THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Essays on the Organizational Economics of the Lobbying Market

Miguel Andres Espinosa Farfan

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor of Philosophy

London, June 2017

Declaration

I, Miguel Andres Espinosa Farfan, certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 56607 words.

Acknowledgements

This thesis is the result of uncountable hours, frustrations, and encouraging comments and moments. Different entities contributed to the completion of this thesis. I classify them in two: internal and external entities.

On the first set, I thank my stubbornness for helping me to be perseverant. I do not see any way to finish such a demanding task without feeling the need to be thinking about the same issues for so many weeks. I thank my dreaming capacity for convincing me that there was something interesting coming up as a result of all this distress. I thank my stoicism for constantly psychoanalyzing me and contributing to my endurance. Finally, I thank my memory, especially the one recalling my successful events, to keep me on track to think that another victorious moment could come with any part of these papers.

On the second set, I thank external factors. Firstly, I thank my belief about other people's belief about me as a scholar. Secondly, I thank people, or what they represented for me in different stages of the PhD life (at this stage, I cannot differentiate between these two). Luis Garicano provided several sharp and insightful comments. Although I thank his comments, I feel in an eternal debt with him for being such as inspiring figure. Realizing that he faced challenges in his own research as well as I did, pushed me to continue. Gilat Levy and John Sutton shared their thoughts about my research in different ways. They helped the thesis to be a better research document and they helped me to become a young researcher.

I acknowledge STICERD's financial support in helping me to buy some data I used in this thesis, as well as Jane Dickson for providing me with patient administrative support. Although tons of PhD fellows provided specific comments in different sections of the dissertation, it would be a crime to not acknowledge Michel, Frank, Giulia and Eddy. Although tons of professors provided specific comments in different sections of this dissertation, I should be deported out of academia if I do not acknowledge Jordi Blanes i Vidal, Jim Snyder, Jr., Catherine Thomas, Glen Weyl.

Not only as a product of the short memory or the salience of the recent past, I feel that Alexia has made a difference on this project. She not only had to sacrifice the pleasure of dinners and lunches to listen to my research, but she spent time giving precise comments on each chapter. More importantly, she was obtuse in her belief about my ability. Independently, of the truth of her beliefs, I will always be in debt to her for keeping me on earth at the same time as letting me have short flights.

As this dissertation seems to be the ultimate short-run outcome of my PhD life, it seems unfair to not acknowledging some people that contributed in different stages. I thank Mark Wilbor, Margaret Bray and Maria Lopez-Uribe for helping in many different ways and circumstances. These past years would have been an eternal and painful loop if I had not counted on the presence of Matteo and Albert. Finally, I thank both the scholars and administrative staff that made my visit to both University of Chicago and Stanford possible. Thus, both have a well-deserved place in my memory.

Although, a dedication has, only for me, a social value, for my mom, without understanding the value of my research, it means everything. I dedicate this thesis to her because I know that without her, I would have not simply got this far from home and from our non-PhD inertia.

Abstract

This thesis contains three chapters examining firms' behaviour and decision making when they seek to influence policies in the US through lobbying activities.

The first chapter studies the main trade-off that firms face when they face the decision to integrate or outsource knowledge workers. The chapter proposes a model that predicts that firms requiring large firm-specific skills, or low levels of issue-specific skills, or facing a large number of transactions will integrate as opposed to outsource the service provider. Using a newly collected dataset on the US federal lobbying industry, I conduct firm-fixed effect estimations and I find strong evidence supporting the theoretical predictions. To provide further empirical evidence, I exploit a quasi-experiment that the Oil and Gas industry faced: The BP oil spill. The spill increased the issue-specific skills needed to conduct advocacy activities and in line with the theory developed in the chapter, I show that the affected industry started using more external, as opposed to internal lobbyists after the oil spill.

The second chapter studies the effect of a technological upgrade on firms' vertical integration decision. I use the model proposed in the first chapter to show that a technological shock, introduced by the Open Government Act decreased the cost of acquiring issue-specific skills, which in turn, made firms less likely to outsource. Then, I use structural models to measure the magnitude of this technological effect and conduct counterfactual exercises to study the influence that the regulation had on the industry.

The third chapter studies the relationship between lobbying expenditures and market structure. I show that less and no more concentrated industries spend more on lobbying. To explain this empirical puzzle, I propose a theoretical model that includes the level of excludability in the payoffs. I provide empirical evidence that firms in less concentrated industries tend to lobby for more excludable goods and I show that including this dimension can explain the empirical puzzle. To provide causal evidence, I use national-level mergers that change city-level market structures. Collecting a new data set of city-level lobbying expenditures, I show that controlling for the level of excludability in the payoffs, more concentrated industries spend more on lobbying efforts.

Contents

List of tables

List of Figures

Chapter 1

Sourcing of Expertise and the Boundaries of the Firm:The Case of Lobbyists[1]

Abstract

This chapter proposes and tests a theory of vertical integration with knowledge workers. Outsourcing allows firms to solve hard problems at the cost of transmitting firm-specific knowledge. By hiring someone internally, firms save on these communication costs, with the downside of incurring costs of acquiring knowledge. Exploiting the increasing returns to the use of knowledge implies conducting easy and frequent activities in-house and harder and less frequent tasks in the external market. The economy saves communication costs when firms with large firm-specific knowledge conduct activities in-house. I confirm the empirical validity of this theory using data from a knowledge-intensive industry: US Federal Lobbying. First, I validate the main theoretical predictions using client fixed-effects estimations with information at both the industry and bill levels. Second, I exploit the 2010 BP oil spill as an exogenous increase in the difficulty of the lobbying activities for the oil and gas extracting industry, and I show that it led to a disproportionate increase in the use of external lobbyists for the affected industry.

[1] I am deeply indebted for the precise and insightful comments made by Jordi Blanes i Vidal, Luis Garicano, Gilat Levy, Jim Snyder, Jr., John Sutton, Catherine Thomas and Glen Weyl. This project has benefited from several interviews, comments and suggestions made by tens of lobbyists that have decided to remain in the anonymity. I acknowledge the comments and suggestions made by Oriana Bandiera, Heski Bar-Isaac, Giuseppe Berlingieri, Christopher Berry, Matilde Bombardini, Tim Bresnahan, Cheng Chen, John M. de Figueiredo, Ethan Bueno de Mesquita, Matej Drev, Liran Einav, Alessandro Gavazza, Matthew Gentry, Anna Gumpert, Nan Jia, Lisa Kahn, Karam Kang, Peter Klein, Timothy Lambie-Hanson, Jin Lin, Rocco Macchiavello, Alan Manning, Luis Martinez, Guy Michaels, Jeanine Miklós-Thal, Pablo Montagnes, Suresh Naidu, Derek Neal, Steve Pischke, Brian K. Richter, Roberta Romano, Raffaella Sadun, Mark Schankerman, Pasquale Schiraldi, Paulo Somaini, Alois Stutzer, Chad Syverson, Guido Tabellini, John Van Reenen, Rick G. Vanden Bergh, Stephane Wolton, Hye Young You as well as participants at Work in Progress seminars at LSE, UChicago and Stanford University, 14th IIOC, 2016 TADC, 15th IOEA, 20th Annual Conference of the SIOE . A previous version of this paper was awarded the 2016 IOEA First Prize Award. Finally, I am extremely grateful to the comments received from my PhD fellows: Michel Azulai, Alexia Delfino, Tam Hiu Fung, Christos Makridis, Stephan Maurer, Frank Pisch, Oleg Rubanov and Giulia Zane. Financial support from STICERD is gratefully acknowledged.
E-mail: m.espinosa@lse.ac.uk.

"*Clearly there is... a problem of the* **Division of Knowledge** *which is quite analogous to, and at least as important as, the problem of the division of labor. But while the latter has been one of the main subjects of investigation ever since the beginning of our science, the former has been as completely neglected, although it seems to me to be the really central problem of economics as a social science" (von Hayek, 1937, p.49) (emphasis in original).*

1.1 Introduction

Firms' dependence on knowledge workers such as lawyers and managers has increased remarkably in recent years.[2] Firms face a fundamental question when they use these workers: under which circumstances should firms hire them internally and when should they outsource their services? The decision has implications not only for the internal dynamics of the firm and the functioning of markets, but also for other important issues in the economy, such as the distribution of earnings.[3] Thanks to the vertical integration research conducted over the last 80 years, we now have a better understanding of the causes and consequences of the integration decision. However, all of the leading theories have equated the study of the integration decision with the study of incentives.[4] That is, they have focused on the appropriate management of incentive alignment problems. As knowledge is the key input in the production process of knowledge workers and, more broadly, is at the heart of the organizational design problem, a natural alternative approach is to put aside the incentive issue and focus on the acquisition and communication of knowledge. This is the approach that I take in this chapter.

I provide a theoretical framework inspired by Arrow (1974) and Garicano (2000) to guide the empirical analysis of the integration decision with knowledge workers. Firms use knowledge workers to solve problems. Solving problems requires issue-specific and firm-specific knowledge. More frequent problems require less issue-specific knowledge. I refer to a problem with a low (high) level of issue-specific

[2] Berlingieri (2014) shows that the professional business services industry (NAICS codes 54, 55 and 56) exerts the biggest influence on the rest of the economy in terms of forward linkage measures. He also shows that this industry has the largest change in input-output linkages, an increase close to four times in 60 years, and is the sub-industry with the greatest growth. Remarkably, more than 90% of the industry's output is used as an intermediary input in other industries. Goldschmidt and Schmieder (2015) provide empirical evidence on the increase in outsourcing for low-knowledge occupations in Germany.

[3] For examples of these implications for the firm, markets and distribution of earnings, see Holmstrom and Roberts (1998), Holmstrom (1999) and Goldschmidt and Schmieder (2015), respectively.

[4] See, for instance, Williamson (1971, 1979, 1985) and Klein et al. (1978) for the *Rent-Seeking Theory*; Simon (1951) Williamson (1971, 1973, 1975, 1991) for the *Adaptation Theory*; Grossman and Hart (1986), Hart and Moore (1990) and Hart (1995) for the *Property Rights Models;* and Holmstrom and Milgrom (1991, 1994) and Holmstrom (1999) for *Multi-Tasking Models.* According to Gibbons (2005), the first two focus on the ex-post incentive problems and the other two on the ex-ante incentive problem. The first two are associated to the transaction cost economies paradigm.

knowledge as an easy (difficult) problem. In-house workers know more about the firm than external service providers do. This leads to the main trade-off in the economy: bringing someone in-house saves on costs of communicating firm-specific knowledge but requires the firm to pay for the workers' acquired skills. Although outsourcing allows to access to a larger amount of issue-specific knowledge, it implies paying higher communication costs.

I study this trade-off under a simple but powerful insight: there are increasing returns to the use of knowledge, as the cost of acquiring it is independent of its rate of use. The optimal organization of the economy requires that these returns are exploited in-house by solving frequent problems and in the external market by solving infrequent problems for several firms. Thus, firms with frequent problems bring someone in-house, and in-house workers solve routine, easy tasks (as less-knowledge-intensive problems are more frequent). As difficult problems occur infrequently, firms facing these problems cannot exclusively finance a service provider. However, the external market allows firms to join interests and finance the large acquisition of issue-specific knowledge that they need to solve problems. Therefore, firms with infrequent problems outsource the service, and external providers solve rare, harder tasks. This organization of the economy implies that the issue-specific knowledge level differs across the firm and the external market. As earnings are proportional to this type of knowledge, external service providers earn more than in-house staff. Furthermore, external service providers who work for more firms -*more leverage*- have more knowledge, and, as a consequence, higher earnings than providers who work for fewer firms. Finally, the communication costs are minimized when firms with greater firm-specific knowledge solve their problems with in-house workers. Summing up, the main theoretical predictions are that firms with more firm-specific knowledge or that face frequent or easy problems will use in-house instead of external workers.

I test these predictions for a particular knowledge-intensive occupation: US federal lobbyists. Acquiring information on legislative proposals and persuading policy makers are two typical examples of lobbyists' tasks. The latter task occurs less often than the former, as not all legislative proposals affect the firm or simply because policy makers' preferences are inexpugnable. The activities of acquiring information and persuading policy makers differ by their level of knowledge requirements. For instance, several government and private databases facilitate knowledge acquisition by providing updated and detailed information on legislative proposals.[5] In contrast, persuading policy makers requires more knowledge: lobbyists need to know who is the best policy maker to persuade, the best persuasion strategy and so on. Thus, finding out what is happening in Washington is a (relatively) routine, easy

[5] See, for instance, www.govtrack.us, www.congress.gov, www.sunlightfoundation.com, www.opensecrets.org, among other sources, such as the US Senate and US House of Representative web pages.

task, whereas persuading a senator is a less common, harder task. Anecdotal evidence shows that the first activity is conducted with in-house lobbyists, whereas the second is conducted with external advocates.

The unique database that I build for this study is comprised of the universe of US federal lobbying transactions in which I can differentiate not only between in-house and external transactions, but also the specific identity of firms and lobbyists, the period of the transaction, and the main advocacy activities, such as the bills in which the firm is interested. Since it contains transactional-level information and allows me to differentiate the specific identities of all of the agents demanding and supplying advocacy services, the database used in this study overcomes common challenges that researchers face when they study the integration decision.[6] To test the predictions from my theoretical framework, I enrich this database by including the industry of firms, lobbyists' sociodemographic variables, and information about the committees studying the lobbied bills, among other variables.

Despite the richness of this dataset, the main challenge is to measure activities that differ by knowledge requirements and frequency. I argue that variation in the characteristics of the firms' industries and the bills that affect them leads to variation in the frequency and knowledge intensity of the lobbying activities. Firms respond to this variation by choosing in-house or external lobbyists. For the case of the bills, it is useful to understand how they are studied. Bills are sent to congressional committees after their introduction in Congress. Policy makers then, interview witnesses to acquire relevant knowledge about the legislative proposals. Variation in the composition of witnesses provides a rich source of heterogeneity in the types of problems that firms face. I construct as a first measure of knowledge requirements of the lobbying activities an index at the committee-semester level that accounts for the fraction of high-knowledge witnesses.[7] I proxy frequency with the total number of bills studied in the committee-semester combination.

A shortcoming of these measures is that not all the firms lobby for bills. An alternative source from which to construct proxies for these theoretical variables is the *RegData 2.2*, a database that uses text analysis and machine learning algorithms to create regulation measures at the industry-year level.[8] I use two measures from this database: the number of words related to the regulation of the industry and

[6] Berlingieri (2014, 2015) mentions, that it is usually difficult to differentiate between internal and external transactions. Additional challenges that this literature has faced include being able to differentiate the specific identity of the agents in both sides of the market and to have the universe of transactions in the market. Examples of studies in which the demand side and not the supply side is known are Anderson and Schmittlein (1984), Abramovsky and Griffith (2006) and Galdon-Sanchez et al. (2015). On the other hand, Garicano and Hubbard (2007, 2009) are examples of papers with information on the supply side of the market, but a lack of information on the identities of the demand side of the market.

[7] This is similar in spirit to Azoulay (2004), in which the proportion of academic investigators (over the total number of participants in clinical studies) captures the relative importance of knowledge-intense activities.

[8] See Al-Ubaydli and McLaughlin (2015).

number of times that five strings -*measuring the difficulty of the regulation*- appear among these regulation words.

In this chapter, I conduct two sets of empirical exercises with the aforementioned data: Fixed Effects and Difference in Difference Estimations. I first show the validity of the most important theoretical predictions using fixed-effect estimations at the firm and time levels. I find that firms lobbying for bills studied by high-difficulty-index committees or belonging to industries with a more difficult regulatory environment -*large fraction of restraining words over the total number of regulating words*- tend to advocate with external rather than in-house staff. Moreover, firms lobbying for bills studied in committees that receive more bills use internal instead of external advocacy services. The effects are economically relevant. For instance, a one-standard-deviation decrease in the difficulty measure based on the witnesses' knowledge is associated with an increase in the fraction of lobbying reports conducted in-house by 21%.

In the theoretical framework, firm-specific knowledge can also be easily interpreted as industry-specific knowledge. Given the lack of a comprehensive measure at the firm level, I use the total number of regulating words as a proxy for industry-specific knowledge. Firms explain their industry's regulations when they hire external service providers. The larger the number of regulating words in the industry, the longer it takes to explain them and, therefore, the larger are the communication costs. As my model predicts, I find robust empirical evidence that firms belonging to industries with a large number of regulating words tend to conduct their advocacy activities in-house.

These analyses, though, do not provide causal evidence for my theory. To go a step further, I run differences in differences estimations using the British Petroleum (henceforth BP) Deepwater Horizon oil spill of 2010 as an exogenous shock to the difficulty of lobbying activities for the oil and gas extracting industry. I show that firms belonging to this industry decreased their use of in-house staff after the spill. The point estimates are significant and economically relevant. For instance, the affected firms decreased the fraction of transactions made in-house (fraction of in-house lobbyists) by about 26% (21%) after the oil spill.

To tackle concerns over the simplifying assumption that problems have only one type of issue-specific knowledge, I also consider the possibility that firms face two different types of issues. In this case, the main theoretical predictions are robust, but new insights emerge. Service providers can work on one (*specialists*) or two (*generalists*) topics. I show that *generalists* exist due to complementarities across issues, and, although both *specialists* and *generalists* exist in both internal and external markets, *specialists* (*generalists*) are more common in the external (internal) market. I confirm the empirical validity of this prediction for several knowledge-intensive industries.

I also show two additional empirical patterns accounted for by my model. First, I provide evidence of a novel mechanism to explain the fact that large firms tend to be more vertically integrated than smaller firms: small firms tend to face easy (difficult) problems less (more) often than large firms.[9] Second, proxying lobbyists' knowledge level with their work experience as federal lobbyists and any previous work experience in the federal government, I provide empirical evidence that more-knowledgeable lobbyists have both greater leverage and higher earnings.

In the final section of the chapter, I address concerns about the external validity of my results. Acknowledging that each industry has its own peculiarities, I show that there are some similar empirical patterns between lobbyists and other Professional Business Services (henceforth PBS). For instance, I show that, in contrast to the literature which finds that low-skill workers earn more in the internal market, external service providers in both PBS and lobbying industries earn more than in-house workers.[10]

This chapter enriches our understanding of the vertical integration decision in several ways. First, to the best of my knowledge, this is the only manuscript to simultaneously propose and test a vertical integration theory with knowledge workers that does not consider incentive-alignment issues.[11] Second, using newly compiled data, I propose new ways to measure knowledge intensity of workers' activities. Although the lobbying reports database has been used before, I am not aware of an effort to complement it with information at the firm, lobbyist, industry, bill and committees levels.[12] I believe that these data and knowledge-based measures will allow researchers to answer other important questions, either in the incentives literature or as extensions of my theoretical framework. Third, I provide a theoretical framework and new empirical facts to both the Political Economy and Organizational Economics literatures. While the former has focused on the relationship between either lobbyists and policy makers or firms and policy makers, in this chapter, I enrich the understanding of the scarcely studied firms-lobbyists relationship.[13] Furthermore, to the best of my knowledge, the effect of the BP spill on vertical integration patterns has not been studied before. For the Organizational Economics literature, I enrich the understanding of the reasons to

[9]Contrary to the explanation developed in this chapter, Antras and Helpman (2004), de Figueiredo and Silverman (2006), Bombardini and Trebbi (2012), and Kerr et al. (2014), among many others, have associated these patterns by the differences in the cost of outsourcing and running an in-house office.

[10]For evidence on the differences of wages for low-skill workers, see for instance Abraham and Taylor (1996), Dube and Kaplan's (2010) and Goldschmidt and Schmieder (2015)

[11]My paper focuses on the acquisition, use, communication and distribution of knowledge to explain firm boundaries. An alternative approach for instance taken by Henderson and Cockburn (1996) and Nickerson and Zenger (2004) focus on the creation of knowledge and capabilities and its relationship with the firm boundary.

[12]For example, Blanes-i-Vidal et al. (2012) and Bertrand et al. (2014) have used the lobbying reports database.

[13]For example, Hirsch and Montagnes (2015) and Kang and You (2016) point out the scarcely studied firm-lobbyist relationship.

hire knowledge workers internally or externally, the characteristics of in-house and external staff, and the distribution of knowledge and earnings across the firm and external markets.

Related Literature

Although my approach differs sharply from the leading incentive-based vertical integration theories, this chapter finds support for some of the previous predictions but questions the general applicability of other results.[14] The *Transaction Cost Economics* literature predicts that the probability of integration increases with the specificity of the transaction. My results are consistent with this prediction, but the underlying mechanisms are different. In my case, the larger the firm-specific knowledge, the more costly it is to outsource (i.e. larger communication costs) and, therefore, the more likely to use in-house workers. In the existing literature, the more specific the transaction is, the more likely the firm is to face a hold-up problem; thus, in order to avoid it, the activity should be conducted internally.[15] However, as Klein (1988) points out, this literature does not clarify how this mitigation occurs in the case of human assets. Furthermore, in a context in which there is firm-specific knowledge, the hold-up problem can be intensified inside the firm, as workers that have received specific knowledge can attempt to hold their employers up and *vice versa*. Therefore, it seems that the hold-up explanation, is in the best case, incomplete.

Another prediction from this literature is that the more difficult the transaction is, the more likely it is to be vertically integrated.[16] My results show the opposite, as only the external market provides ways to accumulate high levels of knowledge. A final prediction from this literature is due to Williamson (1985): the cost of a hierarchical structure is easier to recover when the transactions are more frequent. This chapter complements his insight by providing both theoretical and empirical support to this idea.[17]

The empirical context of this chapter does not easily translate to either the *Property Rights* or the *Multitasking* theories. With respect to the property rights literature, as Dube and Kaplan (2007) and Bresnahan and Levin (2012) argue, there are no clear theoretical implications when non-human assets are completely nonexistent or irrelevant, as in the case of PBS industries. Clearly, the difference is that the law does not provide control rights over human beings. So, buying a

[14]This paper includes, in Section 1.5.1, an extensive discussion of the leading incentive-based vertical integration theories and their relationship with my theoretical framework.

[15]See, for instance, Klein, Crawford and Alchian (1978), Monteverde and Teece (1982), Anderson and Schmittlein (1984) and Joskow (1988).

[16]See, for instance, Monteverde and Teece (1982), Masten (1984), Anderson (1985) and Tadelis (2002). Empirical papers in this branch of the literature have proposed to measure difficulty of the transaction with its skill requirements.

[17]Strikingly, the issue of the frequency has not received much attention. For instance, two of the most comprehensive and recent literature reviews of the subject, Lafontaine and Slade (2007) and Bresnahan and Levin (2012), neglected the topic.

machine and hiring someone in-house are totally different because the employee can always quit. I interpret this chapter as a useful way to understand the vertical integration decision for the case of human assets. With respect to *Multitasking* models, a standard prediction from this theory has no empirical support in the lobbying context. As the outcome of persuading policy makers is harder to measure than the outcome of investigating the political environment, the *Multitasking* theory predicts that the latter activity should be conducted externally and the former internally. However, anecdotal evidence suggests that this is not the case.

A comment on the *labor economics* literature is warranted. While this literature has recognized the advantages of hiring internally and externally separately, I am not aware of any work connecting these two forces. Some papers argue that one reason to hire employees is to receive returns from firm-specific knowledge, whereas another branch argues that the reason for contracting out is to access greater amounts of knowledge.[18] It seems natural to think that what matters in understanding vertical integration patterns is the relative gain in one with respect to the other. To the best of my knowledge, this chapter is the first manuscript to make the natural empirical and theoretical connections between these two.

Although the economics literature has not focused on knowledge as one of the key inputs in the production process, a field of the Management literature has done so. The starting point in the field is the *Resource-based-view* of the firm. This literature focuses on explaining heterogeneity in firm performance by arguing that these differences are due to firms' different resources or capabilities.[19] The *knowledge-based-view* of the firm argues that knowledge is firms' most important resource or capability.[20] One of the main results from this literature is that knowledge-based resources can differ across firms and produce competitive advantage across firms. This type of advantage explains performance heterogeneity. In this literature, the integration decision is explained by the fact that different pieces of knowledge are complementary, and, therefore, conducting activities through the market can be inefficient. That is, if the knowledge used to produce one activity is useful in producing a different activity, the activities should be conducted in an integrated way.

The focus on knowledge as a way to understand organizational design is the main element that my analysis shares with these theories.[21] The economics literature has studied the integration decision concentrating on how to manage the misalignment

[18]Examples of the first argument are Becker (1964) and Autor (2003). On the other hand, Abraham and Taylor (1996) and Grossman and Helpman (2002a, 2005) among others argue that one of the reasons to contracting out is specialization in the external market. In this literature, the specialization corresponds to high-levels of issue-specific knowledge.

[19]See, for example, Eisenhardt and Martin (2000), Teece et al. (1997), Barney (1991), Peteraf (1993), Wernerfelt (1984) and Penrose (1959).

[20]See, for example, Foss (1996), Grant (1996, 2002), Kogut and Zander (1992, 2000) and Nickerson and Zenger (2004).

[21]For an early discussion of the advantages of focusing on knowledge to understand the firm boundaries problem, see Demsetz (1988).

of incentives. However, the *knowledge-based-view* of the firm, and this chapter, point to the fact that knowledge plays a role in defining firm boundaries even in the absence of any misalignment of interests. In this chapter, contrary to the *knowledge-based view*, knowledge complementarity is not relevant in explaining the integration decision. Instead, firms face problems that differ in their knowledge requirements, and the cost of acquiring or communicating knowledge determines the relative benefits of integration. Thus, although, this chapter focuses on knowledge like the management literature, I study different theoretical mechanisms, propose new empirical measures and use different empirical methodologies to test theoretical predictions.

This chapter is divided into five sections. Section 1.2 presents my theoretical framework. In Section 1.3, I begin by presenting the data used and the institutional context. Then, I present the results of client fixed effect estimations and the relationships among generalists, specialists, firms' size and leverage. I end this section with the BP example. In Section 1.4, I discuss the external validity of my theory with regard to other PBS occupations. Then, I finish this section with a short discussion summarizing the main results from the chapter and proposing further developments.

1.2 Theoretical Framework

This section is divided into two subsections. In the first part, I present the main setting in Section 1.2.1 and the results of the model. To gain some intuition, I first present results for the case in which firms face problems in only one issue in Section 1.2.2 and then I move to Section 1.2.3 where I present results for the case of two issues. The second part discusses the robustness of my theoretical framework in Section 1.2.4 and the relevant related literature in Section 1.2.5. I end up this Section by summarizing the main results of the model in Section 1.2.6. All the proofs can be found in Appendix 1.5.4.

1.2.1 Preliminaries

I consider an economy with a large number $M < \infty$ of ex-ante homogeneous clients (firms) and an infinite set N, of ex ante homogeneous service providers.[22] Clients exogenously receive one problem per unit of time spent in production for each issue: A and B.[23] Service providers, not clients, solve problems.

Demand. Problems differ by the level of issue-specific knowledge requirements (i.e., difficulty of the problem). I denote this level by $Z_i \in [0, 1]$ with $i = A, B$. Within each issue, the problems are ordered by increasing level of difficulty. The random variable Z_i is independent and identically distributed according to a con-

[22] I assume a large number of clients in order to apply Law of Large numbers below. I use the words clients and firms to refer to the individual unit of the demand side.

[23] An issue can be interpreted as a type of general knowledge.

tinuous cumulative distribution function $F_i \geq 0$ with $F_i(0) = 0$, $F_i' = f_i > 0$ and $f_i' < 0$. That is, easy problems are more common. Solutions to problems in both issues are equally valuable to these clients. Clients' payoff function is production minus labor costs, and the normalized value of production is 1 when a service provider solves a problem and 0 otherwise.

Supply. Solving problems requires knowledge. All service providers must learn the easiest (most common) problems before learning the harder (less common) ones, so that the more knowledgeable agents know everything that the less knowledgeable ones do, and more. That is, knowledge is cumulative.[24] Providers are characterized by a vector $(z_A, z_B) \in [0,1]^2$. Service providers with issue-specific knowledge z_i solve any problem in issue i if the difficulty of the problem lies between 0 and z_i. Service providers increase this knowledge at a cost proportional to the size of the interval of knowledge. That is, learning how to solve problems in the interval $[0, z_i]$ costs $c_i z_i$, where c_i is the constant per-period unit cost of acquiring knowledge. This setting captures the fact that harder problems are more costly to learn. As in Rosen (1983), the cost of acquiring knowledge about problems is independent of its utilization. Therefore, the model is characterized by increasing returns to the use of knowledge. Without loss of generality, I assume that the outside option of not working in any market is 0.

Markets. There are four markets in this model- two internal and two external. First, there is an internal market for each issue, each of which is characterized by a one-to-one relationship between the client and the service provider. Second, there is an external market for each issue. In the external market i, each service provider works for $n_i \in \mathbb{R}_+$ clients (i.e., leverage), which is an endogenous variable to the problem.

Communication Cost-Firm-Specific Knowledge. In addition to the knowledge required to solve problems, clients need firm-specific knowledge for production. As in Arrow (1974) and Cremer et al. (2007), firms develop codes within the organization. Therefore, in-house and external service providers differ by the levels of this type of knowledge.[25] I capture this idea with a communication cost in the external market. This cost, denoted as $h_i \in (0,1)$ is the time that external providers spend on each client in addition to the time that it takes to solve problems. Intuitively, the more specific the knowledge that the firms use, the larger is the cost of externally contracting, as it takes more time to communicate that knowledge. There is only one communication cost per issue and this knowledge may differ across issues. A

[24] Knowledge is cumulative within an issue. That is, low levels of knowledge in one issue are not necessary to acquire higher levels of knowledge in another issue.

[25] As anecdotal evidence supporting these knowledge differences across markets, Drutman (2010, p. 43) cites one in-house lobbyist comment about external lobbyists: "I don't believe most folks understand this stuff [the company-specific issues], and it's not worth their time to get up to speed."

way to interpret h_i is with the average firm-specific knowledge of M clients.[26]

Any study on firm boundaries requires, at least implicitly, a definition of what a firm is. In this chapter, a firm is an abstract place that allows agents to develop a common code. That is, it is the place where the owner attenuates communication costs due to differences in firm-specific knowledge across internal and external markets. This concept of a firm differs sharply from the aforementioned vertical integration literature. In the incentive literature, a firm is an abstract place that attenuates renegotiation or monitoring costs or maximizes the surplus via ex-ante investments due to the command or residual control rights of the firm's owner.

Time constraint. Each service provider has a time constraint with total labor supply endowment normalized to 1. As in Garicano (2000), the constraint implies that the expected time for solving problems has to be equal to the total labor endowment. That is, $1 \geq \Pr(problem) \cdot n_i \cdot [1 + \mathbf{1}_{external_i} \cdot (h_i - 1)]$. Let $\Pr(problem)$ be the probability that service providers face problems and n_i the leverage (number of clients) of the service provider for issue i. The leverage is 1 in the internal market and is endogenously determined in the external market. Finally, $\mathbf{1}_{external_i}$ is an indicator function equal to 1 in the external market of issue i and 0 otherwise. I assume that the burden of this cost falls fully on the receiver, as is standard in the literature.[27]

Wages. Service providers receive a constant per-period unit wage compensation w. The total wage compensation is wz if the service provider has knowledge z. Let w_i^j be the wage in the j-th market (j=I for *Internal* and E for *external*) for issue $i = A, B$. The wages are endogenous to the problem.[28]

Summing up, the total number of clients and set of service providers (M, N), the distribution of problems (F_i), the cost of acquiring issue-specific knowledge (c_i), and the average firm-specific knowledge of the economy (h_i) for each issue $i = A, B$ are exogenous parameters in the economy. The endogenous variables are the vector of wages (w_i^j), allocation of clients to each market and levels of acquired knowledge (z_i) and leverage (n_i) for each service provider for each issue.

Timing. First, service providers choose the breadth (market in which they want to work) and depth of the level of issue-specific knowledge. Then, the problems of the clients are realized. Finally, clients are allocated to markets; the markets clear and production takes place. This timing is similar in spirit to Murphy (1986) and Garicano and Hubbard (2007). To decrease the burden of this chapter, I assume that a matching technology allocates clients across markets by maximizing the total surplus of the economy. In section 1.2.4, I show that the main results obtained under

[26] Clearly, more complicated codes can affect in-house staff. An intuitive way to understand this differential result across markets is as the extra cost of acquiring firm-specific knowledge between internal and external service providers.

[27] See for instance, Bolton and Dewatripont (1994) and Garicano (2000).

[28] To simplify the presentation of the main setting, I will introduce the notation of the wages for generalists and specialists in Section 1.2.3.1.

this arrangement are robust to several other alternatives.

Solution. As service providers take decisions before the problems are realized, they do so by maximizing the expected surplus of the client and the service provider given the market they choose to go. Then, the demand is realized and the optimal organization of the economy allocates clients across markets. In this setting, there is a fixed cost of acquiring knowledge that is independent of its utilization. The organization of the economy maximizes the net payoff of all the agents in the economy by exploiting the increasing returns to the use of knowledge. I first characterize the optimal cutoffs that define the firm and the market boundary, respectively. The firm boundary is characterized by the level of difficulty at which the expected benefit of increasing the issue-specific knowledge equals the cost of doing so in the internal market. This is the maximum level of knowledge that the in-house service provider would reach. On the other hand, the market boundary is characterized by the level of difficulty at which the expected benefit of increasing the issue-specific knowledge for n clients equals the cost of doing so in the external market. The number of clients is determined by the external service provider's time constraint. With an infinite labor supply, the wages compensate for the costs of acquiring knowledge and, given a perfectly elastic labor supply, the demand determines the employment level in each market.

In this setting, service providers are ex-ante homogeneous and differ ex-post due to knowledge investments. This allows me to determine the distribution of knowledge as an equilibrium of a vertical integration model and to abstract from innate or pre-market comparative advantage. Under this framework, I learn not only about the service providers' earnings distribution, but also about the distribution of agents across vertical integration possibilities and, within each possibility, the degree of specialization.

