertising-intensive industries even as market size grows large. Similar results are found in a paper by Ellickson (2007), which studies supermarket chains and shows that a small number of firms (4 to 6) capture the majority of sales and investment in distribution system plays a role of endogenous sunk costs in this market. Finally, Berry and Waldfogel (2010) show that in the newspaper industry, where quality is produced by means of fixed cost outlays, average quality increases with market size. On the other hand, in the restaurant industry, where quality is maintained by variable costs, range of qualities increases with market size and each product maintains a small market share.

In all of these papers, however, R&D spillovers are absent and and they do not test the implications of R&D spillovers and their potential control would have on the very notion of endogenous sunk costs, and, consequently on market concentration and competition. Consequently, there is no empirical testing on how the effectiveness of endogenous sunk costs in affecting the market outcome would impact the market concentration and competition.

Unlike the theoretical and empirical literature of Sutton's approach, the corresponding literature on R&D spillovers is rather rich (see, for instance, the comprehensive survey by Hall et al. (2010), and Keller (2004), see also Bloom et al. (2013)).

The theoretical prediction of the effects of spillovers on the incentives to invest in R&D (that is, in the endogenous sunk costs in our setup), however, is controversial. Spence (1984) was the first to demonstrate that knowledge spillovers have disincentivising effects on R&D investment[5]. This is a very intuitive result, but empirical evidence is often opposite: industries which are most likely to suffer from knowledge spillovers (pharmaceutical, IT technologies, etc.) also are among the industries which invest most in R&D.

[5] See also Suzumura (1992); d'Aspremont and Jacquemin (1988); Kamien et al. (1992); Tesoriere (2008) on how spillovers affect the incentives to create R&D joint ventures and other cooperation mechanisms.

Thus, Cohen and Levinthal (1989) conjectured that R&D investment not only generates new knowledge, but also increases a firm's ability to absorb and assimilate information generated by other firms. In their setting higher knowledge spillovers generate higher iincentives for firms to invest in R&D.

Both approaches to the analysis of spillovers effect on R&D incentives have their merit, but the theoretical predictions are inconclusive. For example, De Bondt (1997) reviews different theoretical approaches and concludes that different models often provide opposite results. In general, most existing models[6] show that spillovers increase general level of productivity in the industry (market, sector, economy), but have contradicting results on the effect of spillovers on individual firm incentives to innovate.

Our distinction between the ex ante and ex post spillovers would enable us to reconcile these seemingly opposing predictions. More specifically, whether the spillovers have a disincentivising effect on R&D investment, depends on the nature of giving away spillovers. That is, whether spillovers are exogenous or under the control of the firm. In the latter case, the increase in spillovers may trigger firm's private protection and so increased spillovers lead to increased R&D in equilibrium. If this is not the case, then spillovers display a standard disincentivising effect on R&D investment.

Finally, as for the literature that deals with a firm explicitly managing spillovers (that is, introducing private protection against giving away spillovers), there is usually a choice between applying, for example, patents or opting for the private protection that comes typically in the form of secrecy (see Hall et al. (2014)). Thus public and private mechanisms used to restrain giving away spillovers are considered to be mutually exclusive, but there are at least two reasons why this may be unrealistic. First, empirical evidence shows that i) both public and private protection of innovations are used by firms (Cohen et al., 2000), and ii) that private protection might be more valuable than patents, especially for small firms (see, for example, Leiponen and Byma (2009) and Arundel (2001)). Secondly, innovations usually do not consist of one piece of knowledge, so eventually not everything can be patented and more complicated protection strategies have to be used. As Anton et al. (2006) puts it: "[b]ecause innovations are rarely composed of a monolithic piece of knowledge, a combination of patenting and secrecy is common." Similarly, in digital markets, the combination of copyright and firms' private protection is typical (see Žigić et al. (2013) for the analysis of interaction between the private and public IPR protection,

[6]Models include endogenous growth literature, contestable inventions models, cooperation in R&D and consortia, and models with asymmetric R&D strategies (leaders and followers in inventions).

and the survey of the related literature).

Recall that our extended setup in which firms manage spillovers can be also viewed as a context where both public and private protection interact. In this respect, Atallah (2004) and Henry and Ruiz-Aliseda's (2012) modeling of private protection is somewhat related to ours. Both papers consider costly investment by firms in making the product more complex and more complicated to copy. Henry and Ruiz-Aliseda (2012) model R&D investment as an invention race, with leading firms and several followers, while Atallah (2004) considers R&D as non-tournament cost-decreasing investment, which is more similar to our setup.

Like Atallah (2004), we show that public protection (patents) and private protection act as substitutes: the higher the level of ex ante (or exogenous) spillovers (that is, the lower the level of public protection), the higher incentives to use costly private protection. Contrary to Atallah (2004), however, we show that investment in protection from spillovers is more likely if R&D investment becomes less efficient. The reason for those differences is that we assume that R&D costs and protection costs are interrelated. That is, the decision to protect against spillovers increases R&D costs by some fraction, unlike in Atallah (2004), where the costs of R&D and costs of protection are additively separable in firms' profit function. As a result, if R&D investments are less efficient, firms spend less on R&D, and therefore protection is less costly. Moreover, as a consequence of lower barriers to entry (that is, lower protection costs), more firms would enter the market. Clearly both factors point towards more private protection.

The major difference of our approach from the related literature on knowledge spillover management is, in general, that market structure (number of firms) is not given in our setup but it is, among other things, the outcome of the interaction between the private and public protection from giving away spillovers.

1.3 The Basic Model

1.3.1 Model setup

Much like Sutton (1991) or Sutton (2007), we exploit essentially the same three-stage game setup in our basic model. In the first stage firms decide whether or not to enter the market and the firms that enter incur sunk entry cost, F_0. In the second stage the firms choose sunk investment to set the quality of the product, which we refer to as R&D

investment. Finally, in the last stage, N firms which entered the market simultaneously choose quantities, x_i. The total cost of choosing quality level u_i for firm i is $F_i = F_0 + u_i^\delta$, where u_i is the quality level of good i, F_0 is a setup cost, and $\delta > 1$ is a model parameter that measures the effectiveness of R&D in raising perceived quality. A lower value of δ means that a given level of fixed R&D outlays leads to a greater increase in quality. When δ tends to infinity, R&D investment becomes more ineffective in enhancing quality. We consider R&D investment as an instrument to produce product innovations (product quality), which are valued by consumers. Due to spillovers, those innovations can be simultaneously developed by all firms in the market. Examples of such product innovation could be, for instance, new models or modifications of mobile phones, personal computers, or automobiles. Such spillovers are coined "output spillovers" (Hinloopen, 2000) since the competitors benefit from already achieved innovation ("output") rather than from investment in innovation that would instead imply "input spillovers" (for the distinction and the economic implications of the these two types of spillovers see Amir, 2000; Amir et al., 2003).

Consumers, who are (as in Sutton, 1991, 2007) assumed to be homogenous in valuation of quality, buy a good from firm i, based on the observed quality u_i. A typical consumer's utility function is of the form

$$U = (ux)^\beta z^{1-\beta}$$

where z is the outside good, and x is the "quality" good, $u \geq 1$ reflects the perception of good $x's$ quality.

We start solving the model backward. Each firm offers a single good with quality u_i at price p_i. Now, after observing prices and qualities of all firms, the consumer chooses to buy from the one which has the highest u_i/p_i ratio. For firms to have positive sales in equilibrium, we must have

$$u_i/p_i = u_j/p_j \text{ for any } i \text{ and } j. \tag{1.1}$$

With the given Cobb-Douglas form of utility function, let β be the share of income spent on the "quality" good (for derivations of that see Appendix 1.A.1). Following the notation of Sutton (2007), total spending on "quality good" in the market S is such that $S = \sum_{j=1}^{N}(p_j x_j)$. Note that S is the key parameter that serves as the measure of the market size[7]. Also, we define $u_i/p_i = u_j/p_j = 1/\lambda$, where λ can be interpreted as the price of good x with a *unit* quality. Now, if the price of a good x with a *unit* quality is λ, the price of a good with quality u_i is $p_i = \lambda u_i = Su_i/\sum_{j=1}^{N}(u_j x_j)$[8].

At the last stage of the game, investment in qualities are already sunk, and firms simultaneously choose quantities to be sold to maximize profits. Firm i solves:

$$\max_{x_i}\Pi_i = p_i x_i - cx_i = \lambda u_i x_i - cx_i$$
$$FOC(x_i) : \lambda u_i + u_i x_i \frac{d\lambda}{dx_i} - c = 0$$
$$u_i x_i = \frac{S}{\lambda} - \frac{cS}{\lambda^2 u_i}$$

Summing up first-order conditions for all $i = 1, ..., N$, and rearranging it, we obtain profit expression for firm i, after simultaneous choice of x_i by each firm, as a function of quality choice u_i:

$$\Pi_i = S\left(1 - \frac{N-1}{u_i\sum_{j=1}^{N}(1/u_j)}\right)^2 = S\left(1 - \frac{N-1}{1 + u_i\sum_{j\neq i}(1/u_j)}\right)^2 \tag{1.2}$$

In the second stage, the firm i makes a decision about u_i and faces the following choice: it can stick with the basic quality level described by $u_i = 1$, or invest in R&D and opt for higher quality where $u_i > 1$. In the former case, there is no R&D investment and thus no spillovers from other firms since the basic quality is already known and well established, while in the latter case (setting u_i to $u_i > 1$), the firm i chooses its investment in R&D and also benefits from the R&D of the other firms via knowledge spillovers. We focus on the latter case that turns out to be relevant when market size is "large" enough.

Thus firms choose optimal level of investment in quality u_i, while for consumers, a firm's i perceived product quality would be effectively $u_i^* \geq u_i$. The reason for that are spillovers from other firms in the industry. It is at this stage of the model that we depart from Sutton's original 3-stage game setup and introduce knowledge spillovers to

[7] S can be also expressed as $M\beta y$, where M is the measure of consumers in the market and y is income of a representative consumer.

[8] If we divide S by the total quantity of good x sold (weighed by quality), $\sum_{j=1}^{N}(u_j x_j)$, then $S/\sum_{j=1}^{N}(u_j x_j) = \lambda$

the model. Similar to Spence (1984) and Kamien et al. (1992), we define u_i^* in a linear way as

$$u_i^* = u_i + \sum_{j \neq i} \theta u_j, \tag{1.3}$$

where θ is an industry spillover parameter such that $\theta \in [0, 1)$. We denote the quality choice of the firm i as u_i, and u_j is the quality choice by each of the other $N-1$ firms. So firm's i effective quality comprises from the quality choice u_i of the given firm i, and the fraction θ of the quality choices of other firms, which enter u_i^* through spillovers.

In other words, u_i^* includes both the features and qualities developed by firm i, and some portion of features and qualities developed independently by other firms in the market, and, as discussed in the introduction, the channels via which this transfer of knowledge takes place are reverse-engineering, labor force flows among firms, strategic alliances between firms, knowledge dispersion to competitors through "vertical channel" (supplier-client), etc.

With this definition of spillovers, the profit expression to be used in the second stage now becomes:

$$\Pi_i = S \left(1 - \frac{N-1}{u_i^* \sum_{j=1}^{N} (1/u_j^*)} \right)^2 = S \left(1 - \frac{N-1}{1 + u_i^* \sum_{j \neq i} (1/u_j^*)} \right)^2 \tag{1.4}$$

When firm i makes a decision about u_i, it compares the marginal benefit with the marginal cost of the investment in quality.

The marginal benefit from investing in quality is:

$$\frac{d\Pi_i}{du_i} = \frac{\partial \Pi_i}{\partial u_i^*} \frac{du_i^*}{du_i} + \sum_{j \neq i} \frac{\partial \Pi_i}{\partial u_j^*} \frac{du_j^*}{du_i} \tag{1.5}$$

Now, $\frac{du_i^*}{du_i} = 1$ and $\frac{du_j^*}{du_i} = \theta$ from (1.3). Deriving the expressions for $\frac{\partial \Pi_i}{\partial u_i^*}$ and $\frac{\partial \Pi_i}{\partial u_j^*}$ and imposing the symmetry condition [9], we obtain expression for $\frac{d\Pi_i}{du_i}$:

$$\frac{d\Pi_i}{du_i} = \frac{2S(N-1)^2(1-\theta)}{N^3 u(1+(N-1)\theta)} \tag{1.6}$$

Also, note that $\frac{dF_i}{du_i} = \delta u^{delta-1}$.

As we have argued above, for small market size, the investment in quality does not pay off. The marginal benefit is then lower than marginal costs (that is, $\left.\frac{d\Pi_i}{du_i}\right|_{u=1} \leq \left.\frac{dF_i}{du_i}\right|_{u=1}$), so firms do not invest in quality enhancing R&D and the standard quality $u_i = 1$ prevails in the market equilibrium. As a result, the number of firms is determined by exogenous fixed entry outlays, F_0, and we call this market setup the "exogenous sunk costs regime". As market size increases in this regime, more firms enter the market, the market concentration decreases without limit and, in the absence of endogenous sunk costs, would approach zero as the market size reaches infinity. Beyond a certain critical value of S, (say, $\hat{S}$), however, it may pay off for a firm to deviate and start investing in quality (that is, to set, $u_i > 1$)[10]. Thus, for a large enough market size, profit maximization (with respect to u) *may require* a shift to another, endogenous sunk cost regime that results in quality enhancing investment. This would be exactly the case in our model (when spillovers are not "too large") so for $S > \hat{S}$ a firm chooses $u_i > 1$ by setting $\frac{d\Pi_i}{du_i} = \frac{dF_i}{du_i}$ and this yields:

$$F_i = \frac{2S(N-1)^2(1-\theta)}{N^3\delta(1+(N-1)\theta)} + F_0, \tag{1.7}$$

which gives us optimal investment into quality for each firm in symmetric equilibrium, given N firms entered [11].

Finally, to compute the number of firms entering in the first stage, we impose a zero profit condition (free entry): $F_i = \Pi_i$. Expression for Π_i in symmetric equilibrium

[9]The symmetry condition which simplifies the following expressions is $u_i = u_j = u$, yielding:
$u_i^* = u_i + \sum_{j\neq i}\theta u_j = u + \theta(N-1)u = u(1+(N-1)\theta)$;
$u_j^* = u_j + \theta u_i + \sum_{k\neq i, k\neq j}\theta u_k = u + \theta u + \theta(N-2)u = u(1+(N-1)\theta)$;
and $u_i^* = u_j^* = u^* = u(1+(N-1)\theta)$.

[10]Appendix 1.A.5 provides an example of how to calculate the critical value $\hat{S}$ for certain parameter values.

[11]It is easy to demonstrate that $d\left(\frac{d\Pi_i}{du_i} - \frac{dF_i}{du_i}\right)/du_i < 0$ at $u_i = \frac{2S(N^*-1)^2(1-\theta)}{N^{*3}\delta(1+(N^*-1)\theta)}^{1/\delta}$. Therefore, the second order condition is satisfied.

becomes $\Pi_i = S\left(\frac{1}{N}\right)^2$, with (1.7) we obtain:

$$\frac{2S(N-1)^2(1-\theta)}{N^3\delta(1+(N-1)\theta)} + F_0 = S\left(\frac{1}{N}\right)^2 \tag{1.8}$$

The relation (1.8) above is an implicit equation for the optimal number of firms, N^*, from which we can express N^* as a function of market size S and parameters (F_0, θ, δ).

1.3.2 The lower bound of concentration, spillovers and the effectiveness of R&D investment

The above analysis sets a stage to study how the interplay between spillovers, market size, and the effectiveness of investment in quality improvement affects firms' R&D outlays (that is, endogenous sunk costs) and, consequently, equilibrium number of firms and market concentration under the condition of free entry. The effect of an increase in market size, S, on the endogenous sunk costs (in the absence of spillovers) is already well known, mainly due to the influential work of Sutton (1991; 2001; 2007). Thus, the key insight of Sutton is that an increase in S leads to an escalation of R&D expenditure and so to the restraint of entry of new firms, which in turn results in a rather concentrated market even when the market sizes grows without limit. Here we briefly discuss how basic findings of Sutton change once we allow for spillovers among the firms. In this section we focus on the effects of spillovers on the lower bound of market concentration, while in subsection 1.3.4 we look more closely at the interaction between spillovers and the change in the endogenous sunk costs. We use the Herfindahl index, H as the standard measure of market concentration, which in the symmetric equilibrium assumes the value $H = 1/N^*$. We rewrite (1.8) as:

$$\frac{F_0}{S} = \frac{N\delta(1+(N-1)\theta) - 2(N-1)^2(1-\theta)}{N^3\delta(1+(N-1)\theta)} \tag{1.9}$$

and this, in turn, enables us to state our first proposition.

Proposition 1. *An industry with low spillovers (that is, $\theta < \frac{2}{2+\delta}$), for which endogenous sunk costs matter, will, ceteris paribus, remain highly concentrated as the size of the market increases, while an industry with high spillovers will become fragmented with an increase in S.*

Proof. Label the value of N when S tends to infinity as $N^*_\infty(\theta)$. It is then straightforward to show that for low spillovers $0 \leq \theta < \frac{2}{2+\delta}$, there is a *finite* value of $N^*_\infty(\theta)$ which satisfies the condition (1.9) as S tends to infinity and that $dN^*_\infty(\theta)/d\theta > 0$ on this interval (see Appendix 1.A.3 for the formal proof of this result). Figure 1.1 demonstrates proposition 1 graphically. $\square$

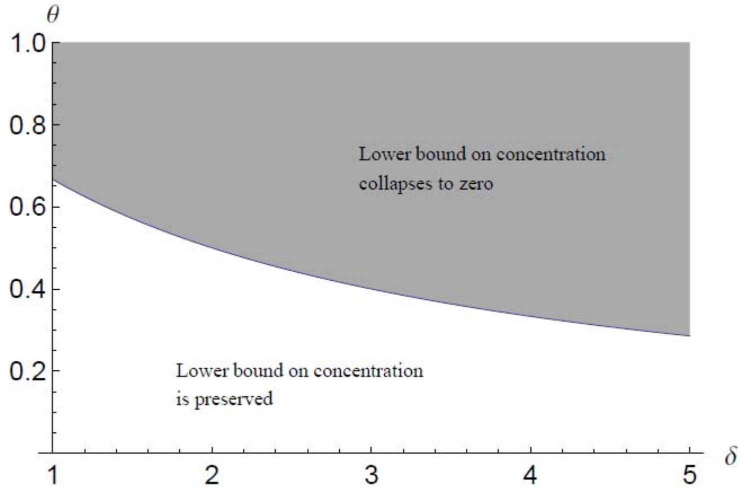

Figure 1.1: Values of parameters θ and δ, where the lower bound of concentration collapses, as market size S approaches infinity.

Thus, much as in Sutton (1991), there is a lower bound of market concentration that is strictly positive when spillovers are low, even if the market size is unlimited. To shed more light on this important result, it would be insightful to compare it with the standard market setup in which there are no endogenous sunk costs, that is, with the situation in which there are only setup costs, F_0. In this exogenous sunk cost regime a number of firms will tend to infinity when F_0 goes to zero (given "very large" market size). In the endogenous sunk costs regime, however, this will not be the case (given that $\theta < \frac{2}{2+\delta}$) even if we set $F_0 = 0$. So the number of firms will be finite (and the lower bound of concentration positive) even if the setup costs are absent and the market size approaches infinity. The reason, as we know, is the escalation of R&D expenditures with the growth of market size, which in turn prevents the entry of new firms. In fact, an alternative and straightforward way to find out the lower bound of market concentration in the endogenous sunk cost regime is to set $F_0 = 0$ and then calculate the number of firms in the free entry equilibrium that will give us exactly $N^*_\infty(\theta)$. This is intuitive because in the absence of the setup costs the maximal possible number of firms would enter under a zero profit condition and this finite number is independent on the market size (given that market size is large enough to support the endogenous sunk cost regime). That, in turn, yields the lower bound of market concentration for $F_0 > 0$ and market size tending to infinity.

For large spillovers, on the other hand, the market outcome is slightly different. When $\theta \geq \bar{\theta} = \frac{2}{2+\delta}$, and as S goes to infinity, it must be that the $N^*_\infty(\theta)$ also goes to infinity, in order for (1.9) to be satisfied. In other words, $N^*_\infty(\theta) = \infty$. Thus the positive lower bound of concentration disappears[12]. The firm's sunk outlays, however, do not vanish once the spillovers reach or slightly exceed the threshold level, $\bar{\theta}$, but become insufficient to block the new entry once the market size increases. Thus there is a kind of hybrid regime in which there are endogenous sunk costs on the one side, but the positive lower bound of concentration vanishes on the other side. Finally, at another, higher threshold of the spillover level, $\tilde{\theta} > \bar{\theta}$, the disincentives effect of spillovers prevails and is so strong that a firm opts for the basic quality by setting $u_i = 1$. The level of $\tilde{\theta}$ depends on the underlying parameters of the model, that is, $\tilde{\theta}(\delta, S, F_0)$ (see Appendix 1.A.6 for the derivation of $\tilde{\theta}$). Much like $\bar{\theta}$, it decreases in δ, yielding the lower critical level of spillovers beyond which there is a exogenous sunk cost regime (for a given S and F_0). In addition, $\tilde{\theta}$ increases in S because larger market size provides motivation to invest more in R&D, so larger spillovers are needed to offset this incentive. Lastly, and also intuitively, $\tilde{\theta}$ increases in fixed entry costs since this increase, ceteris paribus, makes entry more difficult and therefore enhances the incentive to invest in R&D. In order to offset this, the critical level of spillovers $\tilde{\theta}$ has to increase. Finally, it can be shown that $\lim_{S\to\infty} \tilde{\theta} = \frac{2(1+F_0)}{2(1+F_0)+\delta} < 1$, so for spillovers larger than this limit, there is an exogenous sunk cost regime irrespective of market size[13].

We now switch for the moment to another important parameter of the model - the effectiveness of R&D investment, δ. First recall that δ is an inverse measure of R&D effectiveness and this implies that the larger δ is, the lower will be investment in R&D. Consequently, when δ tends to infinity, a firm ceases to invest in R&D in the limit and

[12]Moreover, N^* approaches infinity at a different rate, depending on the value of the spillover parameter, with higher spillovers leading to N^* increasing at a higher speed.

[13]Nocke (2007), however, shows that large spillovers restore an endogenous sunk cost regime in a particular dynamic model, in which firms compete in endogenous sunk outlays on quality, and there is a collusive "underinvestment" equilibrium, initially without spillovers. So firms do not deviate from this equilibrium, because if one of them does, the other firms retaliate and also keep increasing their level of u from then on. Such a permanent escalation in sunk costs is very costly and so not profitable compared to collusion, but the appearance of large spillovers changes it completely. The incentives to deviate is still present, but the escalation of sunk costs will not be so costly because of spillovers. Thus punishment would not be effective and therefore the "underinvestment" equilibrium would no longer be sustained (so only the escalation equilibrium is possible).

continues to adhere basic quality $u_i = 1$ (note that $\lim\limits_{delta \to \infty} u_i = 1$).

Proposition 2. *An industry, in which it would be easy to enhance the (perceived) product quality ("low" δ), would be more concentrated than an industry that has lower R&D effectiveness ("high" δ) given that both industries are exposed to the same "low" level of spillovers (that is, $\theta < \bar{\theta}$) and have the same market size.*

Proof. Note that both the critical levels of spillovers, $\bar{\theta}$ and $\tilde{\theta}$, depend on δ; an increase in δ leads to a fall in $\bar{\theta}$, so the lower bound of concentration falls. By the same token, a rise in δ results in the fall of $\tilde{\theta}$ and so, ceteris paribus, the exogenous sunk regime appears at the lower critical spillover level while the associated market concentration is lower. □

As expected, if R&D investment is not very effective in raising quality (δ is high), firms do not invest much in R&D, and so barriers to entry are lower. In such circumstances a lower level of spillovers is needed for the number of entrants to grow without limit as market size increases, leading the lower bound of concentration to collapse to zero.

To illustrate how spillovers affect market concentration, we provide the numerical example below (using parameter values $\delta = 2$, $F_0 = 2$), and solve the model for the equilibrium N^* for different values of θ. The figure below demonstrates how the equilibrium concentration $1/N^*$ and its lower bound changes for different values of θ as market size S increases. For example, the upper line represents the standard case when $\theta = 0$, (spillovers are zero). For small S (dotted part of that curve), there is an exogenous sunk costs regime. For S high enough, there is an endogenous sunk cost regime, the lower bound of market concentration approaches approximately 0.4, and the equilibrium number of firms is finite. The lowest full line represents the case where $\theta = 0.7$. The lower bound on concentration approaches zero with S going to infinity, and equilibrium number of firms approaches infinity. So, with $\theta = 0.7$ there is an exogenous sunk cost regime for small S, and hybrid regime for $S > 303$. On the other hand, for $\theta = 0.9$ for any S, we obtain an exogenous sunk costs regime: for the assumed parameter values δ and F_0, we obtain $\tilde{\theta} < 0.9$, and we are in the exogenous sunk costs regime for all S. Therefore, our **testable hypothesis 1** states that increasing the market size leads to a decrease in market concentration in the industries with high spillovers, but if spillovers are low, a market size increase would affect fragmentation at a much lower rate (if at all).

As we can see from the parameterized example, the lower bound on equilibrium concentration level ($1/N^*$) decreases with spillovers, and for the values of the spillover parameter $\theta > \bar{\theta} = \frac{2}{2+\delta}$, it completely disappears.

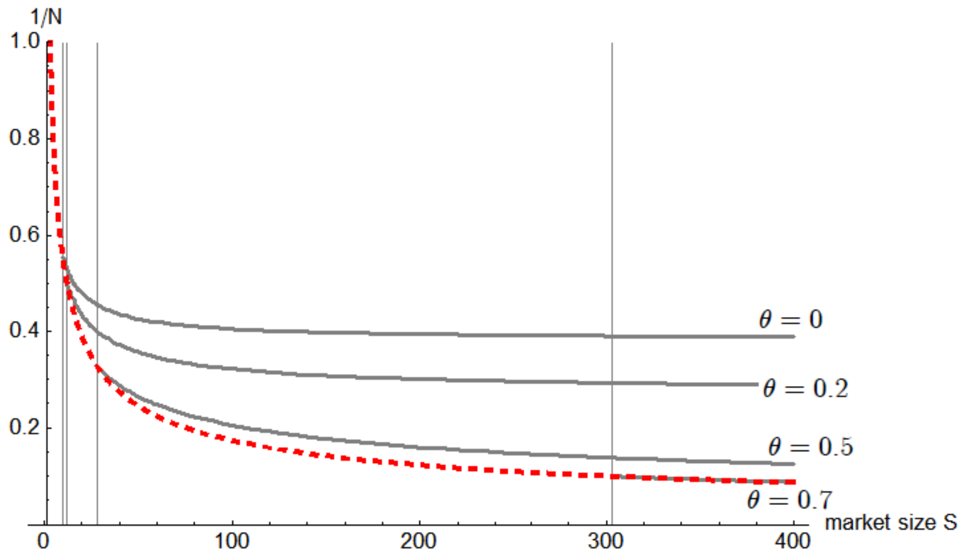

Figure 1.2: The lower bound on the concentration level as a function of market size S, for different spillover parameter θ. The dotted line represents the exogenous sunk costs regime, and vertical lines denote $\hat{S}$ for different θ.

1.3.3 Spillovers and the R&D and profit disincentives

We now demonstrate how the effect of spillovers affects the incentives to invest into endogenous sunk costs. We will express equilibrium expenditures on quality (for different θ) by an individual firm, and by the whole industry, as a function of S. To do that, we use the solution for N from (1.8) and plug it into the expression (1.7) for the endogenous sunk costs of an individual firm. The corresponding figure is below. The higher θ is, the lower is *individual* spending on quality. Thus, for instance, R&D investments in the case of no spillovers ($\theta = 0$), is much higher compared with higher θ values, and for $\theta = 0.7$ endogenous sunk costs investment is close to 1. For low spillovers, the effect of market size S on R&D investment is very significant, so an individual firm's investment into quality F_i increases significantly as market size S grows. However, as spillovers increase, R&D investment remains at the negligible level, and S has a very small or no effect at all on R&D.

The same result holds for the *total* industry investment in quality improvement. Although an increase in spillovers induces entry of new firms, the disincentive effect of increased θ more than offsets it, so the total industry R&D investment falls as well. Thus, we would obtain an analogous graph for the total industry R&D expenditure as that for the firm's individual R&D investment. Therefore, our **testable hypothesis 2** is that the higher knowledge spillovers are, the lower R&D expenditures are, for both an

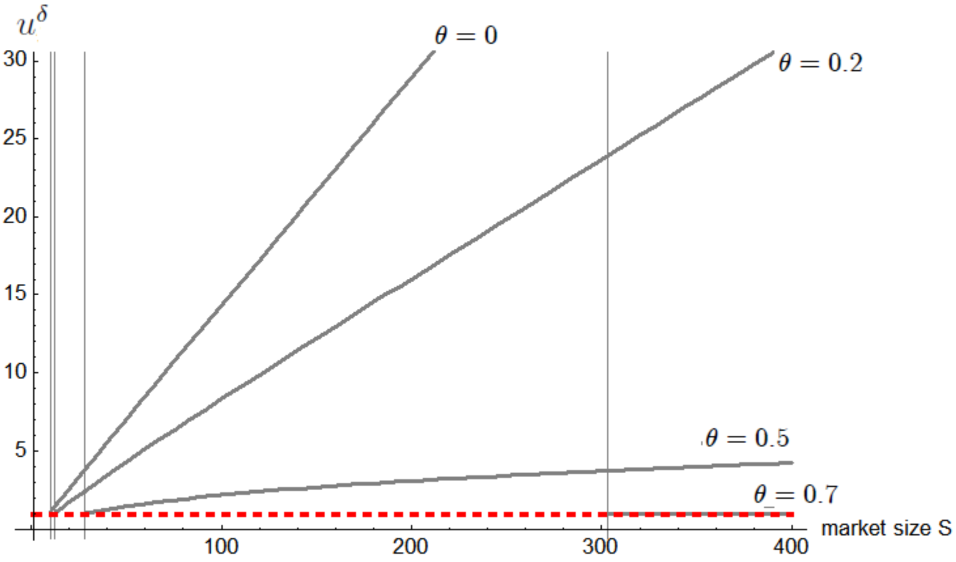

Figure 1.3: A firm's expenditures on quality and market size, for different values of θ. Dotted line represents expenditures in case of exogenous sunk costs regime: $u_i = u_i^* = 1$

individual firm and an industry as a whole, other things being equal. In other words, increasing the size of the market leads to an increase in R&D expenditure, but in the industries with high spillovers this increase in R&D happens at a much lower rate (if at all) than in the industries with low spillovers.

The intuition for the above results, summarized in Figures 1.2 and 1.3, is roughly as follows: the impact of giving away spillovers becomes stronger than its receiving counterpart as the industry spillover parameter rises. Each firm realizes that all other firms will free-ride on its investment, and also it would be optimal to free-ride on the investment of others. Thus the consequence of rising spillovers are decreasing endogenous sunk costs (see Figure 1.3), a larger number of firms entering the industry and, other things being equal, lower market concentration (see Figure 1.2). Once spillovers surpass the threshold of $\overline{\theta} = \frac{2}{2+\delta}$, the disincentives to invest become so strong that the firms reduce their investment in R&D so much that these investment cease to serve as the effective control of entry so that the lower bound of concentrations disappears. That is to say, N^* tends to infinity as market size increases.

Note also the disincentive effect that spillovers exhibit on a firm's profit, which we place in Lemma 1.

Lemma 1. *As R&D spillovers increase, a firm's profits decline. That is,* $\frac{d\Pi_i}{d\theta} = \frac{\partial \Pi_i}{\partial N^*}\frac{dN^*}{d\theta} + \frac{\partial \Pi_i}{\partial \theta} < 0$.

This is in line with the empirical finding by Hanel and St-Pierre (2002) who show that information spillovers negatively affect profits. Interestingly enough, the negative sign here does not come from the direct effect of spillovers since it vanishes ($\frac{\partial \Pi_i}{\partial \theta} = 0$) due to the symmetry in receiving and giving away spillovers. Apparently, the key is in the indirect effect that turns out to be negative. Thus, the equilibrium profit declines in the number of firms while the equilibrium number of firms increases with spillovers due to the mechanism described above (that is, $\frac{\partial \Pi_i}{\partial N^*} < 0$, and $\frac{dN^*}{d\theta} > 0$; see Appendix 1.A.2 for the complete proof).

1.3.4 Endogenous versus exogenous sunk costs regimes: entry and escalation effects

We now aim to study how the interplay among spillovers, market size and free entry affects the firm's outlays on R&D. For that purpose, we decompose the change of endogenous sunk costs (dF_i/dS) into a direct and indirect effect. Thus, $dF_i/dS = (\partial F_i/\partial N) \times (dN^*/dS) + \partial F_i/\partial S$ where the first part $(\partial F_i/\partial N) \times (dN^*/dS)$ stands for "entry effect" while the second part $(\partial F_i/\partial S)$, describes "the escalation effect" [14]. The entry effect is typically negative and tells us what the change would be in the endogenous sunk cost outlays of a firm due to entry of new firms induced by increasing market size[15]. More specifically, the increased size of the market would result in an entry that would in turn negatively affect the investment in R&D, due to the fact that incentives to invest decrease with more firms in the market.

The entry effect in our setup is $\frac{\partial F_i}{\partial N} \times \frac{dN^*}{dS} = \frac{F_0}{S} \times f(N^*(S), \theta, \delta)$, where $f(N^*(S), \theta, \delta)$ is the function of the N^* and the model parameters (see Appendix 1.A.4).

Lemma 2. *The entry effect tends to vanish as the market size goes to infinity and so the escalation effect is the predominant one for the "large" market given that spillovers are low (that, is $0 \leq \theta < \bar{\theta}$).*

As can be seen from the expression (1.A.3), the key determinant of the entry effect is

[14] The "escalation effect" and "entry effect" are derived formally in the Appendix 1.A.4.

[15] It turns out that for very small or zero spillovers this derivative can be positive. As Vives (2008) showed, the entry of new firms has two opposing effects on R&D investment: the direct demand and the indirect price pressure effects that work in opposite directions. The direct demand effect typically dominates the price pressure effect, and R&D decreases with the number of firms. It is possible, however, that the price pressure effect dominates the demand effect so that an increase in the number of firms causes an increase in R&D expenditures (see Appendix 1.A.4 for a more detailed discussion of these points).

the ratio of setup costs to market size, F_0/S, as it defines the "capacity" for additional entry (see Appendix 1.A.4 for more details). When, for instance, F_0/S is low (say, due to large market size) there are already "enough" firms in the market equilibrium so further decreases in this ratio makes room for less and less additional entry. So as S approaches infinity, the ratio F_0/S, goes to zero and at the limit, there is no space for any additional entry because all entry possibilities were exhausted. Thus, the entry effect is of second-order importance for "large markets" (since the ratio F_0/S is then small). Moreover, it would be completely absent when there are no set-up costs, F_0, (that is, when $F_0 = 0$, $dN/dS = 0$).

Recall that the total entry (at the limit) is comprised of either a finite or infinite number of firms, depending on the level of spillovers: for $0 \leq \theta < \bar{\theta}$, there is a finite number of firms when S tends to infinity, while for the spillover such that $\theta > \bar{\theta}$, N goes to infinity as S goes to infinity.

As for the second, escalation effect (unlike the entry effect), it is strictly positive, does not depend on S so it does not vanish at the limit when S tends to infinity, provided that spillovers are low (that, is $0 \leq \theta < \bar{\theta}$).

$$\frac{\partial F_i}{\partial S} = \frac{2(N^* - 1)^2(1 - \theta)}{N^{*3}\delta(1 + (N^* - 1)\theta)} > 0 \tag{1.10}$$

Recall that the escalation effect, (when strong enough!) is at the heart of the non-fragmented market structure and, consequently, a strictly positive lower bound of concentration[16]. This effect, however, monotonically weakens with the rise of spillovers and for spillovers such that $\theta > \bar{\theta}$, an increase in market size leads to an unlimited increase in the number of firms, which, in turn, results in the zero escalation effect at the limit, that is, $\lim_{S\to\infty} \frac{\partial F_i}{\partial S}\Big|_{\theta \geq \bar{\theta}} = 0$ (and, consequently, $\lim_{S\to\infty} \frac{dF_i}{dS}\Big|_{\theta \geq \bar{\theta}} = 0$, given that the entry effect also vanishes at the limit). Finally, at $\tilde{\theta}$ there is a switch to the exogenous sunk cost regime, as we saw, and so the firms cease to invest in R&D ($\frac{\partial F_i}{\partial S} = 0$ and so $dF_i/dS = 0$ for $\theta \in [\tilde{\theta}, 1)$) irrespective of the size of the market.

To conclude, the key factor governing the total change of the sunk costs in the large markets (small F_0/S ratio) is the size of the escalation effect. The non-fragmented market structure appears when spillovers are small (that, is $0 \leq \theta < \bar{\theta}\theta$). Beyond the critical level θ, however, the total effect, dF_i/dS, although positive, becomes "too weak" to hold down

[16] See Etro (2013) for an example where market concentration can even rise with the increase in market size.

the entry of new firms when the market size increases. Thus, as we saw, for $\theta \in [\bar{\theta}, \tilde{\theta})$ there is a kind of hybrid regime: firms do invest in R&D (that is, there are endogenous sunk costs), but, on the other hand, these investments do not escalate when the market size grows. Instead the level of endogenous sunk costs barely changes, but the number of firms increases with the growing size of the market. In other words, there appears, like in the exogenous sunk cost regime, a fragmented market structure whereby market concentration tends to zero. Apparently, spillovers and R&D act as the substitutes: the larger mutual R&D spillovers are the lower the R&D effort needed to achieve the given level of perceived quality, and so firms curb R&D when spillovers rise. Finally, for the spillovers level above $\tilde{\theta}$, $dF_i/dS = 0$, so there are exogenous sunk costs in this case. Note that in the standard, Sutton (1991), setup without spillovers, the exogenous sunk costs regime appears only for a "small" market (when $S < \hat{S}$), while in our setup the exogenous sunk cost regime appears when $\theta > \tilde{\theta}$, irrespective of the size of market.

Note that we could do a similar exercise, and decompose the effect of spillovers on a firm's R&D on its direct and indirect effects, that is, $dF_i/d\theta = (\partial F_i/\partial N) \times (dN^*/d\theta) + \partial F_i/\partial\theta$. Clearly, the direct effect of spillovers is negative due to the prevailing disincentive effect while the indirect effect is also typically negative, given that an increase in spillovers makes entry easier ($dN^*/d\theta > 0$) while the presence of more firms usually induces all firms to restrain their R&D outlays ($\partial F_i/\partial N < 0$) in equilibrium (see Vives, 2008, and Appendix 1.A.4).

1.4 Extended Model: Managing Spillovers

1.4.1 Model setup

Firms need to expect future profits (rents) in order to have incentives to invest in R&D but, as we just noted, increased spillovers have a negative impact on a firm's profit and R&D incentives, so a firm may consider the prevailing "giving away spillovers" to be excessive and may try to curb them. In this light, one typically thinks of patents and copyrights as the means to prevent spillovers and restore the incentives for innovation. Cohen and Levin (1989), however, provide an extensive review of literature on the effectiveness of patenting in different industries, and come to the conclusion that in many industries (machinery, electronics, food processing, etc.) only a negligible share of firms use patents[17]. Instead, firms use other measures to protect R&D investment from spillovers, such as: secrecy, product complexity and ability to learn quickly. As Shenkar (2010) noted "...[L]egal protections have weakened at the same time that codification, standardization, new manufacturing techniques, and growing employee mobility making copying easier". Also, Cohen et al. (2002) demonstrate that secrecy and lead time appropriability mechanisms are more effective than patents in protecting innovations for firms in the USA. Along the same lines, Scotchmer (2006) defines so called private or technical IPR protection as an alternative to legal patents. International trade literature (see, for example Taylor, 1993) refers to physical "masquing" techniques which are used by producers who try to ensure the appropriability of their product innovation.

So we now allow firms to use costly measures to privately protect their R&D investment from giving away spillovers. As we argued in the introduction, one situation when private protection against spillovers may emerge is the case when the adoption of the quality improvements of other firms by firm i is at the edge of IPR violation, and this would especially be the case if the public IPR protection is not possible or, more likely, if it is ineffective (say, due to enforceability problems, high litigation costs, etc.). For instance, in the case when spillovers are realized through reverse-engineering[18], such a costly private protection measure would make the product more complicated to disassemble and copy. Atallah (2004) interprets this prevention of spillovers as any costly activity which enhances the secrecy of the product. If spillovers are realized through the labor force flows between firms, costly private protection measures may mean that companies pay key employees more to prevent them from leaving as, for instance, in Zabojnik (2002), Gersbach and Schmutzler (2003). They interpret the costly prevention of spillovers as extra wages paid to employees so that they do not leave the firm and transfer important information to competitors; and in the case of receiving spillovers - this is the extra wage the firm has to pay to the competitor's workers to be able to hire them.

A somewhat different notion of endogenous spillovers than the one we use here was adopted in the early spillover literature, where endogenous spillovers typically mean that firms deliberately fully or partially share their research output with each other. So firms

[17]Mansfield (1986) shows that patent protection is important almost exclusively for innovations in the pharmaceutical industry: 60% of innovations would not be developed without patent protection; while in other industries (machinery, metal, electrical equipment, instruments, motor vehicles, textiles) only between 0 and 15% of innovations would not be developed without patent protection.

[18]Samuelson and Scotchmer (2002) describe legal issues related to reverse engineering, referring to it as an appropriate and allowed industrial practice, and describe costly measures taken by firms to protect their product from such copying.

cooperate in R&D by setting research joint ventures or research consortia in which they endogenously and cooperatively set both the "giving away" and "receiving" spillovers[19].

Finally, note that unlike in the above literature on cooperation in R&D, the notion of endogenous knowledge spillovers in our context has the meaning of unilaterally (non-cooperatively) curbing the "giving away spillovers".

By decreasing the spillover θ, firm i will also decrease the effective qualities of all other firms, which will in turn have a positive effect on its profits ($\frac{\partial \Pi_i}{\partial u_j^*} < 0$ for all $j \neq i$).

In this section we assume that firms have an option to adopt costly protection against spillovers. Thus, firms are able to restrain the size of spillovers if they find them too large and if this is not too costly to do. For simplicity, we assume that firm i has a choice to decrease spillovers from θ to 0. In this case, the costs would be $F_i = F_0 + \alpha u_i^\delta$, with $\alpha > 1$, where α is a cost shifter that reflects the fact that private protection of quality is costly as compared to costs $F_i = F_0 + u_i^\delta$, when firm i does not prevent spillovers[20]. On the benefit side, if firm i protects its investment from spillovers, its effective quality remains the same (given that no other firm chooses to protect its investment): $u_i^* = u_i + \sum_{j \neq i} \theta u_j$, but the effective quality of all other firms decreases: $u_j^* = u_j + \sum_{k \neq j,\ k \neq i} \theta u_k$, as compared to $u_j^* = u_j + \theta u_i + \sum_{k \neq j, k \neq i} \theta u_k$. We look for the set of parameters which satisfy the conditions for symmetric Nash equilibria, where all firms either simultaneously choose to protect their investment from spillovers, or they do not protect.

The timing of the model is much like in the previous section, with introduction of one further step. In the first stage firms decide whether or not to enter the market. In the second stage, the firms that entered pay sunk entry cost, F_0, and also choose sunk investment in the quality of the product. In the third stage, firms decide whether to protect their investment from spillovers or not. Note that sunk costs investment and protection decisions are taken at different stages (see, for example, Gersbach and Schmutzler, 2003; Atallah, 2004, for similar timing). Such a "sequential" setup implies that firms, while deciding on protection, observe the level of R&D investment of their competitors. Finally, in the last stage, N firms which entered the market simultaneously choose quantities, x_i.

[19]The pioneering article in this sense was Kamien et al. (1992), followed by Poyago-Theotoky (1999), Amir et al. (2003), and Tesoriere (2008). See also De Bondt (1997) for an early survey about the role of spillovers in R&D incentives who, among other things, noted that in reality spillovers are endogenous to a large extent, and possibly interacting with exogenous information leakages.

[20]See Taylor (1993) for the related definition of the cost function of a firm which adopts "masquing" techniques to prevent or make giving away spillovers more difficult.

Such multiple-stage game setting, with this specific timing assumption, is quite restrictive, especially if one would like to use it in more applied research. However, such limitations allow us to derive the market structure completely endogenously, without pre-determining the number of firms competing in the market. If one wishes to relax some of the assumptions (i.e., allow firms to decide on the level of sunk costs investment conditional on the level of protection they select at the same time or before, or allow asymmetry in sunk costs and protection choices), that would require a model where the number of firms (or potential entrants) is exogenously determined at the beginning of the game. For example, a setting with a fixed number of firms in the market and a potential entrant would allow us to model a partially endogenous market structure, but at the same time a setting with a more relaxed assumption would remain tractable.

Proposition 3. *For ex ante spillovers such that $\theta \geq \hat{\theta}(\delta)$, there exists a symmetric "protection" equilibrium. That is, all firms in an industry adopt protection (resulting in zero ex post spillovers) and no single firm has a unilateral incentive to deviate to a "no protection" strategy. If, however, $\theta \leq \hat{\hat{\theta}}(\delta)$ then no single firm has an incentive to unilaterally adopt a protection strategy and thus a "no protection" equilibrium could be sustained as a symmetric equilibrium. Moreover there is no difference between the ex ante and ex post spillovers in this case.*

Proof. First, consider the equilibrium where all firms choose to manage spillovers. Given that all firms have chosen protection (implying that $\theta = 0$), the firms choose the optimal investment level into quality. For this equilibrium to be well defined, the firms, as we know, have to operate in the endogenous sunk costs regime. That is, the market size has to be large enough ($S \geq \hat{S}$) and we assume that this is the case. (For instance, for $\delta = 2, F_0 = 2, \theta = 0$, market size has to be such that $S \geq \hat{S} = 8.5$)[21].

From (1.6), we obtain $\frac{d\Pi_i}{du_i} = \frac{2S(N-1)^2}{N^3 u}$, and $\frac{dF_i}{du_i} = \delta\alpha u^{\delta-1}$. Profit maximization requires that $\frac{d\Pi_i}{du_i} = \frac{dF_i}{du_i}$, and by symmetry assumption, $u^\delta = \frac{2S(N-1)^2}{N^3\delta\ \alpha}$. Much as in the previous section, we have the zero-profit condition,

$$\alpha \frac{2S(N-1)^2}{N^3\delta\alpha} + F_0 = S\left(\frac{1}{N}\right)^2 \tag{1.11}$$

which determines the number of firms that enter the market in the "protection" equilibrium.

[21]For details on deriving $\hat{S}$, see Appendix 1.A.5.

Now, assume that a firm i decides to deviate and stops protecting from spillovers at stage 3. The profit expression for firm i changes to:

$$\Pi_i^D = S\left(1 - \frac{(N-1)}{1 + u_i^* \sum_{j\neq i}(1/u_j^*)}\right)^2$$

where $u_i^* = u_i$, and for all other firms $u_j^* = u_j + \theta u_i$. Now, with this symmetry assumption the profit expression becomes:

$$\Pi_i^D = S\left(\frac{(2-N)\theta + 1}{N + \theta}\right)^2 < S\left(\frac{1}{N}\right)^2$$

The cost expression for a deviant firm i becomes $F_i^D = F_0 + u^\delta = F_0 + \frac{2S(N-1)^2}{N^3\delta\ \alpha}$. Now, if

$$\Pi_i^D - F_i^D = S\left(\frac{(2-N)\theta + 1}{N + \theta}\right)^2 - F_0 - \frac{2S(N-1)^2}{N^3\delta\ \alpha} \leq 0, \tag{1.12}$$

where N is the solution to the implicit equation (1.11), firm i does not have incentives to deviate from symmetric equilibria, where all firms protect against spillovers. This means that for $\theta \geq \hat{\theta}(\delta)$ a symmetric protection equilibrium can be sustained, where $\hat{\theta}(\delta)$ is determined by (1.12) which holds with equality.

To derive $\hat{\hat{\theta}}(\delta)$, consider the equilibrium where none of the firms use protection against spillovers. As in the previous section, (1.7) defines the costs of investment for firm i, and profit is $S\left(\frac{1}{N}\right)^2$. Now, assume that a firm i decides to deviate and starts protecting from spillovers at stage 3. the profit expression for firm i becomes:

$$\Pi_i^D = S\left(1 - \frac{(N-1)}{1 + u_i^* \sum_{j\neq i}(1/u_j^*)}\right)^2$$

where $u_i^* = u_i + \sum_{j\neq i} \theta u_j$, and $u_j^* = u_j + \sum_{k\neq j,\, k\neq i} \theta u_k$. Now, in the symmetric case, $u_i^* = u + \theta(N-1)u$, and $u_j^* = u + (N-2)\theta u$. The profit expression becomes:

$$S\left(\frac{(2N-3)\theta + 1}{N(\theta(N-1)+1) - \theta}\right)^2 > S\left(\frac{1}{N}\right)^2$$

The cost expression for firm i becomes now $F_i^D = F_0 + \alpha u^\delta = F_0 + \alpha\frac{2S(N-1)^2(1-\theta)}{N^3\delta(1+(N-1)\theta)}$. A "no protection" equilibrium exists if:

$$\Pi_i^D - F_i^D = S\left(\frac{(2N-3)\theta+1}{N(\theta(N-1)+1)-\theta}\right)^2 - F_0 - \alpha\frac{2S(N-1)^2(1-\theta)}{N^3\delta(1+(N-1)\theta)} \leq 0, \qquad (1.13)$$

where N is the solution to the implicit zero profit condition (1.8). In order to obtain the critical value of spillovers beneath which this equilibrium exists, we have to equate (1.13) to zero and solve for θ (as an implicit function of R&D effectiveness). This yields $\hat{\hat{\theta}}(\delta)$ implying that for $\theta \leq \hat{\hat{\theta}}(\delta)$ all firms choose a no-protection strategy and this outcome can be sustained as a symmetric equilibrium. In this equilibrium firm i does not have incentives to deviate from symmetric equilibria, in which none of the firms protects against spillovers. □

We demonstrate the solution to inequality (1.12) with a numerical example. For the parameter values ($S = 100, F_0 = 2$), we define such a combination of values of spillover θ and investment cost parameter δ, that (1.12) holds. In the Figure 1.4 below, the dashed area defines the set of parameters δ and θ, for which a "protection" equilibrium will exist. We can see that for high enough spillovers firms will not deviate from protection. There is a critical level of spillovers, $\hat{\theta}(\delta)$, represented by the lower bound of the dashed area in Figure 1.4, above which the protection regime is sustained. On the other hand, the shaded area defines the set of parameters δ and θ, for which (1.13) holds and a "no protection" equilibrium exists (for $S = 100$, $\alpha = \{1.5; 2\}$, $F_0 = 2$). For spillovers

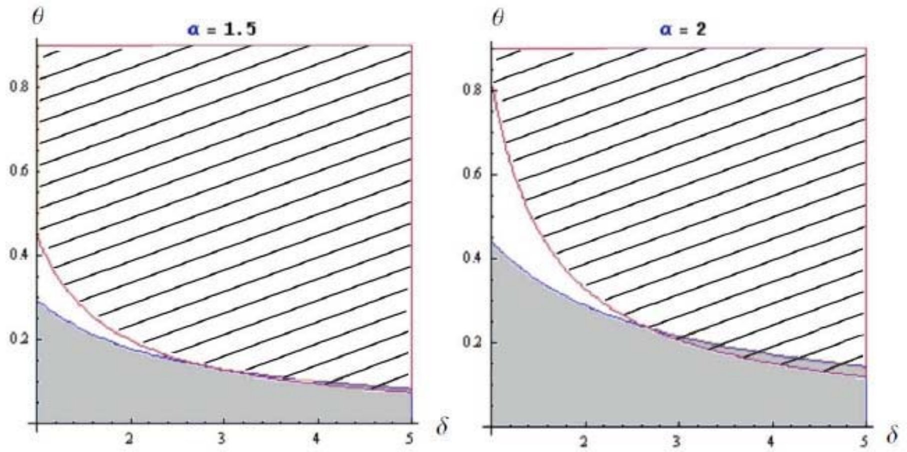

Figure 1.4: "No protection" (shaded, below) and "Protection" (dashed, above) symmetric pure strategy Nash equilibria (parameter values $S = 100, F_0 = 2$)

low enough, firms will not undertake costly protection measures. The reason is that costs of protection (in terms of higher R&D expenditures) do not directly depend on the level of spillovers, but the benefits (in terms of profit gain) do. So if spillovers are low, the benefits of starting protection (or alternatively, the loss of profit because of not protecting) are low compared to incurred costs of protection. Note, further, that the critical values of spillovers depend on the effectiveness of R&D. For high enough values of δ firms would tolerate only very small spillovers and so the range of parameter θ for which a no-protection equilibrium would exist, becomes smaller as δ increases. The reason for that is that the low effectiveness of R&D (high δ), leads to low endogenous sunk costs (see the expression 1.7) and to more firms entering the market. With more firms in the market, the benefit from protection rises, and a firm is willing to undertake it even for small spillovers.

It is insightful to interpret the above story in the context of the interaction between public and private protection. When public protection is "lax enough" (that is, when $\theta \geq \hat{\theta}(\delta)$), it triggers private protection that eliminates the "giving away spillovers". Thus, private and public protection become "complements" to each other once the threshold spillovers level has been reached and the protection equilibrium then occurs. For the ex-ante spillovers that are below a threshold level, firms, however, do not use private protection but rely instead on the (imperfect) public protection that serves as a substitute for the costly private protection, and so there is a no-protection equilibrium outcome. Moreover, the lower the efficiency of R&D investment is (that is, the larger is δ), the easier it is to induce private protection (that is, even relatively strong but not perfect public protection triggers private protection when δ is large).

Much as in the case of protection equilibria, for a "no protection" equilibrium to be well defined, the firms have to invest in R&D (that is, they have to operate in either an endogenous or hybrid regime). Thus, the level of spillovers in the case of a no protection equilibrium has to be beneath the critical level $\tilde{\theta}$ in order to preclude the exogenous sunk cost regime (see Figure 1.4). Moreover, for a "no protection" equilibrium to exist, the necessary condition is that spillovers have to be low enough not to trigger the protection, that is, $\theta < \hat{\theta}(\delta)$[22].

For the values of parameters that are not in the shaded or dashed areas (intermediate θ and low δ values), there is no symmetric equilibrium in pure strategies. For such values of parameters, if all firms protect, there are always incentives for one firm to deviate to "no

[22]Note that for a "no protection" equilibrium to be well defined, $\hat{\theta} < \tilde{\theta}$ has to hold and this is always the case in our setup, if market size S is large enough.

protection". On the other hand, if all firms do not protect, a firm always has incentives to deviate from "no protection" behavior and starts protecting its quality features. Also, there is an area (where dashed and shaded areas intersect) in Figure 1.4 above where both "protection" and "non protection" equilibria exist. That is, if all firms choose to protect their investment from spillovers, each single firm would not prefer to deviate to not protecting; on the other hand, if all firms choose not to protect, each single firm would prefer not to deviate and start protecting.

Also, as α increases from 1.5 to 2, the "no protection" (shaded) area become larger - it is now more costly to protect against spillovers, and a "no protection" equilibrium is more likely, other things being equal. On the contrary, the "protection" (dashed) area shrinks as it becomes more costly to protect.

1.4.2 The lower bound of concentration when spillovers are managed

As Figure 1.4 demonstrates, for different values of θ different equilibria will emerge. In the following analysis we fix cost parameters $\delta = 2$, $\alpha = 2$, $F_0 = 2$, and draw the concentration schedule as a function of market size for different θ (Figure 1.5).

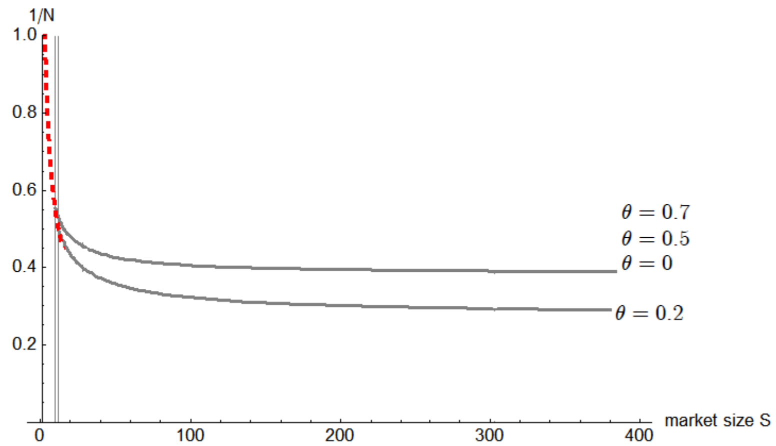

Figure 1.5: Concentration and market size, for different values of θ

For $\theta = 0$ and $\theta = 0.2$, we have "no protection" equilibria, and the concentration versus market size schedule is the same as in Figure 1.2. However, for higher values of θ "protection" equilibria will emerge, meaning that firms will choose to manage spillovers and decrease them to zero. For these cases, the concentration versus market size schedule coincides with the upper curve in Figure 1.2. As a result, we can see that, for high values of spillover parameter θ, the lower bound on concentration remains high and does not decrease to zero if we allow firms to protect from spillovers.

From comparing Figures 1.2 and 1.5, it is clear that the "exogenous sunk costs" regime disappears for large θ, if firms operate in the environment where they can manage spillovers. The reason is that for spillovers larger than $\tilde{\theta}$, and the absence of active protection, firms would operate in an exogenous sunk costs regime and would incur outlays $u = 1$. When, on the other hand, firms could manage spillovers, then such high values of θ would trigger protection, eliminate spillovers, and as a result, firms would choose $u > 1$. This happens because spillover values $\theta > \tilde{\theta}$ which determine the "exogenous" regime fall in the "protection" equilibrium (dashed area on Figure 1.4). That is, for such high θ values firms prefer to manage spillovers and decrease them to zero, and as a result, firms choose $u > 1$.

In the environment where protection against spillovers is feasible, what matters for the choice of endogenous sunk costs is not the level of ex ante spillovers but the level of ex post spillovers that, under our assumption, completely vanish if firms find it optimal to undertake protective measures. Moreover, empirical evidence (see, for example, Mansfield et al., 1981; Mansfield, 1984) demonstrates a strong and positive link between market concentration and the costs of imitation. This finding is consistent with our analysis to the extent to which the high imitation costs reflect the presence and strength of private protection. In other words, the large size of the imitation costs may indicate that the strength of ex-ante spillovers were large enough to trigger private protection and thus make imitation costs high. So the ex-post spillovers are low or zero and that, in turn, results in larger endogenous sunk outlays that hamper the entry of new firms and lead to high market concentration.