In order to gain some intuition, I start by solving the problem when there is only one issue. Then, I move to the case in which clients face problems in both issues, and I consider the service providers' decision to become either a generalist or a specialist.

1.2.2 Benchmark: One Issue

Consider the case with only one issue. The problem is solved recursively: first, I characterize the in-house solution and then do the same for the external market. An intuitive way to think about this problem is that there is a set of service providers that try to increase the probability of solving their clients' problems as much as they can. As these providers are limited by the cost of acquiring knowledge, they cannot solve any problem. Then, a second set of service providers will choose the necessary level of knowledge to solve the problems of clients for whom the first set of providers could not find a solution.

1.2.2.1 *Internal market: Firm Boundary*

The time constraint of the provider $1 = (1 - F(0)) \cdot 1$. That is, the service provider allocates her one endowed unit of time to one client, who faces problems with probability $(1 - F(0))$. The joint surplus in the internal market is $F(z) - cz$. The problem is, thus, to choose the length of the interval of issue-specific knowledge acquired to maximize the expected output. The solution of this problem is characterized by the following first-order condition:

$$f(z^*) = c$$

The level z^* represents the knowledge level at which the marginal benefit of having someone in-house is equal to the cost of acquiring knowledge. The marginal benefit of this problem represents the increase in the probability that a problem will be solved.

1.2.2.2 *External Market: Market Boundary*

The ex-ante payoff function for a client in the external market is $[F(z) - F(z^*)] - w^E z$. The net earnings of the service provider are $nw^E z - cz$, given that the external provider has leverage equal to n clients. Each of these clients draws one problem. External service providers are asked to solve $(1 - F(z^*)) \cdot n$ problems, which they can address in $(1 - F(z^*)) \cdot n \cdot h$ units of time. Therefore, the external service provider's time constraint is represented by $1 = (1 - F(z^*)) \cdot n \cdot h$ and is limited by the in-house provider's issue-specific skill z^*. Since each service provider has one unit of time available, the leverage of the external provider n is implicitly given by the time constraint.

The joint surplus in the external market is $[F(z) - F(z^*)] - \frac{c}{n}z$.[29] Finally, notice that, from the time constraint, I can solve for $\frac{1}{n} = h \cdot (1 - F(f^{-1}(c)))$. Then, the objective function is $[F(z) - F(z^*)] - czh(1 - F(f^{-1}(c)))$. The solution of this problem is characterized by the following first-order condition:

$$f(z^{**}) = ch(1 - F(f^{-1}(c)))$$

The external market allows service providers to acquire higher levels of skills by sharing the costs of acquiring these skills with several clients. Since in this market, for each problem, the marginal benefit (given by the frequency) of bringing someone in-house is lower than the marginal cost, no single client can hire an internal provider. However, the beauty of the market is that it allows several clients to share the acquiring-skills cost of the service provider to the point at which the marginal cost per client intersects the per-client marginal benefit. Comparing the two first-order conditions and applying the implicit function theorem, I state my first result.

[29] As the expected surplus of one client is $[F(z) - F(z^*)] - w^E z$, the total surplus for n clients is $n[F(z) - F(z^*)] - nw^E z + nw^E z - cz$.

Lemma 1.1 *The knowledge of external service providers is larger than the knowledge of internal service providers (i.e., $z^* < z^{**}$), and both knowledge levels are decreasing in c (i.e., $\frac{\partial z^*}{\partial c}, \frac{\partial z^{**}}{\partial c} < 0$). The larger the firm-specific component, the lower the knowledge acquired by the external providers (i.e., $\frac{\partial z^{**}}{\partial h} < 0$).*

This result simply states that service providers differ by their level of issue-specific skills according to their breadth, and in turn, this knowledge is always strictly decreasing in the cost of acquiring knowledge. As in-house staff own the code within the organization, they are not affected by changes in firm-specific knowledge. However, external service providers acquire a lower level of issue-specific knowledge when the firms have more specific codes within their organizations. I see this result as a corollary of Becker and Murphy (1992). They claim that the degree of specialization (depth of issue-specific knowledge) is limited not only by the extent of the market, but also by the costs of *coordinating* specialized workers. In this chapter, the relevant *coordination* cost occurs between clients and external service providers and is given by the firm-specific knowledge. The next results provide a characterization of the optimal way to allocate clients to markets.

Lemma 1.2 *There are two cutoffs of knowledge levels z^* and z^{**}. Clients with sufficiently easy (very frequent) problems (i.e., $z \leq z^*$) go to the internal market, while clients with intermediate levels of difficulty (medium-level of frequency) (i.e., $z^* < z \leq z^{**}$) hire external service providers. Finally, clients that face hard (infrequent) problems (i.e., $z > z^{**}$) do not hire any service provider.*

Figure 1.1 shows these two cutoffs. The left shaded (green) area represents the activities for which the marginal benefit of conducting the activities in-house is larger than or equal to the marginal cost. The cut-off z^* represents the firm-boundary, whereas the cut-off z^{**} denotes the market boundary. The white (intermediate) area represents the activities in which clients outsource the service, whereas the right, dark (black) area denotes clients that do not use service providers. The organization of the economy gives rise to consulting by exception, whereby in-house workers deal with the most common problems, and external service providers deal with the less frequent and harder problems. The next result provides some intuition for when clients differ by their level of firm-specific knowledge. The expected (ex-post) surplus is the surplus in the economy before (after) the problems have been realized.

Lemma 1.3 *The expected surplus in the external market is maximized when clients with high firm-specific knowledge are allocated to the internal market. The ex-post surplus in the external market can be increasing or not in the clients' firm-specific knowledge h.*

The basic intuition behind the ex-ante surplus result is that the economy saves communication costs when clients with high firm-specific knowledge avoid the external market and conduct their activities in-house.[30]This is straightforward and comes directly from applying the Envelope theorem to the objective function of the external market. For the ex-post case, the main trade-off in the economy is given by the leverage and the wage bill. When firm-specific skills are greater, the time constraint kicks in, and, therefore, external service providers have less leverage. As a consequence, the service provider acquires a lower level of issue-specific knowledge and, therefore, is cheaper to hire. The economy maximizes the ex-post surplus allocating clients with more, rather than less firm-specific knowledge, to the external market when the marginal effect on leverage is larger than the marginal effect on the wage bill. This depends on the specific distribution of problems F and the value of the cost of acquiring knowledge c.

1.2.2.3 *Equilibrium*

Let N_I (N_E) be the number of service providers in the internal (external) market. As there are M homogeneous clients, and each receives one and only one problem independently and identically distributed according to F, this distribution is also useful to represent the total demand in the economy. Intuitively, if there are $\frac{1}{3}$ tasks that are easy, for M large, there will be $\frac{M}{3}$ clients facing easy problems. The equilibrium conditions in this economy are for the internal market:

$$M \int_{0}^{f^{-1}(c)} f(x)dx = N_I$$

This condition establishes that the number of providers that go to the internal market is equal to the number of internal market demanders. And for the external market:

$$\frac{M}{n} \int_{f^{-1}(c)}^{f^{-1}\left(ch(1-F(f^{-1}(c)))\right)} f(x)dx = N_E$$

That is, the total number of service providers needed equals the total number of clients demanding external services divided by the number of clients that each service provider works for. Finally, as all the service providers should get the same ex-ante payoff:[31]

[30] For an early discussion of this, see Monteverde (1995).

[31] Notice that this is without loss of generality. If the individual rationality constraint in the internal market is $w^I z - cz \geq \phi^E$, where ϕ^E is the payoff in the external market, $w^I = \frac{\phi^E}{z} + c$, which implies that the objective function in the internal market is $F(z) - cz - \phi^E$. In this case, the first-order condition does not change. For the external market, the individual rationality constraint implies that $w^E = \frac{\phi^I}{nz} + \frac{c}{n}$. The objective function in this market will be $[F(z) - F(z^*)] - \frac{cz}{n} - \frac{\phi^I}{n}$, which would give us the same first-order condition. The payoffs are $\left(w^I - c\right) z^* - \phi^E =$

$$0 = \left(w^I - c\right) z^* = \left(nw^E - c\right) z^{**}$$

That is, the equilibrium wages are given by $w^E = \frac{c}{n} = \frac{w^I}{n}$ for $z^*, z^{**} > 0$. To characterize the optimal organization of the economy, I solve it as follows: first, internal and external market-clearing conditions give me the number of internal and external service providers. Note that, as $(1 - F(f^{-1}(c)))\, h = \frac{1}{n}$, the external providers' time constraint uniquely determines the number of clients for whom each external provider works. Finally, wages are given by the ex-ante equality payoff condition. Although the net earnings in equilibrium are zero for both types of providers, the earnings for internal providers are cz^*, whereas the earnings for external providers are $n\frac{c}{n}z^{**}$. Thus, the excess of gross earnings for external providers is proportional to the differences in knowledge levels and is given by $c\left[f^{-1}(ch(1 - F(f^{-1}(c)))) - f^{-1}(c)\right] > 0$.

Vertical Integration The fraction of vertically integrated clients is given by:

$$VI = \int_0^{f^{-1}(c)} f(x)dx \Big/ \int_0^{f^{-1}\left(\frac{c}{n}\right)} f(x)dx$$

and using the fundamental theorem of calculus I get:

$$VI = \frac{F(f^{-1}(c))}{F(f^{-1}\left(\frac{c}{n}\right))} = \frac{F(f^{-1}(c))}{F(f^{-1}((1 - F(f^{-1}(c)))\, hc))}$$

Theorem 1.1 *The fraction of vertically integrated clients increases with firm-specific knowledge h and decreases with the cost of acquiring issue-specific skills.*

An increase in firm-specific knowledge h decreases the knowledge level acquired by external service providers z^{**}. This modifies the market boundary without changing the firm boundary. As a consequence, the fraction of vertically integrated clients is larger. The effect of the cost of acquiring knowledge on the integration patterns is not trivial, as it modifies both the firm and market boundaries. The theorem states that the change in the cost affects disproportionately more the internal than the external market. Therefore, a decrease in this cost, increases the integration in the economy.

Comparative Statics Before I explore the case with two issues, I conduct some simple comparative static exercises using the case in which the CDF is the exponential function $F(z) = 1 - e^{-\lambda z}$, as in Garicano (2000), Garicano and Rossi-Hansberg (2006) and Caliendo and Rossi-Hansberg (2012). In this case, the knowledge levels for internal and external service providers are given by:[32]

$\left(nw^E - c\right) z^{**} - \phi^I = 0$, which is equal to $\phi^I - \phi^E = \phi^E - \phi^I = 0$. This implies that $\phi^E = \phi^I$ and $2\phi^I = 0$.

[32] I include the restrictions $\lambda > c$ and $\lambda^2 > c^2 h$, to ensure that $z^*, z^{**} > 0$.

$$z^* = -\frac{1}{\lambda}\ln(\frac{c}{\lambda}) \text{ and } z^{**} = -\frac{1}{\lambda}\ln(\frac{c^2h}{\lambda^2})$$

and it is easy to see that $z^{**} > z^*$.[33] The levels of earnings in the economy are:

$$w^I z^* = -\frac{c}{\lambda}\ln(\frac{c}{\lambda}) \text{ and } nw^E z^{**} = -\frac{c}{\lambda}\ln(\frac{c^2h}{\lambda^2})$$

It is easy to see from these equations that external service providers earn more than in-house staff due to their additional issue-specific skills. From the time constraint, it is easy to see that the number of clients that a particular external provider has is given by $n = \frac{\lambda}{hc}$, whereas the total number of suppliers in equilibrium is:

$$N_I = M\frac{(\lambda - c)}{\lambda} \text{ and } N_E = Mhc^2\frac{(\lambda - ch)}{\lambda^3}$$

Total clients are N_I in the internal market and $\frac{M}{\lambda^2}(c\lambda - c^2h)$ in the external market. Finally, the fraction of vertically integrated clients is given by:

$$VI = \frac{\lambda(\lambda - c)}{\lambda^2 - c^2h} = \frac{n(\lambda - c)}{n\lambda - c}$$

Once I have characterized the equilibrium objects, I can obtain some comparative static results. Table 1.1 summarizes the signs of these effects. Unfortunately, some of these effects do not have a clear sign and, thus, depend on the specific values of the parameters.

The first row gives the comparative static results when I change the firm-specific knowledge component h. The internal providers' knowledge does not change; however, as h is larger, the external providers' time constraint is more restrictive, and their knowledge decreases. As the time constraint binds, a change in some other variable must compensate for this increase in the specific component h. In this case, a decrease in leverage. Finally, the lower the level of external service providers' knowledge, the smaller is the mass of clients that can solve problems in the external market and, as a consequence, the lower the vertical integration fraction.

The second row shows the comparative static results when I change the cost c. The effects of this variable on the firm and market boundaries are derived in Lemma 1.1. As the leverage is defined by the time constraint, a decrease in the cost c makes it more likely that service providers face problems; therefore, each of them can work with fewer clients.[34]

[33]Notice that given that $\lambda > c$ and $h \in (0, 1)$, $\lambda > ch$, which implies that $n > 1$, and given that $z^{**} = -\ln(\frac{c^2h}{\lambda^2})\left(\frac{1}{\lambda}\right) = -\ln(\frac{c}{n\lambda})\left(\frac{1}{\lambda}\right)$, I have that $z^{**} > z^*$.

[34]The comparative static exercises for the parameter of the CDF, λ, show that $sign(\frac{\partial z^*}{\partial \lambda}) = sign(\ln(\frac{c}{\lambda}) + 1))$, $sign(\frac{\partial z^{**}}{\partial \lambda}) = sign(\ln(\frac{c^2h}{\lambda^2}) + 2))$ and $sign(\frac{\partial VI}{\partial \lambda}) = sign(\lambda^2 + hc^2 - 2ch\lambda)$. The signs of these functions depend on the specific values of the parameters.

1.2.3 Problem with two Issues

1.2.3.1 *Internal Market: Firm Boundary*

The ex-ante objective function for the client is $\pi^S(z_A, z_B; c_A, c_B) = F_A(z_A) + F_B(z_B) - w_A^I z_A - w_B^I z_B$, as the client faces problems in both issues, A and B. With two issues, there are two relevant cases to look at: one in which the client uses two providers, each of whom works on one issue; and another in which the client uses only one person to deal with both issues. A service provider that works on only one issue is called a specialist, whereas a provider working on more than one issue is a generalist.

Hiring Two Specialists

As the time constraint for a service provider working on issue $i = 1, 2$ is $1 = (1 - F_i(0)) \cdot 1$, the solution of this problem is given by:

$$f_i(z_i^*) = c_i$$

Hiring One Generalist

If the client hires only one person, it is necessary to include the time allocation decision to maximize the joint surplus. Let $t_A \in [0, 1]$ be the fraction of time that the service provider spends on issue A. In this case, the objective function for the client is:

$$t_A F_A(z_A) + (1 - t_A) F_B(z_B) - w^I [t_A z_A + (1 - t_A) z_B]$$

where w^I is the wage for the generalist internal service provider. The service provider's earnings depend on the time she spends on each issue. She will receive a payment $t_A w^I$ for each level of knowledge acquired in issue A and $(1 - t_A) w^I$ for each unit of knowledge in issue B. The time constraint implies that the service provider has one unit of time to allocate a fraction t_A to problems of issue A and a fraction $(1 - t_A)$ to problems of issue B. Each of these problems will occur with probability $1 - F_i(\widehat{z_i})$, where $\widehat{z_i} \geq 0$ is an endogenous object that represents the minimum issue-specific knowledge problem that the service provider will face. As the time constraint is $1 = t_A(1 - F_A(\widehat{z_A})) + (1 - t_A)(1 - F_B(\widehat{z_B}))$, t_A can be expressed as follows:

$$\frac{F_B(\widehat{z_B})}{[F_B(\widehat{z_B}) - F_A(\widehat{z_A})]} = t_A$$

Lemma 1.4 *The optimal time allocation is characterized by a corner solution. The provider should give all her time to the activity she is more likely to face. That is, if type i problems are more likely to occur, then the provider should not devote any time to activity $j \neq i$.*

The previous lemma states that the time constraint that I have used cannot characterize generalists in the internal market. Rosen (1983) provides a solution to the existence of generalists: complementarities among tasks. With complementarities, the service provider spends less time solving the same number of problems or, alternatively, more problems in the same time. As the service provider's time constraint determines the time allocation, a natural way to include complementarities is with a constant $\theta > 1$ as follows:[35]

$$1 = \theta \left[t_A(1 - F_A(\widehat{z_A})) + (1 - t_A)\,(1 - F_B(\widehat{z_B})) \right]$$

and solving for t_A: [36]

$$\frac{(1-\theta) + \theta F_B(\widehat{z_B})}{\theta \left[F_B(\widehat{z_B}) - F_A(\widehat{z_A}) \right]} = t_A^*$$

Proposition 1.1 *There is a range of θ such that $t_A \in (0,1)$. For this range of values of θ, generalists do not exist for $\widehat{z_A} = \widehat{z_B} = 0$.*

This result states that the inclusion of complementarities allows the existence of generalists and that these service providers are never allocated to solve the most frequent problems for both issues in the internal market. Figure 1.2 provides more intuition on this result. Plugging t_A^* into the objective function, I get:

$$\pi^G\left(z_A, z_B; t_A^*, c_A, c_B\right) = t_A^* \left[F_A\left(z_A\right) - F_B\left(z_B\right) \right] + F_B\left(z_B\right) - c_A z_A - c_B z_B$$

The first-order conditions for the levels of issue-specific knowledge for each issue are:

$$[z_A] : t_A^* f_A\left(z_A\right) - c_A = 0 \implies z_A^g = f_A^{-1}\left(\frac{c_A}{t_A^*}\right),$$

$$[z_B] : -t_A^* f_B\left(z_B\right) + f_B\left(z_B\right) - c_B = 0 \implies z_B^g = f_B^{-1}\left(\frac{c_B}{1 - t_A^*}\right)$$

where z_i^g denotes the level of issue-specific knowledge that a generalist in the internal market will reach on issue i. Notice that for $t_A \in (0,1)$, $z_A^g < z_A^*$ and $z_B^g < z_B^*$. That is, if there are generalists in the internal market, they know less about each issue than the internal service providers working on the firm boundary. Let $z_A^{=}, z_B^{=}$ be the set of maximum values for which a client prefers to have two

[35] To understand this better, assume that there is one time unit of endowment to solve a number of problems x, and each problem requires t units of time. That is, the time constraint is $1 = tx$, which implies that we need $\frac{1}{x}$ units of time to produce x units. With complementarities, $1 = \theta t x$, and then one needs fewer units of time $\frac{1}{\theta x} < \frac{1}{x}$ to produce the same quantity x. A larger θ implies a larger level of complementarities.

[36] It is easy to see from this equation that $sign(\frac{\partial t_A^*}{\partial \theta}) = sign(F_A(\widehat{z}_A) - F_B(\widehat{z}_B))$. That is, if the provider is more likely to face problems of issue B than of issue A, the larger the level of complementarities between issues A and B, the larger the fraction of time spent on issue A.

specialists rather than one generalist; that is,

$$z_A^{=}, z_B^{=} \in \max \left\{ z_A, z_B | \pi^G \left(z_A^g, z_B^g; t_A^*, c_A, c_B \right) + \varepsilon = \pi^S \left(z_A, z_B; c_A, c_B \right) \right\}$$

For the sake of understanding, suppose that $z_j^* \geq z_i^* \geq z_j^= \geq z_i^=$, where z_j^* is the firm boundary of issue i.

Proposition 1.2 *There are at most two cutoffs (excluding the firm boundary) in the internal market for each issue. For issue j, the relevant cutoffs are $z_j^=$ and z_i^*. Clients with $z_j \leq z_j^=$ hire two specialists; clients with $z_i^* \geq z_j \geq z_j^=$ hire a generalist; and clients with $z_j \geq z_i^*$ hire one specialist.*

Corollary 1.1 *If $(z_A^=, z_B^=) = (z_A^*, z_B^*)$, there is only one cutoff. In this case, the internal market contains only specialists. If $(z_A^=, z_B^=) < (z_A^*, z_B^*)$, clients with $(z_A^=, z_B^=) \geq (z_A, z_B)$ will hire two specialists, and clients with $(z_A^=, z_B^=) \leq (z_A, z_B) \leq (z_A^g, z_B^g)$ will hire one generalist.*

Figure 1.2 shows the inclusion of the levels $z_A^=, z_B^=$ with $z_A^= < z_B^=$. In this figure, clients with $z_B, z_A \leq z_B^=$ use two specialists. For $z_B^= \leq z_B, z_A \leq z_A^*$, clients use one generalist. Finally, clients with $z_B^* \geq z_B > z_A^*$ use only one specialist in the internal market. Here, the main forces at play are the frequency and difficulty of problems. Intuitively, a single service provider can solve both types of problems if they are easy enough; however, as easier problems are more frequent, the single service provider's time constraint kicks in and restricts the provider to solving the most frequent problems for both issues.

To sum up, generalists may exist because of complementarities. The internal market may have, at most, two different cutoffs. If there is no cutoff, then there are only specialists. If there is at least one cutoff, there are both specialists and generalists. In the case of one cut-off, specialists solve the most common problems, while generalists solve less common problems. With two cutoffs, some specialists solve the most common problems and other specialists solve the less common problems. The generalists solve the medium-frequency problems of one market and the less frequent problems of the other market.

1.2.3.2 *External Market*

Hiring Two Specialists

The ex-ante profits for a client in the external market i are $(F_i(z_i) - F_i(z_i^*)) - w_i^E z_i$, and the net earnings of the service provider are $n_i w_i^E z_i - c_i z_i$, given that the external provider works with n_i clients. The service provider's time constraint is $1 = (1 - F(z_i^*)) \cdot n_i \cdot h_i$; then, the joint surplus is $[F_i(z_i) - F_i(z_i^*)] - c_i z_i (1 - F(z_i^*)) h_i$. The solution of this problem is characterized by the following first-order condition:

$$f_i(z_i^{**}) = c_i(1 - F(z_i^*)) h_i$$

Hiring one Generalist

The time constraint for a generalist in the external market is given by:

$$1 = \theta \left\{ t_A^E \left[\left(1 - F_A\left(\widetilde{z}_A\right)\right) n_A^g h_A \right] + \left(1 - t_A^E\right) \left[\left(1 - F_B\left(\widetilde{z}_B\right) n_B^g h_B\right) \right] \right\}$$

That is, a generalist spends a fraction t_A^E of her time dealing with problems of issue A. As in the internal market, $\widetilde{z}_i$ is an endogenous variable representing the easiest problem that the provider will face. The service provider has to allocate her available time to n_i^g clients, each with h_i industry-specific knowledge (communication costs). This time constraint is the same as that of an in-house generalist when $n_i^g = h_i = 1$ for $i = A, B$. Solving for t_A, I get:

$$\frac{1 - \theta \left[\left(1 - F_B\left(\widetilde{z}_B\right)\right) n_B^g h_B \right]}{\theta \left[\left(1 - F_A\left(\widetilde{z}_A\right) n_A^g h_A\right) - \left(1 - F_B\left(\widetilde{z}_B\right) n_B^g h_B\right) \right]} = t_A^{E*}$$

Lemma 1.5 *Generalist external service providers can only exist for the combinations of* $n_A^g h_A$ *and* $n_B^g h_B$ *such that* $\widetilde{z}_A \in \left(Q_A\left(1 - \frac{1}{\theta n_A^g h_A}\right), Q_A\left(1 - \frac{n_B^g h_B}{n_A^g h_A}\right)\right)$ *and* $\widetilde{z}_B \in \left(z_I^B, Q_B\left(1 - \frac{n_A^g h_A}{n_B^g h_B}\left(1 - F_A\left(\widetilde{z}_A\right)\right)\right)\right)$ *or* $\widetilde{z}_B \in \left(Q_B\left(1 - \frac{1}{\theta n_B^g h_B}\right), 1\right)$, *where* Q_i *is the quantile function of the* i*-th issue.*

Notice the trade-off in the combinations of $n_A^g h_A$ and $n_B^g h_B$. When $n_A^g h_A$ increases, the lower and upper bounds for $\widetilde{z}_A$ increase. That is, the generalist external provider can handle harder problems in topics A. However, this increase in the leverage in issue A decreases the maximum level of difficulty in issue B that the provider can handle. Once I have characterized the service provider's time constraint, I can solve the problem for both the clients and the service provider. The main difference here with respect to the internal market is that the acquisition of knowledge will depend on $n_A^g + n_B^g$ clients. Let w_i^g be the wage of the generalist in the external market for issue i.[37] The joint surplus for the external service provider and her clients is:

$$t_A^{E*} n_A^g F_A\left(z_A\right) + \left(1 - t_A^{E*}\right) n_B^g F_B\left(z_B\right) - \left[w_A^g t_A^{E*} n_A^g z_A + w_B^g \left(1 - t_A^{E*}\right) n_B^g z_B \right]$$

In this problem, there are n_i^g clients looking for solutions to issue i problems. The total cost of this joint surplus is given by the wage bill of the knowledge acquired to solve type i problems, times the total number of clients of issue i and the time invested in type i problems. The first-order conditions imply that:

$$z_A^g = f_A^{-1}\left(\frac{c_A}{t_A^{E*} n_A^g}\right),$$

[37] As it will be clear below, I cannot guarantee at this stage that I can obtain a single wage for generalists in the external market. This is different from the internal market case, as the knowledge for both issues can be represented by a cutoff level in only one issue.

$$z_B^g = f_B^{-1}\left(\frac{c_B}{(1-t_A^{E*})\,n_B^g}\right)$$

where z_i^g denotes the level of issue-specific skills of the generalist in the external market i-th.

Lemma 1.6 *If $t_A^{E*} = 1$, then the generalists' knowledge is the same as the specialists' in the market boundary $z_A^g = z_A^{**}$. Furthermore, $sign(t_A^{E*} - \frac{n_A^s}{n_A^g}) = sign(z_A^g - z_A^{**})$, where n_A^s is the leverage of specialists in market boundary for market A.*

The first part of this result states that when generalists tend to spend all their time on one issue, they converge to have the same knowledge level as specialists' in the market boundary. The second part of the result states that external generalists acquire less knowledge on issue A than external market specialists located in the market boundary if the fraction of time spent on issue A is lower than the ratio of the leverages of specialists and generalists. So, here again, as above, there is a positive relationship between leverage and knowledge.

Lemma 1.7 *For a given issue, external providers always acquire more knowledge than any type of internal providers.*

This result is similar to Lemma 1.1, and implies that external service providers have higher earnings.

Table 1.2 shows all the possible combinations of vertical integration and external contracting that clients can face. Figure 1.3 shows an alternative way to present this result. In the figure, left shaded areas (green) represent the regions of vertical integration for each issue. The black areas show the range of values in which clients prefer not to hire any service provider. Finally, the white area shows the regions in which clients prefer to contract externally.

1.2.3.3 *Equilibrium*

The total number of employed service providers is $N_A^I + N_B^I + N_A^E + N_B^E$, where N_i^k denotes the number of service providers in market k=internal or external for issue $i = A, B$.

Internal Markets

Suppose that $z_j^* \geq z_i^* \geq z_j^= \geq z_i^=$. Then, the internal market-clearing condition for topic i is:

$$M\int_0^{z_i^*} f_i(x)dx = N_i^I \text{ and } M\left[\int_0^{z_j^*} f_j(x)dx - \int_{z_j^=}^{z_i^*} f_j(x)dx\right] = N_j^I$$

and the clearing condition for the external markets is:

$$\frac{M}{n_i}\int_{z_i^*}^{z_i^{**}} f_i(x)dx = N_i^E \text{ , and}$$

$$M\left[\int_{z_j^*}^{z_j^{**}} f_j(x)dx/n_j + \int_{\widetilde{z}_j}^{z_j^g} f_j(x)dx/n_j^g\right] = N_j^E$$

Finally, as the service providers are ex-ante equal, and there is an infinite supply, they get the same expected earnings; that is, in the internal market, the payoff of specialists facing problems that appear too often is $0 = \left(w_i^I - c_i\right) z_j^= = \left(w_j^I - c_j\right) z_j^=$, and this should be equal to the payoff of the generalists $w^I\left[t_A^* z_i^* + (1-t_A^*)\, z_i^*\right] - c_A z_i^* - c_B z_i^*$, which must be equal to that of specialists working on the least frequent problems $\left(w_i^I - c_j\right) z_i^* = \left(w_j^I - c_j\right) z_j^*$. In the external market, specialists earn $\left(n_i w_i^E - c_i\right) z_i^{**} = \left(n_j w_j^E - c_j\right) z_j^{**}$, whereas generalists earn $\left(w_A^g t_A^{E*} n_A^g - c_A\right) z_A^g + \left(w_B^g\left(1 - t_A^{E*}\right) n_B^g - c_B\right) z_B^g$. This implies that the wages in equilibrium are $w_i^I = c_i$, $w^I = c_A + c_B$, $w_i^E = \frac{c_i}{n_i}$, $w_A^g = \frac{c_A}{t_A^{E*} n_A^g}$ and $w_B^g = \frac{c_B}{\left(1-t_A^{E*}\right)n_B^g}$. Finally, the earnings in the internal market are $c_i z_j^=$ for specialists with frequent problems, $c_i z_i^*$ for specialists with infrequent problems and $(c_A + c_B)\, z_i^*$ for generalists. In the external market, the earnings are $c_A z_A^g + c_B z_B^g$ for generalists and $c_i z_i^{**}$ for specialists.

Finally, the following result shows that the relationship between vertical integration and the firm-specific levels found in Theorem 1 still holds in the case of two issues.

Lemma 1.8 *The level of vertical integration in the industry is decreasing in the firm-specific knowledge levels h_i and h_j.*

1.2.4 Alternative Arrangements

The following cases are alternative arrangements for the economy when there are one or two issue-specific knowledge levels. To simplify the discussion, I present the results only for the case of one issue. The main point here is that changing some assumptions of the model gives the same key predictions: clients with easy or frequent problems solve their problems in-house, whereas clients with harder or infrequent problems ask for help in the external market.

1.2.4.1 Multiple Layers in the External Market

Above, I have considered a situation in which there is only one layer in the external market. The problem with several layers is similar and is also solved recursively. The solution of the problem for the i-th layer is given by:

$$z_i^{**} = f^{-1}\left(ch(1 - F(z_{i-1}^{**}))\right)$$

with $n_i = \frac{1}{h(1-F(z_{i-1}^{**}))}$.

Lemma 1.9 *For any two layers in the external market, i and j, with $i < j$, $z_j^{**} > z_i^{**}$ and $n_j > n_i$.*

This result states the positive relationship between levels of knowledge and leverage of clients. I provide preliminary empirical evidence in Section 1.3.3.3. There are at least two different interpretations for this result. First, service providers with more knowledge can solve problems for a larger set of clients. Second, a larger set of clients can finance the service provider's acquisition of more knowledge.[38] The implications for earnings are clear. External service providers in layer j earn $n_j w_j^E z_j^{**}$, where w_j^E represents the wage for the j-layer in the external market. This means that for two layers, j and i with $j > i$, $n_j w_j^E z_j^{**} > n_i w_i^E z_i^{**}$. Service providers in layer j earn more than i as they acquire a higher level of knowledge. The equilibrium condition in the internal market is the same as before; however, for the case of the external market, the total number of external service providers required is $M\left[\sum_i \left[\frac{F(z_i^{**})-F(z_{i-1}^{**})}{n_i}\right]\right]$.

Lemma 1.10 *Let $z \sim Exp(\lambda)$, $n_i = \left(\frac{\lambda}{ch}\right)^i$ and $z_i^{**} = -\ln\left[\frac{1}{hn_{i+1}}\right]\left(\frac{1}{\lambda}\right)$. In this case, the difference in knowledge between any two layers i and $i-1$ is given by the constant $z_i^{**} - z_{i-1}^{**} = \frac{1}{\lambda}\ln(\frac{\lambda}{ch})$; the total number of required layers to cover the entire external market area is given by the ceiling function of $\frac{1-z^*}{z_i - z_{i-1}}$, $\left\lceil \frac{1-z^*}{z_i-z_{i-1}} \right\rceil = \frac{\lambda+\ln(\frac{c}{\lambda})}{\ln(\frac{\lambda}{ch})}$. The difference in the leverage of the external providers is given by $n_i - n_{i-1} = \left(\frac{\lambda}{ch}\right)^{i-1}\left(\frac{\lambda}{ch} - 1\right)$, which is increasing in i. Finally, the cost that each client pays in the external market decreases at an increasing rate as leverage increases. That is, $\frac{c}{n_i} - \frac{c}{n_{i+1}} = \left(\frac{\lambda}{ch} - 1\right) / \left(\frac{\lambda}{ch}\right)^{i+1}$, which is decreasing in i.*

Figure 1.4 represents the case with three layers in the external market. The level z_i^{**} represents the level of knowledge in the i-th layer of the external market. This level is characterized by the marginal condition $f(z_i^{**}) = \frac{c}{n_i}$.

1.2.4.2 Clients receive a continuum of problems

Consider a situation in which all the clients receive a continuum of problems in the interval $[0, 1]$ from F- an extension that can be easily included in my setting. In this case, all the clients will hire in-house providers for the most frequent tasks and external providers for the least frequent tasks. In this economy, all the clients will be the mixed-type; that is, they will go to both internal and external markets.

1.2.4.3 Clients with in-house and external providers

Consider a situation in which the joint surplus of n clients and $n+1$ service providers is maximized, and each client receives only one problem. Here, again, the clients will

[38] This positive relationship between knowledge and the number of clients is due to the supermodularity of z and n in the objective function. To see this, notice that the objective function is $\underset{z_i}{argmax}\left[F(z_i) - F(z_{i-1}^*)\right] - \frac{cz_i}{n_i}$. Then, the cross-derivative of this function with respect to both z_i and n_i is $\frac{c}{n_i^2} > 0$.

hire someone in-house for the most common activities and an external provider for the least common activities. There are two natural settings: in one, the joint surplus is maximized over the external provider's level of knowledge once the knowledge of the internal providers has been decided. In the other, the joint surplus is maximized by simultaneously choosing the knowledge levels for both internal and external service providers. I call the first situation the sunk in-house investment and the second one the flexible in-house investment situation.

Sunk in-house investment

In this case, the joint surplus is given by $nF(z) - cz - ncz^*$ with the time constraint for the external provider equal to $1 = (1 - F(z^*))nh$. Notice that, in this case, the production of all the clients depends on the external and not the internal service provider's knowledge. The external service provider's knowledge maximizes the joint surplus where:

$$f(z^{**}) = c(1 - F(z^*))h$$

This level corresponds exactly to the level of the external provider's knowledge found in Section 1.2.2.2.

Flexible in-house investment

In this case, the joint surplus is given by $nF(z_e) - cz_e - ncz_i$, where z_e (z_i) represents the knowledge level of the external (internal) provider. As in the sunk in-house investment, the maximum difficulty level of problems that can be solved for any of the n clients is given by the knowledge of the external provider. In this case, the external provider's time constraint is equal to $1 = (1 - F(z_i))nh$, and the first-order conditions are given by the following equations:

$$[z_i] : f(z_i) = \frac{1}{hz_e},$$

$$[z_e] : f(z_e) = c(1 - F(z_i))h$$

Although I have not been able to precisely compare the in-house and external knowledge levels with z^* and z^{**}, Proposition 3 in Garicano and Rossi-Hansberg (2006) suggests that in-house workers acquire less knowledge than they would absent the flexible in-house investment, since in-house workers substitute learning for asking (i.e., $z_i < z^*$ and $z_e > z^{**}$).

The relevant point here is that in-house workers solve the most common problems whereas external service providers solve the least common problems, as $z_i < z_e$ and $f' < 0$. Since their knowledge levels differ, external service providers earn more than in-house staff.[39]

[39] For the case of the exponential function, $z_i = z_e + \log\left(\frac{ch}{\lambda}\right)$, which implies that $z_e > z_i$ as $\lambda > ch$.

1.2.4.4 Matching clients and providers at the beginning

I have considered a situation in which clients are allocated to the market in which they can find solutions to their problems once the service providers have made knowledge investments. Alternatively, clients and service providers can be randomly matched before the problems are drawn; they maximize the joint surplus, and then the problems are realized. Consider the initial matching of $M_1 < M$ clients with service providers. Let M_2 be the number of clients that are not initially matched. The joint ex-ante surplus is given by $F(z) - cz$ with a first-order condition equal to $f(z^*) = c$. Then, the problems are realized. Among the M_i clients with $i = 1, 2$, a fraction $F(z^*)$ will have problems that can be solved in-house, and a fraction $1 - F(z^*)$ will have problems that need to be addressed in the external market. Therefore, $(M_1 + M_2)\, F(z^*) = M\ F(z^*)$ will find solutions in the internal market, while $M\ [1 - F(z^*)]$ will go to the external market. Therefore, assuming one layer and leverage equal to n, the number of required in-house providers will be $M_1 + M_2 F(z^*)$ and the number of required external service providers will be $\frac{M[1-F(z^*)]}{n}$. The leverage depends on the external service provider's time constraint, which in this case is $n = \frac{1}{h[1-F(z^*)]}$. For the equilibrium characterization, the total number of service providers and clients in the external market is the same, and the difference in the total number of clients in the in-house market is given by $M_1 F(z^*)$. With ex-ante homogeneous service providers and infinite supply, wages compensate for the cost of acquiring skills (i.e., $w = c$). The predictions in this case are almost identical to those in the case considered above: in-house workers solve the most common problems which are the easiest ones, while external service providers solve most difficult problems and, as a consequence, acquire more knowledge and earn more.

Finally, I present two simple results for the case in which clients have problems in two periods.

1.2.4.5 Two Consecutive Periods

For the case of one issue, once the clients have hired someone in-house, they may face a difficulty level in the future that the internal provider cannot manage. The probability that, in the first period, they hire someone in-house is $F(z^*)$, and the probability that, in the second period, that provider cannot manage the difficulty level is $F(z^*)(1 - F(z^*))$. Therefore, a fraction of the clients $F(z^*)^2$ will use internal providers in two consecutive periods; a fraction $(1 - F(z^*))^2$ will use external providers in both periods; and a fraction $2F(z^*)(1 - F(z^*))$ will use both types. The following result compares the probability of finding clients using only the internal market versus using only the external market.

Remark 1.1 *After two periods, the fraction of clients using in-house staff exclusively is larger than the fraction of clients using external providers exclusively if*

$c > f\left(Q\left(\frac{1}{2}\right)\right)$, where Q represents the quantile function.

For the case of two issues, I have an additional result. The probability that in the first period they have hired someone in-house for topic i-th is $F_i(z_i^*)$ and the probability that they need an external provider for issue i is $(1 - F_i(z_i^*))$.

Lemma 1.11 *If $F_j(z_j^*) \geq F_i(z_i^*)$, it is more common to find that clients with in-house providers in issue i in the first period hire external service providers for issue i in the second period.*

1.2.5 Discussion

1.2.5.1 Building Blocks

Rosen (1983) elaborates on the intuition that the return to the investment in knowledge is increasing in its rate of utilization because investment costs are fixed. Becker and Murphy (1992) propose that communication costs limit the extent of the market. Garicano (2000) builds upon these two previous intuitions by suggesting that the key trade-off of the organization occurs between the costs of communicating and acquiring knowledge. Using Garicano's approach, I think of the economy as an organization and transform the first (second) layer of the organization into the firm (market) boundary. My model is close to that of Garicano and Hubbard (2007), who focus on the role of hierarchies in the organization of human-capital intensive production. This chapter is different because I focus on the vertical integration problem and not on hierarchies and their relationship with the size of the market. Murphy (1986) and Garicano and Hubbard (2007) explains the existence of generalists with market uncertainty. My approach is closer to Rosen (1983), who explains generalists with complementarities. Furthermore, in my model, clients always have problems in two issues, whereas in Murphy (1986) and Garicano and Hubbard (2007), the models allow the possibility of not having problems in a given issue.

1.2.5.2 Ex-ante and Ex-post differences

The assumption of ex-ante homogeneous service providers is not realistic, but it simplifies the burden of my analysis. The important point, which is fully developed in Garicano and Rossi-Hansberg (2006), is that the introduction of ex-ante differences accentuates the sorting in the economy. The initially more knowledgeable people have comparative advantage in the external market, so they go to that market, and, in equilibrium, they end up acquiring even more knowledge than in the case of ex-ante homogeneous agents. The less knowledgeable agents go to the internal markets. Therefore, the differences (between in-house and external) in knowledge patterns hold for both cases: ex-ante homogeneous and heterogeneous service providers.

1.2.5.3 Other Skills

Knowledge may not be the only skill required to solve problems. For instance, Blanes-i-Vidal et al. (2012) and Bertrand et al. (2014) conclude that both connections and knowledge are important skills for lobbyists. My model can accommodate this view in two different ways. First, the choice variable of the service providers contains both social and human capital. So, what I call knowledge or skills in the model can be interpreted as an index that comprises both types of capital. This is certainly valid, as acquiring a network of policy makers is independent of the rate of use; thus, making connections has increasing returns to scale to their use. Second, one can think of the level of connections as an ex-ante difference among service providers (i.e., family networks), which, therefore, does not matter in an ex-ante homogeneity setting.

1.2.6 A Summary of the Main Predictions

In this section, I summarize the main results of the theoretical section:

1. Activities that occur often (or have low issue-specific knowledge requirements) are conducted in-house, whereas activities that occur less often (or are more difficult) are outsourced;

2. The economy saves communication costs with external service providers when clients with high firm-specific knowledge conduct their activities in-house;

3. Service providers' knowledge levels differs according to the market they are in. External providers have more knowledge than internal service providers. This implies that the levels of earnings differ across internal and external markets proportionally to the difference in the level of knowledge acquired. As a consequence, external providers earnings are larger than their counterpart internal ones;

4. The possibility of sharing the cost of acquiring knowledge with several clients allows external service providers to acquire more knowledge. Each of these clients is not able to form a profitable one-to-one match with the service providers, as the clients face problems that occur too infrequently. However, the beauty of the external market is that it allows clients to join interests and pay the learning costs for harder (non-frequent) problems;

5. In the case of one or two issues, the fraction of vertically integrated clients is increasing in firm-specific knowledge, decreasing in the cost of acquiring knowledge c and ambiguous in the parameters of the distribution of problems F;

6. When there are two issues, generalists and specialists exist in both internal and external markets. Generalists exist due to the presence of complementarities among issues. In the internal market, there can be, at most, three types of service providers. Specialists work either for the most or the least frequent issues faced in the internal market, whereas generalists solve the intermediate or less frequent problems.

1.3 Empirics

This section is divided into four subsections. In Section 1.3.1, I present the data and the institutional context. Section 1.3.2 presents the fixed effect estimation results at both the client and the transaction levels. Then, in Section 1.3.3, I provide some empirical evidence for other results derived from the theoretical section. Finally, in Section 1.3.4, I use the BP oil spill event to provide causal evidence on the effect of difficulty on vertical integration decisions.

1.3.1 Data and Institutional Context

The Lobbying Disclosure Act of 1995 (henceforth LDA) requires lobbyists to register and to report on their lobbying activities to the Senate Office of Public Records (henceforth SOPR). According to the Act, lobbying activity is defined as contacts with officials, including background work performed to support these contacts.[40] Two types of registrants are required to report under the LDA: external and internal lobbyists. External lobbyists, who work for lobbying firms, take on lobbying responsibilities for a number of different clients and, under the LDA, they are required to file a separate report for each of their clients. Internal lobbyists are *self-filing* organizations that conduct in-house lobbying activities. Both types of registrants are required to report good-faith information every three months. Up until the end of 2007, they were required to report these estimates biannually.[41]

The starting unit of observation is a lobbying report. Each SOPR report not only contains the name of the client and individual lobbyists, but also specifies the House(s) of Congress and federal agencies contacted, as well as the bills in which the client was interested.[42] Clients can have more than one report in a given period, as they can use both internal lobbyists and one or more groups of external lobbyists. The lobbying reports dataset starts at the first semester of 1999 and finishes with the second semester of 2014. It contains 44,039 clients and 56,759 lobbyists.

[40] The LDA defines a *lobbyist* as a person spending 20% or more of her time engaged in lobbying activities.

[41] For the sake of a better comparison, for most of the estimations, I will focus on semester-level time variation.

[42] Although, lobbying reports contain information on expenditures, I have not considered this variable in most of my analysis, as in-house reports include *non-disaggregated* expenditures not directly related to advocacy activities (i.e. office space).

1.3.1.1 Bills, Committees and RegData

Bills are legislative proposals that can be introduced at any time while the Congress is in session by any member of either house.[43] After introduction, the bill is referred to the appropriate committee or committees, based on the committees' jurisdiction, which is defined by congressional rules.[44] In the House of Representatives, this referral is controlled by the Speaker, following the advice of the House Parliamentarian; in the Senate, it is managed mainly by the Senate Parliamentarian on behalf of the presiding officer of the Senate. Parliamentarians are nonpartisan officials that provide technical assistance and expertise on the legislative procedure of the Congress. They serve for several years; indeed, there have been only five parliamentarians in each house since 1928.[45]

The objective of the committees is to study bills and consider whether or not to send them for further action. The committees are divided into sub-committees that have a narrower jurisdiction in the topics. The initial stage of this study process consists of public hearings at which committee or sub-committee members invite witnesses with the purpose of gathering relevant information. Witnesses are either specialists on the topic or people affected by the matter. They represent different views on the topic and can have different backgrounds, such as government, academia or business. Committees that manage more technical subjects require witnesses with more experience or higher education levels.[46]

I have web-scraped the name and title or occupation of all the witnesses in all the reported congressional hearings since 1999. I have classified these occupations into two groups: high and low levels of knowledge requirements. Titles that include PhD, professor or senior manager are classified as high, whereas all other occupations are classified as low. Examples of low-knowledge occupations are farm owners, farm producers, assistant secretaries and average citizens with an interest in the issue. Ideally, as the lobbying reports provide information on the lobbied bills, I would like to consider knowledge-intensity measures at the bill level. However, this is not feasible. First, not all the bills are studied in congressional committees, and some of these meetings study a set of bills that have a common topic. Second, focusing on the information at the bill level does not provide straightforward measures for the frequency variable. Third, as lobbyists can choose the specific bills they lobby for, the inclusion of a measure at the bill level can bias the estimates due to double causality. As a consequence, I aggregate the information at the committee-semester

[43] Kang (2016) is one of the few papers using both SOPR data and information of the bills. She focuses on only a small subset of bills and ignores the richness of the information contained in the Congressional Committees. However, she advances the understanding on the lobbying industry by structurally estimating the returns to lobby when the unit of observation is the *policy* rather than the bill.

[44] Most of the bills go to only one committee. There are 16 standing committees in the Senate and 20 in the House of Representatives.

[45] The Senate established this figure since 1937.

[46] For more information, see Sachs (1999), Sullivan (2007) and Heitshusen (2015).

level. This ensures that all the bills will receive knowledge-intensity and frequency measures, and, more importantly, I avoid obvious double causality problems in my estimations.

In order to capture a comprehensive measure of knowledge requirements, I create a difficulty index at the committee-semester level using principal components analysis. The index uses as an input the number of sub-committees and the knowledge intensity of these committees measured by the fraction of witnesses with high-knowledge occupations over the total number of witnesses. A committee with many sub-committees will tend to deal with issues that need more specialized study than will committees with few sub-committees. A committee with a larger fraction of high-knowledge witnesses will tend to deal with greater knowledge requirements issues. Proxying knowledge requirements using the difficulty index, or simply the fraction of high-knowledge witnesses, gives the same qualitative results.

A drawback of this methodology is that not all clients lobby for bills. On average, across semesters, 50% of the clients report advocacy activities for bills. In order to overcome this problem, I propose an additional measure of knowledge intensity using two variables based on Al-Ubaydli and McLaughlin's (2015) RegData 2.2. This database uses text analysis and machine learning algorithms to create regulation intensity measures at the industry-year level based on the Code of Federal Regulations (henceforth CFR). The CFR is an annual codification of rules made by executive and federal government agencies.[47] RegData classifies industries at the NAICS four-digit code levels for the period 1999 to 2014.[48] For each industry-year combination, I use two variables from this database: 1) the number of words related to the regulation of the industry;[49] and 2) the number of restriction strings related to the industry. The latter variable counts the number of times that any of the following five strings appears among the regulating words: *shall, must, may not, prohibited*, and *required*.

Firm-specific knowledge (i.e., h), which is the source of communication costs with external service providers, can also be easily interpreted as industry-specific knowledge. Given the lack of a comprehensive measure at the firm level, I use the total number of regulating words as a proxy for industry-specific knowledge. The intuition is that when clients hire external service providers, they explain the regulations in their industry. The larger the number of regulating words, the more costly it is to explain it to the external service providers and, therefore, the larger the

[47] These rules come from two main resources: congressional bills that become laws and regulations made by federal Agencies. The code is divided into 50 titles representing broad subject areas in federal regulations, such as Agriculture, Energy, Banking and Public Health.

[48] RegData also classifies industries at the NAICS three- and two-digit code levels. All the results presented in this paper are robust to the definition of an industry.

[49] Proxying regulation with these types of measures is not completely new. Coffey et al. (2012) and Mulligan and Shleifer (2005) proxy the extent of regulation of the whole economy with the number of pages and size of digital versions of regulations, respectively. The innovation of this database is to include regulation measures at the industry and not at the economy-wide level.

communication costs. In the database, external service providers have, on average, 5.9 clients in each semester belonging to 3.4 three-digit code industries. Anecdotal evidence shows that in a given period, lobbyists conduct lobbying activities for a given topic for different industries. Over time, they tend to lobby for different industries, but when they have clients from the same industry, they lobby for different activities.

I use the fraction of the number of restriction words over the total number of regulating words to proxy for the difficulty of the lobbying activities. Clients belonging to industries with a larger fraction of regulating words will face a tougher regulatory environment. In order to match the RegData information with the lobbying reports, I conducted extensive data work to detect the industry of the clients using ORBIS, COMPUSTAT and other web sources such as the client's webpage. I also include information at the bill and committee levels using web scraping techniques. I extract data from several web sources such as the Policy Agendas Project (hereafter PAP) and the Congressional Bills Project (henceforth CBP).

1.3.1.2 Validation of $f' < 0$ and proxy for communication cost

Figure 1.5 shows that more difficult activities tend to occur less frequently. The LHS part of the figure shows the frequency function of the committee-knowledge requirements' index, whereas the RHS part of the figure shows the frequency of the fraction of restriction words.[50] The bottom line from this figure is that one of the main assumptions of the model has empirical validity. The decreasing pattern of the figure holds across all the time periods of my database. An important implication of this graph is that there is an empirical monotonic relationship between frequency and difficulty. Therefore, it is enough to estimate the effect of frequency (difficulty) to know the effect of difficulty (frequency).

I conduct validation exercises for the proxy of communication costs using empirical proxies of industry specificity from the displaced workers' literature and the trade literature.[51] Both of these sources provide a non-comprehensive cross-section of empirical measures for industry specificity. To the best of my knowledge, there is no available panel dataset on industry specificity with which to conduct this validation exercise.

[50] Notice that in my model, each firm faces the same distribution of problems; then, by the law of large numbers for identically and independent distributed observations, this probability is reflected in the whole economy. Suppose that 100 clients are facing the following distribution function of problems: 2/3 have easy problems and 1/3 hard problems. If these problems are iid, after all the firms get their problems, 2/3 of the firms will have easy problems and 1/3 will have hard problems. Therefore, the density of the problems and the density of the types of clients or industries facing these problems is the same.

[51] For the displaced workers' literature, see, for instance, Jacobson et al. (1993), Carrington (1993), Neal (1995), Parent (2010) and Couch and Placzek (2010). For the trade literature, see Rauch (1999) and Nunn (2007). In Sub-section 1.5.2, I focus on the results from Jacobson et al. (1993) and Couch and Placzek (2010) for the first type of literature, whereas I use the data of Nunn (2007) for the trade literature.

The displaced workers' literature shows that workers who switch industries following displacement have significantly larger earnings losses than workers that remain in the same industry after displacement. I take this intuition one step further. If the earning losses from switching industries proxies for industry-specific knowledge, the level of the losses may proxy for the level of specificity. That is, industries in which workers suffer more from leaving the industry will be those with higher levels of industry-specific skills.

The trade literature provides an alternative data source. Nunn (2007) constructs a measure of relationship specificity at the industry level using information on whether the inputs are sold on an organized exchange or are reference priced in trade publications. An input is relationship-specific if the value of the input in a buyer-seller relationship is similar inside and outside the relationship. If the input is sold on an organized exchange, the market is thick (many buyers and sellers), and, as a consequence, the input is not relationship-specific. A similar intuition applies for the case in which the input price appears in trade publications. Therefore, an intuitive measure of industry specificity is the value of inputs that are neither bought and sold on an exchange nor reference priced. In sub-section 1.5.2 in the Appendix, I show that the total number of regulating words is significantly correlated with both the displacement and trade literature measures.

1.3.2 Main predictions

There are three key predictions from my model: Clients use in-house lobbyists to solve frequent or easy problems or when their industry-specific knowledge is greater.

1.3.2.1 Estimations

I run the following two sets of fixed effects estimations:

$$VI_{ijnt} = \beta_j + \gamma_t + \eta X + \varepsilon_{ijnt}$$

$$fr_{jnt} = \beta_j + \gamma_t + \eta X + \varepsilon_{jnt}$$

where i indicates a transaction, j a client, n an industry and t a time period, which can be either a semester or a year. The unit of observation in the first regression is a lobbying transaction, whereas it is a client-period in the second regression. At the transaction level, VI_{ijnt} is a dummy variable that takes the value 1 when the transaction is conducted internally and 0 otherwise. At the client level, I use fr_{jnt} to denote the fraction of internal lobbying transactions that the client has. I obtain similar results when I replace this variable with the fraction of in-house lobbyists. I control for both client (β_j) and time fixed effects (γ_t). I include client dummies to control for mean differences in the dependent variable across clients and time dummies γ_t to control for the dependent variable growth common to all clients. Intuitively, I justify the inclusion of client fixed-effects as there may be some client-level omitted characteristics, such as size, labour union status, geographical variation or relationship with politicians and the federal government, all of which can cofound the vertical integration decision.[52] Unfortunately, I do not have comprehensive measures of these variables for my dataset. I use time fixed effects, as there may be some time-varying changes in the dependent variable as a result of the Financial crisis, changes in market uncertainty, lobbying regulation among other reasons.[53] For most of these exercises, I cluster the standard errors at the client level. I also report results for alternative specifications in which the client fixed effects are replaced by industry fixed effects. The idea behind including industry-fixed controls is that there may be inherently important differences across industries in terms of regulation difficulty and knowledge requirements.[54] In this case, I cluster the standard errors at the industry level.

The key independent variable in these estimations, X, represents frequency, difficulty or industry-specific knowledge. For the frequency variable, I present here the results on the committee-based knowledge requirement measures. I obtain qualitatively similar results when I use the alternative frequency measure. For the difficulty measure, I present both sets of results: the one based on the knowledge requirements of the committees and the fraction of restraining words over the total number of regulating words. The variation in the committee based measures is at the transaction level, whereas it is at the industry-year level for the measures based on RegData. As a given transaction can have more than one bill, the first measure of the difficulty index (frequency variable) of the transaction is the weighted average over all the committees' knowledge requirements' indexes (frequencies).[55]

1.3.2.2 Some Descriptive Statistics

Table 1.3 shows some examples of the four-digit industries with the largest and smallest fraction of restraining words over the total number of words as well as the industries with the most and least total number of regulating words. The left-hand side of the table shows the industries organized by the level of knowledge requirements, while the right-hand side shows industries organized according to the

[52]For examples of the relevance of size of the client, see Section 1.3.3.2 of this chapter; for union status, see Abraham and Taylor (1996); and on geographical variation, see Chinitz (1961) and Autor (2003).

[53]For the effect of the Financial Crisis on integration patterns, see Knudsen and Foss (2014) and Chapter 2 from this dissertation; for economy-wide related time varying patterns, see Abraham and Taylor (1996); and for the effect of policy changes, see Chapter 2 of this dissertation.

[54]See, for instance, Helper (1991) in the intrinsic cross-industry differences on the propensity to outsource.

[55]Let b_{cjt} be the total number of bills sent to the sub-committee c being lobbied for client j at period t. Let f_{ct} be the frequency (number of bills) sent to committee c at period t. Then, the measure for the client at period t is $\sum_c f_{ct} \left(\frac{b_{cjt}}{\sum_c b_{cjt}} \right)$. Similar calculations apply for the case of the difficulty index.

total level of regulating words. Both columns use 2014 data. Similar rankings are obtained for different time periods.

Table 1.4 shows the mean, standard deviation and total number of observations for the main variables in my analysis: VI, fr, frequency, two difficulty measures and industry-specific knowledge. The unit of observation used to construct this table is the client-semester for all the variables except VI, which is constructed at the transaction level. The number of observations for the frequency and the first measure of difficulty is smaller than for the other variables, as only a fraction of clients lobby for bills. For the sake of interpretation, I normalize all the independent variables by their standard deviation in the following econometric exercises.

1.3.2.3 Frequency

Table 1.5 shows the results when I use the frequency variable proxied by the average number of introduced bills across all the lobbied bills. This table is organized as follows: The last two columns control for client-fixed effects, and the first column controls for industry fixed effects. The first and third columns control for semester fixed-effects and the even columns for industry·year fixed-effects. The last set of fixed-effects are intended to detect time variation within an industry. That is, the level of integration and regulation can evolve differently across industries over time (see, for instance, the BP case in Section 1.3.4.2). The results are divided into two sections. The top panel shows the results at the transaction level and the bottom panel shows the estimations at the client level. In this table, all of the coefficients of interest are statistically significant, and their signs are consistent with the theoretical framework. The results imply that an increase of one standard deviation in the number of introduced bills is associated with an increase in the probability of vertical integration by an amount between 1.1 and 1.6 percentage points. Similarly, an additional standard deviation in the number of introduced bills is associated with an increase in the fraction of in-house reports between 1.4% and 1.9%. I consider these effects economically relevant. For instance, in the latter case, as the mean of the dependent variable is 18%, a one standard deviation increase in the frequency variable increases the fraction of reports made internally between 7% and 11%.

1.3.2.4 Difficulty

Table 1.6 shows the results for the first difficulty measure based on the knowledge requirements of the committees dealing with the bills for which the clients lobby. Table 1.7 shows the results when I proxy the level of difficulty by the ratio of restraining words to the total number of words regulating an industry. As my model predicts, the greater the difficulty, the lower is the level of vertical integration. In these tables, all of the coefficients are significant at least at the 10% level. The results imply that an increase of one standard deviation in the fraction of restraining words is associated with a decrease in the probabillity of conducting advocacy activities with in-house staff by an amount between 2.2 and 2.7 (1.4 and 2.1) percentage points. Similarly, an additional standard deviation in the difficulty index (fraction of restraining words) is associated with a decrease between 3.3% and 4% (1.5% and 3.5%) in the fraction of in-house reports. When the dependent variable is at the transaction level, this corresponds to a decrease in the probability of about 16 to 20 (10 to 15) percentage points. When the dependent variable is the fraction of in-house reports, this ratio decreases by about 18% to 22% (8% to 19%).

1.3.2.5 Industry-Specific Knowledge

Table 1.8 shows the results when I proxy the level of industry knowledge with the number of regulating words of the industry. The table shows that all the coefficients are statistically significant and positive, as the theoretical section predicts.[56] In particular, a one standard deviation increase in the total number of words regulating an industry is associated with 0.4 or 0.8 percentage points in the probability of vertical integration and 0.3% or 0.8% in the fraction of in-house reports. These effects are smaller than those calculated for the cases of frequency and difficulty. They represent a decrease in the probability of integration of three to six percentage points and 2% to 4% on the fraction of internal reports.

1.3.3 Other Predictions

In this section, I focus on three empirical patterns accounted for by my theoretical section. First, I show that there are both generalists and specialists in internal and external markets. However, generalist lobbyists tend to be in the internal market and external service providers tend to be specialists. Second, I present an alternative way to explain differences in the vertical integration patterns by the size of the clients. I show that the density function of the problems differs by clients' size. Then, I present some empirical evidence on the matching patterns between clients' size and level of specialization of the in-house specialists. Third, I show that the more knowledgeable external service providers earn more and work with more clients.

1.3.3.1 Generalists and Specialists

In order to analyze the first prediction, I focus on lobbyist-level data. The SOPR data allow me to separate the lobbyists into two subgroups based on whether or not they are in-house lobbyists for the full sample period.[57] The categories I use in this section are as follows: 1) internal lobbyists (56.1%), who have **always** lobbied as

[56] This result is consistent with that of de Figueiredo and Kim (2004). Using 150 contacts with the Federal Communications Commission, the authors show that firms use their own employees to lobby for issues with a high degree of firm-specific information. The authors acknowledge that one possible mechanism to explain this pattern is with communication costs.

[57] In this section, I focus on the lobbyists for whom I was able to get sociodemographic information based on *lobbyist.info*. Further information on this database will be provided below.

in-house lobbyists; and 2) external lobbyists (43.9%), who are intermediaries that have **never** lobbied as in-house lobbyists.[58] Using lobbyists' issue assignments, I construct a Herfindahl concentration index (hereafter HHI) for lobbyists and, following Bertrand et al. (2014), categorize them into two possible corner solutions: generalists and specialists. A lobbyist is a generalist if more than 25% of her assignments are never on the same issue, whereas she is a specialist if at least 25% of her assignments are on the same issue.[59]

Table 1.9 shows the main descriptive statistics of this exercise.[60] For the LHS section, the last row shows the average HHI across both internal and external lobbyists, respectively. Lobbyists working on a larger number of issues have a lower value in the index. The table shows that external lobbyists have a larger average HHI than in-house lobbyists. The first two rows show the distribution of generalists and specialists across markets. Consider the row of generalists: among 15,760 lobbyists classified as generalists, 78% are in-house lobbyists, while 22% are external lobbyists. Table 1.9 also shows that almost 60% of the specialists are external lobbyists. The right-hand side section shows the fraction of in-house and external lobbyists that are generalists and specialists. For instance, among the sample of external lobbyists, 79.1% are specialists and 20.9% are generalists.[61] This table shows two broad patterns: generalists tend to work as in-house lobbyists, whereas external lobbyists tend to be specialists. These patterns are robust to different ways to define generalists and specialists. Figure 1.6, which shows the HHI values for both internal and external lobbyists, reveals two main patterns. First, among the lobbyists with HHI larger than 0.5 (or any other cutoff above 0.25), there is a larger percentage of external than internal lobbyists. Second, the fraction of internal lobbyists with HHI larger than 0.5 (or any other cutoff above 0.25) is lower than

[58] An additional possible categorization across lobbyists is the mixed-type lobbyist, who is an intermediary that has worked in both markets. There are 25,001, 19,526 and 3,057 internal, external and mixed-type lobbyists, respectively. For the sake of brevity and concreteness, I focus in this section on only the first two types of lobbyists. However, the inclusion of the third type of lobbyist does not change the main patterns presented in this section. Although external lobbyists can work for several clients, some of these lobbyists work for a single client. That is, they are *de facto* internal lobbyists. Although I recognize this is an interesting topic for further research, I have neglected this issue in the theoretical application, as it does not seem to be empirically relevant. For instance, if I restrict the sample to lobbyists working at least ten years, less than 2% of the external lobbyists work for only one client.

[59] Notice that the definitions of specialists and generalists are not exhaustive. There may be lobbyists who do not match any of these classifications. For instance, a lobbyist working on four issues in four periods with the following allocation: first period: 20, 40, 20, 20; second period: 40, 20, 20, 20; third period: 20, 20, 40, 20; and fourth period: 20, 20, 20,40. This lobbyist is not a specialist because she did not work on a given issue at least 25% of the time in each period. On the other hand, she is not a generalist because she spent 40% of her time on at least one issue in one period.

[60] The raw numbers of lobbyists in the internal market are: 12,293 generalists, 9,115 specialists and 3,594 that do not match any definition. For the external market, there are 3,467 generalists, 13,335 specialists and 2,724 that do not match any definition.

[61] For the calculations of the right-hand side of the table, I consider only the lobbyists that were classified as specialists or generalists. That is, each column adds up to 100.

the fraction of external lobbyists with HHI larger than 0.5. A broad pattern that emerges from this figure is that internal lobbyists tend to be more concentrated among lower values of HHI, whereas external lobbyists have both more dispersion in the values and a larger fraction of lobbyists with large HHI values. Taking this evidence together, this suggests that internal lobbyists tend to be generalists, while external ones tend to be specialists.[62]

A possible concern from these results is that a lobbyist can erroneously be classified as a specialist if she has worked only few times. Table 1.10 shows that the patterns highlighted in this section are robust once I control for the tenure of the lobbyists.

1.3.3.2 Size of the Clients

In this section, I investigate the relationship between the size of the client and vertical integration status. Although previous studies in the political economy literature find that large firms tend to lobby more than small firms, I am not aware of any empirical study relating size and vertical integration status in the lobbying context.[63] Grossman, Helpman, and Szeidl (2006) theoretically predict that the most productive or larger firms outsource. Girma and Gorg (2004) and Jabbour (2013) find empirical evidence for this prediction. On the contrary, Antras and Helpman (2004) assume that fixed costs under vertical integration are higher than in the case of outsourcing, and, therefore, most productive and larger firms vertically integrate. Supporting this prediction, Abraham and Taylor (1996) and Hortacsu and Syverson (2007) provide empirical evidence that smaller firms tend to contract out more.

These papers focus on the fixed costs either of integration or outsourcing to explain the relationship between the client's size and integration. Although I acknowledge this mechanism, I propose an alternative channel: the variability in the integration decision among small and large clients can be due to the differences in the relative frequency of knowledge-intensity levels of the problems they face. In this section, I first show that large clients tend to be more vertically integrated, and then I provide statistical evidence on differences in the distribution of the type of problems they face.

Large Clients tend to be more vertically integrated In order to analyze this question, I focus on client-level data. Using ORBIS and COMPUSTAT databases, I

[62] As anecdotal evidence supporting the intuition behind the frequency of transactions and the patterns on generalists and specialists across markets, Drutman (2010, p 49) cites one in-house lobbyist's personal interview about generalists and specialists: "*Most of us in the office are generalists. On new issues like energy, we didn't know the concerns, so we need specialized talent. We're a lean shop here, and we're not going to hire an energy expert, so we go to the consultant who can offer a percentage of time for issue expertise. One of our consultants knows the Energy and Commerce Committee very well, so we hire them to explain what the issues are.*"

[63] Examples of papers studying clients' size and lobbying status are Ansolabehere et al. (2002), de Figueiredo and Silverman (2006), Richter et al. (2009), Bombardini and Trebbi (2012) and Kerr et al. (2014).

obtain business activity information such as sales and employees for 15,939 clients. The SOPR data allow me to separate the clients into two subgroups based on whether or not they exclusively used in-house lobbyists for the full sample period. The categories I use in this section are as follows: 1) internal clients (8.3%) that have **always** lobbied with in-house lobbyists; and 2) external clients (91.7%) that have **never** lobbied with in-house lobbyists.[64] I use two variables to proxy for the size of the client: sales and employees. Table 1.11 shows the main descriptive statistics for these two variables discriminated by internal and external clients. From this table, it is clear that larger clients tend to vertically integrate, whereas smaller clients tend to outsource. Figure 1.7 shows the firm size distribution by type of client.[65] I proxy size with both the logarithm of sales (left-hand side) and logarithm of the number of employees. For the sake of space, I do not present other results confirming these patterns, such as the previous fixed effect estimations controlling by size, mean differences *t-test* and Kolmogorov-Smirnov equality of distribution tests. The bottom line is that internal clients are larger than external clients. This pattern is robust to the way that I proxy size, as other variables such as gross revenue and total assets provide the same results.