This gives us **testable hypothesis 3**, that industries which are characterized by high ex ante spillovers would not become fragmented as the size of the market increases, provided that firms use protection measures against knowledge diffusion. So the positive lower bound of concentration is preserved in this case. In fact, our theoretical prediction is also consistent with the empirical evidence that some industries, which have high spillovers incur substantial sunk costs while, at the same time, displaying a high level of market concentration (even at the global level)[23]. The underlying mechanism is, however,

[23]Aschhoff et al. (2013, see Tables 16, 37-39) extensive description of innovation activities by German firms demonstrates that the top 5 sectors (out of total 20) by R&D intensity are Pharmaceuticals, Electronics, Motor and Vehicles, and Telecommunications. Firms in those sectors much more often than firms in other sectors, use competitors as an important source of information for innovation projects.

different than the one based on a learning and absorption capacity story (as in Cohen and Levinthal, 1989). In our case, the simultaneous existence of (ex ante) spillovers and high R&D lies in the curbing of ex post spillovers rather than in a firm's ability to assimilate and exploit information generated by other firms.

1.4.3 The lower bound of concentration and effectiveness of sunk costs in raising quality

An important insight of the endogenous market structure literature is that a "higher" effectiveness of R&D investment, captured by lower δ in our setup, implies a more concentrated market structure. This happens because when δ is low, firms find it more attractive to deviate upwards in their R&D spending and so the equilibrium level of R&D is higher and the number of active firms is lower. However, introducing endogenous protection from spillovers makes this negative relation between δ and F_i non monotonic.

Proposition 4. *When firms in an industry have an option to manage spillovers, the relationship between the market concentration and effectiveness of R&D may become non-monotonic if there is the switch from no-protection to the protection equilibrium once a "large enough" value of δ has been reached. Consequently, the change in the lower bound of concentration is also non-monotonic in δ.*

In Figure 1.6 below we demonstrate how an equilibrium firm's R&D spending F_i changes as δ increases in the setting with endogenous protection against spillovers. In order to see this, let us first assume that both δ and ex ante spillovers were initially low (that is, $0 < \theta < \bar{\theta}$ while θ is "slightly" above unit). As δ increases, investment into quality is less attractive, and equilibrium sunk costs F_i decrease. However, at some level of δ, firms will start to protect against spillovers, which will make R&D investment into quality once again more attractive, irrespective of the high value of δ. The reason for this is, as indicated above, that decline in cost effectiveness results in lower equilibrium sunk costs that in turn invites entry of new firms and makes protection more attractive. As a result, equilibrium R&D spending F_i "jumps up", producing the discontinuity in the relationship between δ and R&D.

The importance of information from competitors means that firms in those sectors experience high ex ante knowledge spillovers. At the same time, those firms also use secrecy and other informal methods of protecting intellectual property. This would lead to low ex post knowledge spillovers in those sectors, and explains why sectors like Pharmaceuticals, Electronics, Motor and Vehicles, and Telecommunications also have high R&D intensity, despite the possible disincentivising effect of ex ante knowledge spillovers.

In this setting with protection against spillovers, R&D investment does not necessarily decrease in δ. For instance, R&D investment F_i for $\theta = 0.4$ (right) is not lower than for $\theta = 0.2$ (left) for $\delta \geq 1.75$ as would be the case with exogenous spillovers[24].

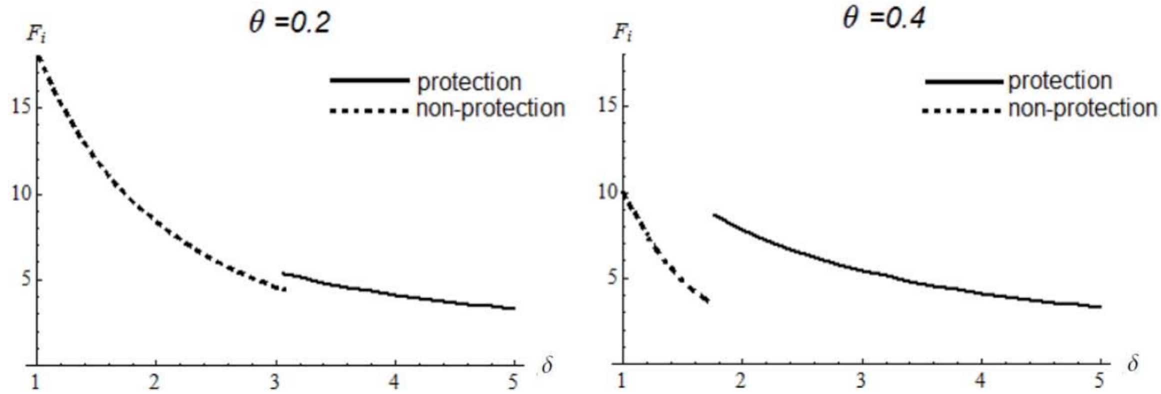

Figure 1.6: Individual firm's sunk costs F_i (R&D investment) as a function of δ for $\theta = 0.2$ (left) and for $\theta = 0.4$ (right)

As for the relationship between δ and equilibrium number of firms (concentration) is also non-continuous and is driven by similar logic. A detailed description of the relationship between equilibrium number of active firms N^* and the cost parameter δ is provided in Appendix 1.A.7, as well as corresponding graphs.

Therefore, our **testable hypothesis 4** relates R&D expenditures (and market concentration) to the effectiveness of those expenditures, in the sense that lower effectiveness of R&D expenditures (higher δ) would not lead to lower endogenous sunk costs if the firm starts to use costly protection against spillovers for higher values of δ. Moreover, if ex ante spillovers are high, firms are more likely to use protection against spillovers even for effective R&D expenditures (low δ). This would lead to even higher R&D expenditures (because of curbed spillovers and low δ), and an even more concentrated market structure.

1.5 Conclusion

We used a simple version of the Sutton (2007) model that illustrates the economic consequences of endogenous sunk costs on firms' entry and market concentration and extended it by allowing for spillovers stemming from firms' investment in product quality. In the

[24] In the setting with exogenous spillovers (Figure 1.3), we showed that the higher is spillover, the lower is R&D investment.

first part of our paper, we assume that spillovers are exogenous to the firms. As expected (in this setup), spillovers tend to decrease the lower bound of market concentration and when strong enough completely eliminate it. Eliminating the lower bound does not, however, imply immediate elimination of a firm incurring endogenous sunk costs nor, consequently, a switch to the exogenous sunk costs regime. We showed that for an intermediate range of spillovers, $\theta \in [\overline{\theta}, \tilde{\theta})$, firms do invest in R&D although the market concentration becomes fragmented as market size grows (that is, there is no positive lower bound). Finally, for very large spillovers, $(\theta > \tilde{\theta})$, firms do not invest in quality improvement due to very strong disincentive effects of spillovers. In other words, potential leakages of information in the industry are so large that firms refrain from R&D so that spillovers do not materialize. The effectiveness of R&D investment plays a key role in determining both threshold levels of spillovers beyond which the lower bound disappears, $\overline{\theta}$, and beyond which the exogenous sunk cost regime appears, $\tilde{\theta}$. The testable hypothesis, given an exogenous spillovers assumption, is that market concentration and its lower bound will be lower when spillovers are higher and the effectiveness of investment in raising quality is lower.

In the second part we allow firms to protect their investment against spillovers, if it would be optimal for them. We focus on symmetric pure strategy Nash equilibria, where all firms either protect their investment or not. As a result, for different values of parameters different equilibria may arise. For low values of θ a "no protection" equilibrium exists while for high enough values of θ and δ a "protection" equilibrium occurs. Contrary to the case of exogenous spillovers, we show that ex ante spillovers may lead to a more concentrated market structure due to the possibility of firms' private protection from spillovers, and this represents another testable hypothesis. It also suggests the related testable hypothesis that a lower effectiveness of raising quality (higher δ) does not lead to lower endogenous sunk costs, and, consequently, to lower market concentration if it triggers firms to use costly protection against spillovers.

We also show that the framework in which firms manage spillovers can be viewed as the situation where both public and private protections are present and that, in turn, enables one to study the economic interaction of the two forms of protection.

Finally, it is worth stressing that our paper is related to three separate topics in industrial organization literature: i) innovation and R&D incentives, ii) market structure and iii) knowledge spillovers. While the related literature usually studies all three notions separately, commonly assuming market structure and R&D spillovers as exogenous parameters, we put forward a theoretical model that simultaneously and endogenously determines the equilibrium values of all three features under considerations.

1.A Appendix 1

1.A.1 Derivation of demand for quality good

Consumers' maximization problem is:

$$\max_{x,z}\ (ux)^{\beta} z^{1-\beta}$$
$$s.t.\ px + p_0 z \leq I$$

With the budget constraint satisfied with equality, $x = \frac{I-p_0z}{p}$ and the utility function becomes $(ux)^{\beta}z^{1-\beta} = \left(u\left(\frac{I-p_0z}{p}\right)\right)^{\beta} z^{1-\beta}$. FOC with respect to z are:

$$FOC(z):\ \beta\left[u\left(\frac{I-p_0z}{p}\right)\right]^{\beta-1} z^{1-\beta}\frac{p_0u}{p} = \left[u\left(\frac{I-p_0z}{p}\right)\right]^{\beta}(1-\beta)z^{-\beta}$$
$$p_0z = (1-\beta)I$$
$$px = \beta I$$

1.A.2 Derivation of $\frac{d\Pi_i}{d\theta}$

Direct effect of θ on Π_i : $\frac{\partial\Pi_i}{\partial\theta}$.

Taking the expression for profit (1.4), it can be seen that θ enters Π_i *directly* only in the expressions for u_i^* and u_j^*. Therefore,

$$\frac{\partial\Pi_i}{\partial\theta} = \frac{\partial\Pi_i}{\partial u_i^*}\frac{du_i^*}{d\theta} + \underbrace{\frac{\partial\Pi_i}{\partial u_j^*}\frac{du_j^*}{d\theta} + \frac{\partial\Pi_i}{\partial u_k^*}\frac{du_k^*}{d\theta} + ...}_{N-1 \text{ of them}},\ j\neq i, k\neq i.$$

From (1.4) we calculate partial derivatives:

$$\frac{\partial\Pi_i}{\partial u_i^*} = 2S\left(1-\frac{N-1}{1+u_i^*\sum_{j\neq i}(1/u_j^*)}\right)\frac{N-1}{\left(1+u_i^*\sum_{j\neq i}(1/u_j^*)\right)^2}\sum_{j\neq i}(1/u_j^*)$$

and

$$\frac{\partial\Pi_i}{\partial u_j^*} = 2S\left(1-\frac{N-1}{1+u_i^*\sum_{j\neq i}(1/u_j^*)}\right)\frac{N-1}{\left(1+u_i^*\sum_{j\neq i}(1/u_j^*)\right)^2}\left(-\frac{u_i^*}{(u_j^*)^2}\right)$$

$\frac{du_i^*}{d\theta} = \sum_{j\neq i}u_j$ and $\frac{du_j^*}{d\theta} = u_i + \sum_{k\neq i,k\neq j}u_k$. With symmetry assumption, $\frac{\partial\Pi_i}{\partial\theta} = 2S\left(\frac{1}{N}\right)\left(\frac{N-1}{N^2}\right)\frac{N-1}{u}\times(N-1)u - 2S\left(\frac{1}{N}\right)\left(\frac{N-1}{N^2}\right)\frac{1}{u}\times(N-1)u\times(N-1) = 0$

Indirect effect of θ on Π_i : $\frac{\partial\Pi_i}{\partial N^*}\frac{dN^*}{d\theta}$.

$\frac{\partial\Pi_i}{\partial N^*} < 0$ from (1.4). In order to find $\frac{dN^*}{d\theta}$, we use zero-profit condition (1.8). It defines N^* from the implicit function $G = F_i(N^*,S) + F_0 - \frac{S}{N^{*2}} = \frac{2S(N-1)^2(1-\theta)}{N^3\delta(1+\theta(N-1))} + F_0 - \frac{S}{N^{*2}} \equiv 0$. From implicit function theorem, $\frac{dN^*}{d\theta} = -\frac{\partial G/\partial\theta}{\partial G/\partial N^*}$.

First, $\partial G/\partial\theta = \dfrac{-2S(N^*-1)^2}{\delta N^{*2}\left(1+\theta(N^*-1)\right)^2} < 0$. Second, $\partial G/\partial N^* > 0$ (See Appendix 1.A.4 for formal derivation of $\partial G/\partial N^*$) and as a result, $dN^*/d\theta > 0$.

Therefore, the indirect effect of spillovers on profit is negative: $\frac{\partial \Pi_i}{\partial N^*}\frac{dN^*}{d\theta} < 0$

1.A.3 Proof of proposition 1: derivation of $N^*_\infty(\theta)$

We can rewrite the zero-profit condition (1.8) as:

$$\frac{F_0}{S} = \frac{N(1+(N-1)\theta)-(2/\delta)(N-1)^2(1-\theta)}{N^3(1+(N-1)\theta)} \tag{1.A.1}$$

Now, if $S\to\infty$, the left part of the expression above decreases to zero. Then the right part will be equal to zero in two cases:

(a) for $\theta \geq \frac{2}{2+\delta}$, only if $N\to\infty$. The denominator of the right hand side expression is a polynomial of degree four (always positive), and the numerator is an always positive polynomial of degree two (for $\theta \geq \frac{2}{2+\delta}$ and $N>0$). Therefore, as $S\to\infty$ and $\frac{F_0}{S}\to 0^+$, the right hand side expression approaches zero only if $N\to\infty$.

(b) for $\theta < \frac{2}{2+\delta}$, the expression in the numerator of (1.A.1) can be equal to zero for a finite value of N. We demonstrate this with the following observation. The numerator of the right hand side of (1.A.1) is a polynomial of degree two, which describes an inverse parabola (for $\theta < \frac{2}{2+\delta}$). At $N=1$, the derivative of the numerator is positive. Therefore, this polynomial represents an increasing (at $N=1$) inverse parabola, which reaches its maximum at $N>1$, then decreases and crosses horizontal axes at $N>1$. This numerator is divided by a positive polynomial of degree four, which guarantees that the crossing point of the right hand side expression with horizontal axes is determined by the numerator. Solving

$$N(1+(N-1)\theta)-(2/\delta)(N-1)^2(1-\theta)=0 \tag{1.A.2}$$

provides value of $N^*_\infty(\theta)$ (the upper bound on the number of firms).

Therefore, $\theta\in[0,\frac{2}{2+\delta})$ represents an endogenous sunk costs regime region of spillovers, where the number of firms is limited by a finite number as $S\to\infty$.

In fact, an alternative and straightforward way to find this upper bound on the number of firms in the endogenous sunk cost regime is to set $F_0=0$ and then calculate the number of firms in the free entry equilibrium that will give us exactly $N^*_\infty(\theta)$.

1.A.4 Derivation of entry and escalation effects and their limits when S tends to infinity

The marginal effect of an increase in the market size on a firm's endogenous sunk costs, $F_i(N^*,S)$ is:

$$\frac{dF_i(N^*,S)}{dS} = \underbrace{\frac{\partial F_i(N^*,S)}{\partial N^*}\frac{dN^*(S)}{dS}}_{\text{"entry effect"}} + \underbrace{\frac{\partial F_i(N^*,S)}{\partial S}}_{\text{"escalation effect"}}$$

1. The *"escalation effect"*:

With $F_i(N^*,S)=\frac{2S(N-1)^2(1-\theta)}{N^3\delta(1+\theta(N-1))}$ we obtain: $\dfrac{\partial F_i(N^*,S)}{\partial S} = \dfrac{2(1-\theta)(N^*-1)^2}{\delta N^{*3}\left(1+\theta(N^*-1)\right)} > 0$

2. The *"entry effect"*:

Using the expression for $F_i(N^*,S)$, and zero profit condition $\frac{F_0}{S} = \frac{N^*\delta(1+(N^*-1)\theta)-2(N^*-1)^2(1-\theta)}{N^{*3}\delta(1+(N^*-1)\theta)}$, we obtain *"entry effect"*: $\dfrac{\partial F_i(N^*,S)}{\partial N^*}\dfrac{dN^*(S)}{dS}$.

The sign of the effect depends on the signs of two parts. First, consider $\dfrac{dN^*(S)}{dS}$. In order to derive explicitly $\dfrac{dN^*(S)}{dS}$, we use a zero-profit condition (1.8). It defines N^* from the implicit function $G = F_i(N^*, S) + F_0 - \frac{S}{N^{*2}} = \frac{2S(N-1)^2(1-\theta)}{N^3\delta(1+\theta(N-1))} + F_0 - \frac{S}{N^{*2}} \equiv 0$.

From the implicit function theorem, $\dfrac{dN^*(S)}{dS} = -\dfrac{\partial G/\partial S}{\partial G/\partial N^*}$;

(a) First, $\partial G/\partial S = \dfrac{2(1-\theta)(N^*-1)^2}{\delta N^{*3}\left(1+\theta(N^*-1)\right)} - \dfrac{1}{N^{*2}} = \dfrac{-2(1-\theta)(N^*-1)^2 + \delta N^*\left(1+\theta(N^*-1)\right)}{-\delta N^{*3}\left(1+\theta(N^*-1)\right)}$. To simplify notation, we denote the numerator of this derivative as $Z = -2(1-\theta)(N^*-1)^2 + \delta N^*\left(1+\theta(N^*-1)\right)$. Using the derivations in Appendix 1.A.3, Z is positive if $\theta > \frac{2}{2+\delta}$. If $\theta < \frac{2}{2+\delta}$, we have that Z describes an inverse parabola in N, which crosses horizontal axis at $N > 1$. This means that, Z is decreasing in N, for $N > 1$. However, note that Z/δ is the same as expression in (1.A.2), which defines the upper bound on the number of firms N^*_∞. As a result, Z is positive for $N \leq N^*_\infty$. Therefore, $\partial G/\partial S < 0$.

(b) Second, $\partial G/\partial N^* = \dfrac{-2S(N^*-1)(1-\theta)(N^*-3+(N^*-1)(2N^*-3)\theta)}{\delta N^{*4}(1+\theta(N^*-1))^2} + \dfrac{2S}{N^{*3}} =$
$= \dfrac{-2S(N^*-1)(1-\theta)(N^*-3+(N^*-1)(2N^*-3)\theta) + 2S\delta N^*(1+\theta(N^*-1))^2}{\delta N^{*4}(1+\theta(N^*-1))^2}$.

Below, we show that the numerator of $\partial G/\partial N^*$ is always positive:

$$-2S(N^*-1)(1-\theta)(N^*-3+(N^*-1)(2N^*-3)\theta) + 2S\delta N^*(1+\theta(N^*-1))^2 > 0$$

Dividing by $-2S$:

$$(N^*-1)(1-\theta)(N^*-3+(N^*-1)(2N^*-3)\theta) - \delta N^*(1+\theta(N^*-1))^2 < 0$$

Further, we substitute $2(1-\theta)(N^*-1)^2$ instead of $\delta N^*(1+\theta(N^*-1))$, using expression (1.A.2):

$$(N^*-1)(1-\theta)(N^*-3+(N^*-1)(2N^*-3)\theta) - 2(1-\theta)(N^*-1)^2(1+\theta(N^*-1)) < 0$$

Dividing by $(N^*-1)(1-\theta)$:

$$(N^*-3+(N^*-1)(2N^*-3)\theta) - 2(N^*-1)(1+\theta(N^*-1)) < 0$$

$$N^*-3+(N^*-1)(2N^*-3)\theta < 2(N^*-1)(1+\theta(N^*-1))$$

Opening the brackets:

$$N^*-3+2N^{*2}\theta-5N^*\theta+3\theta < -2+2N^*+2\theta-4N^*\theta+2N^{*2}\theta$$

and simplifying, we obtain

$$-1+\theta < N^*(1+\theta)$$

which always holds. Therefore, $\partial G/\partial N^* > 0$.

As a result, we have that $dN^*/dS > 0$.

Now, we consider the first part of "entry effect": $\dfrac{\partial F_i(N^*, S)}{\partial N^*}$. Typically, as the number of firms increases, each firm has fewer incentives to invest into quality: $\dfrac{\partial F_i(N^*, S)}{\partial N^*} < 0$. In our setup this is also the case unless spillovers are zero or rather small.
$\dfrac{\partial F_i(N^*, S)}{\partial N^*} = \dfrac{-2S(N^*-1)(1-\theta)(N^*-3+(N^*-1)(2N^*-3)\theta)}{\delta N^{*4}(1+\theta(N^*-1))^2}$. The sign of this derivative depends on the sign of $N^*-3+(N^*-1)(2N^*-3)\theta$. For $\theta > \theta_c$, where $\theta_c = \frac{3-N^*}{(N^*-1)(2N^*-3)}$, we have $N^*-3+(N^*-1)(2N^*-3)\theta > 0$ and $\dfrac{\partial F_i(N^*, S)}{\partial N^*} < 0$. On the other hand, for $\theta < \theta_c$, we have

$\dfrac{\partial F_i(N^*,S)}{\partial N^*} > 0$

Vives (2008) assigns the general ambiguity of the sign of $\dfrac{\partial F_i(N^*,S)}{\partial N^*}$ to the two opposing effects: i) the direct demand (or size) effect, and ii) the indirect (or elasticity) price pressure effects. The direct demand effect predicts that, for a given market size, if more firms enter, the residual demand of a firm will decline, and a firm has fewer incentives to invest in R&D. The elasticity of residual demand, however, will increase, and this will have a positive effect of the R&D incentives because with higher elasticity of demand (that a firm faces) it is optimal for a firm to expand output and that, in turn, makes the investment in R&D more attractive. The latter describes the second, indirect price pressure (elasticity) effect. Note that the direct demand effect typically dominates the price pressure effect, and R&D decreases with the number of firms. However, as we have just seen, it is possible that under certain circumstances (zero or very small spillovers in our case) it is the other way around (see Vives, 2008, for a comprehensive discussion on this point).

As already noted, this *"entry effect"* $\frac{\partial F_i(N^*,S)}{\partial N^*} \times \frac{dN^*}{dS}$ is of second order importance for a large market. Thus, the direct *"escalation effect"* $\frac{\partial F_i(N^*,S)}{\partial S}$ dominates, and governs the effect of S on endogenous sunk costs. In our setup, the *"entry effect"* turns out to be:

$$\frac{\partial F_i(N^*,S)}{\partial N^*} \times \frac{dN^*}{dS} = \\ = \frac{F_0}{S} \frac{(N^*-1)(\theta-1)(N^*-3+1(N^*-1)(2N^*-3)\theta)}{(N^*(4-N^*+\delta)+6\theta-3-2N^*(6+N^{*2}+\delta-N^*(4+\delta))\theta+(N^*-1)^2(N^*(2+\delta)-3)\theta^2} \tag{1.A.3}$$

Finally, let us consider what happens to the *"escalation effect"* and *"entry effect"* as S approaches infinity.

First, it is straightforward to see from the expression above that the limit of *"entry effect"* $\frac{\partial F_i(N^*,S)}{\partial N^*} \times \frac{dN^*}{dS}$ is zero. This holds irrespective of the ratio F_0/S, provided that we are not in exogenous sunk costs regime region[25].

The limit of the *"escalation effect"*, on the other hand, does not approach zero for $\theta < \bar{\theta}$: $\lim_{S\to\infty,\ \theta<\bar{\theta}} \dfrac{\partial F_i(N^*,S)}{\partial S} = \lim_{S\to\infty,\ \theta<\bar{\theta}} \dfrac{2(1-\theta)(N^*-1)^2}{\delta N^{*3}(1+\theta(N^*-1))} = \dfrac{2(1-\theta)(N^*_\infty-1)^2}{\delta N^{*3}_\infty(1+\theta(N^*_\infty-1))} > 0$. However, for $\theta \geq \bar{\theta}$, with S approaching infinity we also have that N^* approaches infinity, so $\lim_{S\to\infty,\ \theta\geq\bar{\theta}} \dfrac{\partial F_i(N^*,S)}{\partial S} = \lim_{S\to\infty,\ \theta\geq\bar{\theta}} \dfrac{2(1-\theta)(N^*-1)^2}{\delta N^{*3}(1+\theta(N^*-1))} = 0$.

1.A.5 Exogenous sunk costs region

As already mentioned, u is defined on the domain $[1,\infty)$, which means that for some range of $S < \hat{S}$ firms will be choosing $u=1$. In order to determine $\hat{S}$, we first have to derive the equilibrium number of firms for $u=1$. Zero-profit condition becomes: $1^\delta + F_0 \equiv S/(N^*)^2 = \Pi_i$. Then, $N^* = \sqrt{\frac{S}{1+F_0}}$ if $u=1$. Now, we have to find maximum S, for which $\left.\frac{d\Pi_i}{du_i}\right|_{u=1} \leq \left.\frac{dF_i}{du_i}\right|_{u=1}$. Substituting $N^* = \sqrt{\frac{S}{1+F_0}}$ and $u=1$ in the expressions $\frac{d\Pi_i}{du_i}$ and $\frac{dF_i}{du_i}$, we obtain:

$\dfrac{2S(1-\theta)\left(\sqrt{\frac{S}{1+F_0}}-1\right)^2}{\left(\sqrt{\frac{S}{1+F_0}}\right)^{3/2}\left(1+\theta\left(\sqrt{\frac{S}{1+F_0}}-1\right)\right)} \leq \delta$. Solving this expression for S, we obtain $\hat{S}$

Numerical computations demonstrate that for $\theta=0.2$, $F_0=2$ and $\delta=2$, we obtain $\hat{S}=12$. This means that for the market size below $S=12$, firms are choosing a "basic" quality level $u=1$, and the number N^* of firms in the market is $\sqrt{\frac{S}{1+F_0}}$. Similarly, for $\theta=0.5$ we have that $\hat{S}=27$, and for $\theta=0.7$ we obtain $\hat{S}=303$. But for $\theta=0.9$ there is no threshold $\hat{S}$. This means that for such high spillovers, for any market size S, firms choose the "basic" level of quality $u=1$, and there are no endogenous sunk

[25] Clearly, decomposition on *"entry"* and *"escalation effect"* makes sense only when we have endogenous or hybrid regime.

costs. We define the region $S < \hat{S}$, where only exogenous sunk costs play a role, as the "exogenous sunk costs region". The dotted line in Figure 1.2 demonstrates the "exogenous sunk costs region" for different θ.

1.A.6 Derivation of $\tilde{\theta}$

$\theta > \tilde{\theta}$ defines a region where spillovers are so high that it does not pay-off to have positive expenditures on quality, and firms stick the to basic quality level $u_i = 1$.

In order to determine $\tilde{\theta}$, we first have to derive the equilibrium number of firms for $u = 1$. From zero-profit condition: $1^\delta + F_0 \equiv S/(N^*)^2 = \Pi_i$, we have $N^* = \sqrt{\frac{S}{1+F_0}}$ if $u = 1$. Now, substituting $u = 1$ and $N^* = \sqrt{\frac{S}{1+F_0}}$ into $\left.\frac{d\Pi_i}{du_i}\right|_{u=1} \leq \left.\frac{dF_i}{du_i}\right|_{u=1}$, we find that this condition holds for $\theta > \tilde{\theta}$, where

$$\tilde{\theta} = \frac{2F_0^2 + 2S + 2 - \sqrt{F_0+1}\sqrt{S}(\delta+4) + F_0(2S - 4\sqrt{F_0+1}\sqrt{S} + 4)}{2F_0^2 + 2S + 2 - \sqrt{F_0+1}\sqrt{S}(\delta+4) + F_0(2S - 4\sqrt{F_0+1}\sqrt{S} + 4) + S\delta} \tag{1.A.4}$$

Therefore, as spillovers are high, $\theta > \tilde{\theta}$, firms are choosing the "basic" level of quality $u = 1$, and we are in the "exogenous sunk costs regime".

In general, the higher market size S is, the higher has to be the spillover threshold $\tilde{\theta}$ for which the market switches to an exogenous sunk costs regime: $\frac{d\tilde{\theta}}{dS} > 0$. However, it is easy to notice that $\tilde{\theta}$ is always below 1. For S approaching infinity, we obtain $\lim_{S\to\infty} \tilde{\theta} = \frac{2(1+F_0)}{2(1+F_0)+\delta}$. This means that for any S, if spillovers are $\theta > \frac{2(1+F_0)}{2(1+F_0)+\delta}$, we obtain an exogenous costs regime. This limit of $\tilde{\theta}$ is decreasing in δ: if increasing quality is very costly, the market will be characterized by an exogenous sunk costs regime even for smaller spillovers. On the other hand, this limit of $\tilde{\theta}$ is increasing in F_0: if fixed entry costs are high, the market will be characterized by an exogenous sunk costs regime only for very high spillovers.