Density of Problems differ by Clients' size. Assume that there are two types of clients indexed by $l = L, S$ facing problems with the same knowledge-acquiring cost technology. The internal market surplus for client type l is $F_l(z) - cz$ with firm boundary given by $z_l^* = f_l^{-1}(c)$. As the total mass of integrated clients type l is $\int_0^{z_l^*} f_l(x)\,dx$, there are more integrated clients type L if $F_L(z_L^*) > F_S(z_S^*)$.

Lemma 1.12 *If* $z_l \sim Exp(\lambda_l)$ *for* $l = L, S$, $\lambda_L > \lambda_S \Leftrightarrow F_L(z_L^*) > F_S(z_S^*)$.

This remark implies that if the density of problems for two sets of clients is exponential, the set of clients with a larger parameter rate λ will have a higher level of integration in the market.

I apply this result to the data. I define a large (small) firm in two different ways. In the first case, a firm is large (small) if its sales are above (below) the median value of the sales for the whole sample of companies. In the second case, a firm is large (small) if its sales are above (below) the 75th (25th) percentile of the sample of firms. Similar definitions apply when I proxy size with the number of employees. For each set of firms, large and small, I calculate the density of problems based on congressional committees' knowledge measures introduced in Section 1.3.1.1. To

[64] There are only 355 clients that used both internal and external lobbyists for this period. The mean value of their sales is $2.026 million and the average number of employees is 362.5. These percentages are very similar when I take the full sample of clients, so at least from an aggregate point of view, the matching process of ORBIS and COMPUSTAT with the business information does not change the weights that each category (internal vs external) of client has.

[65] Both figures use Kernel Epanechnikov estimation methods. The bandwidth used for the LHS figure is 0.1591 whereas it is 0.1104 for the RHS figure. The number of employees is normalized to thousands of units with a maximum value equal to 2.2 million employees. The value of sales is normalized to millions of units with a maximum value of 420,016 million.

estimate these densities, I use all the bills that they lobbied for. For each density, I estimate the parameter rate λ by Non-Linear Squares. The main result is that all of these estimations give the same results: λ_L is statistically and significantly larger than λ_S.

So the bottom line is that large and small clients have different densities of problems. Large clients have a larger mass among problems with low-knowledge intensity. Therefore, among the internal clients, a larger fraction consists of large clients. Given this result and Corollary 1.1, I expect to find that larger vertically integrated clients will have a larger fraction of internal specialists than smaller internal clients will.[66] Figure 1.8 is a scatter plot between the HHI of lobbyists and size of the clients and provides empirical evidence on the previous intuition. Internal specialists tend to be with the largest clients and the generalists working internally tend to be hired by smaller internal clients.

1.3.3.3 Earnings, leverage and knowledge

Lemma 1.9 predicts that external service providers with more knowledge have both greater leverage and a higher level of earnings. In this section, I provide empirical evidence in support of these predictions. In order to conduct this exercise, I use information from the *lobbyist.info* database, which contains information on lobbyists' sociodemographic characteristics, such as work experience in federal governmental institutions. A lobbyist is a *Revolving Door Lobbyist* if she worked in federal agencies, the White House or Congress before becoming a lobbyist. Table 1.12 presents descriptive statistics of lobbyists with federal government experience discriminated by in-house and external lobbyists.

I proxy the level of earnings with the average earnings per semester over all the reports for the last year that the lobbyist appears in the database.[67] I proxy the level of knowledge in two different ways: years of experience as a federal lobbyist and whether or not the federal lobbyist is a revolving-door advocate. The former variable takes values between one and 16 years. The leverage is measured by the average number of clients that the external lobbyist had across all the reports in the last year of lobbying activity. To facilitate the interpretation of the results, I consider the quantiles (and not the actual number) of leverage.[68]

Figure 1.9 shows the main relationships for these variables. The left-hand side shows the relationship between earnings and experience, while the right-hand side shows the relationship between experience and leverage. In this figure, dashed lines

[66] Notice that, from Corollary 1, there are two regions where internal specialists solve problems: the most and the least frequent problems. The best way to relate this intuition to the following empirical pattern is to think that it is more common for internal specialists to solve very frequent problems rather than the least frequent problems.

[67] I deflate the earnings by the CPI with constant prices in 2009.

[68] The upper and lower bounds of the quantiles are as follow: first, one client; second, between one and three clients; third, between three and five clients; fourth, between five and eleven clients; and, finally, the last quintile is for more than eleven clients.

denote *revolving-door lobbyists*. The LHS section shows that lobbyists with more years of experience as federal lobbyists earn more. Furthermore, for the same number of years of experience as federal lobbyists, *revolving-door lobbyists* earn more than *non-revolving-door* lobbyists. On average, across years of experience, lobbyists with previous federal government experience earn $21,300 more than the other lobbyists. The figure also shows some divergence: as the lobbyists accumulate more experience, the returns to being a *revolving-door* lobbyist increase. An alternative way to read the figure is to ask how many more years of experience a *non-revolving-door* lobbyist needs in order to get the same earnings as a *revolving-door* lobbyist. A *revolving-door* lobbyist with nine years of federal lobbying experience earns $85,396, and a *non-revolving-door* lobbyist with 11 years of experience earns $85,642. That means that, on average, the *revolving-door* experience is worth two years of federal lobbying experience. However, these returns change with the number of years of lobbying experience. The RHS section shows that lobbyists with more years of federal lobbying experience work for more clients. On average, *revolving-door* lobbyists require less federal lobbying experience than *non-revolving-door* lobbyists to work for the same number of clients. For instance, a *revolving-door* lobbyist is in the third quintile when she has 5.5 years of experience, whereas a *non-revolving-door* lobbyist requires 8.28 years of experience to be in the same quintile.

Summing up, Figure 1.9 provides evidence of the positive relationship between the level of earnings and experience, and between experience and leverage. As experience proxies for knowledge, I take these results as supporting empirical evidence of Lemma 1.9.

1.3.4 Causal Inference: The BP Oil Spill

My theory predicts that when lobbying activities are more difficult, clients use external service providers. I use the 2010 British Petroleum (BP) Deepwater Horizon oil Spill as a quasi-experiment to explore the validity of this prediction.

1.3.4.1 Interpretation of the event

On April 20, 2010, high-pressure methane gas from the Deepwater Horizon oil well rose into the drilling rig, where it ignited and exploded. This explosion led to the burning and sinking of the Deepwater drilling rig. It was followed by a massive offshore oil spill in the Gulf of Mexico that was considered the largest environmental disaster in US history.[69]

By providing anecdotal evidence, I argue in this section that the spill increased the difficulty of lobbying activities not only for BP, but also for other companies in the oil and gas extracting industry. I summarize this evidence with three set of

[69] For more information see, http://www.theguardian.com/environment/2012/nov/28/epa-suspends-bp-oil-spill. Scientists estimated the total amount of released oil was close to 4.9 million barrels (http://www.uscg.mil/foia/docs/dwh/fosc_dwh_report.pdf).

examples: the reactions of Congress, the federal Government and (potential) voters. In the next subsection, I provide empirical evidence that the skill-requirements of lobbying activities increased disproportionately for the extracting industry.

Congress. The oil industry has a strong and long-standing relationship with Republican policy makers.[70] The pressure for more regulation of the oil and gas extracting industry was so strong after the event that even Republican congressmen were proposing bills that would negatively affect the oil industry. For example, in May 2010, Roy Blunt (R-MO), who was among the top three money recipients from the industry for the period 2009-2010, introduced the bill H.R. 5356 with the purpose of increasing the cap on liability for economic damages resulting from an oil spill.[71] The increase in the cap had a clear expected cost to firms at risk of having oil spills, as they would bear a greater responsibility for damages. Furthermore, several bills were introduced after the event with the aim of making the oil and gas extraction business more difficult.[72] Some of these bills did not have any precedent.

Federal Government. In June 2010, President Obama created a nonpartisan national commission to provide a deeper understanding of the BP oil spill. The commission investigated the causes of the spill and released a final report in January 2011, concluding that not only could the well blowout have been prevented, but that it was an example of the failure of risk-prevention practices within the entire extracting industry.[73] The commission agreed that deepwater exploration has intrinsic risks, but that it is the responsibility of both the industry and the regulatory agencies to restructure the way that business is being done to improve safety throughout the industry. Thus, the report blamed both oil extracting firms and regulators- the former for being irresponsible in its safety practices, as confirmed by the evidence on 79 well-control accidents between 1996 and 2001, many due to negligence by oil and gas extracting firms;[74] the latter for lagging behind on regulating the real risks associated with deepwater drilling. Therefore, this report had two main effects: it intensified the stigma associated with oil drilling and increased the pressure on the government to improve the regulatory oversight of the industry. In addition, in May 2010, President Obama ordered a delay on the issuing of new offshore drilling leases until it was clear whether tougher regulation was needed,

[70] For instance, among the political campaign contributions that have come from companies in the industry, 76% have been given to Republican candidates. The total campaign contributions for the period 1990 to 2016 were US$182,188,234.

[71] http://maplight.org/content/oil-spill-response-bills.

[72] For instance, the bill H.R. 5436 proposed to prohibit issuing permits for any deepwater drilling in the Gulf of Mexico; the bill H.R. 5222 proposed to suspend exploration and production activities in the outer continental shelf until the investigation of the BP oil spill concluded. In the Senate, bills S.3763 and S.3643 both proposed to implement new technology and improve safety surrounding offshore energy production. Bills S.338 of 2011 and S.598 of 2013 proposed to prohibit royalty incentives for deepwater drilling.

[73] https://www.gpo.gov/fdsys/pkg/GPO-OILCOMMISSION/pdf/GPO-OILCOMMISSION.pdf

[74] Ibid.

and Secretary of the Interior Ken Salazar suspended deepwater offshore drilling on the Outer Continental Shelf for a period of six months.[75]

Voters. Policy makers care about what voters want, and voters in the US reacted strongly to the BP oil spill. Barrage, Chyn and Hastings (2014) show that stations selling BP-branded combustibles manifested an important effect on both prices and volume of sales. They interpret this finding as a shift in demand away from BP as a way to punish the company's bad practices. Second, there was an active response from the people to boycott the oil extracting industry and especially BP. In the months following the spill, there were dozens of protests. Interviews with protesters show that the aim of the boycott was not specifically aimed to affect BP's reputation, but also to show public discontent with the operation of the industry without adequate safeguards.[76] Facebook groups such as *Boycott BP* and online petitions such as the ones produced by a consumer advocacy group *Public citizen* asked policy makers for tougher regulation of the extracting industry. Greenpeace's spokesman Phil Radford asked publicly to ban all offshore oil drilling forever.[77]

To sum up, the federal Government and voters publicly blamed regulators and firms in the oil extracting industry. Policy makers reacted by proposing legislation to increase regulation of the industry. Consequently, the firms in the industry faced a tougher environment. In the next subsection, I empirically show that these firms reacted by using external lobbyists more heavily.

1.3.4.2 Empirical Strategy and Results

I argued in the previous subsection that different channels increased the difficulty of the lobbying activities of the oil and gas extracting industry (and BP in particular) after the BP oil spill. I interpret this shock in two different ways: first, lobbying activities similar to those that the oil companies were conducting before the oil spill became more difficult; and second, the oil spill response from different regulatory agencies was to propose bills that had never or very rarely been proposed before. Therefore, my model predicts that either the low frequency or the difficulty of the lobbying activities implies more outsourcing.

There are two natural candidates for the treated observations: BP and other firms conducting activities similar to BP's. The primary activities of BP are categorized in 2007 NAICS codes 211111 (Crude Petroleum and Natural Gas Extraction). BP agreed to pay £18.7 billion for caused damages and, as a consequence, the company sold off $38 billion in assets from 2010 to 2012.[78] This change in the structure

[75] For more information see Johnston, Nicholas and Nichols, Hans (1 May 2010)." Obama Says New Oil Leases Must Have More Safeguards." Bloomberg. and https://www.doi.gov/news/pressreleases/Interior-Issues-Directive-to-Guide-Safe-Six-Month-Moratorium-on-Deepwater-Drilling.

[76] Wheaton, Sarah (2 June 2010). "Protesters gather at BP stations." *The New York Times.*

[77] Phil Radford (24 May 2010). "[BP]resident Obama: Where Does BP Begin and Obama End?". The Huffington Post.

[78] For more information see, NY Times Article.

of the firm can confound its vertical integration decision. As a consequence, I use as a treated group all the firms conducting lobbying activities that belong to 2007 NAICS codes 211111, *excluding* BP. As a control group, I include all the firms that belong to the Oil and Gas industry, excluding the code above. Examples of this control group are codes 3251 (Basic Chemical Manufacturing), 324191 (Petroleum Lubricating Oil and Grease Manufacturing) and 324199 (All Other Petroleum and Coal Products Manufacturing). These exclusions leave me with 52 clients in the treated group and 218 in the control group. Examples of the treated group are Chevron, Exxon Mobil, Phillips 66, Shell and Devon. For these estimations, I include neither code 213111, Drilling Oil and Gas Wells, nor code 213112, Support Activities for Oil and Gas Operations, as they are strongly linked to the affected industry. For this exercise, I focus on quarterly reports, which are mandatory since 2008. The period of analysis runs from the first quarter of 2008 to the second quarter of 2014.

Figure 1.10 shows the time-series patterns of reports made internally for these two groups. The graph shows two main patterns. First, the treated and control groups have similar increasing trends before the oil spill. This is confirmed by statistical exercises such as leads and lags estimations and *t-test* results. Figure 1.11 presents the coefficient estimates of the leads and lags exercise. For the control group, there does not seem to be a change in the increasing trend around the second quarter of 2010. This increasing pattern is aligned with facts presented in Chapter 2 from this dissertation, in which the fraction of internal reports began to increase after 2007 for all of the industries. Second, there is a significant drop in the fraction of internal reports for the treated group in the same quarter of the BP oil spill. The fraction evolves with a flat trend after that, and it does not show any tendency of going back to the levels prior to the spill.

The knowledge-requirement variables used in the previous section support the story in which the oil and gas extracting industry companies, and not the proposed control group, received a shock to the difficulty of lobbying activities. Figure 1.12 provides empirical validation of this. The LHS section shows the time-series patterns of the fraction of high-knowledge witnesses (over the total of witnesses) for the treated and control groups. The RHS section shows the yearly series on the ratio of restraining words to the total number of regulating words for these two groups. Both parts of Figure 1.12 show the same two broad patterns. First, as an additional empirical confirmation of the validity of the control group, the time series for this group do not provide any suggestion of an empirical shift around 2010. Second, the difficulty of the lobbying activities for the treated group increased around the BP oil spill. The fraction of high-knowledge witnesses being interviewed in the committees dealing with bills lobbied by oil and gas extracting companies started increasing in the first semester of 2010. This series reaches its peak in the first semester of 2012 and then starts decreasing again. Qualitatively similar results emerge when I consider the difficulty index instead of the fraction of high-knowledge witnesses. When I measure difficulty with the ratio of restraining words, it is clear that since 1999, the ratio of restraining words had been decreasing, but in 2010, the fraction started increasing.

I now estimate the effect of the BP oil spill on the fraction of internal reports. Clearly, one can use the spill as an instrumental variable to the knowledge measures and then estimate the effect of an exogenous change in these measures on vertical integration measures. Results not shown here provide empirical validity of the effect on the BP oil spill to the integration measures through knowledge-intensity proxies. In this section, I present results on the direct effect of the spill on the fraction of integration patterns because I am more interested in testing the validity of my model than in measuring the effect of knowledge measures on integration patterns. In addition, the knowledge measures do not provide enough quarter-level observations, implying that conducting an IV-Dif-Dif estimation would need to aggregate the data at such a level that the oil spill timing would be contaminated, and the richness of the integration patterns at the quarter level would be lost.

I regress vertical integration measures on client and quarter fixed effects, and an interaction of the post-spill period with an indicator for whether the client belongs to the same sub-industry as BP:

$$v_{it} = \gamma_i + \theta_t + \delta \left(T_i \cdot P_t \right) + \varepsilon_{it}$$

Here, v_{it} is a measure of vertical integration made by the i-th client at period t. I consider two measures of vertical integration: fraction of in-house reports and fraction of lobbyists working in the internal market. γ_i (θ_t) represent client-level (quarter-level) fixed effects. I focus on two groups: treated and non-treated. The variable T_i takes the value of 1 if client i belongs to the same sub-industry as BP (NAICS 211111) and 0 otherwise. The variable P_t takes a value of 1 for all the quarters since the second quarter of 2010 until the end of the sample and 0 for periods before the oil spill. Table 1.13 shows the main results from these exercises. To control for possible autocorrelation at the client-level, I estimate all the regressions, clustering the standard errors at the client level. The results show that clients belonging to the same sub-industry as BP decreased the use of in-house lobbyists, compared to a similar group of firms that was not affected by the oil spill. This is shown by the negative and significant coefficient measuring the interaction term. The decrease in the fraction of internal reports is substantial. Given that the treated group made, on average, 19% of their reports internally before the spill, the point estimate represents a 26.3% decline in the fraction of reports made internally. Regarding the fraction of internal lobbyists, the average for the treated group before the spill was 24.7%. This represents a decrease of about 21% in the fraction of in-house lobbyists for the clients in the oil and gas extracting industry.

1.3.4.3 Robustness Checks

The oil spill may have increased the number of problems faced by the treated group. Given the in-house staff's time constraint, an increase in the number of problems will mechanically increase the need for external service providers. Thus, the use of external service providers will be explained by the time constraint of in-house staff and not by a change in difficulty. A way to test the validity of this story is to see whether there were changes in the absolute number of in-house lobbyists. A decrease in the use of in-house lobbyists would be inconsistent with the time constraint explanation. Table 1.14 provides evidence that clients in the treated group decrease the use of internal staff and increase the use of external lobbyists. An additional way to control for the change in the number of problems that the clients faced is to run the same econometric specifications as the baseline estimation but including either the total number of reports or the total expenditures of the client as an additional control variable. Table 1.15 presents coefficient estimates once one includes the total number of reports as an additional explanatory variable. The main results presented in the baseline estimations still hold.

Given that the decrease in the fraction of in-house reports is due to a decrease in the in-house staff and an increase in the use of external staff, there is a question one can ask about this pattern. Is the decrease in the use of in-house lobbyists due to a demand or supply shift? For instance, a demand side example is that given that the problems are harder, the affected firms decided not to use in-house lobbyists, as these advocates cannot handle difficult problems. A possible supply side story is that there was a stigma associated with working as an in-house lobbyist for the affected firms and therefore in-house lobbyists quit. These two channels, supply, and demand can explain the decrease in the use of in-house lobbyists but, importantly, they have different effects on the equilibrium prices. If the demand effect dominates, the decrease in the quantity is accompanied by a corresponding decrease in the equilibrium payment, while the shift in the supply will increase the equilibrium payment. In results not presented here, I show that the equilibrium payment in the internal market decreased, supporting the idea that the change is explained by demand and not supply channels.

A second concern is that there may have been negative spillovers to other firms in the oil and gas industry. Although Figures 1.10 and 1.12 do not seem to show any change for the control group around the event, in Table 1.16, I present additional econometric estimations using different control groups and I show that the main results presented in this section hold.[79] A third concern is that the variation in the

[79] The control group industries I use are Retail Sales, Real Estate and Casino and Gambling. These groups were selected according to the pre-oil spill similarity of the trends of the fraction of

interaction variable is not at the client but the industry level. To control for this type of autocorrelation, I show that the main results are robust when I two-way cluster the standard errors (*industry·quarter* and *client*) or, as in Barrage et al. (2014), I aggregate the data at the client-period level (before and after). Table 1.17 and 1.18, show that the results of these alternative two exercises are qualitatively similar to the baseline estimation.

A final concern is that the empirical patterns may be consistent with a change in the stakes at play. For instance, clients may respond by outsourcing the service when they face bills that can affect them more heavily -*as the most knowledgeable lobbyists are in the external market-*. In results not presented here, I provide anecdotal evidence that this is not the case. I show some examples in which bills that could enormously affect the clients were studied in congressional hearings with a small (or null) fraction of high-knowledge witnesses and, consistent with my theory were lobbied using in-house lobbyists. I also provide evidence of bills in which the stakes were low and were studied with a large fraction of high-knowledge witnesses and lobbied with external lobbyists.

1.4 Final Discussion

This section is divided into two parts. In Section 1.4.1, I give preliminary evidence on the external validity of my results for other PBS industries. In Section 1.4.2, I briefly summarize this chapter and propose ways to extend it.

1.4.1 Connection with other PBS Industries

One of the main weaknesses of using lobbying data to make inferences about the behavior of PBS industries is that the advocacy industry may be very different from all the others. Although I acknowledge that each industry has its own specificities, I argue in this section that there are some broad similar patterns between the lobbying industry and other PBS industries. As a consequence, some of the main results from the lobbying data can be applied to such industries.

In order to do that, I use data from the Occupational Employment Statistics (henceforth OES) program.[80]This program from the Bureau of Labor Statistics produces employment and wage estimates annually for over 800 occupations in the United States, sampling over 200,000 non-farm business establishments every semester.[81] In this section, I focus on national occupational estimates for specific industries for the period 2002 to 2014.[82]

internal reports.

[80]Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics, [date accessed: 15 November 2015] www.bls.gov/oes/.

[81]More detailed information can be found at http://www.bls.gov/oes/oes_ques.htm.

[82]In 2002, the OES survey switched from the SIC industry classification system to the NAICS system. As a result, there have been changes in industry definitions. As the web page says: "*For example, under SIC the industry "grocery stores" included their retail establishments, warehouses,*

1.4.1.1 Data

An occupation is defined according to the Office of Management and Budget (henceforth OMB) Standard Occupational Classification (henceforth SOC) system while an industry is defined with the four-digit NAICS classification. I focus on five different occupations: lawyers, managers, IT personnel, accountants and lobbyists. Table 1.19 summarizes the input used for this exercise.[83] As Abraham (1988) and Dube and Kaplan (2010), I define a service provider as external if the service provider works in the NAICS industry primarily concerned with that occupation. For instance, NAICS 5415 corresponds to firms that specialize in computer services. Therefore, the fraction of IT workers (i.e. code 15-10) on the NAICS industry 5415 will be the fraction of IT personnel working in the external market. On the other hand, the rest of the computer specialists (working in all the other industries) will be the fraction of in-house IT personnel. It is important to note that the vertical integration share of lobbyists may not exactly match the empirical patterns using the lobbying reports database, as the only available way to identify lobbyists using the SOC is to use four-digits NAICS codes (5418), which, unfortunately, includes other occupations, such as advertising services.

1.4.1.2 Patterns

I focus here on three empirical patterns: time series patterns of vertical integration; wage differences between internal and external service providers; and the fraction of generalists and specialists working in-house and externally.

Time-Series Patterns Figure 1.13 shows the time series of vertical integration and total employment patterns by occupation. The solid line shows the fraction of employees working in-house. The dashed line shows the total number of employees

transportation facilities, and administrative headquarters. Under NAICS, the four establishment types would be reported in separate industries. Only the retail establishments would be included in the NAICS industry for "grocery stores." " (Bureau of Labor Statistics, U.S. Department of Labor, Occupational Employment Statistics.) The year 2002 also coincides with other two relevant changes in the survey: reference months for the OES survey and mean wage estimation methodology. The reference months for the OES survey were changed from October, November, and December to May and November in order to reduce seasonal influences. Second, the method of calculating mean wages was changed for occupations with any workers earning above $70 per hour in order to remove a downward bias in mean wage estimates. In order to avoid problems of mis-classification and the time-series comparisons, I focus my analysis on the period 2002 to 2014. There was an additional change in the definitions of occupations in 2010. I take into account that the occupations' definitions before and after this change were not affected.

[83] The titles for the occupational codes are: **Lawyers:** 23: Legal Occupations and I exclude 23-1023: Judges, Magistrate Judges and Magistrates. **Managers:** 11: Management Occupations. **IT Personnel:** 15-10 Computer Programmers and System Analysts before 2010 and 15-11: Computer Occupations after 2010. **Accountants:** 13-2011: Accountants and Auditors. **Lobbyists** 27-3031: Public Relations Specialists. The titles of the NAICS codes are as follows. For **lawyers** 5411: Legal Services. **Managers:** 5511: Management of Companies and Enterprises **IT Personnel:** 5415 : Computer Systems Design and Related Services. **Accountants** 5412: Accounting, Tax Preparation, Bookkeeping, and Payroll Services. **Lobbyists:** 541820: Public Relations Agencies.

in both the internal and external markets.[84] Three patterns emerge from this figure. First, there is a broad decreasing trend in the fraction of in-house employees and an increasing trend in the total number of employees across all five occupations. As total employment proxies for market size, this pattern can be easily accounted by Stigler's (1951) intuition. Note that for all the occupations except the lobbying industry, there is a decrease in the employment level around 2010, but then there is a recovery afterwards. Second, Figure 1.13 does not show any change in the integration levels around 2008 for all the occupations, excluding lobbyists. Chapter 2 provides a throughout discussion of the effect of the Financial Crisis on integration patterns. Third, around 2007, the lobbying market had two differences with respect to the other four occupations. On the one hand, in 2008, the total number of lobbyists started decreasing. This is in sharp contrast with other occupations as they display a broadly increasing trend for the studied period. On the other hand, there is an increase in the fraction of in-house lobbyists after 2008 that does not occur in the other occupations. Chapter 2 proposes an explanation on the lobbying time series patterns.

Wages One of the main predictions from my theoretical section is that external service providers acquire more knowledge and, as consequence, have higher earnings. In this section, I test whether external PBS providers have higher earnings levels. Table 1.20 shows the mean and five different percentile values for hourly wage by occupation across all years. For each occupation, I calculate separately the wage statistics for the internal and external market service providers.[85] In this table, the differences in wages across internal and external service providers are only *t-test* statistically significant for mean and percentiles values of 50, 75 and 90. The main message from this table is that the wages of external providers are higher than the wages of in-house employees. This is especially true for medium to top earners. I interpret this exercise as suggestive evidence that external providers acquire more knowledge than internal providers do. However, these results should be interpreted with caution, as I do not have data to control for compensating or demand-side rent differentials. In results not presented here, I show that we can obtain the same patterns when I consider annual instead of hourly earnings.[86]

[84] In units of one million employees.

[85] I deflate wages by the CPI with constant prices in 2009. To construct these values, I weight industries by levels of employment.

[86] These patterns are in sharp contrast with the findings for low-skill occupations proposed by Abraham and Taylor (1996), Dube and Kaplan (2010) and Goldschmidt and Schmieder (2015). For instance, Dube and Kaplan (2010) finds that for the case of janitors and security guards, the wages in the external market are lower than in the internal market. They conclude that this difference is not due to compensating or unobserved skills differentials, but rent differentials. That is, *low-rent industries' firms are more likely to outsource.* My interpretation is that, given that these occupations are low-skill in nature, there are no significant differences in issue-specific knowledge across in-house and external service providers. Therefore, saving wage costs is the first-order concern in the integration decision.

Generalists and Specialists In this section, I focus on specific occupations in which I can detect patterns between generalists and specialists. I use the description of the occupation based on the 2000 occupational classification system. I classify a sub-occupation as a generalist if the description of the occupation includes several and diverse tasks.[87] For instance, code 11-1021: General and Operation Managers describes the activities as:

> Plan, direct, or coordinate the operations of companies or public and private sector organizations. Duties and responsibilities include formulating policies, managing daily operations, and planning the use of materials and human resources, but are too diverse and general in nature to be classified in any one functional area of management or administration, such as personnel, purchasing, or administrative services. Include owners and managers who head small business establishments whose duties are primarily managerial. [88]

On the other hand, an example of an occupation that can be categorized as a specialist is 11-3042: Training and Development Managers, which has the following description:

> "Plan, direct, or coordinate the training and development activities and staff of an organization."[89]

I focus only on Managers and IT personnel as the other occupations did not have a clear way to identify the level of specialization of their sub-occupations. The chosen sub-occupations are summarized in Table 1.21.[90] Table 1.22 shows the percentages of generalists and specialists by level of vertical integration and occupation. The main message from this table is that the great majority of generalists are in the internal market and that there are specialists in both markets. Table 1.23 shows the percentages of internal and external employees by level of specialization and occupation. This table shows that the majority of internal employees are generalists, whereas the majority of external service providers are specialists, just as in Table 1.9. I take this piece of evidence as a first step towards the development of a comprehensive vertical integration theory of knowledge workers.

[87] I include sub-occupations where the list of tasks are difficult to categorize in an unified job activity. For instance, code (15-1071), "Network and Computer Systems Administrators," describes activities as: "*Install, configure, and support an organization's local area network (LAN), wide area network (WAN), and Internet system or a segment of a network system. Maintain network hardware and software. Monitor network to ensure network availability to all system users and perform necessary maintenance to support network availability. May supervise other network support and client server specialists and plan, coordinate, and implement network security measures.*" Exclude "Computer Support Specialists" (15-1041).

[88] http://www.bls.gov/soc/2000/soc_a1c1.htm.

[89] http://www.bls.gov/soc/2000/soc_a3e2.htm.

[90] For the complete list of occupations, see BLS web page.

1.4.2 Concluding Remarks

The aim of this chapter has been to explore the way in which a non-incentive-based theory advances our understanding of the integration decision of knowledge workers. The central point relates to the fact that the acquisition of knowledge is independent of its rate of use, and, therefore, exploiting these increasing returns requires conducting frequent activities in-house and infrequent tasks for several clients in the external market. When easy activities are more frequent, external staff acquires higher levels of knowledge than their in-house counterparts. Finally, as clients pay communication costs in the external market due to firm-specific knowledge, the economy saves communication costs if the clients with more firm-specific knowledge conduct their activities in-house.

Using bill- and industry-level measures, I confirm the model's main prediction using fixed-effect estimations. To tackle causality, I use the BP spill as a quasi-experiment that increased the difficulty of lobbying activities for the oil and gas extracting companies. As more-difficult activities are less frequent, only external service providers can solve harder problems. As a consequence, the oil spill increased the outsourcing of lobbying services for the affected companies.

Abstracting from the current application, the analysis underscores the potential of non-incentive integration theories to explain how firms use knowledge workers. My findings can be extended in several directions. First, by interacting the ideas developed in this chapter with the existing literature on organizational and labor economics, our understanding of integration with knowledge workers will be richer. For instance, it seems natural to add search and matching frictions, dynamic problems and moral hazard issues to my setting. How does knowledge across markets differ when there are matching frictions? What is the optimal organization of the economy when the productivity of lobbying activities varies across workers, and these returns determine the difficulty of the problems faced in the future? What does the organization of the economy look like when service providers can haggle, reduce effort, and pretend to know more than they do, but it is also in their interest to exploit the increasing returns from knowledge acquisition?

Second, in this chapter, I have focused on understanding the causes rather than the consequences of vertical integration.[91] I believe that this has to be the first step in properly assessing the consequences. Two interesting questions should be explored in future research: 1) The market joins interests for a number of clients. What are the implications for the economy when the clients' interests are not totally aligned? How does the knowledge economy solve conflicts of interest among clients? 2) What are the welfare effects of firms integrating with external service providers that serve a representative group of the population?

[91] For recent papers on the consequences of the vertical integration see for example Hortacsu and Syverson (2007), Atalay et al. (2014) and Goldschmidt and Schmieder (2015).

1.5 Appendix

1.5.1 Discussion of the Model and Related Literature

The integration decision literature has focused on the problem of incentives. In this chapter, I focus on the integration decision with knowledge-intensive workers. These workers differ from other types of workers because in the production process physical assets are irrelevant and knowledge is the key input. As a consequence, I take the natural approach of leaving aside the problem of the incentives and I focus on the use and communication of knowledge. This is not to say that the incentive problem is not important. I do believe we have gained great insights from the literature but I argue in this chapter that are strong forces with empirical support not related to incentives that can increase our understanding on the integration decision.

In this section I provide an overview of the relationship between the existing literature and this chapter. I will explain under which scenarios some of the existing incentives-based integration theories are not appropriate in the knowledge-workers context and to what extent my results contrast or confirm previous theoretical predictions. As the Rent-seeking and Adaptation theories study similar forces I comment on both theories jointly under the Transaction Cost Models (henceforth TCE) Here I briefly discuss TCE, Property Rights and Multitasking Models.

Transaction Cost Models (TCE) The main insights from this literature come from Coase (1937), Williamson (1971, 1975, 1979 and 1985).[92] The main predictions of these theories are that there is more vertical integration when the transactions involve more specific investments or when the transactions are more complex or frequent.[93]

Hold-Up. Klein, Crawford and Alchian (1978), Monteverde and Teece (1982), Anderson and Schmittlein (1984) and Joskow (1988) hypothesized that an increase in firm-specific investments would increase the likelihood of vertical integration, as the hold-up costs are higher. The main idea is that parties in a transaction make investments that have greater value inside than outside the relationship. This specificity implies that the parties are locked in ex-post. A possible solution to this lock in is writing contracts. However, in an incomplete contract world the parties have incentives to engage in opportunistic behavior ex-post, which in turn creates a hold-up problem. The literature predicts that the hold-up issue is solved or mitigated by bringing the assets that produce the specific investments in-house. As in Klein (1988), it is not clear how the opportunistic behavior is mitigated when we think in human assets. Furthermore, in a context where there is client-specific

[92] For a recent survey, see Tadelis and Williamson (2012).

[93] Although, TCE comments on the effects of the uncertainty on the integration decision, I have neglected it. A possible extension of this paper may consider the inclusion of it.

knowledge and agents differ among internal and external markets by their degree of this knowledge, the hold-up problem is intensified inside the firm as workers that have received firm-specific knowledge can attempt to hold their employers up and vice versa. My model predicts that the larger the firm-specific knowledge the more likely it is that integration will occur. This is not because clients avoid the hold-up problem as in TCE but because clients save communication costs with external providers.