1.A.7 Comparative statics with respect to δ for the number of active firms in the market N^*

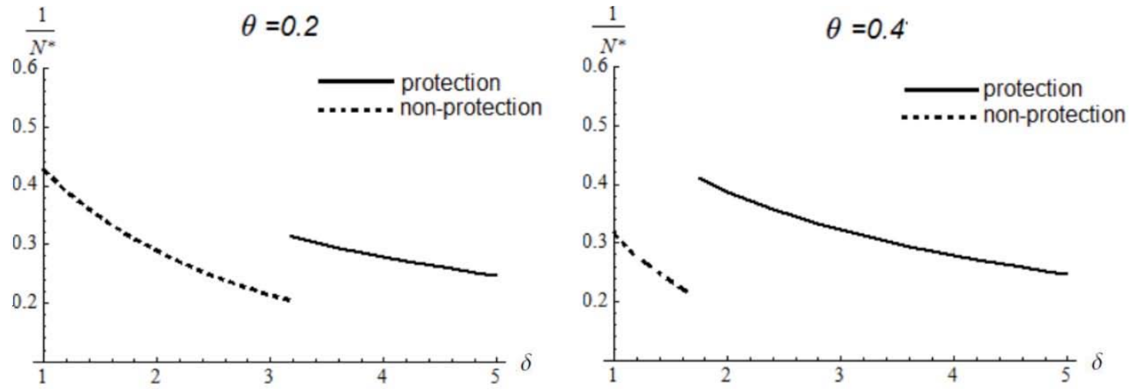

Figure A1: Equilibrium market concentration as a function of δ for $\theta = 0.2$ (left) and for $\theta = 0.4$ (right)

From Figure A1 below, we can split the range of δ values roughly into three parts:

(1) for low δ, when the investment in R&D is more effective, industries with *higher spillovers* θ will have a higher number of active firms in the market N^* and *lower concentration* (dashed line in Figure A1 (right) is below the dashed line in Figure A1 (left) for $\delta \leq 1.75$);

(2) for intermediate values of δ, industries with *higher spillovers* θ will now have a lower number of active firms in the market N^* and the *higher concentration* (solid line in the Figure A1 (right) is above the dashed line in Figure A1 (left) for $\delta \in [1.75, 3]$). This happens because, for higher spillovers θ, firms are already using protection against spillovers, which makes the investment into quality more attractive, increases endogenous sunk costs, and prevents entry;

(3) for high values of δ *the number of firms is the same* for all values of spillovers θ (solid lines in Figure A1 (left) and (right) coincide for $\delta \geq 3$). This happens because, for high enough δ, for any values of spillovers firms find it beneficial to protect against spillovers in equilibrium, which limits effective spillovers to zero and makes equilibrium number of firms equal for all exogenous values of θ.

Chapter 2

Endogenous Sunk Costs, Spillovers and R&D incentives: An Empirical Analysis

2.1 Introduction

In this Chapter, I focus on the role of knowledge spillovers in markets with endogenous sunk costs, and empirically investigate how these spillovers affect the link between market size and market concentration. One of the key predictions in the endogenous sunk costs literature is that an increase in the market size leads, ceteris paribus, to the escalation of endogenous sunk costs (e.g., R&D expenditures) that, in turn, makes entry of new firms difficult or even absent (or negative). Thus, market concentration may not decrease with an increase in market size, but may actually remain unchanged or even increase. In other words, the relation between the market concentration and size of the market is not monotonic in such markets[1]. However, in Chapter 1 of the thesis, we show that the presence of high knowledge spillovers significantly increases the probability that the market will be fragmented, regardless whether endogenous sunk costs are important in the market. In Chapter 2 I provide new empirical evidence of the relationship between R&D incentives and knowledge spillovers, based on firm level survey data. The results contribute to the debate on the disincentivising effect of knowledge spillovers on the incentives of individual firms to invest in R&D. I am able to demonstrate consistently negative (disincentivising) effect of knowledge spillovers on R&D expenditures for different empirical specifications.

[1] On the other hand, in the markets where only the exogenous sunk costs matter, there is a clear positive monotonic relationship between fragmentation (or concentration) and the market size.

The outline of this Chapter is as follows. Subsection 2.2 gives an overview of the empirical literature which tests endogenous sunk costs model by Sutton (1991), and empirical literature which addresses the role of knowledge spillovers related to incentives to invest in R&D. Subsection 2.3 presents the hypotheses to be tested. Subsection 2.4 describes the empirical model. Subsection 2.5 describes the data and Subsection 2.6 presents the estimation results. Subsection 2.7 concludes.

2.2 Related Literature

This empirical research is related to two broad issues and to two strands of the literature. The first issue stems from the theory of endogenous sunk costs (in the form of R&D investments), and their relationship to firm size, market structure and concentration. The second topic deals with the issue of knowledge spillovers, appropriability and R&D incentives. In what follows, I will describe the current stand of the literature on both topics, and how this paper is related to each of them. This paper makes an attempt to build a bridge between these two most broad issues studied in the empirical literature related to R&D incentives.

Among many empirical papers which test the predictions of the endogenous sunk costs theory, my paper is the first to estimate the effect of knowledge spillovers in the setup of endogenous sunk costs theory, and the effect of those spillovers on market size and fragmentation relationship. This paper does not provide a direct test of the endogenous sunk costs theory. Rather, by establishing and testing the role of knowledge spillovers in the endogenous sunk costs model, I am able to show that basic predictions of the endogenous sunk costs theory cannot be refuted by the data.

2.2.1 R&D, firm size and market structure

Empirical and theoretical literature in this field originates from Schumpeterian hypothesis that firm size and monopolistic position in the market are detrimental to a firms' incentives to innovate. Many years of empirical research established some important robust findings (Cohen, 2010): (a) the effect of firm size on R&D expenditures is positive and, to a large extent, proportional; and (b) there is no strong evidence that market concentration is related to R&D incentives, and compared to the industry effects (which explain more than 30% of the variation in R&D intensity), market concentration might explain only a negligible percentage of the variation in R&D intensity. The lack of conclusive evidence on the relationship between R&D and market concentration should not not be surprising, as the relationship most likely runs both ways. More importantly, innovation, market entry, and market concentration are co-determined by such factors as unobserved technological opportunity and innovation appropriability, as shown in Geroski (1991, 1994).

It was the theoretical setup of Sutton (1991; 2001) that allowed for the empirical research to depart from the limitations of the Schumpeterian framework. First of all, Sutton's endogenous sunk costs framework adopts a "bound approach", predicting the set of possible market structure outcomes, not one-to-one relationship between concentration and innovation. Second, and more importantly, Sutton (1991; 2001) determines a set of core, exogenous factors, which simultaneously drive R&D expenditures and market structure in R&D intensive industries: demand, homogeneity of submarkets (product substitutability) and technological opportunity.

Below I summarize some general predictions of the endogenous sunk costs theory with relationship to R&D incentives, concentration and market size. The theory differentiates sharply between exogenous and endogenous sunk costs markets ("Type I" and "Type II" markets respectively, following Schmalensee (1992)). In the Type I markets, market size does not affect sunk costs. Therefore, as the market grows more firms enter the market, and it is likely to become more fragmented. On the other hand, an increase in the market size for the endogenous sunk costs markets (Type II markets) leads to an increase in the R&D expenditures (or other sunk costs) by the market incumbents, which creates barriers to entry for other firms, and the market is more likely to remain concentrated.

Thus, the predictions about the market size and market concentration are: i) an increase in the market size in the exogenous sunk costs markets does not influence R&D investment, but leads to new firm entry and decreased market concentration; ii) in the endogenous sunk costs markets an increase in the market size leads to an increase in R&D investment, and consequently limits or prevents the entry of new firms.

In the literature review section 1.2 of Chapter 1 we have described several papers which empirically test and confirm the predictions of endogenous sunk costs theory for different markets: Dick (2007) (banking), Matraves (1999) (pharmaceutical), Bronnenberg et al. (2005) (consumer package goods), Ellickson (2007) (supermarket chains), Berry and Waldfogel (2010) (newspaper and restaurant), Marin and Siotis (2007) (chemical industry). The empirical approach of this chapter does not focus on a single market or industry, as do the studies mentioned above. Instead, I test the predictions of endogenous sunk costs theory using the cross-sectional survey data of German firms, where every observation represents a different firm i, and every observation characterizes the market conditions of the main product of firm i (fragmentation, competitive pressure, etc). Therefore, this empirical approach treats every observation as a separate market. This type of aggregate studies pool observations across industries and have the advantage of using larger number of observations compared to industry-level studies. The limitation of such an approach, however, comes from the fact that the effect of market size on R&D is assumed to be the same across industries and markets.

Most related to my approach is a paper by Coscollá-Girona et al. (2011), which tests an escalating effect of market size on R&D expenditures. The authors use panel data of Spanish manufacturing firms (1990-2006), and the empirical evidence shows that the market size variables positively and significantly influence the incentives to carry out product and process innovations.

2.2.2 "Two faces of R&D"

In Chapter 1 of this thesis we added the knowledge spillovers into the simple endogenous sunk cost model by Sutton (2001) and analyzed how the presence of spillovers affects the market structure and concentration as well as a firm's incentives to invest in R&D and its action to prevent information leakage ("giving away spillovers"). We distinguish between the ex ante and ex post spillovers in the sense that the former represents potential information leakages while the latter takes into account a firm's possible reaction to and prevention of such leakages.

Innovation appropriability and knowledge spillovers should naturally form the central questions in any empirical research on R&D incentives. As correctly noted by Cohen (2010), innovation itself is embedded in a product, process or service, therefore, is readily available for all competitors to imitate and reproduce.

There is substantial empirical evidence of a quick leak of information and knowledge in many industries, and on the sources of such information leakage. In the literature review section of Chapter 1 we have already mentioned several sources of such information leakage. Below I continue to describe some empirical evidence on knowledge spillovers in more detail. For example, Caballero and Jaffe (1993), and Henderson and Cockburn (1996) use firm-level data and find significant knowledge spillovers in several industries. Audretsch and Feldman (1996b) interpret R&D spillovers as knowledge externalities which arise from clustered geographical location of firms. According to the empirical model presented in their paper, innovations tend to cluster in geographical space, even after the model accounts for the clustered location of production units (see also Audretsch and Feldman, 1996a). Depending on the size of those spillovers, for some industries clustering innovations in geographical space is more beneficial than it is for others. Ellison et al. (2010) attempt to answer the question of what drives the geographical concentration of industries. The authors find, among other things, that coagglomeration patterns in the manufacturing industries facilitates intellectual spillovers ("ideas").This provides evidence that mutual knowledge spillovers are present in the industries, and producers take this fact into account (see also Shenkar, 2010).

Another example of how spillovers might be realized is through input suppliers. Consider the close relationship between an innovating firm and its suppliers. Such a vertical relationship may result in those suppliers becoming more qualified and hence more attractive as partners to an innovating firm's competitors, potentially enabling these competitors to free ride on the R&D investments made by the innovating firm (Mesquita et al., 2008). In other words, all partners of the supplier may benefit from the supplier's learning in relation to a specific firm that initially invests in the improved product quality (due to specialized inputs requirements, for example). It is reasonable to expect that some (if not the majority) of those partners would be competitors in the final product market. Although the risk of such knowledge dispersion can be reduced by exclusive partnership arrangements, this may not be sufficient to completely prevent knowledge spillovers to competitors. Therefore, this mechanism describes the "vertical channel" of knowledge dispersion between firms (Javorcik, 2004) that could later on evolve in the horizontally linked spillovers where each firm benefits from the spillovers of other firms.

Empirical studies in Schumpeterian tradition, which I analyze above, conjecture that firms with an ex post oligopolistic power in the market have more incentives to invest in innovations. Therefore, it was natural for Spence (1984) to conjecture that knowledge spillovers might have a disincentivising effect on R&D investment: his model demonstrates the trade-off between the spillovers decreasing individual firm incentives for R&D and increasing overall productivity in the industry. This is a very intuitive result, but existing and very rich empirical evidence is inconclusive: while Hanel and St-Pierre (2002) show that information spillovers negatively affect profits, Cohen et al. (1987) find that appropriability can have positive or negative effect on R&D intensity, depending on the industry. Those industries which are most likely to suffer from knowledge spillovers (pharmaceutical, IT technologies, etc.) are also among the industries which invest most in R&D.

Cohen and Levinthal (1989) offer an "absorptive capacity" explanation for this inconclusive empirical evidence. They conjecture that R&D investment not only generates new knowledge, but also increases a firm's ability to accept and assimilate information generated by other firms. The authors refer to this concept as "the two faces of R&D". In their theoretical model higher knowledge spillovers generate higher incentives for firms to engage in R&D investment.

Both approaches to the analysis of spillover effects on R&D incentives have their merit, but later theoretical results are inconclusive on the net effect of knowledge spillovers and R&D incentives. For example, De Bondt (1997) generalizes different theoretical approaches and concludes that different models often provide opposite results. In general, existing models[2] agree that spillovers increase the general level of productivity in an industry (market, sector, economy), but disagree about the effect of spillovers on the incentives of individual firms to innovate. In addition, knowledge spillovers increase the incentives to cooperate in R&D investments. In such situations, the empirical evidence could shed some light on which effect of spillovers on R&D is stronger: positive knowledge dispersion and productive creativity; or disincentivising effects, including the motivation to free ride on the investments of others and thereby save money. However, empirical literature in general lacks reliable data (measures of spillovers) and methodological approaches to estimate spillover effect on R&D, and also provides inconclusive results. Below I will argue that these inconclusive results are to a large extent driven by the lack of proper proxies for knowledge spillovers and by a possible two-way causality between R&D investment and some measures of knowledge spillovers used in the empirical research.

Usually the measure of information flows (patents citations) or the accumulated stock of knowledge (number of patents) are used as a measure of spillovers. However, as noted by De Bondt (1997), knowledge spillovers are to a large extent endogenous, determined by the specific appropriability mechanisms prevailing in the industry, and those (endogenous) knowledge spillovers to some extent interact with the (exogenous) information leakages and channels. Therefore, the biggest problem of the empirical approach is to find correct measures for spillovers, controlling for the possible appropriability instruments used by

[2]Models include endogenous growth literature, contestable inventions models, cooperation in R&D and consortia, and models with asymmetric R&D strategies (leaders and followers in inventions).

firms.

In what follows, I will describe some of the empirical approaches used, and the conclusions which were derived about R&D and spillover relationship. Survey-based approaches are most interesting, using similar data structure to that in this Chapter. Levin's (1988) paper uses survey data of 650 US managers about R&D investments in their companies and knowledge spillovers. As the measure of spillovers, the author uses the effectiveness of various "learning methods" used by the firms: hiring the employees of innovating rivals, personal communication with rivals' employees, reverse engineering, their own R&D investments, etc. The author claims that if the manager responds indicating that these "learning methods" are relatively effective, then the spillovers are also high in that environment. Results obtained by Levin are as follows: the higher the knowledge spillovers (as defined in the paper), the higher is the R&D intensity[3]. No disincentivising effect of spillovers was demonstrated. In order to explain why these results ware obtained, this paper further analyses R&D intensity and spillover measures across industries. For example, computer and electronics industries have higher R&D intensity and also higher spillovers, compared to other less research intensive industries. Levin attributes this to specific technologies used in each industry, and to the fact that it is hard to measure the effect of spillovers in such a heterogeneous pool of industries, where R&D and spillovers play very different roles.

Cohen and Levinthal (1989) test their model with an empirical survey data (1975-1977, almost 2,000 business units in the U.S. manufacturing sector), and are able to demonstrate that there is a positive direct effect of spillovers on R&D incentives, which supports their theoretical conclusion that a firm's own investment in innovation helps to absorb the innovation developed by others. On the other hand, higher spillovers are associated with lower innovation appropriability, and lower appropriability leads to lower R&D incentives. Overall, the net effect of spillovers on R&D proves to be negative. Cohen and Levinthal (1990) use the same data, but apply more rigorous estimation techniques. Their results also suggest that there is a positive absorption incentive associated with spillovers, and it could be strong enough to overcome the negative appropriability effect of knowledge spillovers on R&D incentives.

Similarly, a paper by Cohen and Walsh (2000), using a Carnegie Mellon Survey on Industrial R&D, estimates the effect of spillovers on R&D. As a measure of spillovers the paper uses the importance of information from competitors in developing a firm's own

[3]R&D investment divided by total sales.

innovations. The authors correctly note that the measure of spillovers is not exogenous in their data, because it is highly influenced by the appropriability mechanisms used by the firms. Again, knowledge spillovers have two offsetting effects on a firm's investment in R&D. On the one hand, high spillovers decrease appropriability and decrease incentives to invest in R&D. On the other hand, spillovers will influence R&D incentives directly, either positively (if information from competitors is complementary to a firm's own R&D), or negatively (if information from competitors is a substitute for a firm's own R&D). Empirical evidence in Cohen and Walsh (2000) demonstrates a strong complementarity between a firm's own R&D and information from competitors.

An opposite effect is demonstrated in Kim and Marschke (2005), who build the model of labor flows, patenting incentives, and R&D intensity. They are able to show that higher (research) labor mobility increases incentives to patent innovations created by firms, and decreases their R&D incentives. The authors find empirical evidence that supports their findings using a cross-sectional firm level dataset and demonstrate that higher labor mobility is associated with lower R&D intensity.

As mentioned above, empirical research of knowledge spillovers and R&D incentives requires reliable data on spillovers prevailing in the industry or in specific firm environment. As noted, using patent data as a measure for spillovers suffers from endogeneity problem[4]. In addition, Mansfield (1986) and Cohen et al. (2000) show that only a small number of inventors rely on patents, and other (more informal) mechanisms of protection are considered more effective in achieving appropriability. Interestingly, Cohen et al. (2000) point out that protection mechanisms are not mutually exclusive and are often used together. Different effectiveness of those mechanisms (as reported in surveys) simply means that firms are using different protective strategies, making some mechanisms their central choice. I conclude that knowledge spillovers are to a large extent firm-specific rather than industry-specific, and the whole set of protective measures has to be used to correctly asses the appropriability conditions in which a firm operates.

[4]Cohen et al. (2000) compute correlations of effectiveness of different protective measures (patents, secrecy, lead time) with firm size. They show that only patent effectiveness is positively correlated with firm size, suggesting that large firms believe that patents are more effective. As large firms are also more likely to spend more on R&D, this implies the knowledge spillovers are endogenous, if measured by patent pool or patent effectiveness as reported by firms.

2.3 Testable Hypotheses

The key point of the expression (1.8) is the relationship between the market size S and the equilibrium number of firms N. For the large enough market size S and low spillovers ($\theta^r < \frac{2}{2+\delta}$) there is a finite number of firms that satisfies (1.8). In other words, an increase in S in this case has no effect on the equilibrium number of firms but affects only endogenous sunk costs that rise with S (see Senyuta and Žigić (2012) for formal derivation of this result). When, however, ex post spillovers are high ($\theta^r > \frac{2}{2+\delta}$) or firms do not incur endogenous sunk costs (that is, $u^\delta = 1$), then increase in market size leads to an increase in the number of firms and to market fragmentation (low concentration) with small or no increase in endogenous sunk costs.

The above considerations lead to the first testable hypothesis:

Testable hypothesis 1: markets/industries where endogenous sunk costs are important (Type II markets) and ex post spillovers θ^r are low, would remain concentrated as the size of the market S increases. On the other hand, when θ^r is high or endogenous sunk costs are not important (Type I markets), market fragmentation increases as market size S increases. It is important to note here that this hypothesis differs from the original empirical strategy used by Sutton (2007), which defines predictions of endogenous sunk costs theory in terms of bounds (for example, other empirical papers using a "bound approach" are Ellickson (2007) and Robinson and Chiang (1996)). Instead, hypothesis 1 tests the non-fragmentation result of endogenous sunk costs theory, but cannot be considered a direct test of Sutton's model. Data limitation is the main reason the "bound approach" cannot be employed in estimation. Subsection 2.6 describes in more detail the limitations of the data, as well as empirical results of non-fragmentation hypothesis testing.

Equation (1.7) serves as the basis for the next testable hypothesis, which is tightly related (corollary) to Hypothesis 1 and deals with the impact of i) spillovers and ii) of market size S on the endogenous sunk costs F_i. **Testable hypothesis 2:** endogenous sunk costs expenditures, regarded as R&D and advertisement expenditures, decline in ex post spillovers but increase in market size. Formally, $\frac{dF_i}{d\theta^r} < 0$ while $\frac{dF_i}{dS} > 0$ (see Senyuta and Žigić, 2012, for derivations).

Hypothesis 2 predicts that increase in market size leads to increase in R&D expenditures. This relationship, however, can be reversed, i. e., market size can be endogenous. For example, Acemoglu and Linn (2004) demonstrate that the relationship between market size and R&D can be two-way: higher R&D would increase willingness to pay for the product, eventually would create new markets for the product, and money-value of market size would increase. The authors propose to use exogenous variation in the market size, not related to sales, but determined by demographic characteristics. In the dataset I also cannot rule out a two-way relationship between R&D and market size. One obvious way higher R&D can lead to an increase in the market size is through the export channel: by increasing their R&D expenditures, firms are able reach foreign markets, to export more, and in this way they increase their market size. To control for the endogeneity problem, I estimate an empirical model for the sample of firms whose export constitutes an insignificant share of sales.

Furthermore, the theoretical analysis in Chapter 1 points to the relation between a firm's decision to use private protection against spillovers on one side and the level of ex ante (or exogenous) spillovers θ and R&D efficiency parameter δ on the other side. **Testable hypothesis 3:** there is a positive relationship between a firm's protective measures and the size of ex ante spillovers and efficiency of the R&D investments.

2.4 Empirical Model

This section of Chapter 2 aims to empirically test three hypotheses that stem from the theoretical model in Chapter 1 and are stated in the previous section. I employ cross-sectional data to test the hypotheses. In this case, each observation contains information on a specific firm (with its own R&D spending, number of competitors, market size, knowledge spillovers experienced), and represents one of the realizations of the parameter values in the model: S, θ, θ^r, δ, N. Therefore, I observe multiple realizations of the theoretical model, and each observation of the parameter is individualized: S_i, θ_i, θ_i^r, δ_i, N_i. I test basic relationships of the theoretical model (between the market size and R&D spending, for example) with this cross sectional data.

To test empirical Hypotheses 1 the following econometric model is used:

$$comp = \begin{cases} 1, & \text{if } y^* \leq \eta_1, \\ 2 \text{ to } 6, & \text{if } \eta_1 < y^* \leq \eta_2, \\ 7 \text{ to } 16, & \text{if } \eta_2 < y^* \leq \eta_3. \\ \text{more than } 16, & \text{if } \eta_3 < y^*. \end{cases} \tag{2.4.1}$$

where *comp* is the categorical variable which measures the number of competing firms in the market, and $y^* = \boldsymbol{X_1}\beta_1 + \varepsilon_1$ is the exact but unobserved dependent variable, $\boldsymbol{X_1}$ is the vector of exogenous explanatory variables, which also includes the vector of industry dummies. Ordered logit regression is used to estimate the model (2.4.1), and this model is estimated separately for different market types[5]. $comp = 2$ denotes the base category with 2 to 6 competing firms in the market, while $comp = 3$ and higher denote categories with more competing firms in the market[6]. Hypotheses 1 makes a prediction about the relationship between one of the regressors in $\boldsymbol{X_1}$ (*Market size*) and *comp* variable, stating that this relationship should be positive and more pronounced in Type I markets (exogenous sunk costs markets) compared to Type II markets (endogenous sunk costs markets).

To test Hypotheses 2, which addresses the relationship between R&D expenditures (sunk costs) and market size and ex post knowledge spillovers, the following model is used:

$$RD = \begin{cases} RD^* & \text{if } ln(RD^*) > q \\ 0 & \text{if } ln(RD^*) \leq q \end{cases} \tag{2.4.2}$$

where q is lower than the logarithm of minimum uncensored value of RD, and the latent variable is $RD^* = exp(\alpha_2 + \boldsymbol{X_2\beta_2} + \varepsilon_2)$ Regression (2.4.2) is estimated with Tobit[7]. This hypothesis predicts a disincentivising effect of ex post spillovers on R&D expenditures.

To test Hypothesis 3 which predicts a positive effect of ex ante spillover on the protection efforts, the binary model is used:

$$protect = \alpha_3 + \boldsymbol{X_3\beta_3} + \varepsilon_3, \tag{2.4.3}$$

where *protect* is a measure of the protective actions taken by the firm. Regression (2.4.3) is estimated using probit and ordered logit regression. Several different variations of the model are estimated, depending on the dependent variable: probit regression is used if the dependent variable is the dummy variable for a specific type of protection (secrecy, patent, etc.), and ordered logit regression is used to account for the fact that the *protect* variable is discrete count variable, and takes a finite number of values (the number of

[5] Market types are: (1) Low R&D intensity and low ex ante spillovers, (2) Low R&D intensity and high ex ante spillovers, (3) High R&D intensity and low ex ante spillovers, (4) High R&D intensity and high ex ante spillovers

[6] There is a small number of observations for $comp = 1$ (single firm on the market, no competitors) and I omit this category from the model as well as the observations where $comp = 1$

[7] About half of the observations in the sample have zero spending on R&D

protective activities undertaken by a firm).

Section 2.5 below describes in detail which variables are included in explanatory variable vectors $\boldsymbol{X}$ in each of the regressions.

2.5 Description of the Variables

Empirical estimation is based on the Mannheim innovation panel dataset MIP (1993-2007). This is the firm-level annual survey of German firms conducted by the Center for European Economic Research (ZEW) and this dataset covers a representative sample of the German manufacturing sector and business related services. In particular, the dataset includes firm level information: sales, costs, operating margin, expenditures on innovation activities (including marketing activities) and R&D, private and public IPR protection measures undertaken by firms; and firms' relevant market information: market structure (number of main competitors), market share, the level of spillovers in the market (direct measure, firms' assessment of market spillovers from different sources). This information is submitted by firms to ZEW in the form of a completed questionnaire on a voluntary basis. In addition, each firm is attributed by ZEW to the particular industry based on a standard three-digit industry classification (in total, 22 industries are represented in the sample[8]).

There are several attributes of this dataset which are worth mentioning. Firstly, the market for the main product (and therefore, the market size) is self defined by the firms. The boundaries of the market are not observed by the researcher from the dataset, but are defined by the firms themselves. I believe that the definition of the "relevant market for the main product" in the dataset is limited by both geographical and industry classification, and is quite detailed, because it is provided by the market participants, but I acknowledge that to the large extent this market definition by the firms cannot be observed and controlled for. An alternative approach previously used in the literature is to classify firms based on 2-digit or 4-digit industry classification, and attribute to every firm relevant market/industry characteristics based on the statistical data for that industry. In addition to significantly limiting the variation of market/industry data (the number of industries is usually very small, even if a 4-digit classification is used), this approach has some serious limitations. 2-digit industry classification is very broad, and it is very likely that firms which belong to the same industry by classification, produce

[8]See Appendix for the detailed description of the industries.

products which differ significantly. On the other hand, 4-digit industry classification is too narrow in the sense that most firms produce several products which often belong to different classes.

The second important feature of the data is that the same market can be represented in the dataset several times, if several firms (operating in the same market) appear in the data. However, because of anonymity it is impossible to determine which are those firms[9].

A third point to note about the dataset is that the cross-sectional version of MIP dataset is used. The time span of the MIP is quite long, but the information collected is not consistent across years. For example, such a crucial variable as characteristics of the market structure is collected only in 2005 (number of main competitors, intensity of competition, etc), data on labor turnover and acquisition and loss of knowledge is collected only in 2006. Many individual observations (firms) drop out of the dataset between years, and new firms appear in the later years, so the dataset is unbalanced. I extract the dataset from the unbalanced panel, and obtain cross sectional dataset, where the number of observations is roughly between 700 and 1000, depending on the variables used in the regression. Detailed Table 2.5 with descriptive statistics is provided in the Appendix, and the description of variables is provided below.

All size-related variables in the dataset are multiplied by a firm-specific random number. This random number is a firm-specific time-invariant constant, i.e. every variable randomized in this way is multiplied by the same number for each firm. This guarantees that firms can no longer be recognized on the basis of absolute values they have reported. This procedure is used on the turnover figures (in millions of DM) and the number of (full-time) employees. The quotient of these two variables (turnover per full-time employee), however, remains unchanged. For the estimation, this randomization procedure is similar to the multiplicative error.

2.5.1 Dependent variables

I use two direct measures of R&D expenditures: spending on R&D (in millions of DM, per year) and R&D intensity (total R&D spendings/total sales) of the firms. Let R&D

[9]Researchers would be able to extract very valuable information if it were possible to determine which firms self-define themselves in the questionnaire as belonging to the same market. In the survey data I use, firms self-determine the relevant market for their main product, and report the characteristics of that market. If several firms reported the characteristics of the same market, it would allow me to avoid the measurement and reporting errors.

intensity for firm i be RDI_i. Then, $RDI_i = \frac{R\&D_i}{Sales_i}$ (as defined by the firm, using accounting information)[10]. Model predictions of the effect of spillovers on R&D intensity are the same as the prediction of the effect of spillovers on R&D expenditures. The database provides information on R&D intensity only. R&D expenditures are calculated by multiplying the *R&D intensity* on the level of *turnover*. As a robustness check, I also use a proxy for R&D expenditures: the number of employees who carry out R&D tasks. This measure is obtained by multiplying the *intensity of personnel R&D allocation* on the number of *all employees*[11].

As the measure of market concentration, I use the *number of main competitors*, as defined by the firm itself. This variable is represented by the categorical variable in the dataset. Four categories are defined in the questionnaires: (1) a firm has no competitors in the main market, (2) $1-5$ competitors, (3) $6-15$ competitors, (4) more than 15 competitors. The firm chooses the category which describes its market structure best[12]. The answers of the firms about the number of their main competitors are grouped in categorical variable $comp_i$: $comp_i$ has value 1 if firm i has *no competitors* (one firm in the market), it has value 2 if firm i has *between 1 and 5 competitors*, it has value 3 if firm i has *between 6 and 15 competitors*, and it has value 4 if firm i has *more than 15 competitors.*

$protect_i$ is another dependent variable in the dataset. The theoretical model predicts that as ex ante spillovers get higher, firms are more likely to use private protection against spillovers (Hypothesis 3). There are two variations of this variable: (i) dummy variable, which takes value 0 if the firm is not using a specific protective measure, and takes value 1 is the firm is using it; (ii) count variable version, which is defined on the range from 0 to 3 (for example, $protect_i = 0$ if firm i is not using any private/public protective measures, and $protect_i = 3$ if firm is using all of the private/public protective measures listed in the questionnaire).