Complexity. Monteverde and Teece (1982), Masten (1984) and Tadelis (2002) argue both theoretically and empirically that the probability of vertical integration increases with the complexity of the transactions. The complexity of the transaction can increase the switching costs for the buyer and as ex-post adaptation can be more costly, the buyer will be better off by conducting the transaction internally. If I interpret complexity of the transaction as the difficulty of the problem, my model predicts exactly the opposite to this literature; the more complex the transactions the more likely they are to be externally outsourced. Intuitively, this occurs in my framework because only external service providers can solve harder problems. While implied by the assumption of a decreasing density function, the model I propose conveys a simple insight: it is the relationship between frequency and complexity and not the degree of complexity that should matter to understand the integration decision.

Frequency. Williamson has used the term frequency in three different contexts.[94] First, Williamson (1991) talks about the frequency of disturbances in the environment concluding that the interaction between asset specificity and frequency will determine the optimal organization form. Second, Williamson (1979) studies the issue of the frequency of the transaction and its relationship with the type of governance that the agents optimally choose. His main conclusion is that the effect of frequency on integration is ambiguous and it depends on the specificity of the investments of the suppliers.[95] Third, Williamson (1985) talks about the frequency of trade among many trading partners. The key idea is that the cost of a hierarchical structure will be easier to recover when the frequency of the transactions is larger.

[94] The relational contracts literature (Baker et al. (1994, 2002)) has used the term to refer to the frequency of trade between specific trading patterns. The theory predicts that repeated interaction can mitigate opportunism, and therefore the incentive to maintain reputation in the external market will make clients more prone to outsource.

[95] In sum, if the investments are not specific the market should prevail but if the investments are idiosyncratic the integration is preferred. There is a natural difficulty to match his ideas with my paper as it is difficult to define the specificity of the knowledge investment of the supplier. If we understand specificity as the type of investments in which the supplier's knowledge loses value outside the relationship, the key factor to study is to which client the supplier goes once she leaves the initial labour relationship. If she goes to a client with more difficult problems, there is no loss in the value of the original knowledge investment. However, if she goes to a client with easier problems, there is a loss in the value of the investment proportional to the difference between the difficulties of the problems faced for both clients. Given this indeterminacy, the knowledge investments are not just specific or not, but they are one or the other, conditional to the new employer.

Although this chapter has intentionally neglected the issue of repeated interaction and uncertainty, it is aligned with the same insight for the third type of frequency and as Williamson (1985), the larger the frequency the more likely to bring someone in-house. This chapter complements his insight by providing both theoretical and empirical support to this idea. Furthermore, he talks separately about the issue of frequency and complexity of the transaction.

Strikingly, the issue of the frequency has not received much attention. For instance, two of the most comprehensive and recent literature reviews of the subject, Lafontaine and Slade (2007) and Bresnahan and Levin (2012) completely neglected the subject. On the other hand, Tadelis and Williamson (2012) mentioned briefly this issue in a footnote to argue that the relationship between frequency and integration is ambivalent.[96]

Property Rights Models Grossman and Hart (1986), Hart and Moore (1990) and Hart (1995) focus on neither contractible nor alienable investments. The key idea in this literature is how the allocation of decision rights to the use of the assets as contingencies appear, can modify the ex-ante investment incentives. Ownership matters because it affects the disagreement point, which in turn affects incentives through the ex-post bargaining. While this literature emphasizes the ownership of "non-human assets" as a tool to exercise power in an incomplete contracts world, it does not have clear implications when this type of assets are completely inexistent or irrelevant as in the case of PBS industries (see for instance Dube and Kaplan (2007) and Bresnahan and Levin (2012)). Clearly, the difference is that the law does not provide control rights over human beings: Buying a machine or hiring someone in-house is totally different as the employee can always quit.

Multitasking Models One of the key predictions of Holmstrom and Milgrom (1991) is that when there are two tasks that differ by the cost of measuring performance, firms bring in-house projects in which the harder to measure task is more important. This prediction is empirically confirmed by Azoulay (2004).

As my empirical application uses lobbyists as knowledge workers, it is useful to discuss two of the main tasks performed by a lobbyist: Investigating the political environment and communicating specialized knowledge to policy makers. Arguably, the second task is harder to measure not only because clients have a hard time assessing the actual knowledge of the lobbyist, but especially because it is hard to confirm the exact message that the lobbyist transmits to the policy maker.[97]

The multitasking literature predicts that the in-house activities should be more about transmitting information to the policy maker than investigating the political

[96] They argue that when there is repeated interaction, integration makes sense as the cost of creating a specialized infrastructure can be recovered. However, in a context where there are reputation effects, which I have neglected, market contracting can be the best solution.

[97] Notice for instance, that the lobbying reports neither provide information on the contacted policy makers nor the specific message transmitted.

environment. However, anecdotal evidence strongly opposes this prediction. External lobbyists tend to interact with policy makers whereas in-house staff tend to investigate the political environment.

A way to conciliate this literature with my approach is that although clients prefer to bring in-house activities that are harder-to measure, these activities tend to be more knowledge intensive. As the market allows service providers to acquire larger levels of knowledge, it is the organization of the market and not the clients' incentives that determine the integration decision.

1.5.2 Validation Exercises

The literature on work displacement has shown strong evidence on the existence of industry-specific skills.[98] These papers show that workers that switch industries following displacement have significantly larger earnings losses than workers that remain in the same industry after displacement. Parent (2010) show evidence of both firm and industry specific skills using the National Longitudinal Survey of Youth and the Panel study of Income dynamics concluding that the industry instead of the firm-specific skills are more important to explain the wage profile in terms of human capital.

One can take this intuition one step further. If the earning losses from switching industries proxies for industry-specific knowledge, the level of the losses may proxy for the level of specificity. That is, industries in which workers suffer more from leaving the industry, will be industries with larger level of industry-specific skills.

Using this intuition I conduct two exercises. I first use estimates of the wage losses by industry of the workers calculated in other papers and I compare to my RegData measures using the 2-digit code classification. Second, I use the displacement workers methodology to estimate earning losses for in-house lobbyists by industry and I compare them with the RegData.

I first use the results from Couch and Placzek (2010). Table **1.A.1** shows the estimated results from Section G from their appendix and average number of regulating words by industry. The second and third column represent the actual values of the coefficients and average number of words, respectively whereas the last two columns give the relative ranking across variables. To improve the accuracy in the comparison of these two results I have used the same time-span that Couch and Placzek (2010) has used: 1993 to 2004.

The bottom line from this table is that there is not a perfect alignment between these two measures but there are some similarities. The correlation across variables is -24% and not significant. The Spearman correlation for the rankings is 26% and not significant. The lack of significance can be due to the weak relationship or the few number of observations. However, I find remarkable the similarity for the rankings. For instance, Manufacturing, Financial/Real Estate and Education Health have the same ranking for both measures. Similarly, the *all other industries* category rank very closely (i.e. 4th and 5th). I obtains similar results when I compare RegData to Jacobson et al. (2003)'s table 2 coefficients .

A problem with these comparisons is that these results are not controlling for the industry that the displaced worker goes. For instance, a large value in these coefficients can be explained by either large values of industry-specificity or by a large fraction of workers going to industries that do not belong to the same or

[98]See for example, Jacobson, LaLonde, and Sullivan (1993), Carrington (1993), Neal (1995) Parent (2010) and Couch and Placzek (2010).

similar sector than the industry in which the worker was displaced from.

Fortunately, appendix H controls by receiver industry, although only for manufacturing and non-manufacturing workers. They estimate that the earning losses for manufacturing workers once they are re-employed in non-manufacturing industries is about 3180.75 after three years. This number is only 1377.26 for the case of workers that are displaced from non-manufacturing industries. Similar results are obtained from Table 3 of Jacobson et al (1993). I interpret this result as manufacturing industries having a larger level of industry-specific skills than other industries. The average number of words for manufacturing industries in RegData is 9.099.640 whereas is only 3.380.321 for other industries. The bottom line is that manufacturing industries have not only more regulating words than non-manufacturing industries but the fraction for both measures between manufacturing and non-manufacturing is about 40%. I take this result as a simple step to externally validate the measures I use in this chapter.[99]

An alternative data source come from the International Economics literature. Nunn (2007) constructs measure of relationship-specificity at the industry level using information on whether the inputs are sold on an organized exchange or are reference priced in trade publications. An input is relationship-specific if the value of the input in a buyer-seller relationship is similar inside and outside the relationship. If the input is sold on an organized exchange, the market is thick (many buyers and sellers), and as a consequence the input is not relationship-specific. A similar intuition applies for the case where the input price appears in trade publications. Therefore, an intuitive measure of industry-specificity is the value of inputs that are neither bought and sold on an exchange nor reference priced.

Table **1.A.2** shows the correlations of this exercise. For better comparability, with Nunn's data , I have used RegData at the 4-digit levels for the year 1997. Similar results emerge when I use alternative years. Nunn's data is originally provided at the 6-digit level. To convert it at the 4-digit level I have simply aggregated all the 6-digit codes by taking their average. Alternative aggregations give similar results. The data considers two different classifications coming from Rauch (1999): conservative and Liberal. The conservative measure tries to minimize the number of 3 and 4 digit commodities that are classified as either organized exchange or reference priced. The liberal measure maximizes these commodities.(for more details, please see Rauch (1999). The contribution of Rauch is to provide these measures by input whereas Nunn provides these measures at the industry level using I/O tables). The bottom line from this exercise is that there is a positive and significant correlation between Nunn measures and the regulating words.

[99] In order to calculate the average number of words per industry I have taken average of total number of words for two-digit industries across all years. The classification used in the table is as follows: Manufacturing: NAICS 11, 21 and 23, Trade: 42, Financial/Real Estate: 52, 53, Prof./Business Services: 51, 54, 55, 56 Edu./Health Services: 61, 62 and All Other Industries: 71.

1.5.3 Tables

	Function			
Parameter	z^*	z^{**}	n	VI
h	0	(-)	(-)	(+)
c	(-)	(-)	(-)	(-)
λ	(?)	(?)	(+)	(?)

Table 1.1: Summary of the Comparative Statics Exercise.

Note: The rows define the parameters to change whereas the columns define the equilibrium object that is subject to the change of the parameter. For instance, the (1,1) element in the matrix denotes the sign of the $\frac{\partial z^*}{\partial h}$.

	$z_B \leq z_B^*$	$z_B^* < z_B \leq z_B^{**}$	$z_B > z_B^{**}$
$z_A \leq z_A^*$	*VI, VI*	*VI,EC*	*VI,NA*
$z_A^* < z_A \leq z_A^{**}$	*EC, VI*	*EC,EC*	*EC, NA*
$z_A > z_A^{**}$	*NA, VI*	*NA, EC*	*NA,NA*

Table 1.2: Predictions from the Model with Two Issues.

Note: VI, EC and NA indicate Vertical Integration, External Contracting and No Activity, respectively. The first (second) coordinate in each cell predicts the action for topic A (B). According to the table there are six types of clients in the problem.

Difficulty 2: Fraction restraining Words			*Communication Cost: Regulating Words*		
NAICS	*Least*	*%*	*NAICS*	*Least*	*K*
1121	Cattle Ranc. and Farming	0.41	1153	Support Activities for For.	1
2372	Land Subdivision	0.6	3141	Textile Furnishings Mills	1
2382	Building Equip. Cont.	0.64	4236	Elect. and Elec. G. Mer. Who.	1.2
5322	Cons. Goods Re.	0.66	3313	Alum. & Alum. Prod. & Proc.	1.8
1133	Logging	0.67	3111	Animal Food Manuf.	4.5
NAICS	*Most*	*%*	*NAICS*	*Most*	*K*
3351	Electric Ligh. Equip. Manuf.	1.53	3361	Motor Vehicle Manuf.	1394
5171	Wired Telecomm. Carriers	1.55	5221	Depository Credit Inter.	1518.4
4471	Gasoline Stations	1.56	5222	Nondepository Credit Int.	1557.9
4881	Supp. Act. for Air Transport	1.61	2211	Elec. Pow. Gen., Trans. & Distr.	1675.4
6219	Other Amb. Health Care Serv.	1.71	3241	Petr. & Coal Products Manuf.	2547.9

Table 1.3: Restraining and Regulating Words Across Industries.

Note: LHS: Most (bottom panel) and least (top panel) four-digit industries by the fraction of restraining words over the total number of words . RHS: Most (bottom panel) and least (top panel) four-digit industries by the total number of words. Data for 2014. The column % reports the fraction of restraining words over the total number of regulating words. The column K reports the total number of regulating words in thousand units.

	Mean	Standard Dev		Obs
VI	0.135	Total	0.342	409515
fr	0.180	Total	0.367	176143
		Between	0.260	
		Within	0.076	
Frequency	359.380	Total	398.675	36477
		Between	268.297	
		Within	279.541	
Difficulty 1 (Committees)	24.65	Total	29.17	36477
		Between	17.66	
		Within	18.91	
Difficulty 2 (Restraining W.)	0.1344	Total	0.0934	176143
		Between	0.0930	
		Within	0.0906	
Industry K. (/100)	3668.272	Total	7365.37	176143
		Between	7817.72	
		Within	730.372	

Table 1.4: Descriptive Statistics.

Note: The table shows the mean, standard deviation and total number of observations for the main variables: V I, fr, frequency, two difficulty measures and industry-specific knowledge. The unit of observation used to construct this table is the client-semester for all the variables except V I, which is constructed at the transaction level. The number of observations for the frequency and the first measure of difficulty is smaller than for the other variables, as only a fraction of clients lobby for bills.

Transaction Level	*Dependent Var: VI*			
Frequency	0.0155***	0.0160***	0.0162***	0.0112*
St. Err.	(0.0032)	(0.0028)	(0.0032)	(0.0057)
Obs	54472			
Adj R^2	0.462	0.398	0.562	0.852
Client Level	*Dependent Var: fr*			
Frequency	0.0182***	0.0189***	0.0192***	0.0144**
St. Err.	(0.0031)	(0.0033)	(0.0032)	(0.0066)
Obs	36,467			
Adj R^2	0.680	0.650	0.680	0.979
	Fixed Effects			
Client			*X*	*X*
Industry	*X*			
Semester	*X*		*X*	
*Industry*Year*		*X*		*X*

Table 1.5: Frequency and Vertical Integration.

Note: An observation is a lobbying transaction (Top Panel) or a client-semester combination (Bottom Panel). The table shows the coefficients for the (first proxy) frequency variable. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors are clustered at the industry level in the first two columns and the client level in the last two.

Transaction Level	*Dependent Var: VI*			
Difficulty 1	-0.0236***	-0.0259***	-0.0271**	-0.0224*
St. Err.	(0.0043)	(0.0053)	(0.0121)	(0.0284)
Obs	54472			
Adj R^2	0.462	0.498	0.662	0.846
Client Level	*Dependent Var: fr*			
Difficulty 1	-0.0362***	-0.0405***	-0.0377**	-0.0330*
St. Err.	(0.0075)	(0.0083)	(0.0186)	(0.0371)
Obs	36,467			
Adj R^2	0.678	0.691	0.743	0.945
	Fixed Effects			
Client			X	X
Industry	X			
Semester	X		X	
*Industry*Year*		X		X

Table 1.6: First Measure of Difficulty and Vertical Integration.

Note: An observation is a lobbying transaction (Top Panel) or a client-semester combination (Bottom Panel). The table shows the coefficients for the first measure of difficulty (based on congressional committees). (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used to run the regression. Standard errors are clustered at the industry level in the first two columns and the client level in the last two.

Transaction Level	*Dependent Var: VI*			
Difficulty 2	-0.0215**	-0.0215**	-0.0142*	-0.0142*
St. Err.	(0.0106)	(0.0109)	(0.0082)	(0.0084)
Obs	297,916			
Adj R^2	0.659	0.659	0.759	0.759
Client Level	*Dependent Var: fr*			
Difficulty 2	-0.0358**	-0.0358**	-0.0158**	-0.0158**
St. Err.	(0.0179)	(0.0177)	(0.0070)	(0.0071)
Obs	176,143			
Adj R^2	0.501	0.501	0.946	0.956
	Fixed Effects			
Client			X	X
Industry	X	X		
Semester		X		X
Year	X		X	

Table 1.7: Second Measure of Difficulty and Vertical Integration.

Note: An observation is a lobbying transaction (Top Panel) or a client-semester combination (Bottom Panel). The table shows the coefficients for the second measure of the difficulty (based on RegData 2.2.). (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors are clustered at the industry level in the first two columns and the client level in the last two.

Transaction Level	*Dependent Var: VI*			
Industry K.	0.0083***	0.0083***	0.0045*	0.0046*
St. Err.	(0.0029)	(0.0027)	(0.0027)	(0.0029)
Obs	297,916			
Adj R^2	0.660	0.660	0.769	0.769
Client Level	*Dependent Var: fr*			
Industry K.	0.0086**	0.0086**	0.0037*	0.0037*
St. Err.	(0.0040)	(0.0041)	(0.0021)	(0.0023)
Obs	176,143			
Adj R^2	0.501	0.501	0.956	0.956
	Fixed Effects			
Client			X	X
Industry	X	X		
Semester		X		X
Year	X		X	

Table 1.8: Industry-knowledge and Vertical Integration.

Note: An observation is a lobbying transaction (Top Panel) or a client-semester combination (Bottom Panel). The table shows the coefficients for the industry-knowledge variable (based on RegData 2.2.). (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors are clustered at the industry level in the first two columns and the client level in the last two.

	In-House	*External*	*Total*	*In-House*	*External*
Generalist (%)	**78**	22	15760	57.4	20.9
Specialist (%)	40.6	59.4	22450	42.6	**79.1**
HHI	0.31	0.50			

Table 1.9: Generalists and Specialists Across Markets.

Note: LHS: Fraction of generalists and specialists across markets. The percentages shown in the generalist (specialist) category are calculated as a fraction of the total number of generalists (specialist). RHS: Fraction of in-house and external staff across classifications of generalists and specialists. The percentages shown in the in-house (external) category are calculated as the fraction of the total number of in-house (external) lobbyists. Last row: Herfindahl index by market. An observation is a lobbyist.

	External+Specialists	
%	*/Specialists*	*/External*
All periods	59.4	79.1
≥5 years	59.1	67.5
≥7.5 years	63.6	65.5
≥10 years	68.5	64.6
≥12.5 years	69.4	60.8
≥15 years	83.8	52.9

Table 1.10: Fraction of External Specialists Over Different Denominators.

Note: An observation is a lobbyist. The second column show the percentage of external lobbyists that are classified as specialists over the total number of specialists. The last column show the percentage of the specialist external lobbyists over the total of external lobbyists. The first row shows the percentages for all the lobbyists in the sample. The next rows only consider lobbyists that have lobbied a given number of years. For instance, the third row gives the percentages for the case we restrict our sample to lobbyists that have worked in the industry at least 5 years. The table shows that the patterns between vertical integration and specialization remain once I control for the number of periods the lobbyists appear. For instance, for any sample of lobbyists at least 60% of the specialists are external lobbyists. On the other hand, the majority of the external lobbyists are specialists for any sub-sample of advocates.

		Internal C.	*External C.*
Sales	**Mean**	**2.184**	**0.395**
(Million Units)	Std. Dev	30.823	6.228
Employees	**Mean**	**470.699**	**151.048**
(Thousands Units)	Std. Dev	5860.7	2181.793
	Obs	1,286	14,298

Table 1.11: Size Statistics by Type of Client.

Note: The first pair of rows provide the mean and standard deviation values for the sales of the clients. This variable is measured in US Million units. The second pair of rows provide the mean and standard deviation values for the number of employees of the clients. This variable is measured in Thousand units. The last two columns provide the descriptive statistics for the total value of Sales and the total number of employees by type of client. A client is internal is she has always lobbied with in-house lobbyists. A client is external if she has never lobbied with in-house lobbyists. An observation is a client.

	In-House	*External*	*Mixed*	*Total*
Honorable	9.09	80	10.91	165
Senate	15.72	64.45	19.83	1412
House	11.7	69.15	19.15	94
White House	10.93	68.17	20.90	311
Aide	17.15	61.41	21.44	1,516
Clerk	6.57	85.86	7.58	198
Counsel	10.91	71.97	17.12	1,063
Experience	31.84	48.67	19.48	12,894

Table 1.12. Distribution of Revolving-door Lobbyists by Integration Decision and Type of Job Occupied.

Note: An observation is a lobbyist. The first column has the job/office title previously occupied. Columns 2 to 4 have the distribution of the types of lobbyists for each office title, and the fifth column has the total number of lobbyists with that job experience. Take for example, the first row. A lobbyist is categorized as honorable (the title for former members of Congress) if she was either a Senator or a Representative before being a lobbyist. In my sample, there are 165 ex-Congressmen lobbyists. Among them, 80% work as external lobbyists, while only 9% of them work as in-house lobbyists. Senate (House) represents past work in the Senate but not as Senator (House Representative). Experience is a dummy equal to 1 if the lobbyist occupied any public office (included the listed in the table). The values reported for this variable correspond to the percentage of lobbyists with experience in a given category of lobbyists.

	Fraction of Internal Reports				*Fraction of Internal Lobbyists*			
$T_i \cdot P_t$	-0.050**	-0.050**	-0.049***	-0.050***	-0.059**	-0.060**	-0.053***	-0.053***
	0.024	0.024	0.007	0.007	0.028	0.028	0.009	0.009
T_i	0.076***	0.077***			0.069***	0.070***		
	0.020	0.020			0.023	0.023		
P_t	0.03**		0.015***		0.026*		0.07	
	0.01		0.003		0.013		0.001	
Obs		3731				3198		
R^2	0.06	0.07	0.92	0.93	0.03	0.05	0.90	0.91
QFE	N	Y	N	Y	N	Y	N	Y
CFE	N	N	Y	Y	N	N	Y	Y

Table 1.13: Results of the BP oil Spill Dif-Dif Estimation.

Note: The LHS provides the point estimates when the dependent variable is the fraction of reports made internally. The RHS provides the point estimates when the dependent variable is the fraction of lobbyists working in-house. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used to run the regression. Standard errors are clustered at the client level. Q.F E. denotes quarter-fixed effects controls and CFE denotes client-fixed effects. An observation is a client-quarter combination.

	Number of Internal Lobbyists				*Number of External Lobbyists*			
$T_i \cdot P_t$	-0.174**	-0.188**	-0.169**	-0.1925**	0.360**	1.405**	0.382**	1.448**
	0.072	0.078	0.069	0.079	0.145	0.679	0.159	0.681
T_i	-0.409***		-0.404***		2.516**		2.517**	
	0.148		0.148		1.223		1.229	
P_t	-0.0231	-0.0950			-0.132	-0.394		
	0.0837	0.122			0.217	0.240		
obs				3,729				
R^2	0.361	0.841	0.363	0.843	0.055	0.826	0.058	0.829
QFE	N	N	Y	Y	N	N	Y	Y
CFE	N	Y	N	Y	N	Y	N	Y

Table 1.14: Evidence on the Absolute Change on the Number of Lobbyists.

Note: The dependent variable is the total number of internal lobbyists in the LHS and total number of external lobbyists in the right hand side. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used to run the regression. Standard errors are clustered at the client level. Q.F E. denotes quarter-fixed effects controls and CFE denotes client-fixed effects. An observation is a client-quarter combination.

	Fraction Internal Reports				*Fraction Internal Lobbyists*			
$T_i \cdot P_t$	-0.052***	-0.035***	-0.052***	-0.036***	-0.062***	-0.035***	-0.063**	-0.036***
	(0.000)	(0.007)	(0.000)	(0.007)	(0.001)	(0.009)	(0.001)	(0.009)
T_i	0.069		0.069		0.048		0.048	
	(0.013)		(0.013)		(0.020)		(0.020)	
P_t	0.030***	0.013***			0.027***	0.005		
	(0.000)	(0.004)			(0.000)	(0.005)		
Reports	0.010	-0.047***	0.011	-0.047***	0.023	-0.057***	0.023	-0.058***
	(0.014)	(0.002)	(0.014)	(0.002)	(0.019)	(0.003)	(0.019)	(0.003)
obs	3,729				3,196			
R^2	0.008	0.934	0.009	0.935	0.013	0.917	0.015	0.918
CFE	N	Y	N	Y	N	Y	N	Y
QFE	N	N	Y	Y	N	N	Y	Y

Table 1.15: BP Estimations Including Controls for Changes in the Demand.

Note: The dependent variable is the total number of internal lobbyists in the LHS and total number of external lobbyists in the right hand side. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used to run the regression. Standard errors are clustered at the client level. Q.F E. denotes quarter-fixed effects controls and CFE denotes client-fixed effects. An observation is a client-quarter combination.

	Fraction of Internal Reports					
$T_i \cdot P_t$	-0.0575**	-0.0555***	-0.0352*	-0.0271***	-0.0257**	-0.0350***
	0.0282	0.00767	0.0192	0.00567	0.0109	0.00447
T_i	0.0377		0.113***		0.185***	
	0.0237		0.0161		0.00916	
P_t	0.0375**		0.0152**		0.00567	
	0.0154		0.00688		0.00480	
obs	2850		6113		4298	
R^2	0.002	0.944	0.027	0.937	0.218	0.897
QFE	N	Y	N	Y	N	Y
CFE	N	Y	N	Y	N	Y
Control	Retail Sales		Real Estate		Casino	

Table 1.16: BP Estimations with Different Control Groups.

Note: The dependent variable is the fraction of reports made internally. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used to run the regression. Standard errors are clustered at the client level. Q.F E. denotes quarter-fixed effects controls and CFE denotes client-fixed effects. An observation is a client-quarter combination.

	Fraction Internal Reports				*Fraction Internal Lobbyists*			
$T_i{\cdot}P_t$	-0.052**	-0.049***	-0.052**	-0.051***	-0.061**	-0.062**	-0.053***	-0.054***
	(0.024)	(0.011)	(0.024)	(0.011)	(0.028)	(0.028)	(0.013)	(0.013)
T_i	0.078***		0.078***		0.072***		0.072***	
	(0.020)		(0.020)		(0.024)		(0.024)	
P_t	0.030***	0.015***			0.026*	0.008		
	(0.012)	(0.004)			(0.014)	(0.005)		
obs	3,729				3196			
R2	0.006	0.926	0.007	0.927	0.003	0.907	0.005	0.908
QFE	N	N	Y	Y	N	N	Y	Y
CFE	N	Y	N	Y	N	Y	N	Y

Table 1.17: BP Estimations Using Two-way Clustering.

Note: Two-way clustering is as follows: First at the industry-quarter level to solve the problem of autocorrelation within a group and then at the client level to solve for the problem of autocorrelation at the client level. The LHS provides the point estimates when the dependent variable is the fraction of reports made internally. The RHS provides the point estimates when the dependent variable is the fraction of lobbyists working in-house. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used to run the regression. Standard errors are clustered at the client level. Q.F E. denotes quarter-fixed effects controls and CFE denotes client-fixed effects. An observation is a client-quarter combination.

	Fraction of Internal Reports		*Fraction of Internal Lobbyists*	
$T_i \cdot P_t$	-0.0501**	-0.0350***	-0.0614**	-0.0354***
	(0.0239)	(0.00721)	(0.0279)	(0.0093)
T_i	0.0668***		0.0458*	
	(0.020)		(0.0236)	
P_t	0.0300**	0.0125***	0.0264*	0.00474
	(0.0109)	(0.00365)	0.0137	(0.00478)
Obs	407		389	
R^2	0.010	0.973	0.007	0.977
QFE	N	Y	N	Y
CFE	N	N	N	N

Table 1.18: BP Estimations Collapsing Dependent Variables.

Note: These estimations, follow the procedure of Barrage et al. (2014). I collapse the dependent variables into averages within two time periods pre and post-spill period. The LHS provides the point estimates when the dependent variable is the fraction of reports made internally. The RHS provides the point estimates when the dependent variable is the fraction of lobbyists working in-house. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used to run the regression. Standard errors are clustered at the client level. Q.F E. denotes quarter-fixed effects controls and CFE denotes client-fixed effects. An observation is a client-quarter combination.

Occupation	**Occupational Code**	**NAICS**
Lawyers	23 (excluding 23-1023)	5411
Managers	11	5511
IT Personnel	15-10[100]	5415
Accountants	13-2011	5412
Lobbyists	27-3031	5418[101]

Table 1.19: Matching Between Occupation and NAICS Codes.

Note: An occupation is defined according to the Office of Management and Budget (OMB) Standard Occupational Classification (SOC) system while an industry is defined with the four-digit NAICS classification. Each row represents a particular occupation, and I define each according to the occupational codes described in the second column. The last column provides the NAICS codes representing firms whose primary business segment is in that occupation. A service provider is external if the service provider works in the NAICS industry primarily concerned with that occupation.

[100] 15-11 after 2010. The definition for this occupation only change the occupational code but not the description of activities.

[101] Ideally I would like to use the code 541820. However, the survey only contains 4-digit NAICS code industries.

Hourly Wage							
Occupation	*In House*	*Mean*	*Percentiles*				
			(10)	*(25)*	*(50)*	*(75)*	*(90)*
Lawyers	1	40.11	17.22	23.12	34.94	51.29	63.87
	0	47.33	15.71	21.59	35.20	65.27	80[102]
Managers	1	47.83	24.16	31.88	42.09	53.63	60.66
	0	58.86	29.54	39.61	51.17	67.24	73.81
IT P.	1	35.07	21.29	26.76	33.95	42.40	51.27
	0	38.25	21.19	27.60	36.27	46.69	58.53
Accountants	1	30.53	18.22	22.29	28.28	36.30	46.19
	0	34.66	17.61	22.23	29.15	40.31	58.56
Lobbyists	1	26.91	14.60	18.60	24.49	32.76	42.81
	0	32.60	15.40	19.14	26.70	38.55	56.82

Table 1.20: Wages Differences by Type of Market.

Note: The table shows the mean and five different percentile values for hourly wage by occupation across all years. For each occupation, I calculate separately the wage statistics for the internal and external market service providers. The differences in wages across internal and external service providers are only t-test statistically significant for mean and percentiles values of 50, 75 and 90. Wages are deflated by the CPI with constant prices in 2009. Industries are weighted by levels of employment.

	Generalists	*Specialists*
Managers (11-)	1021, 3011, 9011, 9012, 9013	2011,2031,3021,3031,3041,3042, 3051, 3061, 3071, 3111, 3131, 9161
IT P (15-)	1032, 1061, 1071, 1081, 1133, 1141, 1142	1011, 1021, 1031, 1111, 1122, 1131, 1132, 1134

Table 1.21: Matching Occupational Codes and Type of Provider.

Note: The table shows the sub-occupations of Managers and I.T. Personnel that were classified as generalists and specialists according to the description of the sub-occupation.

[102] The survey excludes the exact value of hourly wages when the value is equal to or greater than $80.00 per hour or $166,400 per year. This necessarily implies that external lawyers earn more per hour than internal lawyers.

		Internal	*External*	*Total*
Generalists	*Managers*	**0.961**	0.039	2036201
	IT	**0.783**	0.217	866898
Specialists	*Managers*	0.914	0.086	1234908
	IT	0.666	0.334	900321

Table 1.22: Fraction of Generalists and Specialists Across Markets for Management and IT.

Note: The table shows the generalists and specialists by level of vertical integration and occupation. Total is a row. To construct this table, I proceed as follows: First, I compute the yearly total number of employees for each combination of generalist and specialist in each of the markets (internal and external). Then, I average the total employment for each combination across years and compute the proportions shown in the tables.

		Internal	*External*
Generalists	*Managers*	**0.634**	0.424
	IT	**0.531**	0.385
Specialists	*Managers*	0.366	**0.576**
	IT	0.469	**0.615**
Total	*Managers*	3085713	185396
	IT	1278028	489190

Table 1.23: Fraction of Internal and External Service Providers by the Level of Specialization.

Note: Table 20 shows the percentages of internal and external employees by level of specialization and occupation.Total is a column. To construct this table, I proceed as follows: First, I compute the yearly total number of employees for each combination of generalist and specialist in each of the markets (internal and external). Then, I average the total employment for each combination across years and compute the proportions shown in the tables.

1.5.4 Proofs

Lemma 1.1 *The knowledge of external providers is larger than the knowledge of internal providers (i.e.* $z^* < z^{**}$*) and both knowledge levels are decreasing in* c *(i.e.* $\frac{\partial z^*}{\partial c}, \frac{\partial z^{**}}{\partial c} < 0$*). The larger the firm-specific component, the lower the knowledge acquired by the external providers (i.e.*$\frac{\partial z^{**}}{\partial h} < 0$*).*

Proof. Follows from the text. ■

Lemma 1.2 *There are two cutoffs of knowledge levels* z^*, z^{**}*. Clients with sufficiently high frequency problems (i.e.* $z \leq z^*$*) go to the internal market, clients with intermediate levels of difficulty (i.e.* $z^* < z \leq z^{**}$*) hire external service providers. Finally, clients that face very infrequent problems do not hire any service provider (i.e.* $z > z^{**}$*).*

Proof. There are two cut-off levels of knowledge in the economy, z^*, z^{**}. The optimal organization of the economy requires that clients with problems $z < z^*$ go to the internal market whereas clients with $z^* < z < z^{**}$ use external service providers. To see why, assume there are two clients with problems $z' \leq z^*$ and $z^* < z'' \leq z^{**}$. Notice that the total net production for these two clients under this arrangement is $2 - c\left[z^* + \frac{z^{**}}{n}\right]$. Now, lets consider switching these firms across markets. The firm with problem z' will get $1 - c\frac{z^{**}}{n}$ whereas the other firm will get $-cz^*$, as the in-house service provider cannot solve the problem with difficulty z''. As neither other clients nor service providers modify their payoffs under any of these two arrangements, this concludes the proof. Notice that if the production is proportional to the difficulty of the problem instead of being normalized to 1 for the solved problems I get the same result as $z' + z'' - c\left[z^* + \frac{z^{**}}{n}\right] > z' - c\left[z^* + \frac{z^{**}}{n}\right]$.
■

Lemma 1.3 *The expected surplus in the external market is maximized when clients with high firm-specific knowledge levels are allocated to the internal market. The ex-post surplus in the external market with clients with a higher firm-specific knowledge can be larger or lower than the ex-post surplus with clients with a lower firm-specific knowledge.*

Proof. Assume there are $n-1$ clients with firm-specific knowledge h and one firm with firm-specific knowledge $\widehat{h}$. The surplus for these n firms is $n\left[F(z) - F(z^*)\right] - cz$ with time constraint equals to $1 = (1 - F(z^*)) \cdot \left[(n-1) \cdot h + \widehat{h}\right]$. Rearranging this constraint I get $\frac{1}{n} = \frac{(1-F(z^*))h}{(1-F(z^*))h + 1 - (1-F(z^*))\widehat{h}}$. Notice that the joint surplus is $\pi = \left[F(z) - F(z^*)\right] - \frac{c}{n}z = \left[F(z) - F(z^*)\right] - \frac{(1-F(z^*))hcz}{(1-F(z^*))h + 1 - (1-F(z^*))\widehat{h}}$. Using the envelope theorem I notice that $\frac{\partial \pi}{\partial \widehat{h}} < 0$. Notice, that the surplus in the internal market does not depend on the firm-specific knowledge of the clients, therefore allocating clients with high firm-specific knowledge saves communication costs in the external market. Assume there are two clients with $z, z' < z^*$ and the client with problem z has firm-specific knowledge h whereas the other client has firm-specific knowledge h'. If only one client can go to the internal market, given the result above it has to be the one with lower firm-specific knowledge. In alternative cases (either $z < z^* < z'$, $z' < z^* < z$ or $z^* < z, z'$) the surplus is maximized by sending clients with low firm-specific knowledge to the external market. For the second part of the result, notice that the ex-post surplus of n_i clients in the external market is $n_i - cz_i^{**}$. As the internal market is independent of the firm-specific knowledge, the previous term is equal to $\frac{1}{h_i\,((1-F(z^*)))} - cf^{-1}\left(ch_i\left(1 - F(z^*)\right)\right)$. As I increase h_i the number of firms being served in the external market decreases but the issue-specific knowledge reached by the external service providers decreases, making less costly to hire service providers. As service providers ' payoff do not change with the firm-specific knowledge this concludes the proof. ■

Theorem 1.1 *The fraction of vertically integrated clients increases with the firm-specific knowledge h and can increase or decrease with the cost of acquiring issue-specific knowledge.*

Proof. Follows from taking derivatives from the fraction shown in the text. ■

Lemma 1.4 *The optimal time allocation for this problem is characterized by a corner solution. The provider should give all their time to the activity more likely to face. That is, if it is more likely to appear problems type i, the provider should not provide any time to activity $j \neq i$.*

Proof. Notice that these probabilities are exogenous to the client, so t_A is given. The allocation will depend on the relative probability of facing one problem. Notice that if $\widehat{z_A} = 0 \neq \widehat{z_B}$, $t_A = 1$ and if $\widehat{z_B} = 0 \neq \widehat{z_A}$ the fraction spend on issue A is $t_A = 0$. Finally, notice that as t_A should be lower than 1, which implies that $F_A(\widehat{z_A}) \leq 0$, which only holds when $\widehat{z_A} = 0$. ■

Proposition 1.1 *There is a range of θ such that $t_A \in (0, 1)$. For this range of values of θ, generalists never exist for $\widehat{z_A} = \widehat{z_B} = 0$.*

Proof. As t_A is a fraction of the time spend on issue A, $t_A \in [0, 1]$. The conditions to get $t_A \geq 0$ are : 1. $\theta \leq \frac{1}{1-F_B(\widehat{z_B})}$ and $F_B(\widehat{z_B}) > F_A(\widehat{z_A})$ or 2. $\theta \geq \frac{1}{1-F_B(\widehat{z_B})}$ and $F_B(\widehat{z_B}) < F_A(\widehat{z_A})$. Notice that if $\widehat{z_B} = 0$, the first condition implies $0 > F_A(\widehat{z_A})$ which is not possible. For the second condition, $\theta \geq 1$ and $0 < F_A(\widehat{z_A})$ which only occurs when $\widehat{z_A} > 0$. The condition to get $t_A \leq 1$ is: $\theta \geq \frac{1}{1-F_A(\widehat{z_A})}$. Then, condition 1 for $t_A \geq 0$ and condition for $t_A \leq 1$ imply: $\frac{1}{1-F_A(\widehat{z_A})} \leq \theta \leq \frac{1}{1-F_B(\widehat{z_B})}$ and $F_B(\widehat{z_B}) > F_A(\widehat{z_A})$ whereas condition 2 for $t_A \geq 0$ combined with the condition for $t_A \leq 1$ imply: $\theta \geq \frac{1}{1-F_A(\widehat{z_A})} \geq \frac{1}{1-F_B(\widehat{z_B})}$. ■

Proposition 1.2 *There are at most two cutoffs (excluding the firm boundary) in the internal market for each issue. For issue j, the relevant cutoffs are $z_j^{=}$ and z_i^*. Clients with $z_j \leq z_j^{=}$ hire two specialists, clients with $z_i^* \geq z_j \geq z_j^{=}$ hire a generalist and clients with $z_j \geq z_i^*$ hire one specialist.*

Proof. It follows from the text. ■

Corollary 1.1 *If $(z_A^{=}, z_B^{=}) = (z_A^*, z_B^*)$ there is only one cut-off. In this case, the internal market only contains specialists. If $(z_A^{=}, z_B^{=}) < (z_A^*, z_B^*)$ clients with $(z_A^{=}, z_B^{=}) \geq (z_A, z_B)$ will hire two specialists and clients with $(z_A^{=}, z_B^{=}) \leq (z_A, z_B) \leq (z_A^g, z_B^g)$ will hire one generalist.*

Proof. It follows from the text. ■

Lemma 1.5 *Generalists external service providers can only exist for the combinations of $n_A^g h_A$ and $n_B^g h_B$ such that $\widetilde{z}_A \in \left(Q_A\left(1 - \frac{1}{\theta n_A^g h_A}\right), Q_A\left(1 - \frac{n_B^g h_B}{n_A^g h_A}\right)\right)$ and $\widetilde{z}_B \in \left(z_I^B, Q_B\left(1 - \frac{n_A^g h_A}{n_B^g h_B}\left(1 - F_A\left(\widetilde{z}_A\right)\right)\right)\right)$ or $\widetilde{z}_B \in \left(Q_B\left(1 - \frac{1}{\theta n_B^g h_B}\right), 1\right)$ where Q_i is the quantile function of the market i-th.*

Proof. It follows from the text. ■

Lemma 1.6 *If $t_A^{E*} = 1$ then the knowledge of the generalists is the same as the specialists in the external market $z_A^g = z_i^{**}$. Let n_A^s be the leverage in the external market A with knowledge z_i^{**}. Then, $sign(t_A^{E*} - \frac{n_A^s}{n_A^g}) = sign(z_A^g - z_i^{**})$.*

Proof. It follows from the text. ■

Lemma 1.7 *For a given issue, external providers always acquire more knowledge than any type of internal providers.*

Proof. It follows from the text. ■

Lemma 1.8 *The level of vertical integration in the industry is decreasing in the firm-specific levels h_i and h_j.*

Proof. It follows from the text. ■

Lemma 1.9 *For any two layers in the external market, i and j, with $i < j$, $z_j^{**} > z_i^{**}$ and $n_j > n_i$.*

Proof. It follows from the text. ■

Lemma 1.10 *If* $z \sim Exp(\lambda)$*,* $n_i = \left(\frac{\lambda}{ch}\right)^i$ *and* $z_i^{**} = -\ln\left[\frac{1}{hn_{i+1}}\right]\left(\frac{1}{\lambda}\right)$*. In this case, the difference of knowledge between any two layers i and* $i-1$ *is given by the constant* $z_i^{**} - z_{i-1}^{**} = \frac{1}{\lambda}\ln(\frac{\lambda}{ch})$*, the total number of required layers to cover all the external market area is given by the ceiling function of* $\frac{1-z^*}{z_i - z_{i-1}}$*,* $\left\lceil \frac{1-z^*}{z_i - z_{i-1}} \right\rceil = \frac{\lambda + \ln(\frac{c}{\lambda})}{\ln(\frac{\lambda}{ch})}$ *. The difference in the leverage of the external providers is given by* $n_i - n_{i-1} = \left(\frac{\lambda}{ch}\right)^{i-1}\left(\frac{\lambda}{ch} - 1\right)$*, which is increasing in* i*. Finally, the cost that each client pays in the external market decreases at an increasing rate as I increase the leverage. That is,* $\frac{c}{n_i} - \frac{c}{n_{i+1}} = \left(\frac{\lambda}{ch} - 1\right) / \left(\frac{\lambda}{ch}\right)^{i+1}$*, which is decreasing in* i*.*

Proof. It follows from the text. ■

Remark 1.1 *After two periods, the fraction of clients using exclusively in-house staff is larger than the fraction of clients using exclusively external providers if* $c > f\left(Q\left(\frac{1}{2}\right)\right)$*, where* Q *represents the Quantile function.*

Proof. It follows from the text. ■

Lemma 1.11 *If* $F_j(z_j^*) \geq F_i(z_i^*)$*, it is more common to find clients with in-house providers in issue i hiring external service providers for issue i.*

Proof. The probability of seeing client with both internal and external providers in the same issues is $F_i(z_i^*) - (F_i(z_i^*))^2$ and the probability of seeing a client with in-house provider in issue i and external provider in issue j is $F_i(z_i^*) - F_i(z_i^*)F_j(z_j^*)$. As a consequence, it is more likely to see clients with internal and external providers for the same issue if $F_i(z_i^*) - (F_i(z_i^*))^2 \geq F_i(z_i^*) - F_i(z_i^*)F_j(z_j^*)$. This is reduced to $F_j(z_j^*) \geq F_i(z_i^*)$. In the case, that $z_i \sim Exp(\lambda_i)$, this condition is equal to $\frac{c_i}{c_j} \geq \frac{\lambda_i}{\lambda_j}$. ■

Lemma 1.12 *If* $z_l \sim Exp(\lambda_l)$ *for* $l = L, S$*,* $\lambda_L > \lambda_S \Leftrightarrow F_L(z_L^*) > F_S(z_S^*)$*.*

Proof. It follows directly from applying the exponential function to the inequality. ■

1.5.5 Tables of the Appendix

Sector	Coef	Mean	Rancoef	Rancmean
Trade	-3090.05	10130761	5	1
Manufacturing	-10028.9	7929012	2	2
Financial/Real Estate	-9735.77	5557105	3	3
Prof. Business	-12306.1	3601957	1	4
All Other	-6976.43	1055569	4	5
Edu-Health	2600.68	987600.3	6	6

Table 1.A.1: Results from the Displacement Workers' Literature.

Note: The table shows the estimated results from Section G of Couch and Placzek (2010) and average number of regulating words by industry. The second and fourth column represent the actual values of the coefficients and average number of words, respectively whereas the third and fifth column gives the relative ranking across variables. To improve the accuracy in the comparison of these two results I have used the same time-span that Couch and Placzek (2010) has used: 1993 to 2004.

	Lib Diff	Cons diff
Coeff	0.3516**	0.3567**
SE	0.0282	0.0258
Obs	39	39

Table 1.A.2: Results from the International Economics' Literature.

Note: The table shows the correlations between the number of regulating words and the value of inputs that are neither bought and sold on an exchange nor reference priced. For better comparability, with Nunn's data , I have used RegData at the 4-digit levels for the year 1997. Similar results emerge when I use other years. Nunn's data is originally provided at the 6-digit level. To convert it at the 4-digit level I have simply aggregated all the 6-digit codes by taking the average. Alternative aggregations give me similar results.

1.5.6 Figures

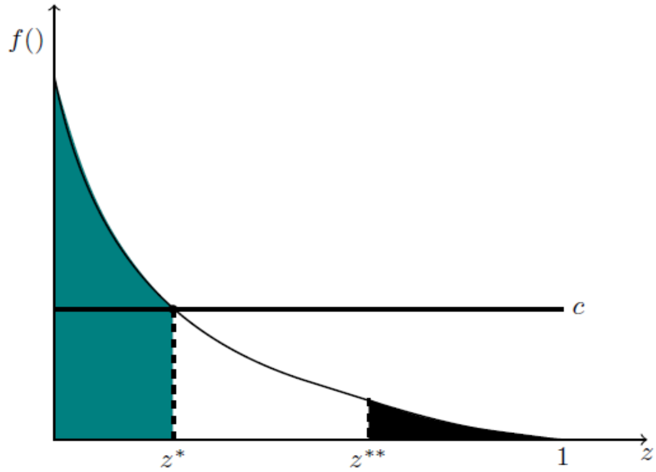

Figure 1.1: Solution with One Issue. Note: The left shaded (green) area represents the activities for which the marginal benefit of conducting the activities in-house is larger than or equal to the marginal cost. The first cut-off represents the firm-boundary, whereas the second cut-off denotes the market boundary. The intermediate (white) area represents the activities in which clients outsource the service, whereas the right hand (black) area denotes clients that do not use service providers.

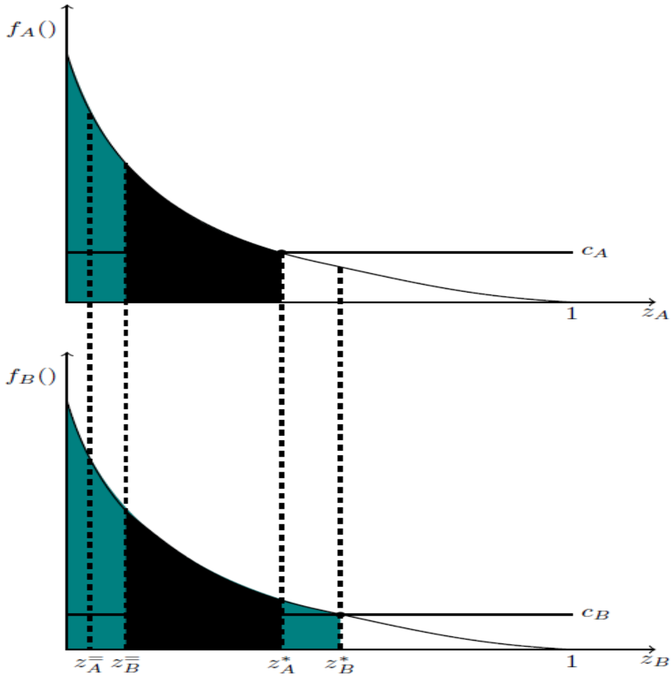

Figure 1.2: Solution with Two Issues and Generalists in the Internal Market. Note: The dark (black) area represents the levels of knowledge that internal generalists acquire.

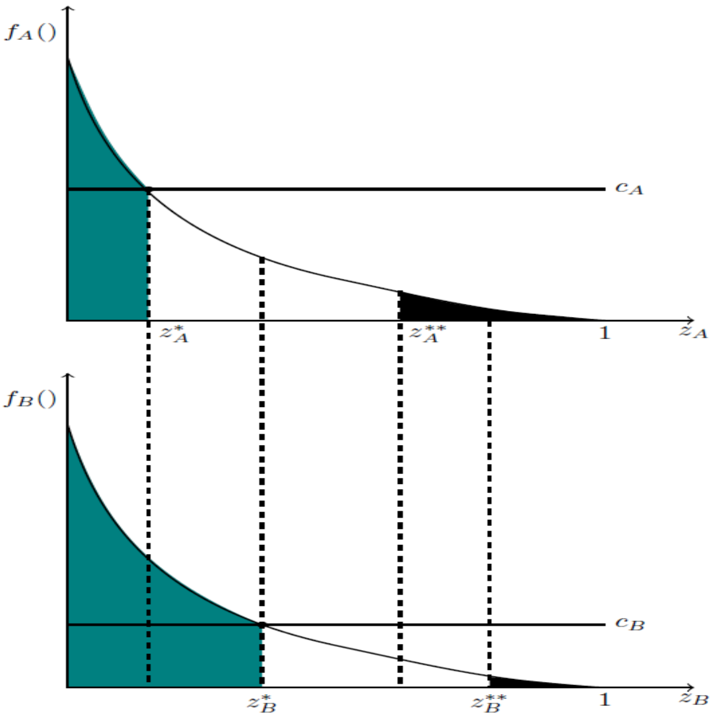

Figure 1.3: Solution with Two Issues. Note: This figure shows 5 out the 6 possibilities discussed in table 2. The case where firms use in-house staff for some issue and leave the problem unsolved for the other issue can be easily included for a low enough cost of acquiring knowledge in issue B.

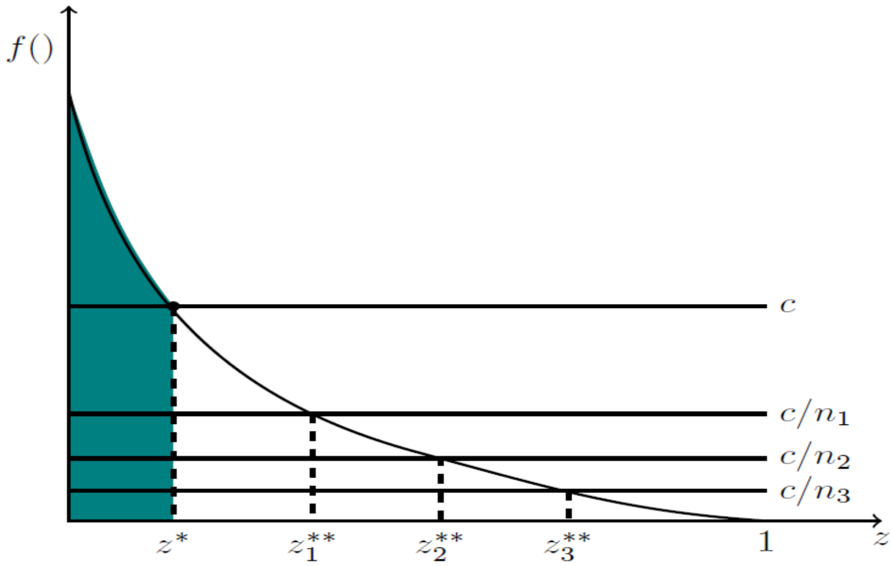

Figure 1.4: Solution to the Problem with Multiple Layers in the External Market: *The Level of Specialization of the External Providers is Limited by the Extent of the Market.*
Note: The figure represents the case with three layers in the external market. The first cut-off represents the firm boundary. Each cut-off in the external market denotes a different layer, implying a different level of knowledge and leverage.

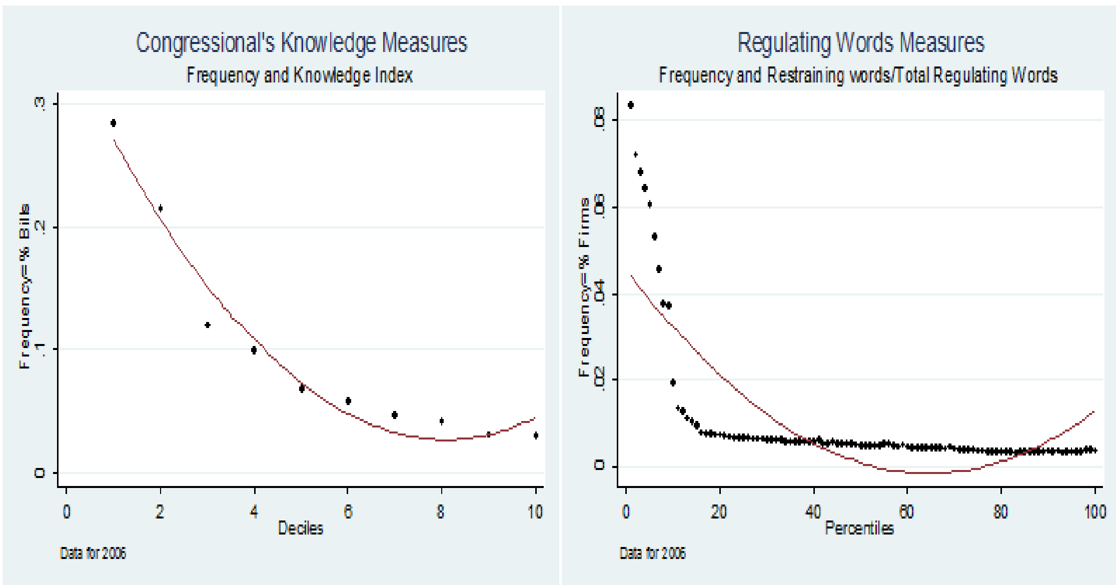

Figure 1.5: Knowledge Proxied by Congressional's and Regulating Words Measures.
Note: LHS: Frequency function of the committee-knowledge requirements' index. The x-axis shows the deciles of the index and the y-axis shows the normalized (fraction over the total) number of bills for each decile. RHS. Frequency function for the fraction of the number of restraining words over the total number of regulating words. The x-axis shows the percentiles of this fraction and the y-axis shows the normalized (fraction over the total) number of firms for each percentile. Similar results emerge when the y-axis is measured in terms of total number of employees or industries. Data for 2006 in both figures. Similar results emerge for different time periods.

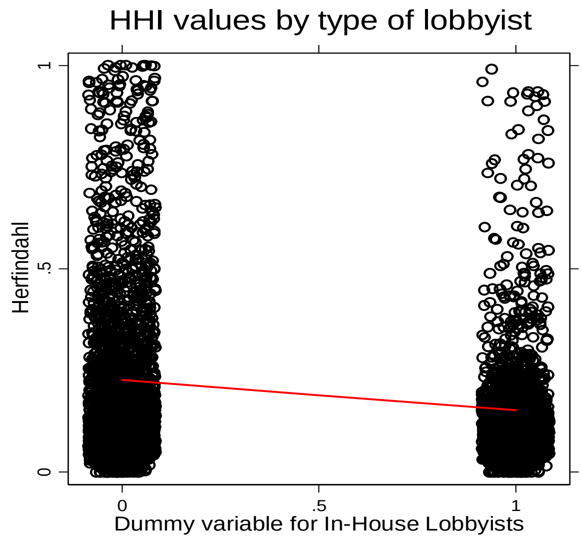

Figure 1.6: Patterns of HHI Across Markets. Note: The value 1 corresponds to internal lobbyists and 0 corresponds to external advocates. The line shows the linear prediction of the HHI on the dummy of internal lobbyists. The fitted regression is $HHI_i = 0.22 - 0.07 \cdot D_i + \varepsilon$ where $D_i = 1$ if lobbyist i is in-house and 0 otherwise. The standard errors are 0.003 and 0.0067 for the constant and slope, respectively. I restrict this analysis to lobbyists that advocate for at least 10 years. Different time restrictions give the same patterns. In this figure, an observation is a lobbyist.

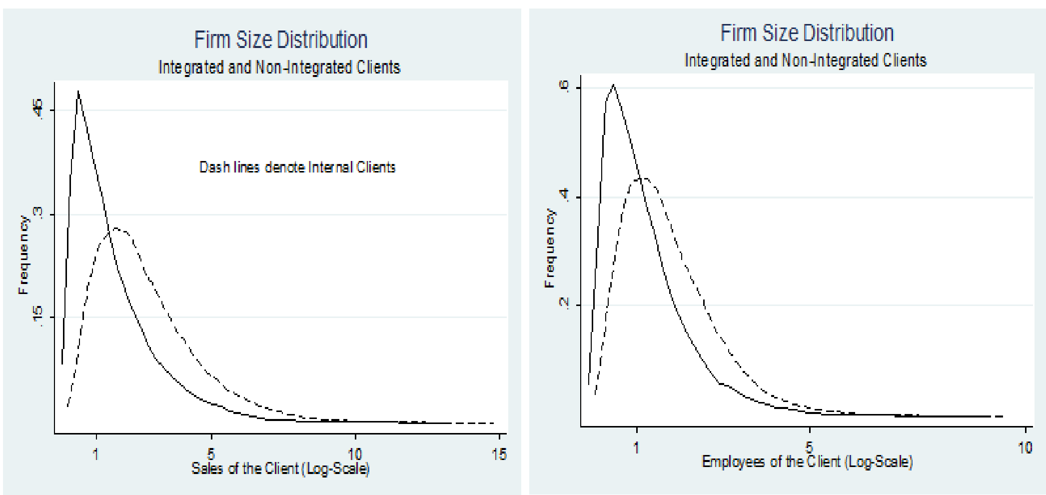

Figure 1.7: Relationship Between Firm Size and Status of Integration. Note: LHS: Firm size distribution by type of vertical integration relationship when the size is proxied by the log of the sales of the client. RHS: Firm size distribution by type of vertical integration relationship when the size is proxied by the log of the number of employees of the client.

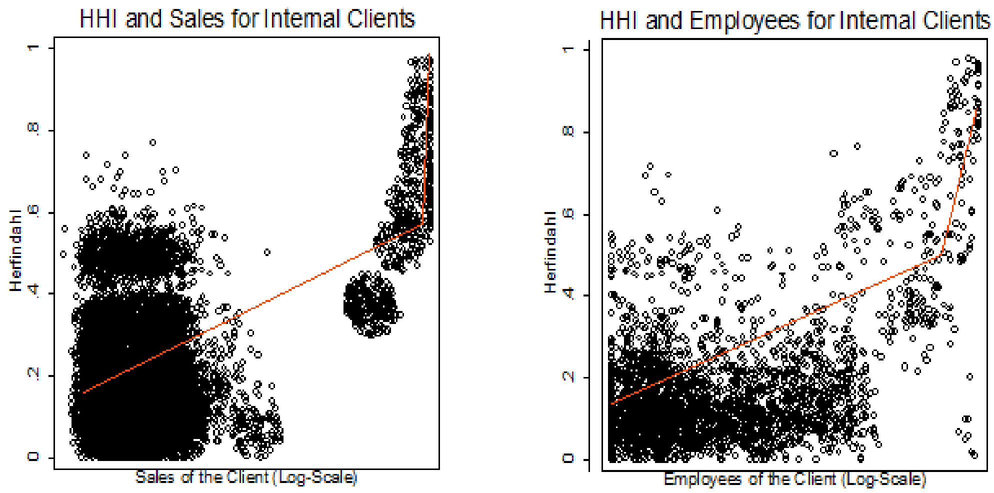

Figure 1.8: Relationship Between Firm Size of Internal Clients and HHI. Note: LHS: Scatter plot of the Herfindahl index and the log of the sales of internal clients. RHS: Scatter plot of the Herfindahl index and the log of the number of employees of internal clients. I restrict this figure to the sample of lobbyists with at least 5 years of lobbying activity. Similar patterns emerge for alternative time spans. An observation is a lobbyist.

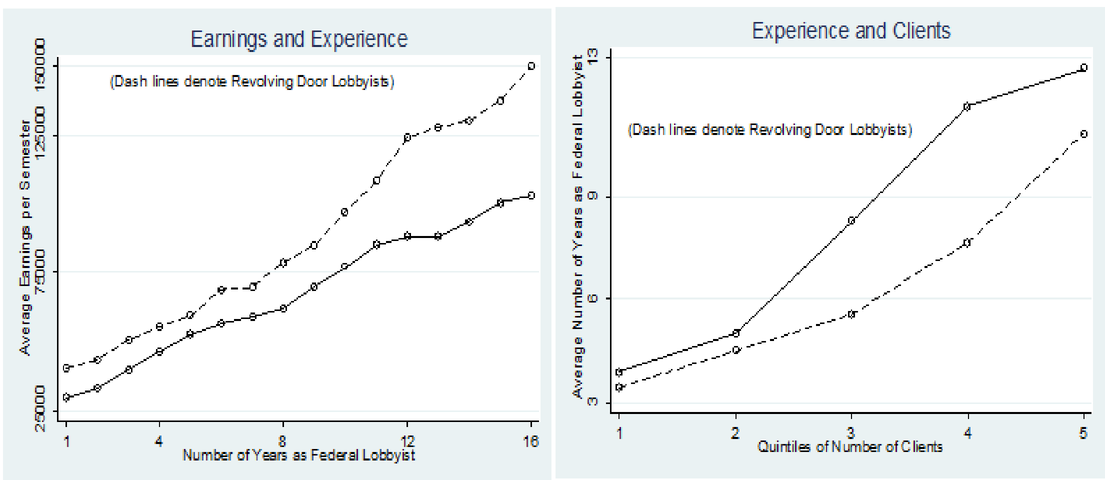

Figure 1.9: Earnings, Experience and Leverage. Note: LHS: Relationship between earnings (y-axis) and experience, as proxied by the number of years as a Federal Lobbyist (x-axis). RHS: Relationship between experience and number of clients. The x-axis reports the quantiles of the distribution of the number of clients. The choice of axis in these figures aim to follow closely my theoretical framework. On the left hand side, I represent the earnings as a function of the years of experience, that is earnings=cz. On the right hand side table, I denote the relationship between experience and leverage.

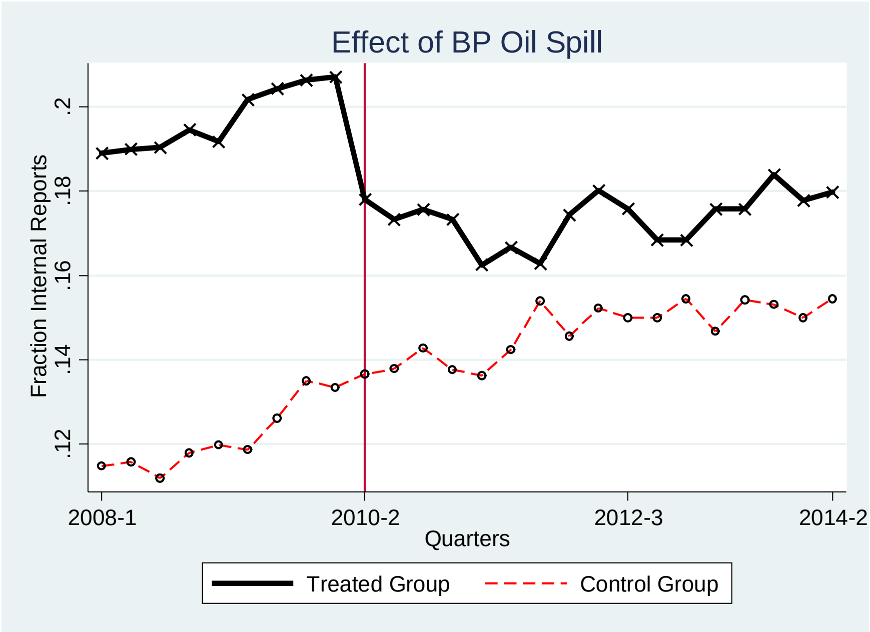

Figure 1.10: BP Oil Spill and Fraction of Internal Reports for Firms in the Oil Industry. Note. The treated group includes all the firms conducting lobbying activities that belong to 2007 NAICS codes 211111, excluding BP. The control group includes all the firms that belong to the Oil and Gas industry, excluding the codes above. An observation is a group- quarter.

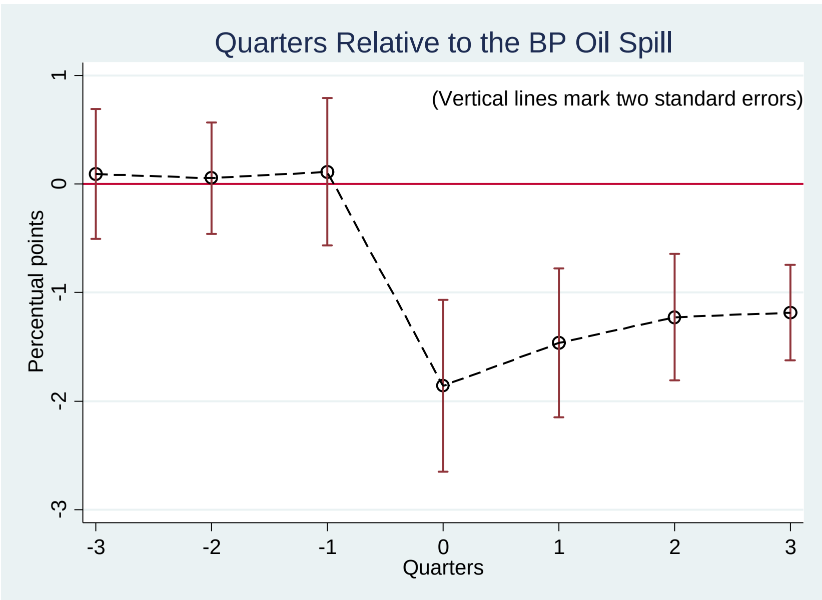

Figure 1.11: Leads and Lags BP Oil Spill Exercise. Note: This figure shows the coefficients of the interaction terms with three leads and three lags. As in the text, the estimation control by quarter and client fixed effects. Standard errors are clustered at the client level. The main take away from this figure is that the leads are very close to 0. That is, there is no evidence for anticipatory effects. This provides further evidence to the figure on the fraction of vertical integration reports for treated and control groups on the common trends assumption. The coefficients of the lags show that the effect tends to be attenuated over time. However, after three quarters the effect remains negative and significant.