[10]Innovation intensities over 0.35 are truncated to 0.35, therefore, values RDI=0.35 are considered outliers and dropped

[11]Variable "intensity of personnel R&D allocation" is provided in the dataset and is calculated as $\frac{\text{employees who carry out R\&D tasks}}{\text{all employees}}$

[12]More detailed information on competitors is believed to be confidential.

2.5.2 Main regressors

2.5.2.1 Measures of spillovers

The dataset contains many variables which can be described as some measures of *knowledge spillovers*. These are firm's estimates of industry/market knowledge spillovers that a firm is experiencing. As explained in the theoretical model, I differentiate between ex ante and ex post knowledge spillovers. The first characterize the environment in which the firm operates (technologies, etc). For example, one can expect that in pharmaceutical and high-tech industries ex ante spillovers will be higher. However, I claim that having very high ex ante spillovers in an industry does not mean that the industry will have very low spending on R&D (because firms will be afraid that their "technological advances" will be stolen, etc.). On the contrary, an environment with very high ex ante spillovers might incentivize all firms to use private protection against spillovers, ex post spillovers would be low, and the industry may experience even higher R&D expenditures. An alternative explanation of positive association between R&D and ex ante spillovers is that in the industries positive absorption effect (described by Cohen and Levinthal (1989)) might dominate, and higher ex ante knowledge spillovers would be associated with higher incentives to invest in R&D.

The model does not discriminate between the two explanations of positive R&D and ex ante spillover relationship (the positive absorption effect of Cohen and Levinthal (1989) and the protection channel described in Chapter 1). Most likely, both effects contribute to the positive relationship. However, I also run a regression where the dependent variable is a level of *private protection* from give away knowledge spillovers. This allows me to investigate whether firms with higher ex ante spillovers are more likely to use more private protection measures. Therefore, it would provide evidence that a protection channel is present in the data and could be driving (to some extent) the positive R&D and ex ante spillover relationship.

Further in this section I present PCA (principal component analysis) of the variables which measures the extent of knowledge spillovers. These include variables which describe the importance of different information sources for the firm ($info1 - info10$), and the extent of use of protection mechanisms ($protect1 - protect8$). These 18 variables are described in detail in the appendix in Table 2.12.

All measures of knowledge spillovers are highly correlated, and that is not surprising (see Table 2.13). The goal is twofold: first, I want to have a single variable which represents spillovers in the database, in order to be able to measure the disincentive effect of spillover on R&D expenditures. Second, I would like to classify the measures of spillovers, that is, to find how they are related. Principal component analysis allows me to perform both. I combine the measures of spillovers (variables) and create 2 new factors - principal components (PC1 and PC2). All 18 variables are combined in such a way that the first new factor (PC1) explains the highest share in the variation of the data. The second factor (PC2) explains the highest share of the remaining variation, and so on. Only two principal components are used in the regression analysis (together they explain more than 50% of the variation in spillover measures data).

Table 2.14 demonstrates factor loadings of different spillover variables on those principal components. In fact, such loading coefficients represent correlation between spillover variables and a specific component. All spillover variables are positively related to PC1, and variables characterizing the importance of information sources have higher correlation with PC1 than variables describing the extent of protection mechanisms. More importantly, second principal component (PC2) is strongly and positively associated with variables describing the extent of protection mechanisms, and negatively but weakly related to the importance of information sources. The third principal component does not demonstrate any clear pattern.

Before starting to characterize possible classification of spillover variables, and therefore, the interpretation of PC1 and PC2, it is interesting to consider the average values for principal components for different industries (Table 2.15). Principal components (exact values for new factor variables were obtained instead of the 18 initial measures of spillovers) are calculated by multiplying the values of factor loadings on the standardized values of initial variables. For example, if factor loading for PC1 is positive and significantly above zero, and the value of the initial variable is also high for a specific observation, the calculated first principal component (PC1) would also be high. The mining industry has the lowest value for PC1 in the analysis. Because all 18 variables have positive loadings on PC1 (are positively correlated with PC1), firms in the mining industry had reported values for spillover measures very much below the mean values. I conclude that for the mining industry, knowledge spillovers are of little importance, as for wholesale, retail, transportation industries and other services. These observations are also confirmed by the fact that average PC2 values for those industries are relatively close to zero.

On the other hand, industries such as producers of different types of equipment, chemicals, machinery, furniture and IT have high values for PC1, meaning that for those industries knowledge spillovers are important in general. However, there are important differences between those industries in the values of PC2. For example, the medical equipment industry, where PC1 was highest among all industries, does not have very high PC2 value, and for the IT industry average PC2 value is negative. Therefore, I consider industries for which PC2 value is low industries which suffer from relatively high ex post spillovers. Moreover, high value of PC1 is attributed to high ex ante knowledge spillovers. The logic is the following. Consider a firm in an industry which demonstrates high PC1. Considering positive and high loadings of all 18 spillover variables, the firm operates in an environment characterized by sizable knowledge spillovers. PC2 has positive and negative correlation with spillover measures, but negative correlation variables are smaller in absolute value. If this firm demonstrates relatively low PC2 value, it means that variables $protec1 - protec8$ (protection mechanisms) are not so important to the firm as variables $info1 - info10$ (information sources), therefore, knowledge spillovers are not curbed by private or public protection measure, and ex post knowledge spillovers remain high.

Principal component analysis allows me to extract ex ante and ex post spillover measures. In the analysis I associate high PC1 values with high ex ante knowledge spillovers θ, and low PC2 values with high ex post spillovers θ^r. In addition, generated PC2 values serve as a basis for dividing the sample in "high θ^r" and "low θ^r" subsamples, which are used in model (2.4.2) estimation. The ability to differentiate between ex ante and ex post spillovers is important. In the regression model (2.4.2) PC1 and PC2 variables were used as the regressors, and these results were also compared with the model in which all 18 variables measuring knowledge spillovers are used instead of PC1 and PC2 variables.

The theoretical model concludes that when one wants to empirically demonstrate the disincentivising effect of knowledge spillovers, proper measure of spillovers a firm is facing has to be considered. The exogenous characteristic of the industry cannot be used as such a measure. Rather, ex post spillovers should be used to estimate their effect on the R&D expenditures. Ex post spillovers will depend on ex ante spillovers, accounting for all the private protection measures used by the firms. For example, if a firm claims that "releasing employees would result in the loss of knowledge" (ex ante spillovers are high), but on the other hand it also claims that "we are paying enough to our employees so that they do not leave the company, and we are not afraid of the knowledge loss", it would mean that ex post spillovers are close to zero for that company, even though the market environment is such that ex ante spillovers are high.

In addition, the distinction between *ex post* and *ex ante* spillovers is in line with Hinloopen (2000) idea of *input* and *output* spillovers. The idea is that input spillovers are related to the knowledge sharing during the process of innovation and research, and output spillovers characterize the extent of final research output appropriability. Similarly to Hinloopen (2000), one can interpret the measure of "ex ante" spillovers as input spillovers, which would not necessarily discourage innovations but rather facilitate R&D cooperation. On the contrary, output (ex post) spillovers characterize final knowledge appropriability and should have disincentivising effect on R&D.

The main predictions about the impact of **ex ante** spillovers and **ex post** spillovers variables on R&D incentives and market fragmentation are summarized by hypotheses 1 and 2.

2.5.2.2 Other regressors: effectivity of investment and market size

The other regressor I control for is the theoretical model parameter δ_i, which measures the effectiveness of R&D in raising product quality for firm i. As the proxy measure of effectiveness of R&D expenditures I use the characteristics of different innovation constraints. The examples of the constraints are: "Innovation costs are too high", "Organizational problems", etc. The measure is given from 0 to 1 (0 if the constraint is not important, and 1 if the constraint is important). Therefore, the higher the measure of constraint, the lower effectiveness of R&D (δ is higher). I use weighted average of different constraints as a measure of effectiveness of R&D.

One of the results in the endogenous market structure literature is that low values of δ (high effectivity of R&D expenditures) imply more concentrated market structure. This happens because when δ is low, firms find it more attractive to deviate upward in their R&D spendings and so the equilibrium level of R&D is higher and the number of active firms is lower. However, introducing endogenous protection against spillovers can make this negative relation between δ and R&D expenditures non-monotonic. For a higher δ, R&D investment is less effective and therefore less attractive, and equilibrium sunk costs decrease. However, theory predicts that with higher δ firms also find it more profitable to protect against existing knowledge spillovers, which will make R&D investment more attractive, irrespective of the increased value of δ. I expect that the effect of δ on R&D expenditures will depend on whether δ is high or low. To account for this non-linear relationship I include the squared value of δ in the regressions (2.4.1)-(2.4.3).

An other regressor is *market size* variable S, calculated from *turnover*, *share of the main product* in turnover, and the *market share of the main product* variables using the following formula: $S = \frac{turnover \times share\ of\ main\ product\ in\ turnover}{market\ share\ of\ main\ product}$. Logarithm of S variable is used in the regressions. Taking logarithm of market size variable helps me to obtain coefficients which are easier to interpret, to ensure a better fit for the linear model, and to tackle the problem of multiplicative measurement error.

2.5.3 Control variables

Variation in R&D expenditures in the sample can be driven not only by the market size and spillovers, but also by different technological research opportunities arising for each firm (Acemoglu and Linn, 2004). In addition to controlling for the differences in effectiveness of R&D (different δ_i) several other controls were used: controls for the industry[13], size of the firm (number of employees), and for the proportion of all employees who have a university degree or other higher education[14], and a dummy variable defining a firm located in East Germany. If variability in R&D across the set of firms is mostly due to the differences in technological opportunities and not in the size of the market, or the level of spillovers, this will be captured by the control variables described in this subsection.

2.6 Empirical Results

2.6.1 Hypothesis 1: Market size and market fragmentation - effect of ex post spillovers

Table 2.1 below demonstrates the idea of decreasing lower bound of concentration with the data. Average market size values (in millions of DM) are calculated for different levels of $C1$, market share of the largest company in the market. The table is constructed separately for Type I (R&D intensity$\leq 1\%$) and Type II (R&D intensity$> 1\%$) markets. If the largest firm in a specific market has a small market share ($C1 = 2.5\%$), which means

[13] *22 industry dummies*, based on the Statistical classification of economic activities in the European Community (NACE)

[14] Proportion of employees with high education is a categorical variable in the sample, and I introduce 9 dummy variables to capture all categories, with the base category being "0% of employees with higher education"

Table 2.1: C1 and Market Size

C1	Type I markets		Type II markets	
	Mean market size	**Number of obs**	**Mean market size**	**Number of obs**
2.5%	416.81	67	608.15	35
7.5%	292.85	23	338.81	32
12.5%	170.93	25	406.45	22
17.5%	740.79	13	774.95	16
25%	142.82	29	371.55	28
40%	77.13	28	203.23	41
65.5%	41.07	18	189.87	26
87.5%	20.03	30	40.55	27
Total	241.23	223	344.65	227

NOTE: the table represents one-firm concentration ratio C1 only for those firms which have reported their market share and identified themselves as the largest firm in the market. Therefore, the dataset misses C1 values for those markets where the firms which participated in the questionnaire have not reported themselves as the largest firms in the market

that the market is very fragmented, the market size is likely to be very large (first row of the table). As the market share of the largest firm in the market increases, the market becomes more concentrated, the average market size gradually decreases[15]. Indeed, the most concentrated markets are also the smallest on average (the last row of the table). More importantly, the average market size is always higher for Type II markets. This observation is consistent with claim that lower bound on concentration for endogenous sunk cost market (Type II) is higher, or that it does not collapse to zero as market size increases.

The structure of data ($C1$ is reported only at intervals, and only by the selected subset of firms which consider themselves to be the largest firms in the market), does not allow me to draw or estimate the concentration lower bound, as in Ellickson (2007) or Robinson and Chiang (1996). To use formal regression analysis, I turn my attention to other measure of market structure - categorical variable for the number of firms competing in the market. Despite the fact that this variable is also measured in the intervals in the database, unlike $C1$ it is reported for all firms in the sample, without regards to their size and relative market share.

In order to test Hypothesis 1, empirical model (2.4.1) is estimated separately for 4 different subsets of firms: Type I with low ex post spillovers θ^r, Type I with high ex post spillovers θ^r, Type II with low ex post spillovers θ^r, Type II with high ex post spillovers

[15]The average market size for $C1 = 17.5\%$ does not follow that pattern. First of all, that could be influenced by the fact that there are some significant outliers for market size in the sample for $C1 = 17.5\%$. However, $C1 = 17.5\%$ level of concentration also has a low number of observations compared to other concentration levels. This could be due to the fact that for some reason firms with market share between 15 and 20% are underrepresented in the sample.

Table 2.2: R&D and Marketing Expenditures and Market Fragmentation

Endogenous Sunk Costs	**Market Fragmentation**			
	single firm	**2-6 firms**	**7-16 firms**	**more than 16 firms**
Share of firms with zero R&D spendings	67%	43%	47%	56%
Share of firms with zero marketing spendings	61%	24%	27%	37%

Table 2.3: Descriptive statistics for the size of the market (in millions of DM per year) for different market structures

Market Size	**Market Fragmentation**			
	single firm	**2-6 firms**	**7-16 firms**	**more than 16 firms**
Mean	109.7	312.1	1 139.7	886.7
Median	25.4	44.1	84.1	65.7
Max	1 307.9	10 454.7	182 921.2	76 652.6
Min	0.48	0.12	0.77	0.17
St.dev.	226.5	1029.3	11940	5975.7
N	24	746	235	179

θ^r. According to Hypothesis 1, for Type I markets, as S increases, the probability that markets have $7-15$ firms rather than have $2-6$ should increase[16]. Before I start with empirical model, consider the characteristics of firms which operate in the markets with different level of fragmentation (Table 2.2).

43% of firms operating in the markets with a small number of firms (2-6 firms), do not spend anything on R&D (24% of firms in this subsample do not spend anything on marketing). However, if I consider more fragmented markets, 47% and 56% spend nothing on R&D (27% and 37% of firms which do not spend anything on marketing). Consistent with the endogenous sunk costs theory, markets with low endogenous sunk costs (marketing and R&D expenditures) have experienced more entry into the market, and as a result those markets are more fragmented.

However, this relationship between R&D expenditures and market fragmentation could be also driven only by the size of the market. It could be that markets with many firms are substantially smaller, and therefore, there is no need for significant spending on R&D and marketing. In the Table 2.3 I describe the market size for different market structures.

From Table 2.3 it is clear that markets with a larger number of firms are in fact larger, which is illustrated by both mean and median. Therefore, low spending on R&D and marketing in fragmented markets cannot be driven by the fact that those markets are

[16]Also, the probability that markets have more than 15 firms rather than $2-6$ should increase

Table 2.4: Descriptive statistics for Type I and Type II markets

	Type 1 markets R&D intensity≤1% mean	std	Type 2 markets R&D intensity>1% mean	std
Share of markets with:				
1 firm	0.04	0.20	0.03	0.17
2-6 firms	0.55	0.50	0.57	0.49
7-15 firms	0.20	0.40	0.20	0.40
Firm size and other characteristics				
Number of employees	91.78	397.49	280.80	2036.48
Sales	29.45	131.09	107.33	940.85
Sales per employee	288.68	196.37	288.41	178.86
Export share	0.10	0.21	0.21	0.27
Geographical reach	0.90	0.83	1.18	0.79
Endogenous Sunk Costs				
R&D expenditures	0.08	0.72	7.50	84.44
R&D expenditures per employee	0.34	0.95	24.70	31.63
Marketing expenditures	0.38	2.35	3.02	66.82
Marketing expenditures per employee	2.83	8.43	5.71	11.42
Market Size				
Market size	388.40	1385.47	634.42	5491.48
Market size per employee	3898.92	5494.96	2795.16	4034.16
Observations	1284		1537	

NOTE: Market Size, Sales and Expenditure variables are in millions of DM (Deutsche Mark), and all per employee variables are in thousands of DM

smaller.

Returning to Table 2.2, if spending on R&D and marketing prevents entry of competitors, then markets with single operating firm should be spending most on the endogenous sunk costs. However, in the subset of markets with a single operating firm 67% of firms do not spend anything on R&D and 61% of firms do not spend anything on marketing. This fact seems to contradict endogenous sunk costs model results, according to which markets with highest concentration should be characterized by a high level of endogenous sunk costs (R&D and marketing expenditures). However, in the sample, markets with single firm are much smaller compared to other market structures (Table 2.3, compare mean, median, maximum). Thus, it is quite likely that monopolists in the markets do not spend much on R&D (and marketing) because the size of the market is *not large enough*, and there is no space for an entrant, even if the incumbent has zero endogenous sunk costs.

Now, I turn to the formal regression analysis of model (2.4.1). Regressions were run separately for the subsample with low and high R&D intensity (RDI≤1% and RDI≥1%), and low and high ex post spillovers θ^r. I consider the subsample with RDI≤1% firms participating in the markets with exogenous sunk costs (Type I), and RDI≥1% I consider as firms with endogenous sunk costs (Type II)[17].

Log of market size S, measure of effectiveness of R&D (δ), number of workers, dummy variables for the share of workers with higher education and industry dummies are included in vector $\boldsymbol{X_1}$ as explanatory variables. Table 2.6 presents the estimation results of ordered logit regression. Consider columns (1) and (2) in Table 2.6. Coefficients for log of market size variable S are positive and significant. This means that increase in the market size increases the fragmentation of the Type I (exogenous sunk costs markets). Importantly, in an exogenous sunk costs market with high ex post spillovers (column (2)) I observe that market size has even stronger positive effect on market fragmentation. On the contrary, columns (3) and (4) show that for Type II markets (with low and high ex post spillovers) increase in the market size does not have a significant effect on the market fragmentation. Both of these results are consistent with Hypothesis 1.

2.6.2 Hypothesis 2: R&D incentives and knowledge spillovers

In the empirical analysis I focus on R&D expenditures as the leading example of the endogenous sunk costs. However, I do the robustness check using marketing expenditures as the other example of endogenous sunk costs, which enables me to draw more general conclusions about endogenous sunk costs and market structure relationship.

Tables 2.7-2.8 show the result of model (2.4.2) estimation. Table 2.9 also presents estimation results of model (2.4.2) with marketing expenditures as the dependent variable. Log of market size S, measure of effectiveness of R&D δ, different measures of knowledge spillovers θ and θ^r, number of employees, dummies for the share of employees with higher education and industry dummies are included in vector $\boldsymbol{X_2}$ as explanatory variables.

In Tobit regression model log of R&D expenditures were used as a dependent variable[18]. Because log transformation of R&D expenditures data was to some extent arbi-

[17]Sutton (1991) uses a similar approach, defining the industries as exogenous or endogenous sunk costs based on the level of spending by firms on advertisement, and also uses a threshold of 1%.

[18]More than 40% of observations have reported zero spending on R&D. This represents a case of data censoring, and Tobit regression was used to solve the problem of sample selection. To use the Tobit model, data on log R&D expenditures has to be normally distributed. However, in addition to being censored at 0, R&D expenditure data is significantly right-skewed (a test for normality rejects that R&D expenditure variable is normally distributed, but it cannot reject that R&D expenditures are log normally distributed). In general, this problem is solved by using the Tobit model on log of expenditure

trary (observations with zero R&D expenditures .were substituted with minimum value of R&D expenditures close to zero in the dataset), and a Poisson model to check the robustness of the Tobit estimation (Santos Silva and Tenreyro, 2011) was also used.

Estimates for the Tobit model are presented in tables 2.7 and 2.8, using different ex post spillover measures. Column (1) does not include any spillover measures in model (2.4.2), columns (2) and (3) include spillover measures PC1 and PC2, and columns (4) in tables 2.7 and 2.8 use all original variables (importance of different information sources and protection measures, described in Table 2.12) which are associated with knowledge spillovers and which were used in principal component analysis.

As discussed in subsection 2.3, there could be a two-way relationship in R&D and market size S variable; that is, the market size variable is endogenous. One way this impact can be realized is through firm excess to the foreign market: only after spending a sufficient amount on R&D it might be possible to start exporting a product, and in this case market size is determined by R&D expenditures. In order to control for this effect, empirical model (2.4.2) was estimated for the sample of firms which operate in the German market only.

In table 2.7 coefficients for the market size S are positive, significant, but different in specifications (1) as compared to (2-4): 1%. increase in market size leads to about 0.46% increase of R&D spending on average, holding other variables constant, if no spillover measures are included in the regression; and 1%. increase in market size leads to about 0.3% increase of R&D spending on average if spillover measures are included.

In tables 2.7 and 2.8 the values of coefficients for the log of market size, $ln(S)$, in columns (4) are similar to those in columns (2) and (3), but different from the coefficients in columns (1), where no spillover variables were included. This suggests a problem of multicollinearity (strong correlation between market size and spillover variables). Indeed, in the regressions in tables 2.7 and 2.8, where all original spillover measures are included instead of principal components (last columns), not all coefficients for spillover-related variables are significantly different from zero, however, most of the coefficients are positive (see Table 2.10). For example, information from inside the company (group of companies), $info1$, is the most important determinant of R&D expenditures among all spillover variables, and the effect of $info1$ is consistent across specifications in tables

data. However, with the Tobit regression model, the censoring threshold would no longer be zero. Lower threshold option becomes ll($ln(y)$), where the threshold $ln(y)$ equals (or is lower) to the logarithm of the minimum uncensored value of R&D expenditures.

2.7 and 2.8. Similarly, knowledge protection using patents, $protect1$, is positively related to R&D spending, and this effect is consistent across specifications. I conclude that coefficients for the sources of information ($info1 - info10$) and for the importance of protection measures ($protect1 - protect8$) are jointly significant in the regressions (4) in tables 2.7 and 2.8. However, because of high correlation between those variables (Table 2.13) it is not possible to determine and compare the effect of each variable separately. Thus, I turn to the regression analysis with principal components (columns (2) and (3)). I use two first principal components in the regressions, which allows me to determine the effect of ex ante and ex post knowledge spillovers, as described in subsection 2.5.2.

If $PC1$ is used as the measure of ex ante spillovers θ in the regression, I observe a positive and significant effect of knowledge spillover on R&D investment. These results simply confirm the empirically observed fact that R&D-intensive industries are also likely to experience high knowledge spillovers. According to Hypothesis 2, ex post knowledge spillovers θ^r have a disincentivising effect on R&D incentives. This result is confirmed for the sample of all firms, but not confirmed for non-exporting firms subsample (Table 2.8), where the coefficient for $PC2$ is not significantly different from zero. Column (3) of table 2.7 shows that increase in $PC2$, which is associated with decrease ex post spillovers, has positive effect on R&D incentives. In other words, if $PC2$ increases[19] R&D expenditures also increase. This confirms the prediction about the disincentivising effect of ex post knowledge spillovers. Results are similar if marketing expenditures are used instead of R&D expenditures as a dependent variable.

2.6.3 Hypothesis 3: protection against spillovers

Subsection 2.3 describes predictions about the effect of ex ante spillover θ, effectivity of R&D (δ) on the costly protection measures undertaken. Regression results are presented in Table 2.11. Model (2.4.3) is estimated using probit (for binary outcomes) and ordered logit (for count data) regression, and vector $\boldsymbol{X_3}$ includes log of market size S, δ and θ, number of employees, share of employees with higher education and industry dummies. Results demonstrate that if R&D investment is less effective in raising quality (δ is higher), the probability of using private protection measures increases. The same is true for higher ex ante spillover θ: as $PC1$ increases, a firm is more likely to seek

[19]This means that success of protection is higher (represented by variables $protect1 - protect8$), ex ante spillovers are lower (represented by variables $info1 - info10$) and as a result ex post spillovers decrease.

protection from spillovers. The effect of market size is not significantly different from zero. Coefficient values are similar for regressions where dependent variable *protect* accounts for all protection measures, formal measures or for private protection measures only (columns 9-11). However, I am interested mainly in ex ante spillover effects on the incentives to use various protection measures. Therefore, predictions about the effect of ex ante knowledge and the effectiveness of R&D investment on the incentives to use protection measures (private and public) are confirmed by the estimates of model (2.4.3).

2.7 Conclusion

This Chapter investigates how incentives to invest in R&D, which are modeled as endogenous sunk costs, respond to knowledge spillovers in the market and market size. I model knowledge spillovers as contribution of other firms $j \neq i$ in the market to the perceived quality of product i. This contribution might stem from such channels as reverse-engineering, labor force flows among firms, strategic alliances between firms, knowledge dispersion to competitors through "vertical channel" (supplier-client), etc. The results indicate that "ex post" knowledge spillovers have a negative and significant effect on R&D investment level, which provides evidence of the disincentivising effect of spillovers. In addition, "ex ante" spillover variables demonstrate a positive relationship with R&D expenditures. These results simply confirm the widely observed empirical fact that R&D-intensive industries are also likely to experience high knowledge spillovers.

In this paper I model R&D investment as endogenous sunk costs, as opposed to other alternatives, such as contestable inventions models, or models of cooperation in R&D and consortia. The endogenous sunk costs model implies a so-called non-fragmented equilibrium: the expansion of the market size beyond a certain threshold does not result in a larger number of firms, and so there is a limit on the entry of new firms, and consequently on the lower bound of market concentration. The dataset allows me to test whether the number of firms in the markets with large R&D expenditures are not responsive to market size increase (Hypothesis 1).

I confirm that in the markets where R&D expenditures are important, as market size increases, the probability that the market will become more fragmented does not increase. This result is consistent with the disincentivising effect of high ex post knowledge spillovers.

Empirical evidence on markets with low R&D expenditures is also consistent with endogenous sunk costs theory. In those markets, as market size increases, the probability that the market will be more fragmented also increases. I conclude that modeling R&D expenditures as endogenous sunk costs is consistent with empirical evidence.

This paper also shows that one should interpret empirical models which estimate the effects of knowledge spillovers on R&D incentives with great caution. Positive (or negative) effect does not necessarily demonstrate that the absorption (or disincentivising) effect is dominant. A researcher must ensure that empirical evidence on the effect of spillovers on R&D incentives is correctly interpreted. For example, in many markets with high R&D expenditures it could be that releasing employees would result in the loss of knowledge. One might interpret this as sign high spillovers, and conclude that R&D expenditures and spillovers are positively related. But such an interpretation fails to account for the fact that firms are also paying enough to their employees so that they do not leave the company. Further, it would mean that knowledge spillovers researcher should consider are low. This example illustrates the necessity to understand what observed knowledge spillovers measure, and how these variables are relevant to testing theory on the spillover-R&D incentives relationship.