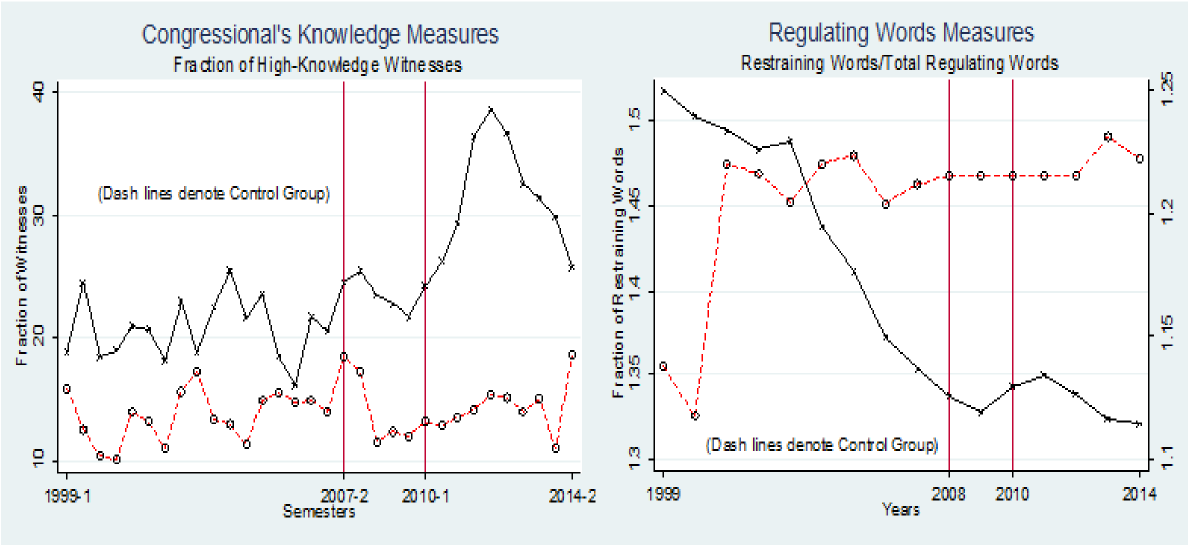

Figure 1.12: Effect of the BP Oil Spill on Knowledge Measures. Note: The LHS figure shows the knowledge measure proxied by the fraction of witnesses. To construct this measure, for each semester I considered all the bills that each of these firms were lobbying. To get a group-level measure, I have taken the weighted average (by number of bills being lobbied) of the fraction of high-knowledge witnesses across all the clients of the group. The vertical lines represent the second semester of 2007 and the first semester of 2010. The RHS figure shows the knowledge measure proxied by the Regulation words variable. The vertical lines represent years 2008 and 2010. An observation in both figures is a group-semester combination. Dashed lines represent the control group.

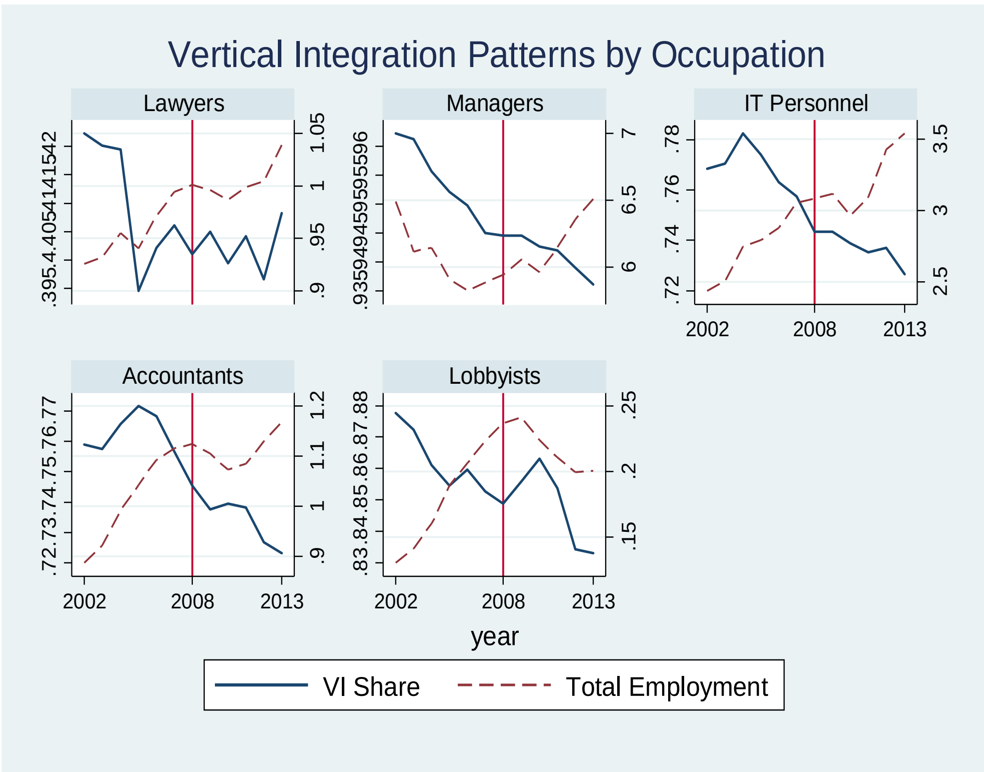

Figure 1.13: Time Series on Total Employment and Fraction of Integrated Clients for Five Knowledge-Intensive Occupations. Note. An observation is a year.

Chapter 2

Technological Change and the Boundaries of the Firm:The Case of Lobbyists[103]

Abstract

This chapter studies the main effects of technological change in vertical integration with knowledge workers. Theoretically, I study two different channels in which technology can modify integration patterns. I show that technologies that decrease the cost of acquiring issue-specific skills increase both the level of vertical integration and the leverage of external providers, however, decrease the earnings inequality of the economy. On the other hand, technologies that decrease the cost of communicating firm-specific knowledge, decrease the integration in the economy but increase both the span of external providers and earnings inequality. I confirm the empirical validity of the prediction of the cost of acquiring skills using data from a knowledge-intensive industry: US Federal Lobbying. To do that, I use the most important policy change in the lobbying industry: The 2007 Open Government Act. I use structural techniques to back out unknown parameters and I show that the introduction of the Act, decreased the cost of acquiring issue-specific skills. Confirming the validity of the theoretical predictions, I show that the Act increased significantly both the use of in-house lobbyists and the number of clients that external lobbyists work for, as well as decreasing sharply the earnings inequality of the economy.

[103] I am deeply indebted for the precise and insightful comments made by Jordi Blanes i Vidal, Luis Garicano, Gilat Levy, Jim Snyder, Jr., John Sutton, Catherine Thomas and Glen Weyl. This chapter is an extended version of the second part of my job market paper. I acknowledge the comments and suggestions made by participants at Work in Progress seminars at LSE, UChicago and Stanford University, 14th IIOC, 2016 TADC, 15th IOEA, 20th Annual Conference of the SIOE.

2.1 Introduction

Despite great and long standing interest in the main determinants of vertical integration, there is no consensus on the reasons for vertical integration in general, and the relationship between technological change and the integration decision, in particular. What are the main consequences of technological change in vertical integration? Would firms be more keen to hire someone in-house or to outsource the service after facing a technological shock? This chapter uses the theoretical model developed in Chapter 1 to, in the light of the theoretical predictions, understand the main effects of technological change on vertical integration.

In Chapter 1, I develop a model that focuses on the acquisition and communication of skills. Focusing on these two channels allows me to differentiate the effects of two types of technologies: skill acquisition and firm-specific communication. Technologies that decrease the cost of acquiring skills will empower in-house workers whereas technologies that facilitate the communication of firm-specific skills will stimulate outsourcing. Intuitively, the first type of technology makes learning cheaper and decreases the need of firms to rely on external service providers, as in-house staff can solve harder problems. On the other hand, cheaper firm-specific communication technologies decreases the comparative value of in-house staff and, therefore, promotes outsourcing.

Technological change has huge impacts in the knowledge economy. For instance, a decrease in the cost of acquiring skills will increase the leverage of external service providers. This comes from the fact that the skills acquired for both in-house and external staff have changed. As a consequence, firms with in-house staff are less likely to ask for help in the external market, so the time constraint of external staff is relaxed. As consultants have also become smarter they can solve a larger set of problems, and therefore they will work for a larger number of clients. On the other hand, a decrease in the cost of trasmitting firm-specific skills relaxes the time constraint of external staff, as they would need less time learning the firm-specific component. Therefore, they can solve problems for a larger set of clients. So, the leverage of external staff is both decreasing in the costs of acquiring issue-specific skills and transmitting firm-specific knowledge.

I argue in this chapter that technology not only has a strong effect on the integration decision but also on important features in the economy, such as earnings inequality. A simple way to measure this variable is with the ratio of the earnings of external staff over in-house providers. As earnings are proportional to the issue-specific skills acquired, as I have shown in Chapter 1, a technological shock that changes the skills acquired also modifies earnings in the economy. Concretely, I show that a decrease in the cost of acquiring issue-specific skills decreases the earnings inequality measure, whereas a decrease in the firm-specific transmission cost has the opposite effect. This result is very intuitive. A decrease in the acquisition cost empowers in-house staff, which increases their earnings proportionally more than the increase in the earnings of external staff. This empowerment is translated, then, into a decrease in the earnings inequality measure. A decrease in the cost of communicating firm-specific skills has the opposite effect. A decrease of this type, empowers external consultants, which make them earn more. Therefore, earnings inequality increases.

Empirical exogenous evidence on these effects is inexistent. To contribute to this debate, I focus on a policy change that decreased the cost of acquiring skills for US federal lobbyists. Empirically, I exploit the most important lobbying regulatory change in the last 20 years, the Open Government Act (henceforth OGA) of 2007, as a quasi-experiment that improved the technological access to advocacy information. I show that, consistent with the intuition above, the technological shock decreased the cost of acquiring issue-specific skills.

Before the Act, lobbyists have to report their advocacy activities by hand and submit it to the Congress. Since then, the lobbyists made these reports electronically. In addition, Sec. 208 of the Act: "*Requires the Secretary of the Senate and the Clerk of the House to: ... make lobbying activity reports available for public inspection over the Internet within 48 hours after such report is filed*". For a lobbyist, an important source of information to conduct her activities is the activities that other lobbyists conduct. For instance, with that information lobbyists can set their advocacy strategies as they understand who is on each side of the debate and how much they care about specific issues. The Act facilitated the access to this information as previously, lobbyists had to conduct a manual search on a specific geographic location. Since the Act, lobbyists could access all lobbyists' reports via an organized database from any location.

In the body of this chapter, I argue that the main effect of the change in the reporting technology was to facilitate the acquisition of information for advocacy activities, reducing the cost of acquiring skills to conduct these activities. Clearly, the main challenge that this chapter faces, is to show that there was an actual decrease in the cost of acquiring skills and that this had an effect on integration patterns. To do that, I use equilibrium conditions from the model developed in Chapter 1. Using these equations and aggregate measures from the lobbying market, I back out some unobservables and I show that under my more conservative estimations, the cost of acquiring advocacy skills decreased in about 20%.

This Act also included other regulatory changes. Importantly, the Act closed some of the channels that lobbyists used to access politicians. Since policy makers acquire knowledge from lobbyists, more-restricted access to them implies an increased need to gather knowledge from other sources. I provide empirical evidence that congressional committees started using high-knowledge witnesses more heavily, and, as a result, the type of problems faced by the firms changed. I decompose the effect of the Act on the economy's vertical integration, showing that around half of the change is explained by a change in the distribution of problems (*difficulty effect*), and about half by the cost of acquiring knowledge (*technological effect*). Recovering the parameters allows me to conduct additional empirical exercises. For instance, I show that if the economy had faced only the *technological effect*, the fraction of firms with in-house lobbyists would have increased by 150%.

Related Literature

There are two branches of literature that are relevant to this chapter. The first branch links characteristics of technological change with integration decisions. Acemoglu et al. (2010) show that technology intensity of downstream industries is negatively correlated to the probability of outsourcing whereas the technological intensity of upstream industries is positively correlated. Bartel et al. (2012) show that firms producing products that require inputs that face rapid technological change will tend to outsource more. Although both of these papers link technology with integration, none of them discuss how information and communication technologies (henceforth ICT) affect skill acquisition or its implications to the vertical integration decision. A second branch of the literature focuses on the effect of technology on internal firm organization. A first set of papers from this branch focuses on this relationship but neglects the main differences on the effects of information versus communication technologies.[104] On the other hand, Bloom et al. (2014) exploit the disaggregation of types of ICT and empirically studies its relationship with firm organization but neglects its distinctive effect on vertical integration. Although this branch of the literature seeks to analyze the effect of different types of technology on internal hierarchies of the firm, no paper has empirically analyzed the effect of ICT on vertical integration patterns.

This chapter is divided into five sections. Section 2.2 presents the theoretical framework. To keep independence across chapters, here I summarize the main setting fully developed in Chapter 1. In this chapter, I focus on presenting some results that will guide the empirical section. In Section 2.3, I briefly present the data used and the institutional context. Section 2.4 starts by explaining the OGA and providing empirical evidence on its consequences. Then, I show the results of the structural estimations. I end this section by discussing alternative approaches to explain the effect of the OGA on integration patterns. Finally, in Section 2.5, I conclude with a short discussion summarizing the main results from the chapter and proposing further developments.

[104] See for example, Acemoglu et al. (2007) and Caroli and Van Reenen (2001).

2.2 Theoretical Framework

2.2.1 Preliminaries

I consider an economy with a large number $M < \infty$ of ex-ante homogeneous clients (firms) and an infinite set N, of ex ante homogeneous service providers. Clients exogenously receive one problem per unit of time spent in production. Service providers, not clients, solve problems.

Demand. As in Garicano (2000), problems differ by the level of issue-specific knowledge requirements (i.e., difficulty of the problem). I denote this level by $Z \in [0, 1]$. The problems are ordered by increasing level of difficulty. The random variable Z is independent and identically distributed according to a continuous cumulative distribution function F with $F'' < 0$. Solutions to problems in both issues are equally valuable to these clients. Clients' payoff function is production minus labor costs, and the normalized value of production is 1 when a service provider solves a problem and 0 otherwise.

Supply. Solving problems requires knowledge. All service providers must learn the easiest (most common) problems before learning the harder (less common) ones. Providers are characterized by a variable $z \in [0, 1]$. Service providers with issue-specific knowledge z solve any problem if the difficulty of the problem lies between 0 and z. Service providers increase this knowledge at a cost proportional to the size of the interval of knowledge. That is, learning how to solve problems in the interval $[0, z]$ costs cz, where c is the constant per-period unit cost of acquiring knowledge. Without loss of generality, I assume that the outside option of not working in any market is 0.

Markets. There are two markets in this model- internal and external. First, there is an internal market, which is characterized by a one-to-one relationship between the client and the service provider. Second, there is an external market, in which each service provider works for $n \in \mathbb{R}_+$ clients (i.e., leverage), which is an endogenous variable to the problem.

Communication Cost-Firm-Specific Knowledge. In addition to the knowledge required to solve problems, clients need firm-specific knowledge in order to produce. In-house and external service providers differ by the levels of this type of knowledge. I capture this idea with a communication cost in the external market. This cost, denoted as $h \in (0, 1)$ is the time that external providers spend on each client in addition to the time that it takes to solve problems.

Time constraint. Each service provider has a time constraint with total labor supply endowment normalized to 1. As in Garicano (2000), the constraint implies that the expected time for solving problems has to be equal to the total labor endowment. That is, $1 \geq \Pr(problem) \cdot n \cdot [1 + \mathbf{1}_{external} \cdot (h - 1)]$. Let $\Pr(problem)$ be the probability that service providers face problems and n the leverage (number of clients) of the service provider. The leverage is 1 in the internal market and is endogenously determined in the external market. Finally, $\mathbf{1}_{external}$ is an indicator function equal to 1 in the external market and 0 otherwise. The burden of this cost falls fully on the receiver.

Wages. Service providers receive a constant per-period unit wage compensation w. The total wage compensation is wz if the service provider has knowledge z. The wages are endogenous to the problem.

Summing up, the total number of clients and set of service providers (M, N), the distribution of problems (F), the cost of acquiring issue-specific knowledge (c), and the firm-specific knowledge of the economy (h) are exogenous parameters in the economy. The endogenous variables are the vector of wages (w), allocation of clients to each market and levels of acquired knowledge (z) and leverage (n) for each service provider.

Timing. First, service providers choose the breadth (market in which they want to work) and depth of the level of issue-specific knowledge. Then, the problems of the clients are realized. Finally, clients are allocated to markets; the markets clear and production takes place.

In the following section, I characterize both firm and market boundaries and I present the main results concerning the effects of technological shocks on leverage, vertical integration and earnings inequality.

2.2.2 Main Results

The problem is solved recursively: first, I characterize the in-house solution and then do the same for the external market. For the internal market, the solution of this problem is characterized by the following first-order condition:

$$f(z^*) = c$$

The level z^* represents the issue-specific skills level at which the marginal benefit of having someone in-house is equal to the cost of acquiring knowledge. The marginal benefit of this problem represents the increase in the probability that a problem will be solved. The level z^* represents the firm boundary.

The objective function in the external market is $[F(z) - F(z^*)] - czh(1 - F(f^{-1}(c)))$. The solution of this problem is characterized by the following first-order condition:

$$f(z^{**}) = ch(1 - F(f^{-1}(c)))$$

The level z^{**} represents the market boundary. The external market allows service providers to acquire higher levels of knowledge by sharing the costs of acquiring issue-specific knowledge with several clients. Since in this market, for each problem, the marginal benefit (given by the frequency) of bringing someone in-house is lower than the marginal cost, no single client can hire an internal provider.

Equilibrium and Leverage in the Economy

The equilibrium conditions in this economy are for the internal market:

$$M \int_{0}^{f^{-1}(c)} f(x)dx = N_I$$

This condition establishes that the number of providers that go to the internal market is equal to the number of internal market demanders. And for the external market:

$$\frac{M}{n} \int_{f^{-1}(c)}^{f^{-1}\left(ch(1-F(f^{-1}(c)))\right)} f(x)dx = N_E$$

That is, the total number of service providers needed equals the total number of clients demanding external services divided by the number of clients that each service provider works for. Finally, as all the service providers should get the same ex-ante payoff:

$$0 = \left(w^I - c\right) z^* = \left(nw^E - c\right) z^{**}$$

That is, the equilibrium wages are given by $w^E = \frac{c}{n} = \frac{w^I}{n}$ for $z^*, z^{**} > 0$. To characterize the optimal organization of the economy, I solve it as follows: first, internal and external market-clearing conditions give me the number of internal and external service providers. Note that, as $(1 - F(f^{-1}(c)))\,h = \frac{1}{n}$, the external providers' time constraint uniquely determines the number of clients for whom each external provider works. Finally, wages are given by the ex-ante equality payoff condition.

Proposition 2.1 *The leverage of the external service providers n, is decreasing in both the cost of acquiring issue-specific skills c and the firm-specific communication cost h.*

This last result is a direct implication of the time constraint. Intuitively, the part related to the cost of acquiring skills c comes from two facts. First, in-house staff can acquire a larger set of issue-specific skills which decrease the demand for external providers. This implies that firms with in-house staff will be less likely to go to the external market to get their problems solved as their staff would be able to solve a larger set of problems. Second, external staff acquire more issue-specific skills, which means that they can solve more difficult problems. These two results combined, imply that the external providers are not only smarter but also that they have a less binding time constraint, which they adjust, in equilibrium, by working for more clients. On the other hand, the impact of the communication cost h, it is straightforward. If it is more costly to transmit firm-specific skills to external service providers, they will have less time to solve problems and therefore, they will adjust their time constraint by working for fewer clients.

Vertical Integration

The fraction of vertically integrated clients is given by:

$$VI = \frac{F(f^{-1}(c))}{F(f^{-1}\left(\frac{c}{n}\right))} = \frac{F(f^{-1}(c))}{F(f^{-1}((1 - F(f^{-1}(c)))\, hc))}$$

Proposition 2.2 *The fraction of vertically integrated clients is increasing in the communication cost h and decreasing in the acquisition cost c.*

These two results are theoretically establishing the intuition that was developed previously in Chapter 1. Technologies that decrease the cost of acquiring skills will empower in-house workers whereas technologies that facilitate the communication of the firm-specific skills will stimulate outsourcing. Intuitively, the first type of technology make learning cheaper and decrease the need of the firms to rely on external service providers, as in-house staff can solve harder problems. On the other hand, cheaper, firm-specific communication technologies decrease the comparative value of in-house staff and therefore, promote outsourcing.

Earnings and Knowledge Inequality

I define knowledge inequality of the economy, as the ratio of the maximum level of issue-specific skills achieved in the external market over the maximum level of this type of skills for in-house staff. Mathematically, this is represented by the following ratio:

$$\frac{z^{**}}{z^{*}} = \frac{f^{-1}(ch(1 - F(f^{-1}(c))))}{f^{-1}(c)}$$

Theoretically, the total earnings for internal and external providers are given by $w^I z^*$ and $nw^E z^{**}$, respectively, where $w^I = nw^E = c$. This implies that the ratio $\frac{z^{**}}{z^*}$ is a representation of both the difference in the issue-specific skills and earnings levels between the market and firm boundaries (i.e., between external and in-house service providers). In other words, this ratio represents the difference between external and in-house service providers.

Proposition 2.3 *The earnings inequality of the economy $\frac{z^{**}}{z^*}$, is decreasing in the communication cost h and increasing in the acquisition cost c.*

This result is a direct implication from the effects of the technology in both firm and market boundaries. First, an increase in the communication cost makes external service providers acquire a lower level of issue-specific skills. As this technology does not affect in-house staff, the earnings inequality measure decreases. On the other hand, a change in acquisition costs affects more the in-house than the external provider. Therefore a decrease in acquisition costs decreases knowledge inequality. Notice, that the knowledge inequality measure is proportional to the inverse of the fraction of vertically integrated clients. Therefore, the effect of a technological shock on inequality is opposite to its effect on integration.

2.3 Data and Institutional Context

The Lobbying Disclosure Act of 1995 (henceforth LDA) requires lobbyists to register and to report on their lobbying activities to the Senate Office of Public Records (henceforth SOPR). According to the Act, lobbying activity is defined as contacts with officials, including background work performed to support these contacts. Two types of registrants are required to report under the LDA: external and internal lobbyists. External lobbyists, who work for lobbying firms, take on lobbying responsibilities for a number of different clients and, under the LDA, they are required to file a separate report for each of their clients. Internal lobbyists are *self-filing* organizations that conduct in-house lobbying activities. Both types of registrants are required to report good-faith information every three months. Up until the end of 2007, they were required to report these estimates biannually.[105]

The starting unit of observation is a lobbying report. Each SOPR report not only contains the name of the client and individual lobbyists, but also specifies the House(s) of Congress and federal agencies contacted, as well as the bills in which the client was interested. Clients can have more than one report in a given period, as they can use both internal lobbyists and one or more groups of external lobbyists. The lobbying reports dataset starts at the first semester of 1999 and finishes with the second semester of 2014. It contains 44,039 clients and 56,759 lobbyists.

2.3.0.1 Bills, Committees and RegData

Bills are legislative proposals that can be introduced at any time while the Congress is in session by any member of either house. After introduction, the bill is referred to the appropriate committee or committees, based on the committees' jurisdiction, which is defined by congressional rules. The objective of the committees is to study bills and consider whether or not to send them for further action. The committees are divided into sub-committees that have a narrower jurisdiction in the topics. The initial stage of this study process consists of public hearings at which committee or sub-committee members invite witnesses with the purpose of gathering relevant information. Witnesses are either specialists on the topic or people affected by the matter. They represent different views on the topic and can have different backgrounds, such as government, academia or business. Committees that manage more technical subjects require witnesses with more experience or higher education

[105] For the sake of a better comparison, for most of the estimations, I will focus on semester level time variation.

levels.

I have web-scraped the name and title or occupation of all the witnesses in all the reported congressional hearings since 1999. I have classified these occupations into two groups: high and low levels of knowledge requirements. Titles that include PhD, professor or senior manager are classified as high, whereas all other occupations are classified as low. In order to capture a comprehensive measure of knowledge requirements, I create a difficulty index at the committee-semester level using principal components analysis. The index uses as an input the number of sub-committees and the fraction of witnesses with high-knowledge occupations over the total number of witnesses. Proxying knowledge requirements using the difficulty index, or simply the fraction of high-knowledge witnesses, gives the same qualitative results.

I propose an additional measure of knowledge intensity using two variables based on Al-Ubaydli and McLaughlin's (2015) RegData 2.2. This database uses text analysis and machine learning algorithms to create regulation intensity measures at the industry-year level based on the Code of Federal Regulations (henceforth CFR). For each industry-year combination, I use two variables from this database: 1) the number of words related to the regulation of the industry; and 2) the number of times that any of the following five strings appears among the regulating words: *shall, must, may not, prohibited*, and *required*.

Given the lack of a comprehensive measure at the firm level, I use the total number of regulating words as a proxy for industry-specific knowledge. The intuition is that when clients hire external service providers, they explain the regulations in their industry. The larger the number of regulating words, the more costly it is to explain it to the external service providers and, therefore, the larger the communication costs. I use the fraction of the number of restriction words over the total number of regulating words to proxy for the difficulty of the lobbying activities. Clients belonging to industries with a larger fraction of regulating words will face a tougher regulatory environment. In order to match the RegData information with the lobbying reports, I conducted extensive data work to detect the industry of the clients using ORBIS, COMPUSTAT and other web sources such as the client's webpage.

2.3.1 OGA

This policy change, signed in 2007 and taking effect in 2008, is the most relevant lobbying industry policy shift of the last two decades.[106] Remarkably, the Act did not include a single regulation discriminating between internal and external

[106]For the period 1999 to 2014, on average, only 3.5% of the bills introduced became law, and among the bills that were approved by both chambers, 50% became law. The OGA was passed by both chambers on August 2, 2007. As a great percentage of the bills at this stage become law, it is intuitive to see OGA's possible effects since the second semester of 2007. More details about the OGA can be found in the first part of the Appendix of this chapter.

advocates. As cited in the OGA, the aim was "to provide greater transparency in the legislative process" on two fronts: disclosure and ethics. Disclosure under the OGA is more strict, as lobbyists must now report their activities electronically and more frequently. Before the Act, they had to make reports every semester and since then, the reports must be made every quarter. Before the Act, lobbyists had to make these reports by hand and submit it to the Congress. Since then, the lobbyists made these reports electronically. In particular, *Sec. 208* of the Act*: "Requires the Secretary of the Senate and the Clerk of the House to: ... make lobbying activity reports available for public inspection over the Internet within 48 hours after such report is filed"*. The main effect of the change in the reporting technology was to facilitate the acquisition of information for advocacy activities.

The OGA also closed some of the channels that lobbyists used to access politicians. One example is given by *subtitle C*, section *533* of the OGA. In this rule, the Act revokes floor privileges and the use of the Members' exercise facilities and parking spaces for some former high-ranking politicians who are registered lobbyists.[107] Another example of a closed channel is given by *subtitle E*, section *552* of the OGA. In this rule, the Act prohibits Senators' staff from having contact with the member's spouse if this spouse is a registered lobbyist. According to a survey conducted among a large sample of lobbyists, 83% of these advocates think that the Act made lobbying activities more difficult.[108] Unfortunately, the survey did not ask lobbyists why or how the difficulty of their activities increased.

2.3.1.1 Consequences and Interpretation of the OGA

Although the OGA requires quarterly reports, in Figure 2.1, I group all the information in terms of semesters for better comparability with the data before 2008. For each semester, I classify clients into two types of contracts: clients that use in-house lobbyists and clients that use only external lobbyists. Figure 2.1 shows the time-series patterns of the fraction of clients using in-house lobbyists.[109] On average, clients with only external lobbyists account for around 72% of the total number of clients in a given semester, while clients using in-house lobbyists represent 28% of the total. The figure presents a declining trend before the OGA. One way to understand this pattern is based on Stigler (1951): in the early phases of the lobbying industry, clients had to be vertically integrated because there were no markets for lobbyists. As the lobbying industry became larger, work that had formerly been done by in-house lobbyists started to be supplied by external lob-

[107] These high-ranking occupations are Senators, former Secretaries of the Senate, former Sergeants at Arms of the Senate and former Speakers of the House.

[108] For more information about the survey, see www.lobbyists.info/HLOGA_Five_Year_Survey.

[109] In this graph, if a client uses both types of lobbyists, it will be considered a client using in-house lobbyists. Similar patterns emerge when I consider the fraction of clients who use **only** internal lobbyists over the total number of lobbyists.

byists.[110] The main takeaway from this figure is the change in the pattern around 2007. After the OGA, the fraction of in-house clients increased from 25% to almost 34%.

Change in the Density of Problems Politicians need knowledge to make informed decisions.[111] They have at least two different sources to acquire it: lobbyists and congressional hearings. If lobbyists have more restricted access to politicians, they need to rely more on other sources to acquire this knowledge. Figure 2.2 shows the density functions before and after the OGA using the congressional knowledge measures introduced in Section 2.3.0.1. For the before–OGA period, I consider three semesters before the OGA was signed, and for the after-OGA period, three semesters since the second semester of 2007. Figure 2.2 provides evidence of a change in the density of problems that the clients face. Bills with low-knowledge requirements become less common, whereas bills with high-knowledge requirements become more common. These differences are confirmed by Kolmogorov-Smirnov tests of equality of distributions.[112] Similar conclusions emerge when I estimate these density functions for different periods around the enactment of the OGA.[113]

To provide additional empirical evidence on the increase in difficulty around the OGA, Figure 2.3 shows the estimated residuals obtained from running the ratio of restraining words over the total number of words on the years considered.[114] The figure shows a remarkable similarity with Figure 2.1. There is a decreasing trend before 2007 and, since then, an increasing trend. For the rest of this section, I focus on the congressional hearings knowledge measures, as I have variation at the semester and not yearly (as with the regulating words) level. However, all the results I present here are robust to the election of the proxies for skills. [115]

[110] Graphical evidence on the increase in the *market* (i.e. number of firms and lobbyists) can be seen in Figure 2.A.1.

[111] See, for instance, Grossman and Helpman (2002b) for an extensive review of the literature in which lobbying activity is seen as a relevant information transmission process from one better-informed party (lobbyists or clients) to less-informed politicians. Bertrand et al (2014) give an empirical argument in favor of this view. They show that politicians listen to lobbyists with opposite political views when they are considered issue-experts.

[112] I conduct this test using all the congressional hearings knowledge measures. The corrected p-value is 0,001, which means that the null hypothesis of equal distributions can be rejected.

[113] In particular, I try four other combinations: 1. Before: 2005-1/2007-1 and After: 2007-2/2009-1; 2. Before: 2006-2/2007-2 and After: 2008-1/2009-1; 3. Before: 2006-1/2007-2 and After: 2008-1/2009-2 and 4. Before: 2005-2/2006-2 and After: 2008-2/2009-2. For all of these combinations, I obtains qualitatively similar results.

[114] I run $z_t = \beta_0 + \beta_1 t + \varepsilon$, where z_t is the difficulty measure at the two-digits levels proxied with the RegData. Similar patterns emerge for other levels of aggregation. I consider the period 1999 to 2014. The graph simply shows the predicted residuals $\widehat{\varepsilon}$ on the years considered.

[115] In addition, Section 2.6.4.1 provides empirical evidence on the differences of the characteristics of the lobbyists entering the market. I observe the lobbyists that started working right after the OGA have more previous work experience in the government. This can be interpreted as suggestive evidence that the increase in the difficulty of the problems is accompanied by an upgrading on the types of lobbyists' skills. This skill-updgrading is significant and small, and therefore cannot provide a full explanation on the observed decrease in the acquisition cost. In Section 2.6.4.2, I also include robust empirical evidence that shows that the difficulty of the problems changed not only for a specific set of issues but for all of them.

The empirical shift in the density function presented above can be seen in my model as follows. Denote B, for the situation before the OGA and A for the situation after the shock. Figure 2.4 shows two examples in which the easy problems (to the left of the intersection of the curves f_A and f_B) become less common and the more difficult problems (to the right of the intersection of the curves f_A and f_B) become more common. Both graphs include the case in which the knowledge-acquiring cost technology is at the level after the Act. There are two relevant situations. First, if the densities cut above the cost, the firm boundary shifts to the right. However, if these densities cut below the cost, the firm boundary moves to the left.

Change in the Cost: Technological Shock Figure 2.5 represents a decrease in the cost of acquiring knowledge. If this cost decreases, the firm boundary moves to the right while if the cost of acquiring knowledge increases, the firm boundary shifts to the left. The change in the observed cost can be due to a change in the technology that facilitates the acquisition of knowledge. As mentioned above, one of the main changes brought by the Act, was to make all the lobbying reports available on the internet within two days of the reporting activity. This change decreased the cost of accessing relevant information. The interpretation for this change is simple. As the technology complements in-house lobbyists and external service providers are time constrained, the technological shock shifted the decision making to be made within the firm as opposed to the external market. This is an empowering effect of technologies that facilitate the acquisition of knowledge.

So at this stage we know that there was an increase in the integration patterns and a change in the density functions. We also know that there seems to be a change in the cost technology. However, it is not clear to what extent each of these components, density and cost of acquiring skills changes, help us to understand the total change in integration patterns. Clearly, the path to follow is to have an estimate of the cost technology. In order to recover these costs, I will use the set of three equations characterizing the equilibrium in the economy. For the rest of the section, I will assume the knowledge requirements of the problems are distributed exponentially. Below, I will discuss on the robustness of this assumption.

2.4 The Effect of the Open Government Act

In this section, I provide maximum likelihood estimations results and using my theoretical framework, I explain how and why the OGA affected the vertical integration patterns of the economy. Then, I present the main results of the counterfactual exercises and a discussion of the robustness in my estimations. I finish this section by presenting a discussion on alternative explanations.

2.4.1 Estimation

2.4.1.1 Setting the Problem

The main results of the structural estimations will be for the case in which the CDF is the exponential function $F(z) = 1 - e^{-\lambda z}$. Here I present the main results for this case but in section 2.4.1.4, I discuss the general applicability of this assumption. In the case of only one issue and $z_B \sim Exp(\lambda_B)$, the solution of the model is as follows. The firm boundary is given by $z_B^* = -\frac{1}{\lambda_B}\ln(\frac{c_B}{\lambda_B})$ while the market boundary is given by $z_B^{**} = -\frac{1}{\lambda_B}\ln(\frac{c_B^2 h_B}{\lambda_B^2})$.[116] The earnings for in-house lobbyists are $c_B z_B^*$ and for external lobbyists $n_B\left(\frac{c_B}{n_B}\right)z_B^{**}$. Clearly, external lobbyists earn more as they acquire a larger amount of issue-specific skills in equilibrium.