2.A Appendix 2

2.A.1 Descriptive statistics

Table 2.5: Sample Characteristics by Industry

Industry	**Market Size**	**R&D expenditures**	**R&D expenditures per employee**	**Marketing expenditures**	**Marketing expenditures per employee**	**Share of Markets with:** single firm	2-6 Firms	7-15 Firms
Mining								
mean	640.97	0.08	2.60	0.06	1.30	0.03	0.64	0.28
sd	2248.04	0.20	6.57	0.13	2.00	0.18	0.48	0.45
count	15.00	24.00	23.00	20.00	19.00	58.00	58.00	58.00
Food, tobacco								
mean	412.51	0.55	4.47	0.61	5.74	0.01	0.59	0.23
sd	670.14	1.40	10.56	1.45	13.11	0.08	0.49	0.42
count	40.00	62.00	61.00	53.00	52.00	142.00	142.00	142.00
Textiles								
mean	327.84	0.43	8.51	0.60	3.99	0.04	0.59	0.20
sd	759.22	0.83	26.83	1.74	7.63	0.21	0.49	0.40
count	39.00	54.00	54.00	46.00	46.00	137.00	137.00	137.00
Wood, paper								
mean	596.14	0.62	8.14	0.38	3.98	0.01	0.57	0.24
sd	1796.83	2.96	24.86	1.40	10.91	0.11	0.50	0.43
count	74.00	107.00	107.00	98.00	98.00	304.00	304.00	304.00
Chemicals								
mean	387.96	3.83	26.86	4.39	13.17	0.01	0.60	0.25
sd	725.24	10.91	47.52	20.24	24.68	0.10	0.49	0.43
count	61.00	76.00	76.00	72.00	72.00	183.00	183.00	183.00
Plastics								
mean	461.32	1.14	7.91	0.68	2.95	0.01	0.54	0.26
sd	1642.27	5.00	13.84	2.45	5.14	0.08	0.50	0.44
count	45.00	69.00	69.00	62.00	62.00	166.00	166.00	166.00
Glass								
mean	137.18	1.54	5.42	0.33	1.74	0.00	0.65	0.23
sd	241.56	5.07	11.14	0.74	2.41	0.00	0.48	0.42
count	37.00	42.00	42.00	44.00	44.00	91.00	91.00	91.00

Continued on Next Page…

Table 2.5 – Continued

Industry	**Market Size**	**R&D expenditures**	**R&D expenditures per employee**	**Marketing expenditures**	**Marketing expenditures per employee**	**Share of Markets with:** single firm	2-6 Firms	7-15 Firms
Metals								
mean	323.89	1.08	6.88	0.33	1.75	0.01	0.53	0.24
sd	701.69	3.50	18.49	1.03	4.94	0.09	0.50	0.43
count	110.00	160.00	160.00	147.00	147.00	384.00	384.00	384.00
Machinery								
mean	293.48	4.52	13.95	0.86	4.06	0.02	0.69	0.19
sd	488.14	15.46	18.96	2.96	7.35	0.15	0.47	0.40
count	82.00	108.00	108.00	90.00	90.00	264.00	264.00	264.00
Electrical equipment								
mean	1763.53	19.94	17.69	3.51	4.89	0.02	0.62	0.23
sd	9727.87	121.20	24.28	16.01	9.67	0.15	0.49	0.42
count	62.00	79.00	79.00	77.00	77.00	218.00	218.00	218.00
Medical, other equipment								
mean	101.86	2.86	28.53	0.63	7.10	0.02	0.72	0.17
sd	184.25	5.79	27.28	1.41	8.68	0.13	0.45	0.37
count	84.00	94.00	94.00	92.00	92.00	229.00	229.00	229.00
Transport equipment								
mean	6492.20	76.90	16.65	1.68	2.48	0.03	0.77	0.13
sd	30906.04	400.69	23.61	5.17	4.69	0.16	0.42	0.34
count	35.00	44.00	44.00	38.00	38.00	115.00	115.00	115.00
Furniture								
mean	556.51	0.83	4.83	0.91	6.57	0.03	0.56	0.28
sd	1173.47	2.18	12.34	2.00	8.15	0.17	0.50	0.45
count	29.00	33.00	33.00	32.00	32.00	72.00	72.00	72.00
Wholesale								
mean	426.40	0.11	5.19	0.31	5.22	0.02	0.64	0.20
sd	1382.18	0.30	17.72	0.77	13.07	0.14	0.48	0.40
count	57.00	77.00	77.00	61.00	61.00	197.00	197.00	197.00

Continued on Next Page…

Table 2.5 – Continued

Industry	Market Size	R&D expenditures	R&D expenditures per employee	Marketing expenditures	Marketing expenditures per employee	Share of Markets with: single firm	2-6 Firms	7-15 Firms
Retail, automobile								
mean	112.83	0.08	6.62	0.23	5.92	0.00	0.71	0.19
sd	221.02	0.29	23.82	0.59	8.50	0.00	0.45	0.40
count	37.00	50.00	50.00	45.00	45.00	98.00	98.00	98.00
Transport, communication								
mean	253.78	1.17	4.99	0.48	3.20	0.07	0.54	0.12
sd	566.34	10.76	18.68	2.11	8.34	0.26	0.50	0.32
count	93.00	173.00	170.00	139.00	137.00	379.00	379.00	379.00
Banking, insurance								
mean	15.96	32.32	17.54	0.00	0.00	0.04	0.46	0.18
sd	.	.	.	.	.	0.19	0.50	0.39
count	1.00	1.00	1.00	1.00	1.00	218.00	218.00	218.00
IT								
mean	171.79	1.34	18.34	0.29	3.75	0.02	0.54	0.22
sd	414.72	3.39	25.47	0.61	4.13	0.15	0.50	0.42
count	58.00	75.00	75.00	76.00	76.00	207.00	207.00	207.00
Technical services								
mean	175.77	1.12	16.92	0.77	2.21	0.03	0.51	0.19
sd	683.39	3.89	33.13	5.05	4.82	0.18	0.50	0.40
count	101.00	180.00	180.00	159.00	159.00	406.00	406.00	406.00
Firm-related services								
mean	86.25	0.08	2.81	0.13	1.90	0.07	0.43	0.14
sd	162.06	0.20	6.28	0.52	3.44	0.26	0.50	0.35
count	37.00	74.00	73.00	67.00	66.00	169.00	169.00	169.00
Other services								

Continued on Next Page…

Table 2.5 – Continued

Industry	**Market Size**	**R&D expenditures**	**R&D expenditures per employee**	**Marketing expenditures**	**Marketing expenditures per employee**	**Share of Markets with:** single firm	2-6 Firms	7-15 Firms
mean	339.74	0.23	4.00	0.10	1.13	0.08	0.42	0.18
sd	1613.08	1.07	17.15	0.42	2.58	0.27	0.49	0.39
count	70.00	117.00	115.00	105.00	102.00	315.00	315.00	315.00
Real estate, renting								
mean	138.37	0.22	12.20	0.11	3.24	0.07	0.59	0.17
sd	171.72	0.73	50.23	0.22	8.64	0.26	0.49	0.38
count	25.00	50.00	44.00	46.00	41.00	98.00	98.00	98.00
Total								
mean	556.81	4.04	11.11	0.82	3.91	0.03	0.57	0.20
sd	5842.72	69.26	25.54	6.09	9.36	0.18	0.50	0.40
count	1192.00	1749.00	1735.00	1570.00	1557.00	4450.00	4450.00	4450.00

NOTE: Market Size and Expenditure variables are in millions of DEM (Deutsche Mark), and Expenditure per employee variables are in thousands of DEM

2.A.2 Regression results

Table 2.6: Market Fragmentation and Knowledge Spillovers

	Type I		Type II	
	$RDI \leq 1\%$ and low θ^r	$RDI \leq 1\%$ and high θ^r	$RDI > 1\%$ and low θ^r	$RDI > 1\%$ and high θ^r
	(1)	(2)	(3)	(4)
ln(S)	0.2886**	0.3368**	0.1517	0.122
	[0.110]	[0.099]	[0.093]	[0.096]
PC1	0.223	-0.2734**	-0.0895	-0.0255
	[0.143]	[0.096]	[0.109]	[0.071]
Number of workers	-0.0004	-0.0036*	0.0002	-0.0001
	[0.001]	[0.002]	[0.000]	[0.000]
δ	0.0343	-0.5325*	-0.7338*	0.0483
	[0.339]	[0.226]	[0.319]	[0.298]
δ^2	-0.4392	0.2223	0.1458	0.1655
	[0.473]	[0.353]	[0.406]	[0.383]
Observations	172	228	190	215

Robust standard errors in brackets
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 2.7: R&D Expenditures and Knowledge Spillovers

	(1)	(2)	(3)	(4)
$PC1$		0.2115**	0.1977**	
		[0.037]	[0.038]	
$PC2$			0.0808+	
			[0.043]	
$ln(S)$	0.4614**	0.3021**	0.3069**	0.2928**
	[0.035]	[0.043]	[0.042]	[0.041]
δ	0.0831	0.3243*	0.2978*	0.2858*
	[0.118]	[0.131]	[0.131]	[0.127]
δ^2	-0.3014+	-0.0041	-0.0143	0.0584
	[0.162]	[0.181]	[0.181]	[0.174]
Number of workers	0.0001**	0.0012**	0.0012**	0.0012**
	[0.000]	[0.000]	[0.000]	[0.000]
18 measures of spillovers	No	No	No	Yes
Observations	595	442	442	442

Standard errors in brackets
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 2.8: R&D Expenditures and Knowledge Spillovers for Firms Which Sell Domestic Market Only

	(1)	(2)	(3)	(4)
$PC1$		0.2465**	0.2468**	
		[0.063]	[0.063]	
$PC2$			0.052	
			[0.087]	
$ln(S)$	0.3066**	0.2126**	0.2130**	0.2617**
	[0.061]	[0.071]	[0.071]	[0.069]
δ	0.0072	0.4551*	0.4445*	0.4875*
	[0.181]	[0.209]	[0.209]	[0.203]
δ^2	-0.3879	-0.0996	-0.1064	-0.0146
	[0.253]	[0.280]	[0.280]	[0.266]
Number of workers	0.0008**	0.0019**	0.0019**	0.0016*
	[0.000]	[0.001]	[0.001]	[0.001]
18 measures of spillovers	No	No	No	Yes
Observations	226	176	176	176

Standard errors in brackets
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 2.9: Marketing Expenditures and Knowledge Spillovers

	(1)	(2)	(3)	(4)
$PC1$		0.1792**	0.1775**	
		[0.030]	[0.030]	
$PC2$			0.0843*	
			[0.039]	
$ln(S)$	0.4585**	0.4056**	0.4064**	0.3861**
	[0.032]	[0.036]	[0.036]	[0.035]
δ	0.1345	0.3483**	0.3210**	0.3118**
	[0.103]	[0.114]	[0.114]	[0.111]
δ^2	-0.2547+	-0.1257	-0.1322	-0.066
	[0.145]	[0.162]	[0.161]	[0.158]
Number of workers	0.0001**	0.0003**	0.0003**	0.0003**
	[0.000]	[0.000]	[0.000]	[0.000]
18 measures of spillovers	No	No	No	Yes
Observations	697	544	544	544

Standard errors in brackets
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 2.10: R&D Expenditures and Knowledge Spillovers (full spillover measures list)

	(1) All firms	(2) Domestic firms
$ln(S)$	0.2928**	0.2617**
	[0.041]	[0.069]
δ	0.2858*	0.4875*
	[0.127]	[0.203]
δ^2	0.0584	-0.0146
	[0.174]	[0.266]
Number of workers	0.0012**	0.0016*
	[0.000]	[0.001]
info1	0.4467**	0.6477**
	[0.085]	[0.131]
info2	-0.2137*	-0.2759*
	[0.084]	[0.133]
info3	-0.0618	-0.1812
	[0.083]	[0.143]
info4	-0.012	-0.038
	[0.088]	[0.155]
info5	0.2996**	0.1921
	[0.106]	[0.186]
info6	0.0299	-0.0153
	[0.102]	[0.188]
info7	0.1494	0.2771
	[0.109]	[0.208]
info8	0.1006	0.1093
	[0.099]	[0.159]
info9	-0.0615	0.0847
	[0.105]	[0.165]
info10	-0.1078	0.0059
	[0.103]	[0.173]
protect1	0.1607*	0.1547
	[0.072]	[0.173]
protect2	0.0017	0.3514+
	[0.078]	[0.187]
protect3	0.1146	0.0562
	[0.163]	[0.504]
protect4	-0.0275	-0.156
	[0.074]	[0.159]
protect5	-0.1477	-0.2312
	[0.118]	[0.306]
protect6	0.1105+	0.0501
	[0.065]	[0.120]
protect7	-0.0243	0.0796
	[0.067]	[0.141]
protect8	0.0973	-0.015
	[0.064]	[0.113]
Observations	442	176

Standard errors in brackets
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 2.11: Protection and knowledge spillovers

	Different types of protection, zero-one data								Number of protection measures, count data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$ln(S)$	0.001	0.0003	-0.0302	-0.0385	-0.0667	-0.0295	-0.0253	-0.0326	-0.0021	-0.0137	-0.0109
	[0.040]	[0.038]	[0.059]	[0.036]	[0.051]	[0.033]	[0.035]	[0.033]	[0.027]	[0.023]	[0.017]
$PC1$	0.4435**	0.2511**	0.2221**	0.2617**	0.2782**	0.3423**	0.2705**	0.3861**	0.3145**	0.2701**	0.2939**
	[0.045]	[0.033]	[0.054]	[0.035]	[0.050]	[0.031]	[0.033]	[0.033]	[0.023]	[0.019]	[0.015]
δ	0.3154*	0.2225*	0.3168+	0.0889	0.2944+	0.0921	0.0753	0.2163*	0.1801*	0.0697	0.1332*
	[0.125]	[0.111]	[0.172]	[0.116]	[0.160]	[0.102]	[0.107]	[0.102]	[0.085]	[0.074]	[0.056]
δ^2	0.4178*	0.4494**	0.4075+	0.0271	0.1543	-0.0873	0.2804+	0.007	0.3031**	0.0243	0.1370+
	[0.174]	[0.158]	[0.247]	[0.159]	[0.225]	[0.142]	[0.148]	[0.141]	[0.117]	[0.104]	[0.078]
Number	0	0.0001	0	0	0.0001	0	-0.0002	0	0	-0.0001	-0.0001+
of workers	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Types of protective measures	Patent design	Registered design	Industry	Trademark	Copyright	Secrecy	Complexity	Head start	Formal measures (1-5)	Informal measures (6-8)	All
Obs	714	693	552	758	612	738	738	738	764	769	769

Standard errors in brackets
+ significant at 10%; * significant at 5%; ** significant at 1%

Table 2.12: Description of variables used for PCA

Variable Name	Description
Information sources ($info1 - info10$)	
$info1$	Importance of information sources from inside the firm or within the group of companies or related companies: 0=not important at a 1=slight importance to 3=great importance
$info2$	...suppliers...
$info3$	...customers...
$info4$	...competitors...
$info5$	...consultancy...
$info6$	...universities...
$info7$	...non-profit research institutions...
$info8$	...trade fairs...
$info9$	...specialists literature...
$info10$	...professional assistance...
Effectivity of different types of protective mechanisms ($effect1 - effect8$)	
$protect1$	Success of patent protection: 1=low, 2=moderate, 3=high, 0 if no patent protection was used
$protect2$	...registered design...
$protect3$	...industry design...
$protect4$	...trademark...
$protect5$	...copyright...
$protect6$	...secrecy...
$protect7$	...complex design...
$protect8$	...head start...

Table 2.13: Correlation between spillover measures

	info1	info2	info3	info4	info5	info6	info7	info8	info9	info10	prot1	prot2	prot3	prot4	prot5	prot6	prot7	prot8
info1	1.00																	
info2	0.64	1.00																
info3	0.74	0.60	1.00															
info4	0.63	0.58	0.71	1.00														
info5	0.39	0.43	0.37	0.45	1.00													
info6	0.46	0.36	0.46	0.43	0.48	1.00												
info7	0.35	0.30	0.38	0.38	0.43	0.69	1.00											
info8	0.59	0.59	0.62	0.61	0.40	0.48	0.42	1.00										
info9	0.59	0.58	0.57	0.58	0.41	0.51	0.44	0.73	1.00									
info10	0.41	0.48	0.40	0.46	0.45	0.36	0.36	0.51	0.58	1.00								
protect1	0.33	0.17	0.31	0.24	0.17	0.38	0.31	0.27	0.21	0.07	1.00							
protect2	0.26	0.17	0.26	0.19	0.12	0.24	0.18	0.23	0.15	0.07	0.57	1.00						
protect3	0.09	0.07	0.07	0.06	0.04	0.07	0.04	0.10	0.08	0.02	0.26	0.34	1.00					
protect4	0.27	0.16	0.24	0.18	0.15	0.24	0.18	0.21	0.16	0.06	0.38	0.37	0.24	1.00				
protect5	0.14	0.09	0.13	0.11	0.09	0.15	0.10	0.14	0.12	0.05	0.26	0.26	0.35	0.34	1.00			
protect6	0.37	0.23	0.37	0.26	0.20	0.39	0.31	0.31	0.28	0.12	0.48	0.33	0.14	0.31	0.23	1.00		
protect7	0.24	0.15	0.23	0.14	0.12	0.22	0.17	0.20	0.19	0.08	0.21	0.15	0.17	0.19	0.20	0.45	1.00	
protect8	0.42	0.25	0.39	0.28	0.20	0.36	0.27	0.32	0.29	0.13	0.42	0.34	0.16	0.36	0.23	0.64	0.51	1

Table 2.14: Principal Component Analysis - Factor Loadings

Variable	First Principal Component	Second Principal Component	Third Principal Component
info1	0.31	-0.10	0.11
info2	0.27	-0.19	0.25
info3	0.31	-0.12	0.10
info4	0.29	-0.18	0.17
info5	0.22	-0.16	-0.08
info6	0.27	-0.06	-0.35
info7	0.24	-0.08	-0.38
info8	0.30	-0.15	0.12
info9	0.29	-0.19	0.08
info10	0.22	-0.24	0.13
protect1	0.20	0.31	-0.04
protect2	0.17	0.33	0.21
protect3	0.09	0.37	0.38
protect4	0.16	0.30	0.15
protect5	0.12	0.34	0.29
protect6	0.22	0.26	-0.35
protect7	0.16	0.26	-0.28
protect8	0.23	0.26	-0.29

Table 2.15: Principal component values by industry

	Principal components mean values by industry		
	pc1	pc2	pc3
Mining	-1.71	-0.00	-0.08
Food, tobacco	-0.07	0.11	0.33
Textiles	-0.15	0.12	-0.15
Wood, paper	-0.19	-0.24	0.39
Chemicals	1.50	0.27	-0.51
Plastics	0.30	0.31	0.31
Glass	0.13	0.00	-0.09
Metals	-0.01	0.07	0.04
Machinery	0.96	0.59	-0.02
Electrical equipment	1.32	0.22	-0.04
Medical equipment	1.71	0.37	-0.54
Transport equipment	1.07	0.42	-0.23
Furniture	0.77	0.67	0.80
Wholesale	-1.22	-0.17	0.28
Retail, automobile	-1.56	-0.14	0.28
Transport, communication	-1.35	-0.15	0.10
Banking, insurance	-0.09	-0.76	0.34
IT	0.66	-0.17	-0.04
Technical services	0.19	-0.11	-0.50
Firm-related services	-0.29	-0.15	-0.04
Other services	-1.24	-0.22	0.03
Real estate	-1.48	-0.18	0.15

Chapter 3

Delegation and Performance[1]

3.1 Introduction

In the field of organizational theory, the choice between vertical and horizontal organizational structures is often considered a crucial one for firms' performance. This choice includes an important trade-off. More horizontal, less vertically integrated organizations prove to be more effective in using "soft" information, compared to more hierarchical organizations. I refer to "soft" information as, to a large extent, non-verifiable and non-transferable to the higher levels of the hierarchy, but the use of this information can increase the effectiveness of decisions taken. For example, if an organization has offices across a wide geographical area, it has to be able to adjust to changing local conditions, to compete successfully and to preserve its regional market shares. In this case, making the organization more horizontal by giving more decision power to the local offices seems to be a reasonable response to fast changing local conditions, which may not be fully observed by headquarters[2]. On the downside, more horizontal organizational structures have higher agency costs: incentives of the local managers rarely coincide with the incentives of the owners. With more authority delegated to the lower levels of the hierarchy, costs and scope of monitoring local offices become higher, and it becomes more difficult to coordinate and implement company-wide policies.

[1]**This Chapter has been previously published in CERGE-EI working paper series: Senyuta, Olena. 2013. Delegation and performance. Economics Institute, Academy of Sciences of the Czech Republic, WP497**

[2]I refer to the lowest level of the organizational hierarchy as local managers, and I refer to the highest level of hierarchy as headquarters, top managers, organization decision-makers or owners.

The problem of authority delegation in organizations has attracted substantial attention from a theoretical perspective, but empirically the importance of delegation remains unestablished. Empirical results suggest that the optimal choice of organizational structure should be considered in the context of a specific industry and market. The problem is the very broad definition of "soft" information, and the fact that organizations operating in different environments can depend on "soft" information to very different extents. Therefore, the benefits of a more horizontal organizational structure could differ substantially across markets. Moreover, the monitoring costs of more horizontal organization may also vary across markets. It is not surprising that different businesses and organizations try to establish their own optimal organizational structure. In such structures, the level of authority delegation could differ substantially across organizations, and also for a single organization over time, as a natural response to changes in the business environment. Therefore, it is important to note that studying the relationship of the organizational structure and performance is limited in many cases to a specific market or industry or even sometimes to the specific organization studied (if that organization represents an exceptional example of organizational structure for an industry).

This paper presents new empirical evidence on the relationship between organizational structure and performance, using the banking industry as an example. Since the 1990s, the banking sector has gone through important changes related to its competitive environment, which have resulted in substantial consolidation in the sector. Thus, large banking institutions have been created with multilevel hierarchical structures, where several managerial layers separate the decision-making agents from the agents who implement those decisions. Advances in informational technologies have played an important role in this consolidation. As information transfer has become cheaper, the costs of having more hierarchical (vertical) organizational structures have also decreased: banks have introduced standardized loan products and have developed the credit-scoring borrower evaluation tool, etc. Changes in market regulation policies have also contributed to the consolidation taking place in the industry (a single-market policy in the European Union, the Riegle-Neal Act in USA, which allows interstate bank mergers). Several studies confirm that decision making in the banking sector has become more centralized, the delegation of control (in lending decisions) has decreased, and the agency costs of vertical distance have also decreased (Berger and DeYoung, 2006). The current trend in the industry is to transfer all decisions on loans to the head offices. This represents the highest level of hierarchy possible in the industry.

While several studies exist which demonstrate that delegation can be used to influence incentives inside an organization, not many empirical studies have investigated how the level of authority delegation influences organizational performance. This paper will address the question whether delegating more authority to the lower levels of a bank hierarchy (local bank offices) leads to an improvement or worsening of performance, both in quantitative and qualitative measures. A simple comparison of the outcome variables (performance) for the local offices with different levels of authority delegated to them might be misleading. The reason is the decision to delegate authority is not randomly distributed among local offices, but is endogenously determined and adjusted by headquarters, based on (possibly unobservable to the researcher) characteristics of the local market (competitive pressure, distribution of market shares, market growth perspectives) and the local branch (experience, monitoring costs, ability of the local office to adapt to local conditions)[3]. Therefore, it is important to account for possible endogeneity of the delegation decisions while estimating the effects of authority delegation on organizational performance.

An important reason why the level of authority delegation in the banking sector might play a crucial role is that in banking, information collected by the local offices is very important for the evaluation of loan applications, which is especially true for small and medium enterprises (Berger and Udell, 2002). Moreover, this information is, to a large extent, "soft" and cannot be costlessly transmitted to the highest levels of a hierarchy. In addition, the banking sector is characterized by strong competition, and local expert knowledge can increase the effectiveness of decision-making. Therefore, this study aims to describe how the level of authority delegation in the banking sector is related to the results of bank activities. This paper also contributes to the literature on relationship banking[4]. More precisely, I provide empirical evidence that the level of lending to an SME significantly depends on the level of authority delegated to the local bank branches. To the extent that the local branch with more authority in lending is more likely to invest in its relationship banking and develop closer relationships with local clients, this paper demonstrates that more authority delegated to the local branch might increase the importance of relationship banking.

The rest of the paper is organized as follows. Section 3.2 reviews theoretical and empirical literature on the effects of authority delegation in the organization. Section 3.3

[3] I refer to those characteristics as a business environment.

[4] For a review, see for example, Boot (2000) and Boot and Thakor (2000).

describes the empirical model specification, characterizes the dataset, and discusses the estimation results. Section 3.4 concludes by discussing how the estimation results could be related to the theories of optimal organizational structure choices.

3.2 Literature Review

Authority delegation has been proven to be an important instrument of shaping managerial incentives. Mookherjee (2006) presents an extensive overview of the literature related to the incentive benefits of delegation. The author concludes that assuming information communication is costly, an upper bound on the size of the message to be communicated exists, and decentralized decision-making can access much more information compared to centralized decision-making. In other words, if communication costs are introduced, delegation is more successful in utilizing "local" information. Aghion and Tirole (1997) build a model where they differentiate between formal authority (the formal right of top managers to make decisions) and real authority (the real control of local managers over decisions due to better information and knowledge of alternatives). The model demonstrates that delegating formal authority to subordinates would be beneficial because it increases their effort and initiative in collecting more information about alternatives, and and subordinates benefit by empowerment and remuneration for their relationship-specific investments. An illustration: local loan managers can exert an effort and obtain "soft" information about borrower — this will help to estimate loan risks more precisely. However, this soft information cannot be transferred to the higher levels the decision-making hierarchy because of its nature[5]. Also, the loan managers' investments in soft information production are not observed by other parties. As long as top-management would not use this information in their decision making, loan managers do not have incentives to generate soft information. Delegating can help to shape the managers' incentives such that they produce loans of higher quality using soft information and exert additional effort to obtain it.

Similarly to this study, Stein (2002) motivates his theoretical model of authority delegation to local managers using the example of the banking sector. The author draws attention to the concentration trend in the industry and associates it with a decline in

[5]Aghion and Tirole (1997) make clear the distinction between hard information, which can be relatively costlessly communicated and verified, and soft information, which to a large extent represents a pure suggestion.

the lending to small businesses by larger banks. The model assumes that local managers have the research advantage for "soft" information about projects; therefore, smaller organizations, which have fewer subordination levels, are more efficient in providing services which are sensitive to "soft" information. Also, holding the size of the organization fixed and changing such characteristics of the environment as the "softness" of information, Stein (2002) demonstrates that having a flatter[6] organization is more advantageous. Moreover, the model shows that increasing the number of managerial layers between decision-makers and local managers leads to an increase in the number of unnecessary bureaucratic procedures, such as the effort spent on "soft" information documentation.

The assumption of costly communication is crucial for the theoretical conclusions described above. In the literature, it is justified by several examples: limited ability (Radner, 1993); coordination costs (Becker and Murphy, 1992); managers' costs of communication (overload); and the costs of learning (Aghion and Tirole, 1997; Garicano, 2000; Stein, 2002; Dewatripont and Tirole, 2005).

There is large empirical evidence for "soft" information usage in relationship to a bank's organizational structure. For example, Sapienza (2002) studies the effect of bank mergers on the credit availability for small businesses. The paper finds that larger banks decrease lending to small firms more than smaller banks. This provides evidence for the relative efficiency of "soft" information usage in small, therefore, less hierarchical banks. Carter and McNulty (2005) provide empirical evidence that small banks have the advantage in small business lending (have a higher net return) using data on the lending activities of US banks from 1993-2001. Berger and Udell (2002) summarize earlier empirical research on the effects of bank mergers on credit availability for small borrowers and relationship lending practices. The authors conclude that empirical evidence exists that shows a reduction in lending to small firms due to bank consolidations, but this lending reduction might not be economically significant in some cases. Whether lending reduction is significant depends on external factors such as market conditions. The authors hypothesize that in some cases, a reduction in relationship lending by larger banks will be accompanied by an increase in relationship lending from other (smaller) banking institutions, and there would be no adverse effect of bank consolidation on loan availability for small firms in the market.

In addition to the studies which implicitly consider larger banks less efficient in "soft" information usage, and therefore conjecture that there should be less information-sensitive

[6] A "flatter" organizational structure has less hierarchical levels.

lending by larger institutions, there are some studies which directly estimate "soft" information usage in the lending decisions by banks. Berger et al. (2005) use firm-level data from the National Survey of Small Business Finance and study the effects of a lending institution's organizational structure on the lending conditions for specific firms. They find that (1) firms communicate with larger banks in a more impersonal way (do not often meet in person but use mail); (2) firms that borrow from smaller banks also have more exclusive and longer lasting relationships with their banking institution; and (3) firms that borrow from larger banks are more credit constrained. Those findings are in line with theoretical predictions that smaller banking institutions are more efficient in providing services that require "soft" information.

Alessandrini et al. (2009) use data on SMEs surveyed by an Italian banking group. The authors find that a larger "functional distance"[7] between local bank offices and the bank headquarters decreases the credit availability for local firms. Liberti and Mian (2009) test the Aghion and Tirole (1997) model's predictions on the use of "soft" information in the loan approval process, using the banking industry, and they use information on corporate loan applications from a large multi-national bank in Argentina. They find that if a loan is approved at the higher levels of the hierarchy, the size of the loan is more sensitive to the "hard" information on the applicant and less sensitive to "soft" information, as compared to loans approved at the lower levels of the hierarchy. Also, this decrease of "soft" information importance is not observed in the data if the information collecting officer is located in the same geographical area as the decision-making loan officer. These results support the view that some part of non-verifiable information is lost as it is communicated to the higher levels of a company's formal hierarchy.

Interesting empirical results are provided in Canales and Nanda (2012), who find that banks with more decentralized organizational structures issue larger loans to firms in part relying on "soft" information compared to centralized banks. However, those decentralized banks also issue smaller loans in a more competitive environment compared to centralized banks. In addition, decentralized banks issue loans to larger firms (therefore, with more "hard" information) in a more competitive environment. The authors conclude that decentralized banks tend to cherry-pick their clients as a response to larger competitive pressure.

As stated above, generating soft information is a costly activity for agents, and agents

[7] "Functional distance" is the distance between the local branch where information on the borrower is collected, and the bank headquarters.

need to be provided with enough incentives to exert this effort. If providing these incentives is too costly for the principal, he may choose not to provide them, there may be more decision centralization, and no soft information would be produced. Why the incentive costs of authority delegation may become more or less important can be explained by the changes in the firm's business environment. Therefore, it would be natural to expect that organizational structure (and the level of delegation) would evolve as a response to the changes in the environment.

Theoretical predictions in this field are ambiguous. For example, Marin and Verdier (2008) construct a general equilibrium model, in which the decision about organizational structure is endogenous and is optimally determined by the firms. The model demonstrates that intermediate levels of competition are associated with the highest levels of authority delegation. Legros and Newman (2008) investigate how other aspects of the business environment — market liquidity and productivity shocks — can influence the choice of optimal organizational structure. The authors demonstrate that positive shocks and the unequal distribution of liquidity will result in less control delegation.

In the banking sector, the centralization of loan decisions due to competitive pressure can be explained by the accompanying growth in technical advances. If information transmission becomes easier (through numerous intra-bank security enhanced networks), the costs of centralization decrease sufficiently. There are not many empirical studies which investigate how organizational structure evolves in response to the changes in the business environment. This is mainly due to a lack of panel or cross sectional datasets which would describe in detail the organizational structure of a company, and to the complexity of this information and its privacy. However, some general conclusions were made by Degryse and Ongena (2007) and Degryse et al. (2009). Studying extensive data on the Belgian banking industry, they discover that tough price competition in the region is usually combined with more hierarchical structures of competing banks.

The problem addressed in this paper is rather different. The choice of organizational structure, which is the level of authority delegation to the local management level, is an optimal decision of the bank, and it is based on the current and future characteristics of the market environment and the market strategy chosen by the bank. In such conditions, it is interesting to measure whether different levels of delegation would result in differences in performance. If yes, and if a more decentralized organizational structure (for example) is associated with better performance, the natural question remaining is why all firms do not choose to use a decentralized organizational structure. There is empirical evidence that the organizational form influences performance even when it is selected optimally. Mullainathan and Scharfstein (2001) demonstrate that vertically integrated producers make different investment decisions compared to non-integrated producers. Krueger (1991) finds that employees in company-owned food chains experience a steeper tenure-earnings profile than employees in franchised chains. Kosová et al. (2013) provide a performance comparison between franchised and company-owned hotels, and find that there are significant differences in performance measures (revenue per room, price, occupancy rate, etc.) between the two structures, but those differences are not significant after the authors endogenize the choice of organizational form.

3.3 Econometric Model

I use the performance of a bank's local branches as the response variable, while the level of authority delegated is the main explanatory variable of interest. I use the heterogeneity in the branches' authority and changes in market conditions for regions to estimate the impact of the authority delegation level on the decision-making process in branches. Branch performance is a function of the level of delegation, controlling for the factors which I describe as market environment variables. The model in the panel setting is as follows:

$$Y_{it} = \alpha_1 + \beta_1 D_{it} + \beta_2 X_{it} + \gamma M_{it} + u_i + \varepsilon_{it}, \tag{3.3.1}$$

where i is the index for branch; t is the time period (month); Y_{it} is quantitative and qualitative data on branch performance; D_{it} is the level of delegated authority for the branch; X_{it} is the vector of controls at the regional level; M_{it} is the vector of unobserved local market and branch conditions in the region at a given time, and possibly correlated with other explanatory variables; u_i is the fixed effect for the region; and ε_{it} is the error term.

The level of delegated authority for a branch is clearly an endogenous variable in my model. It can be illustrated by the following example. When headquarters makes a decision about the level of authority delegated to each branch, they consider X_{it} (observed regional characteristics) and M_{it} (local market and branch conditions unobserved by the econometrician). As a result of estimation (3.3.1) without being able to include M_{it} in the regression, I obtain $\mu_{it} = \gamma M_{it} + \varepsilon_{it}$ in the error term. By the decision process of headquarters, M_{it} and D_{it} are correlated, which means that μ_{it} and D_{it} are correlated in

(3.3.2), and estimates b_1 in the regression

$$Y_{it} = a_1 + b_1 D_{it} + b_2 X_{it} + u_i + \mu_{it} \tag{3.3.2}$$

will be a biased estimate of β_1 from (3.3.1).

To obtain consistent estimates of β_1, I use the instrumental variable vector Z_{it}. Such instruments should be important factors in determining the variation in authority that I observe for different branches, but should not have a direct effect on branches' performance. As the excluded instruments Z_{it}, I use the distance of a regional branch from its headquarters and the number of months the bank was present in the region (duration). I assume that this distance is a proxy for monitoring costs and positively influences the decision to delegate authority, and duration is a measure of how experienced management is with the local market and how much headquarters knows about the local market so that they can delegate more authority to the local level. For Z_{it} to be a valid instrument, I need the distance and duration to be uncorrelated with the performance of branch Y_{it}. This assumption is valid in cases where the local market conditions, local managers' experience, and performance (which determines the outcome variable) are distributed randomly across the country and are not related to the branches' location with respect to headquarters or to how long the bank has been present in the local market. I use the IV estimation to estimate the model with distance and duration as excluded instruments[8].

Thus, I exploit the heterogeneity in the branches' authority and changes in market and branch conditions to identify the impact of regional market characteristics on delegation level for branches, and further, to estimate the effect of the authority delegated on the performance characteristics. More precisely, I am using an instrumental variable regression: In the first stage, I am estimating the authority delegation decision, and in the second stage, the impact of the authority delegated on the branches' performance.

3.3.1 Data

To test the effect of the authority delegated on the performance variables, I combine three monthly panel data sets, which cover 2004-2008:

(1) data on all loan contracts and loan performance, loans for small and medium enterprises (SMEs) and loans for private individuals (PIs) for all branches of one representative bank in a European country. I refer to these as performance variables;

[8] I discuss the validity of instruments further in subsection 3.3.2.

(2) for the same representative bank, monthly data on the level of lending authority delegated to regional branches, separately for SME and PI loans, and other characteristics of the branches. This level of authority is measured as the highest level of the loan size (I refer to this level as "limit" in the text below), for which the branch can make independent decisions without consulting with its headquarters (measured in EUR). The two data sets described above are privately obtained data from a commercial bank which operates in a European country, has a well-developed branch network, serves all types of clients and businesses, and has the second biggest market share in that country;

(3) panel (monthly) data set of market characteristics for a particular region: economic conditions and risks and the level of competition from other lending institutions. This information is obtained from statistical agencies and the financial registry of the unnamed country.

A detailed definition of the variables is provided in Table 3.1 in the Appendix, which divides the variables into two groups: performance and explanatory variables. Each bank-related variable is obtained separately for both SME and PI loans. Further, I shall discuss in more detail how the specific performance measures were generated and why they were selected.

First, I divide the performance measures into two groups: **quantitative** and **qualitative**. Quantitative performance measures include (1) the number of loans (per office) approved by the branch during a period of one month, and (2) the sum of loans (per office) approved by the branch during a period of one month (in EUR). The two measures are expressed in per office terms because each regional branch, which makes the decisions on the loans based on their delegated limits, may have many regional offices, which collect loan applications. Therefore, the branch with a larger office network will naturally have the larger number and amount of loans originated in a specific month. To control for this effect, I measure all quantitative variables in per office terms. Unfortunately, I do not have a measure for specific office size. Therefore, I treat all offices as equal in size.

Qualitative performance measures include (1) the average days the loans approved by a particular branch during a particular month (all offices of the branch) are overdue after 6 months of being on the books, and (2) the percentage of loans that became performing after having been recognized as non-performing during a period of 90 days, and (3) the number of loans that are overdue more than 30 days out of those approved by a particular branch during a particular month (all offices of the branch). In order to generate these performance measures, I created a series of loan pools. The pools consisted of all loans (SME or PI) approved by a specific branch in a specific month. Then, I followed each pool over time (for 3 months, 6 months, etc). After some loans from this specific pool became non-performing, I have calculated the mean days overdue and the quantity of non-performing loans for this pool of loans. This procedure allows me to trace the performance of all loans which were generated in the specific month when the regional branch was experiencing a high or low level of delegated authority.

Figure 3.1 below describes how selected variables vary over time. On the horizontal axis is a time unit (month), and on the vertical are the average values (across the branches) for the main outcome variables (quantitative): the number of generated loans per month, total amount of loans approved per month, and the average limits across regional branches. Again, all the values of outcome variables are calculated per one selling point (office) because there could be more than one office (selling point) operating under each regional branch authority.

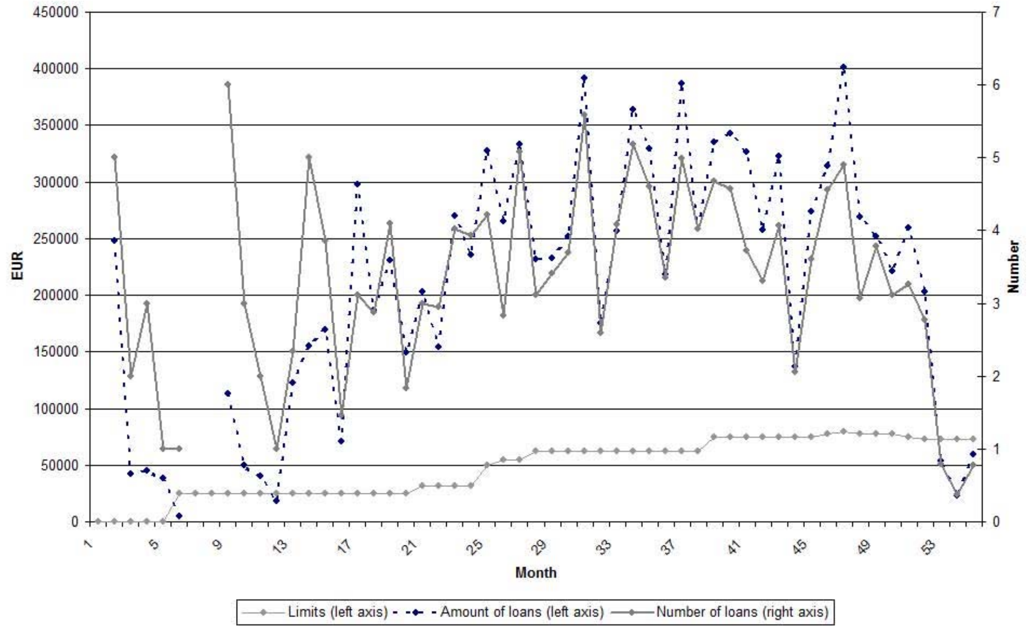

Figure 3.1: Quantitative Performance Measures and Branch Limits Over Time

From Figure 3.1 I infer the average limits increased over time. At the same time, I observe that for the periods with higher limits, I have both a higher average number of loans approved (per office) and a higher average amount approved (per office).

The next two figures describe a similar pattern for the qualitative data (loan performance): the mean number of days the loans approved by the branch are overdue, the percentage of non-performing loans out of those approved by the branch, and the recovery rate. The recovery rate measures what percentage of loans became performing after they had been recognized as non-performing during some period of time (a ninety days' recovery rate is considered here). The observation period is shorter here compared to the quantitative measures because I have to observe the loans for a certain period of time before I can make conclusions about their quality. From the two figures, I conclude that

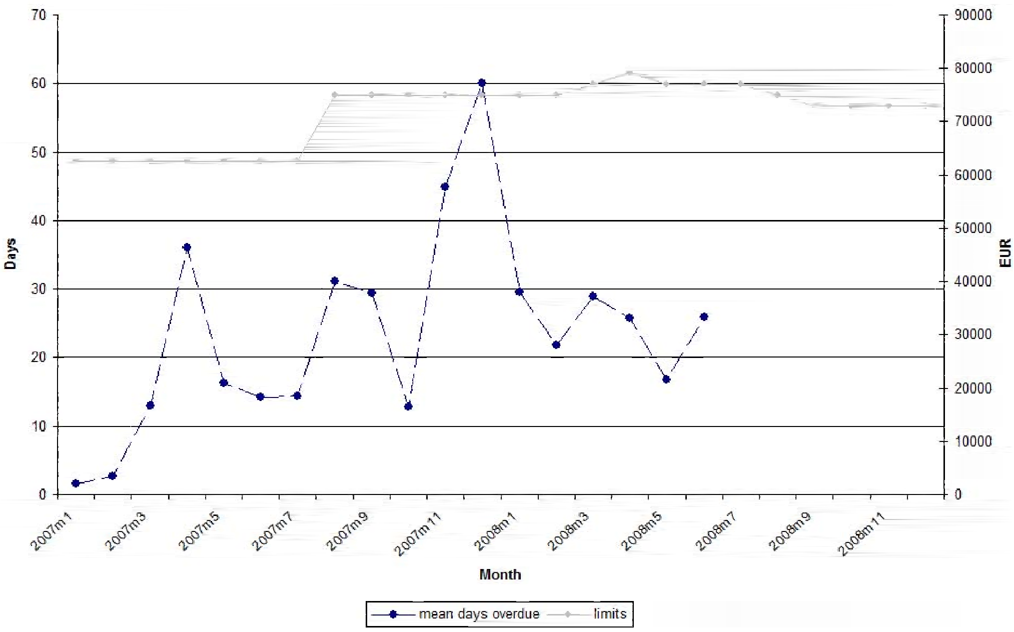

Figure 3.2: Mean Days Loans are Overdue and Branch Limits Over Time

an increasing trend exists in the average days the loans are overdue, and at the same time, average limits also experience an increasing trend. Also, a trend to a decreasing recovery rate is observed, and no clear conclusion can be drawn from the average values of the share of non-performing loans.

With these primitive observations, I could conclude that a positive influence of delegated authority (limits) exists on quantitative performance — more loans are issued. However, there is some evidence that increased authority leads to a worse quality in decisions made — more non-performing loans, longer overdue periods, and lower recovery rates.

Table 3.2 provides the unconditional mean comparison for the observations (for SME loans only) with zero and positive limits. A comparison demonstrates that the level of limits has an important effect on the performance characteristics. Such measures as the

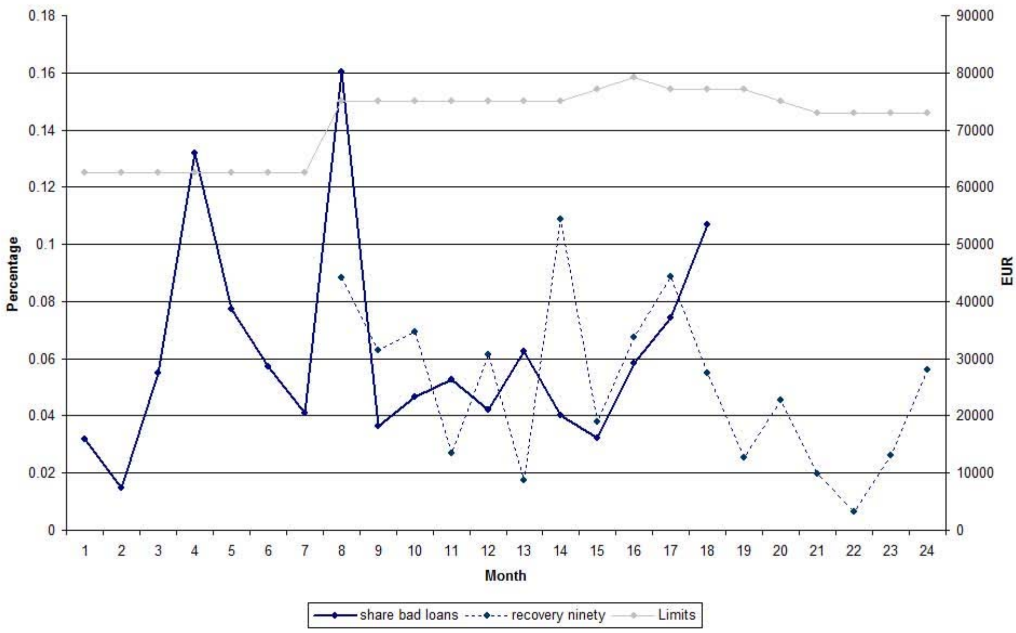

Figure 3.3: Share of Non-performing Loans and Recovery Rate

number of loans originated by the branch and the number of loans originated by the branch per office, the total sum of loans, overdue measures, and non-performing loans measures differ dramatically between the observations with zero and non-zero limits. Interestingly, mean loan size and the recovery measure do not differ between the two groups. This might be considered quite surprising because a branch with high limits is expected to approve larger loans, and this should lead to an increase in the average loan size.

However, such a comparison of means could be misleading because observations with different limits are characterized not only by very different performance measures, but also by very different explanatory variables. For example, observations with non-zero limits have a higher amount of offices for each branch (1.5 versus 4.3), have twice as high a number of competitors in the regional market, have twice as long a presence on the regional market, and are located further from headquarters. Therefore, explanatory variables are closely related to the level of branch limits. As those explanatory variables could be also related to the performance measures, if I do not account for this association, I would over-estimate the effect of limits on performance (a real effect would come from exogenous explanatory variables like market characteristics, but I would incorrectly attribute it to the effect of limits).

The section below will check those predictions about the effect of limits more rigorously. More importantly, I shall test how the level of delegated authority is related to the results of bank activities, controlling for the endogeneity problem, and shall compare the results of the instrumental variable estimation to the simple OLS and FE specifications. I will then draw conclusion regarding the possible bias in the OLS and FE estimates.

3.3.2 Empirical Results

3.3.2.1 Regression Estimates (Quantitative)

Table 3.3 and 3.4 in the Appendix provide the estimates for the quantitative measures (the number of loans and the total amount of loans per office) regressions, both for SME and PI loans. Different specifications were used in the estimation: simple OLS (pooled panel data), fixed effect, and IV specification.

Real Wage Index positively influences performance measures, sign and size of the coefficient does not change in different specifications, which provides evidence that the unobserved heterogeneity and omitted variables that possibly bias the estimates do not correlate with economic conditions in the region. Therefore, those omitted variables are bank-specific characteristics rather than the characteristics of the economic conditions in the region. The results of PI regression are similar to that for SME with respect to the index of the real wage coefficient.

The measure of competitive pressure, the *Number of Competitors* in the region, considerably changes size and sign if I use a fixed-effect specification. This provides evidence that there is a downward bias in the OLS regression if I do not control for unobserved branch effects. Moreover, it is clear that unobserved fixed effects strongly and negatively correlated with the number of competitors variable. Therefore, unobserved fixed effects should be positively related to the outcome variable (quantitative performance) for the bias in the OLS regression to be negative. Establishing that omitted fixed effects are positively influencing performance and negatively related to the number of competitors, it is possible to hypothesize what exactly those fixed, unobserved effects are. They could be related to the market position of the respective bank. The higher the bank's market share, the better the bank is established in a particular region, and the higher loan origination activity is (a quantitative measure). Also, the number of competitors would be negatively related to the strength of the bank's market share (the degree to which the bank is recognized and established). Therefore, not accounting for the unobserved fixed effects brings a downward bias to the coefficient estimates. The results of the PI regression are similar to the SME regression with respect to the number of competitors coefficient.

For the *Number of Offices* variable, both linear and quadratic terms were included in the regression to control for a possible non-linear relationship between the number of offices and quantitative performance. Indeed, if only a linear term is included, the number of offices has positive effect on the quantitative performance, but including the quadratic term makes the coefficient for the linear term negative and positive for the quadratic term. For example, for the SME[9] regression, such coefficients mean that the effect of the number of offices on quantitative performance is described by a quadratic parabola with its minimum at approximately 15. It reflects the fact that the origination of loans is a non-linear technology, so that opening more offices does not necessarily equally bring an additional number of loans. As the number of offices is low (below 15), opening one more office brings fewer loans per office (and a lower amount of loans per office), but later after more offices are opened (above 15), each additional office is associated with a higher per-office quantitative performance. This provides evidence that the bank is working in the region first to establish its clientèle and presence, but with a higher number of offices, the bank can expect to increase its performance.

The *Distance* and *Duration* variables are not related to the performance in any of the regressions. Those variables are present only in no-IV regression because I am using them as the excluded instruments in the IV regressions. I discuss these results in more detail in the subsection 3.3.3 in relation to the discussion of the instrumental variables validity.

The *Limits* variable is the main variable of interest, and it is positively related to quantitative performance in both SME and PI regressions. The economic effect is also not negligible. For example, increasing limits on 1 EUR increases the amount of originated loans per office per month by more than 1 EUR. Taking into account statistics from Table 3.2, regarding the average loan amount originated, this represents quite an important effect of the limits on the quantitative measures. This result suggests that more authority delegation to the regional branch improves quantitative performance.

As explained, branch limits are potentially an endogenous variable because the variable is not chosen randomly by headquarters, but is based on the current and expected

[9] For the PI regression, results are very similar with the only difference being the minimum of the parabola is at 12 offices.

market conditions and strategies. After I have controlled for the endogeneity in the limits variable, the coefficients for limits become higher (but not significant for FE IV specification). Therefore, the bias in the limits variable is negative. In fact, the effect of the limits on the quantitative measures is even more positive. This bias emerges from the fact that some unobserved characteristics (for example, a strong bank position in the market) positively influence the outcome variable, while limits are lower for observations where market position is strong. This devaluates the effect of the limits on the outcome variable because limits are high for those observations where the market position is weak, and this artificially decreases the limits effect on performance.

Results regarding the limits variable are similar for PI and SME regressions. However, it is important to note that the effect of limits is weaker for the SME regression (size of the coefficient). The explanation is the following: Limits delegated to the regional branches are much higher for SME loans because SME loans are also higher in their amount compared to PI loans (Table 3.7). Therefore, the same $1,000$ increase in SME limits has a much lower effect compared to the same increase in limits for PI loans.

3.3.2.2 Regression Estimates (Qualitative)

Further, I analyze the effect of authority delegation on the quality of approved loans. Tables 3.5 and 3.6 in the Appendix present the estimation results for qualitative measures as the outcome variable (mean days overdue, recovery measure, and the number of loans that are more than 30 days overdue).

The *Index of Real Wage* variable has noisier estimates in the quality regressions compared to the previous subsection. Using both PI and SME estimates, I cannot say that the real wage index positively or negatively influences qualitative performance. There could be some evidence that the higher real wage index is associated with a higher probability of loan recovery from a non-performing stage, but this effect is only significant for the SME regression, not for PI.

Number of Competitors is not significant for qualitative SME regressions, reflecting the fact that competitive pressure is not important for the quality characteristics of SME loans. Banks rather compete on the extensive margin (for customers), but not for better customers. However, for PI loans, the number of competitors is negatively related to the recovery rate and increases the number of loans which are more than 30 days overdue. Thus, a more competitive environment decreases the quality of loans to PI.

The *Number of Branch Offices* variable is also not significant for SME loans; there is only limited evidence that the number of offices in the region is significantly influencing quality of PI loans measures. *Distance* and *Duration* are not related to the performance measures in any of the regressions.

The main predictor variable, *Limits*, is positively related to *Mean Days Overdue* measure. Increasing the limits by $1,000$ increases mean days overdue by 0.11 for the OLS SME regression (and by 0.26 for fixed effect regression). For the PI regression, the coefficient is 0.23 days (same size for the FE regression). This effect of limits on the mean overdue days is also economically significant because the average mean days overdue is 2.5 for branches with zero limits and 35 for the branches with positive limits (Table 3.2).

As OLS and FE specifications suggest, more authority leads to lower quality decisions: an increase in the mean days loans are overdue. A comparison of the results of IV estimates with non-instrumented regressions suggests that the endogeneity problem influences the coefficient for the branch limits variable. The limits effect on the overdue measure is not significant if I instrument the limits variable. For example, after observing that market perspectives in a particular branch are deteriorating (due to economic conditions, or competitors' behavior), headquarters may decide to delegate more authority to such a branch in order to use all local knowledge potential. Deteriorating market conditions, then, lead to both: a worse qualitative performance and more authority delegated to the region. As a result, if I neglect this endogeneity in the delegated authority, I will obtain a negative relationship between performance and authority, which is, in reality, not there. Therefore, it is important to control for the endogeneity of the limits variable.

Limits also have a positive effect on the number of loans that are *Overdue More Than 30 Days*, out of those loans approved by a particular branch during a particular month (in all offices of the branch). For example, if limits are increased by $1,000$, the number of loans that will be overdue more than 30 days increases from 0.008 to 0.011 (for different specifications), and the coefficient is significant for both SME and PI regressions, but correcting for the endogeneity in limits makes the value of the coefficient insignificant.

It is important to note that the effect of limits variable is weaker for the SME regression (the size of the coefficient). This is explained by the fact that such limits are much higher for SME loans (Table 3.7) because these loans also have on average a higher amount. Therefore, the same $1,000$ increase in limits for SME loans has a much lower effect compared to PI loans.

Further, regression results demonstrate that limits have no significant effect on the

Recovery Rate variable. This and the evidence from the above regression discussions illustrate the fact that limits influence quantitative measures in the first place as the level of delegation is positively related to loan-generation activities. On the other hand, higher limits might also lead to a deteriorated quality in the loan pool the branch originates. Correcting for the endogeneity of the limits variable reveals that the quality measures are no longer affected by the limits. Therefore, the regression provides only weak evidence that higher limits lead to a accumulation of loans with a higher default potential in the pool of all originated loans.

The results could be interpreted in the following way. I demonstrated that there is positive effect from the level of delegated authority on quantitative performance measures. Also, evidence is weak for a negative effect from the level of delegation on the quality of originated loans. The bank was not constrained in its delegation decisions but made rather optimal choices in its organizational structure. Therefore, the question stated at the beginning of this study remains: if more delegation positively influences performance, why does the bank not choose the optimal level of delegation, which would lead to a best outcome? This paper suggests that there might be a trade-off between quantitative and qualitative performance characteristics. While the bank is able to increase its loan generation with higher limits, there is also some (weak) evidence of loan quality deterioration. Therefore, the bank would optimally change the level of delegation to its regional branches in order to find the balance between the two effects. A more aggressive (market share expansion) market strategy could be associated with the bank providing higher limits to local branches, and the loss-avoidance strategy could be associated with more centralized decision making (lower limits to the local branches).

3.3.3 Instruments Validity

The first stage regression estimates for *SME loans* are provided in the Appendix, Table 3.8 (panel 1 for SME loans). These results confirm that the authority delegated (the level of limits) to the branch for SME loans is strongly positively related to the duration variable (how long the bank was present in the particular market), but the effect of distance from the branch to headquarters is positive and not significant. In fact, being present in the market for one more month increases the level of limits the branch receives by approximately $1,200 - 1,500$ EUR for SME loans and $500 - 600$ EUR for PI loans. The first-stage regression accounts for approximately 40% of the variation in limits for

SME loans (but much less for PI loans). The F-test of excluded instruments also confirms that they are strongly related to the limits variable (they are not weak instruments).

In addition to this, a valid instrument must not be correlated with the outcome variable in my model. For pooled OLS specification, Hansen J statistics (J-stat=0.0427, p-value=0.8362) confirm that the excluded instruments are orthogonal to the error term in the regression of main interest and are correctly excluded from the main regression.

In the FE IV specification, I use only one excluded instrument (duration) because the distance variable is time-invariant and becomes part of the unobserved fixed effects in the FE IV specification. In such a situation, I cannot check for instrument validity by applying the usual test of over-identifying restrictions to the duration variable. I check the relationship between the branch duration and performance measures by including the duration variable in the OLS and FE regressions as additional explanatory variable (Table 3.3 and 3.5). If distance and duration have a direct effect on the performance of the branch, I would expect these variables to be significant in each of the estimated main equations. However, in all cases, the effects of distance and duration on performance are statistically insignificant. Moreover, including these additional exogenous variable does not change the coefficients for other estimated variables at all in the regression. This confirms my conjecture that the impact of distance and duration on performance work only through the "delegation channel". In other words, distance and duration themselves do not have direct effects on the branches' performance, if I control for the level of delegated authority.

The first-stage regression estimates for *PI loans* are provided in Appendix, Table 3.8 (panel 2 for PI loans). These results confirm that the authority delegated (the level of limits) to the branch for PI loans is strongly related only to the duration variable, and the distance variable is not significantly related to the limits variable. The F-test of excluded instruments confirms that they are jointly strongly related to the limits variable, but in the FE specification, the F-value is quite low (3.66), and it is a boundary decision whether excluded instruments are strongly related to the limits variable (instruments might be weak for the FE specification, but are not weak for the pooled OLS specification). For the pooled OLS specification, Hansen J statistics (J-stat=10.72, p-value=0.001) do not allow me to conclude that the excluded instruments are orthogonal to the error term in the regression of main interest and are correctly excluded from the main regression (instruments are not valid for a pooled regression specification).

For the PI loans regression, the proposed instrumental variable *distance* does not affect the level of limits, and the duration variable remains strongly correlated with the limits variable. However, the distance variable cannot be excluded from the main equation because it is related to the outcome variable. The conclusion is that when making the decision on the level of limits for PI loans, the bank uses different considerations other than monitoring costs and local experience. The IV specification for PI loan regressions suffers from either a weak or not orthogonal instrumental variables problem. Therefore, it casts doubt on whether the procedure for correcting for the omitted variable bias in PI loan regressions was helpful and whether the results are conclusive.

3.4 Conclusion

This paper considers the problem of delegated authority using an example from the banking sector. Recently, competition in the banking sector has become more severe, and as a result, the banking sector has undergone many important changes in its organizational structure — decision making has become more centralized, and the delegation of control in lending has decreased. However, many studies demonstrate that delegation of loan decisions is often used by the banks to compete for customers and facilitate lending.

Regression estimates show that more authority delegated has a positive effect on quantitative measures of performance; however, it might also decrease the quality of decisions made. Different regression specifications were employed to control for the endogeneity of the delegation variable. For some qualitative performance regressions, after controlling for the endogeneity of the delegation level, these effects became insignificant, and for the quantitative performance measures, controlling for the delegation endogeneity made the estimates economically more significant. Therefore, results show that not controlling for the endogeneity problem might lead to false conclusions about the relationship between performance and authority.

Moreover, estimation results show that the validity of instruments remains an important problem for some of the regression estimates. Employed instruments (excluded) proved to be strong predictors of the delegation decision for the SME loans but are weak instruments for the PI loans delegation decision. This could reflect the fact that the bank uses different considerations other than monitoring costs (distance variable) and local market experience (duration variable) when assigning the loan limits to the PI loans as compared to SME loan limits. With no evidence of instrument validity, the consistency of estimates and the instrumental variable specification estimates for the case of PI loans might be questioned.

Regression estimates demonstrate that there is a positive effect of the level of authority delegated on the performance measures. Also, there is only weak evidence of a negative level-of-delegation effect on the quality of originated loans. This result is consistent with the bank's optimal authority delegation behavior as a response to changes in local market conditions. The bank was not constrained in its delegation decisions, and it made optimal choices for its organizational structure. This study shows that there might be a trade-off between the quantitative and qualitative performance characteristics. Therefore, the results are consistent with the following bank behavior: Headquarters optimally balance the level of delegation to the regional branch as a response to local market conditions. A more aggressive (market share expansion) market strategy could be associated with the bank providing more delegation to the local branches, and a loss-avoidance strategy could be associated with bank's decision-making being centralized (lower limits to the local branch).

3.A Appendix 3

5 Appendices

Table 3.1: Definition of Variables

Performance variables	
Number of loans	The number of all loans approved by the branch (all offices of the branch) during a period of one month
Amount of loans	The sum of all loans approved by the branch (all offices of the branch) during a period of one month (in EUR)
Mean loan size	Average size of the loan approved by the branch (all offices of the branch) during a period of one month (in EUR)
Number of loans per office	The number of loans per office approved by the branch during a period of one month
Amount of loans per office	The sum of loans per office approved by the branch during a period of one month (in EUR)
Mean days overdue	The average days loans (approved by a particular branch during a particular month, all offices of the branch) are overdue after 6 month of being on the books
Thirty plus	The number of loans that are overdue more than 30 days out of those approved by a particular branch during a particular month (all offices of the branch)
Recovery ninety	Percentage of loans that became performing after they were recognized as non-performing during a period of 90 days
Number of bad loans	Number of loans that became non-performing during a period of one month
Share of bad loans	Share of loans that became non-performing during a period of one month
Amount of bad loans	Total face value of loans that became non-performing during a period of one month
Share of the amount of bad loans	The share of the face value of loans that became non-performing during a period of one month
Explanatory variables	
Limits	The highest level of loan size, for which the branch can make an independent decision, without consulting headquarters (in EUR)
Number of offices	The number of offices the regional branch has in the region
Number of competitors	The number of other banks in the region
Real wage index	Real wage index in the region
Duration	The number of months the bank is present in the regional market
Distance	The (driving) distance from headquarters to the regional branch

Table 3.2: Comparison of Means for the Observations with Zero and Positive Limits

Variable	Observations with zero limits (20% of all obs.)	Observations with positive limits (80% of all obs.)	Difference in means is significant: *** (1%); **(5%); *(10%)
Number of loans	3.15 (2.37)	13.05 (11.4)	***
Amount of loans	256570.1 (259979)	925222.9 (897779)	***
Mean loan size	78214.44 (62622.8)	72804.93 (45929.9)	-
Number of loans per office	2.58 (1.96)	3.56 (2.3)	***
Amount of loans per office	215450.3 (225399)	250884 (199659)	-
Mean days overdue	2.31 (8.51)	31.51 (52.47)	***
Thirty plus	0 0	0.35 (0.83)	***
Recovery ninety	0.08 (0.25)	0.05 (0.12)	-
Number of bad loans	0.14 (0.52)	1 (1.55)	***
Share of bad loans	0.03 (0.09)	0.06 (0.09)	***
Amount of bad loans	13570.49 (48668.4)	102725.7 (203071)	***
Share of the amount of bad loans	0.03 (0.13)	0.07 (0.12)	**
Limits	0 0	102760.7 (37705)	*
Number of offices	1.49 (1.15)	4.3 (3.4)	***
Number of competitors	32.93 (7.01)	65.62 (32.48)	***
Index of real wage	100.9 (7.04)	100.77 (5.53)	-
Duration	8.03 (6.2)	18.16 (12.73)	***
Distance	413.13 (176.45)	540.45 (202.29)	***

Table 3.3: Quantitative Performance Measures for SME

	Number of approved loans (per office)				Sum of loans approved (per office)			
Variable	OLS	FE	IV	FE IV	OLS	FE	IV	FE IV
Limits[10]	0.0203***	0.0171***	0.0337***	0.0275	1206.2874***	1228.0067***	519.54315	2571.5921
	[0.0022]	[0.0033]	[0.0096]	[0.0141]	[205.44955]	[318.71793]	[894.50433]	[1355.3822]
Index of real wage	0.0479**	0.0515***	0.0489**	0.0524***	3243.186*	3666.8983**	3199.044*	3774.2971**
	[0.0145]	[0.0135]	[0.015]	[0.0137]	[1389.7309]	[1291.9851]	[1402.018]	[1316.1033]
Number of offices	-0.663***	-0.8046***	-0.8007***	-0.7664***	-39910.67***	-62346.504***	-31811.376*	-57421.056***
	[0.1007]	[0.1318]	[0.1678]	[0.1025]	[9618.2668]	[12595.807]	[15649.125]	[9867.8143]
Number of offices squared	0.0237***	0.0217**	0.0288***	0.0193***	1609.3899**	1581.4546*	1318.7987	1275.3809*
	[0.0058]	[0.0067]	[0.0076]	[0.0056]	[553.24065]	[643.93763]	[712.77628]	[541.40643]
Number of competitors	-0.0086**	0.3487***	-0.0112**	0.3213**	-680.20522*	29128.725***	-591.86241	25593.309*
	[0.0033]	[0.086]	[0.0037]	[0.1104]	[314.62491]	[8220.924]	[347.2635]	[10624.498]
Distance	0.0003				-29.951444			
	[0.0005]				[44.488134]			
Duration	0.0135	0.0162			-471.87726	2084.9759		
	[0.0122]	[0.0223]			[1163.0556]	[2129.4779]		
Observations	567	567	567	567	567	567	567	567
R-squared	0.221	0.2026	0.1666		0.1147	0.1350	0.0966	
R-squared adjusted	0.2112	0.1596	0.1592		0.1036	0.0883	0.0886	
Number of id		24		24		24		24

Robust standard errors in brackets
*** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$

Table 3.4: Quantitative Performance Measures for PI (private individuals)

	Number of approved loans (per office)				Sum of loans approved (per office)			
Variable	OLS	FE	IV	FE IV	OLS	FE	IV	FE IV
Limits[11]	0.3331***	0.1968***	0.6315***	0.2117	8011.6471***	4820.5108***	13199.1044***	19212.3758**
	[0.0266]	[0.0237]	[0.1667]	[0.2965]	[621.8245]	[550.6499]	[3712.7085]	[10193.4606]
Index of real wage	0.5381***	0.5261***	0.5414***	0.5261***	12613.2757***	12748.6331***	12641.672***	12783.2738***
	[0.11]	[0.0846]	[0.122]	[0.0846]	[2570.8346]	[1964.1564]	[2716.4296]	[2908.6684]
Number of offices	-7.9527***	-10.3517***	-9.6196***	-10.3333***	-152166.5059***	-239799.9032***	-177142.6739***	-221982.7963***
	[0.7616]	[0.832]	[1.3422]	[0.6272]	[17806.8519]	[19307.7666]	[29894.931]	[21563.863]
Number of offices squared	0.3094***	0.3679***	0.4043***	0.3695***	5859.0109***	7655.8755	7423.2696***	9137.5949***
	[0.0449]	[0.0429]	[0.0775]	[0.065]	[1049.0755]	[995.0093]	1726.0597]	[2234.4776]
Number of competitors	0.1802***	2.2187***	0.1863***	2.1574	3177.1056***	75755.1383***	2989.3632***	16712.9661
	[0.0246]	[0.5291]	[0.0281]	[1.5578]	[576.1669]	[12278.8299]	[625.1437]	[53555.6718]
Distance	0.013***				114.7477			
	[0.0035]				[82.2966]			
Duration	0.0748	0.0068			2176.9695	6571.1355**		
	[0.0932]	[0.1359]			[2179.6108]	[3153.2771]		
Observations	601	601	601	601	601	601	601	601
R-squared	0.4271	0.565	0.2929		0.3581	0.5003	0.2809	
R-squared adjusted	0.4203	0.5429	0.2869		0.3506	0.4749	0.2749	
Number of id		24		24		24		24

Robust standard errors in brackets
*** p< 0.01, ** p< 0.05, * p< 0.1

Table 3.5: Qualitative Performance Measures for SME

Variable	Mean days overdue OLS	FE	IV	FE IV	Recovery ninety OLS	FE	IV	FE IV	Thirty plus OLS	FE	IV	FE IV
Limits[12]	0.1091*	0.2598**	-0.1216	1.4775	0.0001	-0.0004	-0.0004	0.0099	0.0008**	0.0005	-0.0005	0.0201
	[0.0572]	[0.1304]	[0.2538]	[1.9637]	[0.0002]	[0.0008]	[0.0008]	[0.0171]	[0.0003]	[0.0007]	[0.0013]	[0.0298]
Index of real wage	0.466	0.4751	0.3621	0.6451	0.003**	0.0023*	0.0027*	0.0024	0.0006	-0.0002	0.0004	0.0031
	[0.3766]	[0.3787]	[0.3954]	[0.478]	[0.0015]	[0.0012]	[0.0016]	[0.0015]	[0.0022]	0.0022]	[0.0023]	[0.0062]
Number of offices	4.2345	2.6097	6.0124	9.0554	-0.0095	-0.0071	-0.0041	0.0051	-0.0068	-0.0075	0.0076	0.1
	[2.9574]	[4.394]	[4.6623]	[8.1963]	[0.0084]	[0.0129]	[0.0154]	[0.0337]	[0.0181	[0.0259]	[0.0287]	[0.1388]
Number of offices squared	0.0229	-0.0049	-0.0152	-0.7681	0.0003	0.0004	0.0002	0	0.0009	0.0006	0.0005	-0.0112
	[0.2151]	[0.2728]	[0.2367]	[1.1589]	[0.0004]	[0.0005]	[0.0005]	[0.0012]	[0.0012]	[0.0015]	[0.0014]	[0.0171]
Number of competitors	0.0629	-2.6155	0.0769	-7.8563	0.0006	-0.0097	0.0003	-0.0266	-0.0006	-0.0292	-0.0007	-0.1406
	[0.1148]	[4.166]	[0.1093]	[11.5055]	[0.0004]	[0.0139]	[0.0005]	[0.0276]	[0.0006]	[0.0243]	[0.0006]	[0.1996]
Distance	0.0047				0				0			
	[0.0124]				[0]				[0.0001]			
Duration	-0.364	0.8514			-0.0004	-0.0021			-0.0016	0.0085		
	[0.3505]	[1.234]			[0.0011]	[0.0028]			[0.0018]	[0.0066]		
Observations	381	381	381	381	311	311	311	311	325	325	325	325
R-squared	0.1161	0.0526	0.0769		0.031	0.0315	0.0092		0.026	0.017	.	
R-squared adjusted	0.0995	-0.0256	0.0645		0.0087	-0.0609	-0.007		0.0045	-0.076	.	
Number of id		24		24		24		24		24		24

Robust standard errors in brackets
*** p< 0.01, ** p< 0.05, * p< 0.1

Table 3.6: Qualitative Performance Measures for PI (private individuals)

	Mean days overdue				Recovery ninety				Thirty plus			
Variable	OLS	FE	IV	FE IV	OLS	FE	IV	FE IV	OLS	FE	IV	FE IV
Limits[13]	0.2374***	0.2287*	-0.008	0.4364	0.0006**	0.0002	0.0027	0.0013	0.0111***	0.0064**	0.0131	0.5631
	[0.0788]	[0.1324]	[0.3812]	[0.8038]	[0.0003]	[0.0004]	[0.0031]	[0.0019]	[0.0028]	[0.0033]	[0.0112]	[5.6102]
Index of real wage	0.2386	0.233	0.2212	0.2363	0.0002	0.0001	0.0002	0	0.0134	0.0133	0.0132	-0.0198
	[0.2783]	[0.2784]	[0.281]	[0.2787]	[0.0013]	[0.0013]	[0.0014]	[0.0013]	[0.0103]	[0.0093]	[0.0104]	[0.3439]
Number of offices	4.5359*	1.6966	4.4619	2.3075	-0.0223***	-0.0032	-0.0306	-0.006	-0.3293***	-0.657***	-0.3785***	-0.2503
	[2.409]	[3.2738]	[3.3988]	[2.5387]	[0.0081]	[0.0122]	[0.0194]	[0.0095]	[0.0824]	[0.1083]	[0.1129]	[3.5464]
Number of offices squared	0.0083	0.1151	-0.0128	0.072	0.0008*	0.0002	0.0013	0.0004	0.0169***	0.0314***	0.0187***	-0.0279
	[0.1765]	[0.199]	[0.2141]	[0.2143]	[0.0005]	[0.0006]	[0.001]	[0.0005]	[0.0056]	[0.0062]	[0.0066]	[0.5753]
Number of competitors	0.0263	0.5019	0.0715	0.2943	0.0002	-0.0248**	0	-0.0313***	0.0052**	0.1419*	0.0065***	-1.2486
	[0.0835]	[2.2701]	[0.081]	[2.6619]	[0.0004]	[0.0116]	[0.0005]	[0.0118]	[0.0025]	[0.0849]	[0.0023]	[14.3981]
Distance	0.0118				0				0.0005			
	[0.0094]				0				[0.0003]			
Duration	-0.3768	0.1943			0.0009	-0.0018			-0.0063	0.0195		
	[0.3008]	[0.7614]			[0.0011]	[0.003]			[0.0087]	[0.0223]		
Observations	578	578	578	578	549	549	549	549	410	410	410	410
R-squared	0.1128	0.0424	0.0944		0.0354	0.0452	.		0.1097	0.1692	0.1037	
R-squared adjusted	0.1019	-0.0082	0.0864		0.0229	-0.0082	.		0.0942	0.1081	0.0926	
Number of id		24		24		24		24		24		24

Robust standard errors in brackets
*** p< 0.01, ** p< 0.05, * p< 0.1

Table 3.7: Statistics for the Limits Variable for SME and PI Loans

	Mean	Median	St. dev	Min	Max
PI Limits> 0 (48.33% of all observations)	46,089.34	50,000	17,088.07	2,500	100,000
SME Limits> 0 (49.62% of all observations)	96,183.21	100,000	36,691.5	50,500	200,000

Table 3.8: First Stage Regressions

	SME Limits		PI Limits	
Variables	OLS	FE	OLS	FE
Index of real wage	-0.0726 [0.2858]	-0.0799 [0.1747]	-0.0002 [0.1696]	-0.0024 [0.1491]
Number of offices	9.7836*** [1.9346]	-3.6659** [1.6965]	4.0664*** [1.1631]	-1.238 [1.4652]
Number of offices squared	-0.3723*** [0.1127]	0.2278*** [0.0866]	-0.2854*** [0.0682]	-0.103 [0.0754]
Number of competitors	0.2133*** [0.0641]	2.6313** [1.1062]	0.091** [0.0378]	4.1025*** [0.9164]
Distance	0.0134 [0.0091]		0.0012 [0.0054]	
Duration	1.1023*** [0.2346]	1.5518*** [0.2802]	0.5837*** [0.1418]	0.4566* [0.2387]
Observations R-squared	567 0.4140	567 0.3680	601 0.1652	601 0.1099
F test of excluded instruments	F(2,560) = 15.02	F(1,538) = 30.68	F(2,594) = 9.30	F(1,572) = 3.66
Hansen J statistics	0.0427 (p = 0.8362)		10.7222 (p = 0.0011)	

Robust standard errors in brackets
*** p< 0.01, ** p< 0.05, * p< 0.1

Bibliography

Acemoglu, D., Linn, J., 2004. Market size in innovation: Theory and evidence from the pharmaceutical industry. The Quarterly Journal of Economics 119 (3), 1049–1090.

Aghion, P., Tirole, J., 1997. Formal and real authority in organizations. Journal of Political Economy 105 (1), 1–29.

Alessandrini, P., Presbitero, A. F., Zazzaro, A., 2009. Banks, distances and firms' financing constraints. Review of Finance 13 (2), 261–307.

Amir, R., 2000. Modelling imperfectly appropriable R&D via spillovers. International Journal of Industrial Organization 18 (7), 1013–1032.

Amir, R., Evstigneev, I. V., Wooders, J., 2003. Noncooperative versus cooperative R&D with endogenous spillover rates. Games and Economic Behavior 42 (2), 183–207.

Anton, J. J., Greene, H., Yao, D. A., 2006. Policy implications of weak patent rights. Innovation Policy and the Economy 6, 1.

Arundel, A., 2001. The relative effectiveness of patents and secrecy for appropriation. Research policy 30 (4), 611–624.

Aschhoff, B., Baier, E., Crass, D., Hud, M., Hünermund, P., Köhler, C., Peters, B., Rammer, C., Schricke, E., Schubert, T., Schwiebacher, F., 2013. Innovation in Germany - results of the German CIS 2006 to 2010. Background report on the Innovation Surveys 2007, 2009 and 2011 of the Mannheim Innovation Panel. Tech. Rep. 13-01, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Atallah, G., 2004. The protection of innovations. Cirano working papers, CIRANO.

Audretsch, D. B., Feldman, M. P., 1996a. Innovative clusters and the industry life cycle. Review of Industrial Organization 11 (2), 253–273.

Audretsch, D. B., Feldman, M. P., 1996b. R&D spillovers and the geography of innovation and production. American Economic Review 86 (3), 630–40.

Becker, G. S., Murphy, K. M., 1992. The division of labor, coordination costs, and knowledge. The Quarterly Journal of Economics 107 (4), 1137–60.

Behringer, S., 2014. Network effects, spillovers and market structure. The Manchester School 82 (2), 143–159.

Berger, A. N., DeYoung, R., 2006. Technological progress and the geographic expansion of the banking industry. Journal of Money, Credit and Banking 38 (6), 1483–1513.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., Stein, J. C., 2005. Does function follow organizational form? Evidence from the lending practices of large and small banks. Journal of Financial Economics 76 (2), 237–269.

Berger, A. N., Udell, G. F., 2002. Small business credit availability and relationship lending: The importance of bank organisational structure. Economic Journal 112 (477), F32–F53.

Berry, S., Waldfogel, J., 2010. Product quality and market size. The Journal of Industrial Economics 58 (1), 1–31.

Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81 (4), 1347–1393.

Boot, A., 2000. Relationship banking: What do we know? Journal of Financial Intermediation 9 (1), 7–25.

Boot, A. W. A., Thakor, A. V., 2000. Can relationship banking survive competition? Journal of Finance 55 (2), 679–713.

Bresnahan, T. F., 1992. Sutton's sunk costs and market structure: Price competition, advertising, and the evolution of concentration.

Bronnenberg, B. J., Dhar, S. K., Dubé, J.-P., 2005. Endogenous sunk costs and the geographic distribution of brand shares in consumer package goods industries. Tech. rep., Citeseer.

Caballero, R. J., Jaffe, A., 1993. How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. NBER Working Papers 4370, National Bureau of Economic Research, Inc.

Canales, R., Nanda, R., 2012. A darker side to decentralized banks: Market power and credit rationing in SME lending. Journal of Financial Economics 105 (2), 353–366.

Carlton, D. W., Oct 2005. Barriers To Entry. NBER Working Papers 11645, National Bureau of Economic Research, Inc.

Carter, D. A., McNulty, J. E., 2005. Deregulation, technological change, and the business-lending performance of large and small banks. Journal of Banking & Finance 29 (5), 1113–1130.

Cohen, W. M., 2010. Chapter 4 - fifty years of empirical studies of innovative activity and performance. In: Hall, B. H., Rosenberg, N. (Eds.), Handbook of The Economics of Innovation, Vol. 1. Vol. 1 of Handbook of the Economics of Innovation. North-Holland, pp. 129 – 213.

Cohen, W. M., Goto, A., Nagata, A., Nelson, R. R., Walsh, J. P., 2002. R&D spillovers, patents and the incentives to innovate in Japan and the United States. Research Policy 31 (8), 1349–1367.

Cohen, W. M., Levin, R. C., 1989. Empirical studies of innovation and market structure. In: Schmalensee, R., Willig, R. (Eds.), Handbook of Industrial Organization, 1st Edition. Vol. 2. Elsevier, Ch. 18, pp. 1059–1107.

Cohen, W. M., Levin, R. C., Mowery, D. C., 1987. Firm size and R&D intensity: A re-examination. Journal of Industrial Economics 35 (4), 543–65.

Cohen, W. M., Levinthal, D. A., 1989. Innovation and learning: The two faces of R&D. The Economic Journal 99 (397), pp. 569–596.

Cohen, W. M., Levinthal, D. A., 1990. Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly 35 (1), pp. 128–152.

Cohen, W. M., Nelson, R. R., Walsh, J. P., 2000. Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not). Tech. rep., National Bureau of Economic Research.

Cohen, W. M., Walsh, J. P., 2000. R&D information flows, appropriability and r&d intensity. In: Spivack, R. (Ed.), Papers and proceedings of the Advanced Technology Program's International Conferences on the Economic Evaluation of Technological Change. Vol. NIST Special Publication 952. Washington: USGPO, pp. 22–29.

Coscollá-Girona, P., López, P. B., Llopis, A. S., Barrachina, M. E. R., Apr 2011. Competitive pressure determinants and innovation at the firm level. Working Papers. Serie EC 2011-02, Instituto Valenciano de Investigaciones Económicas, S.A. (Ivie).

d'Aspremont, C., Jacquemin, A., 1988. Cooperative and noncooperative R&D in duopoly with spillovers. American Economic Review 78 (5), 1133–37.

De Bondt, R., 1997. Spillovers and innovative activities. International Journal of Industrial Organization 15 (1), 1–28.

Degryse, H., Laeven, L., Ongena, S., 2009. The impact of organizational structure and lending technology on banking competition. Review of Finance 13 (2), 225–259.

Degryse, H., Ongena, S., 2007. The impact of competition on bank orientation. Journal of Financial Intermediation 16 (3), 399–424.

Dewatripont, M., Tirole, J., 2005. Modes of communication. Journal of Political Economy 113 (6), 1217–1238.

Dick, A. A., 2007. Market size, service quality, and competition in banking. Journal of Money, Credit and Banking 39 (1), 49–81.

Ellickson, P. B., 2007. Does Sutton apply to supermarkets? The RAND Journal of Economics 38 (1), 43–59.

Ellison, G., Glaeser, E. L., Kerr, W., 2010. What causes industry agglomeration? evidence from coagglomeration patterns. American Economic Review 100 (3), 1195–1213.

Etro, F., 2013. Some thoughts on the Sutton approach. Journal of Economics, 1–15.

Garicano, L., 2000. Hierarchies and the organization of knowledge in production. Journal of Political Economy 108 (5), 874–904.

Geroski, P., 1994. Market Structure, Corporate Performance, and Innovative Activity. Clarendon Press.

Geroski, P. A., 1991. Entry and the rate of innovation. Economics of Innovation and New Technology 1 (3), 203–214.

Gersbach, H., Schmutzler, A., 2003. Endogenous spillovers and incentives to innovate. Economic Theory 21 (1), 59–79.

Hall, B., Helmers, C., Rogers, M., Sena, V., 2014. The choice between formal and informal intellectual property: A review. Journal of Economic Literature 52 (2), 1–50.

Hall, B. H., Mairesse, J., Mohnen, P., 2010. Chapter 24 - Measuring the returns to R&D. In: Hall, B. H., Rosenberg, N. (Eds.), Handbook of the Economics of Innovation, Volume 2. Vol. 2 of Handbook of the Economics of Innovation. North-Holland, pp. 1033 – 1082.

Hanel, P., St-Pierre, A., 2002. Effects of R&D spillovers on the profitability of firms. Review of Industrial Organization 20 (4), 305–322.

Henderson, R. M., Cockburn, I. M., 1996. Scale, scope, and spillovers: The determinants of research productivity in drug discovery. RAND Journal of Economics 27 (1), 32–59.

Henry, E., Ruiz-Aliseda, F., 2012. Innovation beyond patents: Technological complexity as a protection against imitation. Tech. rep., CEPR Discussion Papers.

Hinloopen, J., October 2000. More on subsidizing cooperative and noncooperative R&D in duopoly with spillovers. Journal of Economics 72 (3), 295–308.

Javorcik, B. S., 2004. Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages. American Economic Review 94 (3), 605–627.

Kamien, M., Muller, E., Zang, I., 1992. Research joint ventures and R&D cartels. American Economic Review 82 (5), 1293–306.

Keller, W., 2004. International technology diffusion. Journal of economic literature, 752–782.

Kim, J., Marschke, G., 2005. Labor mobility of scientists, technological diffusion, and the firm's patenting decision. The RAND Journal of Economics 36 (2), pp. 298–317.

Kosová, R., Lafontaine, F., Perrigot, R., 2013. Organizational form and performance: Evidence from the hotel industry. Review of Economics and Statistics 95 (4), 1303–1323.

Krueger, A. B., 1991. Ownership, agency, and wages: An examination of franchising in the fast food industry. The Quarterly Journal of Economics 106 (1), 75–101.

Legros, P., Newman, A. F., 2008. Competing for ownership. Journal of the European Economic Association 6 (6), 1279–1308.

Leiponen, A., Byma, J., 2009. If you cannot block, you better run: Small firms, cooperative innovation, and appropriation strategies. Research Policy 38 (9), 1478–1488.

Levin, R. C., May 1988. Appropriability, R&D spending, and technological performance. American Economic Review 78 (2), 424–28.

Liberti, J. M., Mian, A. R., 2009. Estimating the effect of hierarchies on information use. Review of Financial Studies 22 (10), 4057–4090.

Mansfield, E., 1984. R&D and innovation: some empirical findings. In: Griliches, Z. (Ed.), R&D, Patents, and Productivity. NBER Chapters. National Bureau of Economic Research, Inc, pp. 127–154.

Mansfield, E., 1986. Patents and innovation: an empirical study. Management science 32 (2), 173–181.

Mansfield, E., Schwartz, M., Wagner, S., 1981. Imitation costs and patents: an empirical study. Economic Journal 91 (364), 907–18.

Marin, D., Verdier, T., 2008. Power inside the firm and the market: A general equilibrium approach. Journal of the European Economic Association 6 (4), 752–788.

Marin, P. L., Siotis, G., 2007. Innovation and market structure: An empirical evaluation of the'bounds approach' in the chemical industry. The Journal of Industrial Economics 55 (1), 93–111.

Matraves, C., June 1999. Market structure, R&D and advertising in the pharmaceutical industry. Journal of Industrial Economics 47 (2), 169–94.

Matraves, C., Rondi, L., 2005. Product differentiation, turbulence and sustainability of competitive advantage. Tech. rep., working paper, March.

Mesquita, L. F., Anand, J., Brush, T. H., 2008. Comparing the resource-based and relational views: knowledge transfer and spillover in vertical alliances. Strategic Management Journal 29 (9), 913–941.

Mookherjee, D., 2006. Decentralization, hierarchies, and incentives: A mechanism design perspective. Journal of Economic Literature 44 (2), 367–390.

Mullainathan, S., Scharfstein, D., 2001. Do firm boundaries matter? American Economic Review 91 (2), 195–199.

Nocke, V., 2007. Collusion and dynamic (under-) investment in quality. The RAND Journal of Economics 38 (1), 227–249.

Poyago-Theotoky, J., 1999. A note on endogenous spillovers in a non-tournament R&D duopoly. Review of Industrial Organization 15 (3), 253–262.

Radner, R., 1993. The organization of decentralized information processing. Econometrica 61 (5), 1109–46.

Robinson, W. T., Chiang, J., 1996. Are Sutton's predictions robust?: Empirical insights into advertising, R&D, and concentration. The Journal of Industrial Economics 44 (4), pp. 389–408.

Samuelson, P., Scotchmer, S., 2002. The law and economics of reverse engineering. Yale Law Journal, 1575–1663.

Santos Silva, J. M., Tenreyro, S., 2011. Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. Economics Letters 112 (2), 220–222.

Sapienza, P., 2002. The effects of banking mergers on loan contracts. The Journal of Finance 57 (1), 329–367.

Schmalensee, R., 1992. Sunk costs and market structure: A review article. The Journal of Industrial Economics, 125–134.

Scotchmer, S., 2006. Innovation and Incentives, 1st Edition. Vol. 1. The MIT Press.

Senyuta, O., Žigić, K., 2012. Managing spillovers: an endogenous sunk cost approach. Economics Institute, Academy of Sciences of the Czech Republic.

Shenkar, O., 2010. Copycats: How Smart Companies Use Imitation to Gain a Strategic Edge. McGraw-Hill Education.

Spence, A. M., 1984. Cost reduction, competition, and industry performance. Econometrica 52 (1), 101–21.

Stein, J. C., 2002. Information production and capital allocation: Decentralized versus hierarchical firms. The Journal of Finance 57 (5), 1891–1921.

Střelický, J., Žigić, K., 2011. Intellectual property rights protection and enforcement in a software duopoly. CERGE-EI Working Papers, The Center for Economic Research and Graduate Education - Economic Institute, Prague.

Sutton, J., 1991. Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration. The MIT Press.

Sutton, J., 2001. Technology and Market Structure: Theory and History, 1st Edition. Vol. 1. The MIT Press.

Sutton, J., 2007. Market structure: Theory and evidence. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization. Vol. 3. Elsevier, Ch. 35, pp. 2301–2368.

Suzumura, K., 1992. Cooperative and noncooperative R&D in an oligopoly with spillovers. American Economic Review 82 (5), 1307–20.

Taylor, M. S., 1993. Trips, trade, and technology transfer. The Canadian Journal of Economics 26 (3), 625–637.

Tesoriere, A., 2008. A further note on endogenous spillovers in a non-tournament R&D duopoly. Review of Industrial Organization 33 (2), 177–184.

Vasconcelos, H., 2006. Endogenous mergers in endogenous sunk cost industries. International Journal of Industrial Organization 24 (2), 227–250.

Vives, X., 2008. Innovation and competitive pressure. Journal of Industrial Economics 56 (3), 419–469.

Žigić, K., Střelický, J., Kúnin, M., 2013. The interaction between private and public ipr protection in a software market: A positive and normative analysis. CERGE-EI Working Paper Series (490).

Zabojnik, J., 2002. A theory of trade secrets in firms. International Economic Review 43 (3), 831–855.