Remark. As the firm boundary (i.e. z_B^*) is a decreasing function in the cost of acquiring knowledge c_B (i.e. $\frac{\partial z_B^*}{\partial c_B} < 0$), the total number of clients using in-house lobbyists (i.e. $M_B\frac{(\lambda_B - c_B)}{\lambda_B}$) increases when the cost c_B decreases.

Finally, the fraction of vertically integrated clients is given by:

$$VI_B = \frac{n_B(\lambda_B - c_B)}{n_B\lambda_B - c_B}$$

This fraction is increasing (decreasing) in the communication costs h_B (acquisition cost c_B) and it can be increasing or decreasing in λ_B.[117] Finally, the total number of internal clients I_B, total number of external clients E_B, and the time constraint of the external lobbyists, characterize the equilibrium in the economy:

$$M_B \int_0^{f_B^{-1}(c_B)} f_B(x)dx = I_B = \frac{M_B}{\lambda_B}(\lambda_B - c_B)$$

$$M_B \int_{f_B^{-1}(c_B)}^{f_B^{-1}\left(\frac{c_B}{n_B}\right)} f_B(x)dx = E_B = \frac{M_B c_B}{\lambda_B}\left[1 - \frac{1}{n_B}\right]$$

$$\left(1 - F_B(f_B^{-1}(c_B))\right)h_B = \frac{1}{n_B} = \frac{c_B h_B}{\lambda_B}$$

As there is not a clear way to aggregate the industry knowledge of the economy h, I have decided to back it out from the equations. However, when I calculate the parameters assuming that industry knowledge is the average number of words regulating all the lobbying industries, the main results do not change.

[116] I include the restrictions $\lambda_B > c_B$ and $\lambda_B^2 > c_B^2 h_B$, to ensure that $z_B^*, z_B^{**} > 0$.

[117] The fraction is decreasing in the communication costs as $h_B \propto n^{-1}$.

2.4.1.2 Estimation

In the system mentioned above, I can infer I_B, E_B and n_B from the data by taking the average of each variable across the three semesters prior to the OGA.[118][119] However, there are four unknowns: λ_B, c_B, h_B and M_B.[120] As I need an additional degree of freedom to be able to solve the system, I calculate λ_B by maximum likelihood, and then solve the system analytically. Here, I use all of the congressional hearings with available information on witnesses three semesters before the OGA. For a congressional hearing i in period B (before OGA), I measure $z_{i,B}$ with the fraction of witnesses with high knowledge. Alternative ways to measure the skill variable with a more comprehensive index do not qualitatively change the results presented in this section. Assuming independent and identically distributed observations, the likelihood function is:

$$f_B(z_{1,B}, z_{2,B}, ..., z_{O,B}|\lambda_B) = \prod_{i=1}^{O_B} f(z_{i,B}|\lambda_B) = L\left(\lambda_B|\mathbf{z}_B\right) = \prod_{i=1}^{O_B} \lambda_B e^{-\lambda_B z_{i,B}}$$

where O_B is the total number of observations before the OGA. With the estimate of λ_B, I get an estimate of the other unknowns in the system. In particular, the system can be re-written as follows:[121]

$$c_B = \frac{\widehat{\lambda}_B E_B n_B}{n_B\left(E_B + I_B\right) - I_B},\ h_B = \frac{\widehat{\lambda}_B}{n_B c_B} \text{ and } M_B = \frac{\widehat{\lambda}_B I_B}{\widehat{\lambda}_B - c_B}$$

I follow this methodology for both the periods before and after the OGA. Table 2.1 shows the input variables used, the estimated coefficients $\widehat{\lambda}_B$ and the estimated unknowns.[122] The top panel in the table shows the main inputs of the exercise. The fraction of vertically integrated clients increases by 1%, which is about 7% of the fraction before the OGA. Although there is an increase in the total number of both external and internal clients, the latter variable increases by about three times as much as the former.

The estimated coefficient for the parameter of the distribution before the OGA

[118] The leverage n_B is simply the average ratio of the total number of external clients over the total number of external lobbyists.

[119] This system needs to impose the restriction that: $\lambda_B > c_B$. Notice that if $\lambda_B > c_B$, I_B is positive, and given that $h_B \in (0,1)$, $\frac{1}{n_B} < 1$ by the time constraint. This, in turn, implies that E_B is positive.

[120] M_B cannot be confused with the total number of firms in the economy. The right interpretation of this parameter is the number of firms willing to lobby. I have decided to abstract from the decision of whether or not to lobby, as several papers, such as Salamon and Siegfried (1977), Bombardini (2008) and Kerr et al. (2014), have studied this problem, and I think that this additional feature does not add content to the main insights I am studying.

[121] Notice that the original constraint $\lambda_B > c_B$ implies $\lambda_B > \frac{\lambda_B E_B n_B}{n_B(E_B+I_B)-I_B}$, which is true as far as $I_B\left(n_B - 1\right) > 0$. Therefore, as far as $n_B > 1$, $\lambda_B > c_B$, $\lambda_B > c_B h_B$ and $M_B, c_B, h_B > 0$.

[122] Notice that in this system, $c_B = \lambda_B \cdot K$, where K is $\frac{E_B n_B}{n_B(E_B+I_B)-I_B}$. Then, h_B and M_B have a unique value for different λ_B's. That is, $M_B = \frac{\lambda_B I_B}{\lambda_B - c_B} = \frac{\lambda_B I_B}{\lambda_B - \lambda_B K} = \frac{I_B}{1-K}$ and $h_B = \frac{\lambda_B}{n_B c_B} = \frac{1}{n_B K}$.

is 1.14, and after the OGA it is 0.91. This shows a decrease in the estimated parameter of about 20%. Notice that this has a simple interpretation in terms of the first moment of the distribution. The mean of the distribution in the case of exponentially distributed observations is equal to $\frac{1}{\lambda}$. As $\widehat{\lambda}_A < \widehat{\lambda}_B$, the average knowledge requirements after the OGA are larger than the average of these requirements before the OGA. As larger knowledge requirements are associated with more difficult advocacy activities, a decrease in λ supports the statement of the lobbyists that the difficulty of their lobbying activities increased.

The estimated cost of acquiring knowledge c decreases by 20%. This is one of the main findings of the section and provides evidence on the effect of the use of internet on the reporting of the lobbying activities. While the coefficient for industry-specific knowledge h, and the total number of (potential) clients M also change, they do not seem to be large. The estimation shows that the communication cost only decreased by 0.04 percentage points. This small decrease does not seem significant and presents evidence that the OGA did not modify the communication cost. On the other hand, we observe a slight increase in the total number of clients in the industry of about 4.7%. Decomposing this number we observe that most of the new clients start using in-house rather than external lobbyists.

A final theoretical prediction is also tested in this exercise. A decrease in the cost of acquiring knowledge empowers in-house staff. This makes it less likely that firms with in-house lobbyists will outsource. As external lobbyists have also become smarter and their time constraint is not binding (due to the decrease in demand from clients that now can pay an in-house staff), they adjust in equilibrium by working for more clients. Therefore, a decrease in the cost of acquiring issue-specific skills increase the leverage of the external lobbyists. This leverage was about 2.33 before the OGA and it became 2.36 after the OGA. So, we see an increase in this variable, just as the theory predicts, however not a very significant one. The reason for this small change is that the OGA also changed the density of the problems in the economy. The span of control is increasing in the rate of the exponential function λ. As some channels were closed, the activities became harder, and therefore, we observe a decrease in λ, with the consequent effect on the decrease in the leverage of the economy. In section 2.4.1.3, I conduct counterfactual exercises examining the effect of the technological shock on the leverage of the economy, holding constant the difficulty level of the problems faced by the firms in the economy.

Knowledge Inequality I include an empirical estimate of the knowledge inequality measure, that is the ratio of the maximum level of issue-specific skills achieved in the external market over the maximum level of this type of skills for in-house staff. Mathematically, for the period before the OGA, this is represented by the following ratio $\frac{z_B^{**}}{z_B^{*}} = \ln(\frac{c_B^2 h_B}{\lambda_B^2}) / \ln(\frac{c_B}{\lambda_B})$. Theoretically, the total earnings for internal and external providers are given by $w_B^I z_B^*$ and $n_B w_B^E z_B^{**}$, respectively, where

$w_B^I = n_B w_B^E = c_B$. This implies that the ratio $\frac{z_B^{**}}{z_B^*}$ is a representation of both the difference in the issue-specific skills and earnings levels between the market and firm boundaries. Table 2.1 shows a decrease in the knowledge inequality in the economy. Decomposing the change of this fraction by the numerator and denominator shows that the market boundary increases by 1.8%, whereas the firm boundary increases by 8.5%. That is most of the change in the inequality measures are due to a larger set of issue-specific skills from in-house staff.

All of the exercises presented in this section are for two periods, before and after OGA. Alternatively, I have conducted these estimations for each year of my database. Figure 2.6 shows the results of the evolving knowledge inequality of the economy for the period 1999 to 2014. The figure shows an increasing pattern in the inequality until 2007 and since then, a decreasing pattern.

Decomposing the Effect Using these recovered unknowns, one can decompose the exact contribution of each of the parameters in the total change of the fraction of vertically integrated clients. The total change of this fraction can be decomposed as:

$$dVI_t = \underbrace{\frac{\partial VI_B}{\partial \lambda_B} d\lambda_t}_{Demand} + \underbrace{\frac{\partial VI_B}{\partial c_B} dc_t}_{Supply=Acquiring\ Costs} + \underbrace{\frac{\partial VI_B}{\partial h_B} dh_t}_{Industry\text{-}S.\ Knowledge}$$

where $\frac{\partial VI_B}{\partial x_B}$ is the change evaluated with the parameter estimates before the OGA, $x_B = \lambda_B, c_B$ or h_B and $dx_t \approx x_A - x_B$.[123] Table 2.2 shows the percentage contribution of each of these components: demand, costs and communication costs. I estimate each of these contributions for the case in which I use the estimated coefficient $\widehat{\lambda}$ and both its lower and upper limits. The table shows that the industry-specific knowledge does not contribute significantly to explaining the total change in the integration patterns of the industry.[124] This result is confirmed when I use the RegData to proxy for industry-specific knowledge. Table 2.2 shows that about a half of the change around the OGA is explained by a demand shift and about half by a change in the costs c. Overall, these results confirm the relevance of the channel of the frequency of problems but emphasize the importance of supply-shift channels that can explain the vertical integration patterns.

To sum up, the following interpretation is consistent with the results provided here. The Act closed lobbying channels and decreased the cost of acquiring knowl-

[123] These derivatives are $\frac{\partial VI_B}{\partial \lambda_B} = \frac{c_B\left(\lambda_B^2 - 2c_B h_B \lambda_B + h_B c_B^2\right)}{\left(\lambda_B^2 - c_B^2 h_B\right)^2}$, $\frac{\partial VI_B}{\partial c_B} = -\frac{\lambda_B\left(h_B c_B^2 - 2c_B h_B \lambda_B + \lambda_B^2\right)}{\left(c_B^2 h_B - \lambda_B^2\right)^2}$ and $\frac{\partial VI_B}{\partial h_B} = \frac{\lambda_B(\lambda_B - c_B)c_B^2}{\left(\lambda_B^2 - c_B^2 h_B\right)^2}$.

[124] I run $h_t = \beta_0 + \beta_1 t + \varepsilon$, where h_t is the Industry-specific knowledge measure at the two-digit levels proxied with the RegData. In order to construct it, I have taken the average number of regulating words across all industries. The period used is 1999-2014. The predicted errors show a small decrease in the industry knowledge measure between 2006 and 2008. This supports the fact that the calculated h in Table 14 decreases, but not significantly. Results are available upon request. These patterns are robust to the use of any NAICS codes industry definition.

edge. The closed lobbying channels made the lobbying activities more knowledge-intensive as politicians needed to rely more heavily on congressional hearing witnesses to acquire knowledge. This is confirmed by Figure 2.2 and theoretically this translates into a change in the density of the problems. The Act also decreased the cost of acquiring knowledge as a new technology to acquire relevant information for advocacy activities was implemented. As a consequence, given that the distribution of problems shifted and the technology of reporting decreased the cost of acquiring knowledge, the OGA caused an increase in the fraction of clients using in-house lobbyists.

An advantage of using structural methodologies is that once one recovers the underlying parameters of the economy, one can conduct counterfactual exercises. In the next subsection, I exploit the information on the recovered parameters.

2.4.1.3 Counterfactual Exercises

For simplicity, I call the change in the distribution, the *difficulty effect* and the change in the estimated cost of acquiring issue-specific skills, the *technology effect*. Here, I answer two simple questions. First, what would have been the main outcomes for the economy if the distributions of the problems had not changed and if there had been a change in the cost of acquiring knowledge. That is, what happens if one observes only the *technology effect*. Second, assuming that the cost technology had remained constant, what would have happened if one had observed a change in the density of problems. That is, what happens if one observes only the *difficulty effect*. For all of these analyses, I assume that the industry-knowledge measure remains at the level before the OGA.

Figure 2.7 provides a useful way to see the logic behind the counterfactual exercises. The graph represents, for each period before and after, the density function of problems and the cost of acquiring knowledge. There are four intersections between the density functions and the costs of acquiring knowledge, each of which represents a different firm boundary level. Point (1) represents the intersection that defines the firm boundary before the OGA and point (4) the intersection defining the boundary after the OGA. Point (2) holds fixed the density function of problems before the OGA but changes the cost of acquiring knowledge from c_B to c_A. Point (3) represents the intersection of the acquiring cost technology before the OGA (i.e., c_B) with the density function of problems after the OGA (i.e., f_A). The first counterfactual exercise aims to see the changes brought about by the *technology effect*, which is represented by the changes due to a movement between points (1) and (2). The second counterfactual exercise analyzes the change due to the difficulty effect-that is, changes brought about by moving from point (1) to point (3).

For these exercises, I focus on three objects: fraction of vertically integrated clients, $I/(E+I)$; leverage in the economy n; and $\frac{z^{**}}{z^{*}}$, which is a measure of both knowledge and earnings inequality. For the exponential case, the leverage in the economy is increasing in the rate λ and decreasing in the cost technology c. However, the effect of a change in the cost c or the density function parameter λ in the fraction of integrated clients and the knowledge inequality is not trivial. For instance, the *technology effect* increases the number of internal clients, but it can increase or decrease the number of external clients. The effect on the internal clients is due to the fact that it is now less costly to have in-house staff, and, as a consequence, the firm boundary shifts to the right.

For the case of the external market, there are two countervailing effects. On the one hand, it is less costly to acquire knowledge for all service providers. This effect will move the market boundary to the right, and the fraction of clients that used to leave their problems unsolved but now use external service providers is given by $F(z_A^{**}) - F(z_B^{**})$. On the other hand, in-house service providers can solve problems for more clients; therefore, external service providers will see decreased demand due to clients that have in-house lobbyists. This area is represented by $F(z_A^{*}) - F(z_B^{*})$, and, therefore, the total fraction of external clients will change by $F(z_A^{**}) - F(z_B^{**}) - [F(z_A^{*}) - F(z_B^{*})]$, which is not always positive.[125]

Table 2.3 shows the main results of these counterfactual exercises. The *technology effect* increases the fraction of vertically integrated clients and the leverage of the economy, and decreases the knowledge and earnings inequality measures. On the other hand, the *difficulty effect* decreases both the fraction of integrated clients and the leverage of the economy, and increases the knowledge inequality in the economy.

The *technology effect* increases the fraction of vertically integrated clients from 14.8% to 36.8%, an increase in 150%. Intuitively, the decrease in the calculated cost of acquiring issue-specific skills changes the firm boundary and makes external service providers less likely to face problems. As a consequence, the time constraint is relaxed, and the service providers can work with more clients. Confirming this logic, Table 2.3 shows that the leverage increases by 26%. Finally, a decrease in the cost c increases both the market and firm boundaries, and, depending on the specific parameters, the knowledge inequality in the economy can increase or decrease. In this case, it decreases because the effect on the firm boundary is larger than the effect on the market boundary.

If the parameter of the distribution of problems decreases from 1.14 to 1.07, the fraction of vertically integrated clients decreases from 14.8 to 6.13, a decrease of 59%. For this combination of values, the density functions cut below both costs c_B and c_A. As a consequence, the decrease in the rate λ, shrinks the firm boundary. This means that external service providers are more likely to receive problems from clients. Given the time constraint, the increase in the probability of receiving problems,

[125] This fraction is equal to $\frac{c_A}{\lambda_A^2}(\lambda_A - c_A h_A) - \frac{c_B}{\lambda_B^2}(\lambda_B - c_B h_B)$, and it is not clear whether or not this expression is positive, as it depends on the specific values of the parameters of c, h and λ before and after the OGA.

decreases the number of clients that external service providers have.

2.4.1.4 Robustness in the estimations

One may wonder to what extent the main results from these exercises depend on the assumption of the exponential distribution and, more generally, on assuming a specific functional form. I conduct two exercises to alleviate these concerns: 1) I change the functional form assumption; and 2) I estimate the parameters of the model by non-parametric methods.

First, I run the same exercises when the problems are distributed according to a Pareto distribution. This distribution is an ideal candidate as its density function is strictly decreasing as assumed in the theoretical framework. In this case, the density function $f(z)$ is given by $\frac{\alpha z_m^\alpha}{z^{\alpha+1}}$ with parameters $\alpha, z_m > 0$, where z_m is the minimum possible value of z. Using maximum likelihood estimation, I recover the parameters to conduct similar exercises as above.

I calculate the contribution of demand, supply and industry-specific channels on the change in the vertical integration patterns, and I find that the cost component change explains between 55% and 60% of the change in the patterns. I also conduct the counterfactual exercises on the *technology* and *difficulty effects* and get similar conclusions. In particular, the *technology effect* increases the integration patterns twofold, increases the leverage, and decreases the knowledge inequality in the economy. On the other hand, keeping the value of the cost technology constant a decrease in the parameter α (*difficulty effect*) decreases the vertical integration in the economy.[126] To sum up, the use of the Pareto distribution provides qualitatively similar results.

I also conduct a simple non-parametric exercise. The estimation here is more challenging, and I focus only on recovering an estimate of the cost of acquiring issue-specific skills. The idea is summarized as follows. The fraction of vertically integrated clients before the OGA is given by $\frac{F_B(z_B^*)}{F_B(z_B^{**})} = \frac{F_B(f_B^{-1}(c_B))}{F_B\left(f_B^{-1}\left(\frac{c_B}{n_B}\right)\right)}$. As I observe a set of z_B, I can construct the density $f_B(z_B)$ by kernel methods. Once I have constructed this density, I can have an estimate of the inverse density $f_B^{-1}(z_B)$ using the inverse transform method. Then, I construct $F(f_B^{-1}(x))$ by defining $F(x) = \frac{1}{o_B}\sum \mathbf{1}\{z_B \leq x\}$, where o_B is the total number of observations before the OGA and $\mathbf{1}\{z_B \leq x\}$ is an indicator function. Finally, I find c_B by solving $\frac{F_B(x)}{F_B\left(\frac{x}{n_B}\right)} - \frac{I_B}{(E_B+I_B)} = 0$ for x. I follow this methodology also for the period after the OGA and find two main results. The calculated density functions and the costs of acquiring knowledge differ before and after the OGA. In particular, the

[126] Notice that a decrease in α for the Pareto distribution is similar to a decrease in λ for the exponential function, as in both cases, the cut with the y-axis (i.e. $Pr(Z = z)$) is increasing in the value of the parameter. A decrease in both parameters imply that easy activities became less common and more difficult activities became more common. Furthermore, the mean value of the Pareto distribution is decreasing in α. This implies that a decrease in α increases the mean knowledge requirements of the lobbying activities.

cost decreases, and the density functions cut below both estimated costs. These exercises confirm the robustness of the main results obtained for the case in which problems are distributed exponentially.

In this section, I have focused on the effect of closing lobbying channels and technological shocks on integration patterns in the industry. Potentially, the observed integration patterns can also be explained by the 2008 Financial crisis, alternative theoretical mechanisms or simply, by other regulations of the OGA not related to knowledge-requirement changes. Next section presents evidence that all of these considerations do not seem to play a key role in explaining the empirical patterns outlined here.

2.4.2 Alternative Explanations

In this section, I discuss alternative explanations to the change in the vertical integration patterns *circa* 2007. I start by presenting empirical evidence supporting the view that the Financial crisis did not have a strong effect on the integration patterns. Then, I discuss alternative ways to interpret the technological shock of the OGA as well as the possible effects of additional regulation changes brought by the Act.

2.4.2.1 2007-2008 Crisis

The timing of the OGA coincides with the Financial Crisis. In this section I argue this is not a key determinant to explain the change in vertical integration patterns. First, according to Figures 1.13 from Chapter 1, there are not changes in integration patterns for other Professional Business Services occupations aside from lobbyists around the OGA. Second, Knudsen and Foss (2014) show that firms in Norway started outsourcing more non-core activities since 2007 due the economic recession. The idea is that the Crisis decreased the demand for all the firms, which in turn decreased the frequency of needed non-core activities transactions. Therefore, the probability of outsourcing increased. As advocacy is considered an ancillary activity, Knudsen and Foss (2014) predict a decrease in the fraction of vertically integrated clients. This means that if the Crisis has an effect on the lobbyists' integration patterns it will be the opposite to the one I see around the OGA in figure 2.1.

To add to this preliminary evidence, I explore whether demand or supply shocks affect the vertical integration patterns in the lobbying industry. Similar to Moreira (2016) and based on the Bureau of Economic Analysis data, I construct economic shocks at the industry-year level using supply chain information to capture the variation in demand from downstream buyers. I proceed as follows: **1**. I construct an annual industry by industry sales matrix using the Use and Make input-output tables.[127] **2**. I construct shocks for each industry in each year using the annual

[127] This matrix gives me information on total sales of the industry discriminated by buying (demanding) industry. The Make-table has dimensions industries-commodities and the Use-table

industry-specific nominal Gross Domestic Product (GDP) as the primary business cycle indicator.[128] To calculate the shocks I apply the Hodrick-Prescott filter using a smoothing parameter of 6.25.[129] I capture industry-specific shocks as the deviation of the real GDP from its trend. **3**. I construct a weighted average (by percentage of sales) of demand shocks for industry n simply aggregating the shocks for all the industries that industry n sells in a given year t (i.e. s_{nt}). Then, I estimate the following equations:

$$VI_{ijnt} = \beta_j + \gamma_t + \eta_z s_{nt} + \varepsilon_{ijnt}$$

$$fr_{jnt} = \beta_j + \gamma_t + \eta_z s_{nt} + \varepsilon_{jnt}$$

The key independent variable of this exercise is s_{nt}. This is a demand shock received by the n-th industry at t-th period. A negative and significant η_z suggests that a positive shock in demand faced by firms from j-th industry decreases the level of vertical integration of j-th industry. As the Financial crisis can be interpreted as a negative demand shock, a negative coefficient $\eta_z < 0$ can explain the change in the vertical integration patterns around 2007.

I also use two additional shocks. First, I consider the case where there is only a shock to the industry (supply shock). Second, I include a shock that include both the demand and the supply shocks (economy shock). For each of these variables, I calculate the independent variable in terms of levels, lags, logarithmic levels and lag of the logarithmic level. The main result from these exercises is that none of these economic shocks are statistically different from 0. That is, I do not find empirical evidence that neither demand, supply or economy shocks modify the vertical integration patterns of the lobbying industry. Therefore, it is hard to argue that the Crisis affected the integration patterns around the OGA. All these results are available upon request.[130]

2.4.2.2 Alternative Mechanisms: Monitoring Costs

One channel that can explain vertical integration patterns is monitoring costs. When firms require monitoring of the activities of their staff, vertical integration

has dimensions commodities-industries. In order to get the symmetric industry-industry matrix, I normalize the Make-table by the total of each column and then I multiply it by the Use-table.

[128] I deflate it with industry-specific producer prices at 2009 prices.

[129] For these estimations I deflate the nominal GDP with industry-specific producer prices at 2009 prices. The smoothing parameter has been chosen following Ravn and Uhlig (2002).

[130] Intuitively, one can rationalize this result as follows. A negative economic shock to clients increase the probability of outsourcing for ancillary activities as the demand for their products have decreased. On the other hand, the negative shock increase the frequency of lobbying activities, as policy advocacy seems to be a way to overcome negative economic shocks. Consequently the probability of outsourcing lobbyists decreases. Therefore, the non-significant effect of the Crisis over the lobbying integration patterns can be rationalized if the latter effect is compensated with the former one.

can be a solution for situations characterized by costly monitoring. For instance, as discussed in Williamson (1975), governance (integration) provides the opportunity for firms to monitor more accurately and economically their employees. The advantage that governance offers is attenuated or inexistent in the external market. A direct implication from this framework is that a decrease in the cost of monitoring, will increase outsourcing as the relative value of in-house staff has decreased. That is, if one of the main advantages of having someone in-house is to reduce monitoring costs, when these costs are reduced, some firms that were in the margin (indifferent between in-house and external advocate) should be more prone to outsource the service.

A possible way to interpret the technological shock brought by the OGA is a decrease in the firms' monitoring costs. The Act facilitate firms to not only observe what their lobbyists were doing but also, and more importantly, what other lobbyists working for firms with similar interests were doing. Therefore, for a firm that can see the performance of lobbyists working for other organizations, it is easier to detect, for instance, moral hazard due to information asymmetry with their employees. If this is the right interpretation, we would expect to see that the Act produced an increase in the outsourcing of the economy. Clearly, Figure 2.1 provides evidence against this intuition as it is clear that the level of integration in the economy increased rather than decreased once the Act took place. Therefore, even if monitoring costs are important in this industry, it does not seem correct to interpret the OGA as a technological upgrade that reduced the costs associated to monitoring the activities of the lobbyists.

2.4.2.3 Other Changes brought by the OGA

Change in Mandatory reports Before the OGA, firms had to report lobbying activities if they were spending at least 10.000 US dollars per semester. As the frequency of the reports changed since OGA and firms have to make reports every quarter, the cut-off spending level also changed. Since the first semester of 2008, firms have to report lobbying activities if they spend at least 5.000 US dollars per quarter.

If a firm spends any amount between \$5.000 and \$10.000 before and after the OGA in a quarter, the database would show the client was not conducting lobbying activities before the OGA and started doing it after the OGA. Furthermore, if these clients were conducting in-house activities, the change in the mandatory reports can explain the increase in vertical integration in the industry. In Table 4, I provide evidence that this does not seem to be the case. I show that among all the clients that started lobbying once the OGA took effect, less than 5% of the clients had expenditures close to \$10.000 (up to 20.000). Clearly, there is an additional alternative in which clients were spending less than \$10.000 before OGA and since OGA took effect, they started spending more than \$10.000. This makes sense if firms feel that once they are engaged in public reporting, they have incurred the lobbying stigma cost (independent of the actual lobbying expenditure) and as a consequence they may want to increase the lobbying returns by increasing the lobbying expenditures. Although the data does not allow me to reject this possibility, private interviews with lobbyists and interest groups strongly reject this alternative.[131]

Reporting Incentives Fines for compliance failure increased with the OGA. This change may affect the incentives to hide lobbying activities. If it is more likely that clients with in-house lobbyists hide their activities, an increase in compliance failure's punishment will raise the total number of clients reporting in-house activities. Then, an increase in the observed fraction of vertically integrated clients follow. A possible reason to hide lobbying activities is to avoid sharing sensitive information. For instance, using information of 150 lobbying contacts at the Federal Communications Commission, de Figueiredo and Kim (2004) argue that firms tend to use in-house lobbyists for firm-specific topics prone to sensitive-information leakage.[132] Sensitive information can be production secrets, corruption issues and so on. Due to the lack of data it is hard to know to what extent sensitive information leakages is a relevant force to explain integration patterns. However, I provide two arguments supported by personal interviews with lobbyists that suggest that this is not a first-order factor to explain the change in the integration patterns around the OGA.

First, there are differential legal privileges to protect information when firms use in-house or external staff. *Attorney-client privilege* is a legal concept that keeps confidential the information exchanged between clients and attorneys. This concept applies more commonly in the external rather than the internal market. That is, in the former there is an official relationship between the two entities that benefit from this privilege. However, the internal market is characterized by a situation where the *attorney* is not an entity but just an employee, and therefore the privilege does not always apply. Therefore, from a legal point of view, clients will prefer external staff to benefit from the privilege.

Second, the demand from external lobbyists depends on their ability to protect sensitive information from their clients. For instance, Demski et al. (1999) argue that firms related to banking, accounting, consulting, and legal services tend

[131] In addition, one can focus on the firms that were lobbying both before and after the Act. I find that among the switcher firms (firms that were only using in-house (external) lobbyists and after the Act start using only external (in-house) lobbyists) 90% went from using only external lobbyists to only using in-house lobbyists, whereas 10% went from only in-house lobbyists to external lobbyists. Similar percentages appear when I allow different ways to define switching firms. Interestingly, the absolute value of switchers going from the external to the internal market (immediately the Act took effect) is significantly larger than the number of this type of shifters for any other period.

[132] Anderson (1985) argue that large appropriation hazards are associated with a higher probability of integration. A general problem with this literature is that it leaves unexplained why integration does a better job of restoring efficiency than an outsourcing contract.

to process proprietary information which their clients wish to protect. As a consequence, the firm's ability to safeguard and manage information determines its demand. Then, if a client wants to keep sensitive information secret they can do it with external staff. As a matter of fact, anecdotal evidence shows that sensitive information leakage tends to be more common among employees or former ones rather than external service providers.[133] That is, protecting information may be more effective in the external than in the internal market.

There may be one argument that I cannot totally reject given the available data. There may be some clients that think that the cost of public lobbying reporting is high. However, as there are benefits from the lobbying activity, these clients will lobby without reporting it. If the fines for compliance failure increase, the expected cost of hiding increases and therefore it will be more likely that clients start reporting. If it is more difficult to lobby and not publicly report with external than internal lobbyists, the change in the fines brought by the OGA could have increased the observed fraction of internal clients. Personal interviews with both internal and external lobbyists, in particular with some that started lobbying around the OGA show that this is not the case. They think the cost of public lobbying reporting is very low as the advocacy activities are protected by the First Amendment of the US constitution.[134] They argue that is unlikely that clients will avoid lobbying just for this stigma cost, and in fact, they think that the expected cost (fines and probability of detection) of hiding public reporting is larger than the cost associated to the stigma of lobbying.

A final comment on the differences of hiding technologies between in-house and external staff. The OGA made it the responsibility of the US Government Accountability Office (henceforth GAO) to determine the extent to which lobbyists can demonstrate compliance with disclosure requirements.[135] The office releases yearly reports summarizing the main findings about compliance for a stratified random sample of 100 lobbying reports. I have been through all of these reports to correlate non-compliance patterns with lobbyists integration status.

The GAO reports release information about the specific lobbyists and clients

[133]See for instance, Rajan and Zingales (2001), Zabojnik (2002), Baccara and Razin (2007), and Baccara and Bar-Isaac (2008). These papers study the case where either firm's employees or former ones leak crucial information outside the firm. Anecdotically, see the case of Chelsea Manning (former US Army soldier) leaking sensitive military information to WikiLeaks, Mark Felt (former FBI special agent) leaking information about the Watergate scandal and Sky company employee leaking customer data to a rival firm (http://www.information-age.com/technology/security/2129443/sky-employee-leaked-customer-data) I acknowledge this is not proof that employees are more likely to leak information than external contractors as this result may be due to selection. For instance, firms may share the sensitive information **only** with employees. Therefore, leakages cannot come from external providers.

[134]The Amendment does not call it lobby but "to petition the Government for a redress of grievances". The point is that what is a grievance is a matter of debate and firms can always argue over this right using almost any goal they have.

[135]For more information about GAO, see https://www.gao.gov/about/index.html.

under investigation. I have matched all of the names of lobbyists with its integration status at the time of investigation. In results not presented here, I show that both in-house and external lobbyists have been caught under non-compliance cases. Neither group has a noticeably larger fraction of non-compliant lobbyists. A natural way to interpret this finding is that the technologies that lobbyists have to hide information do not differ by the market (internal vs external) they are in. As a consequence, the integration decision should not respond to the demand of hiding information.

To sum up, the change in the integration patterns can be explained through an increase in fines for non-compliance or hiding activities, if it is more likely that clients use in-house lobbyists when they want to hide sensitive information. I first provide two arguments to explain why firms would like to use external rather than internal lobbyists when they want to hide information. Second, I provide suggestive empirical evidence that the hiding technology of in-house lobbyists is not better than external lobbyists. Therefore, given the arguments provided here, it is difficult to argue that the change in the incentives to report can explain the observed, big shift in the integration patterns around 2007.

2.5 Final Discussion

The aim of this chapter has been to understand the way technological change affects the vertical integration decision when firms need to solve problems that can only be tackled by service providers. The central point this chapter wants to make is that there are at least two different types of technological change that have deeply different effects on the knowledge economy in general, and the integration decision, in particular.

Technological shocks that decrease the cost of acquiring skills empower in-house staff and therefore decreases outsourcing in the economy. On the other hand, shocks that facilitate the transmission of firm-specific skills decrease the comparative advantage that in-house staff have and therefore increase outsourcing. To the best of my knowledge, the impact of these two types of technologies on the integration of the economy has not been analyzed before. Luis Garicano has been a pioneer in the discussion on the distinctive effects of these two types of technologies. This chapter borrows from his ideas and provides an empirical setting to test predictions on the effect of changes in the acquisition cost on integration patterns.

I have found strong evidence supporting the claim that the policy change studied in this chapter decreased the cost of acquiring issue-specific skills and that this could explain the strong increase in integration patterns in the economy. The exercises developed in this chapter use equilibrium conditions for the case of only one issue. Extending this methodology to the case of multiple issues seems very challenging. The main problem consists in that the burden of the estimations would be very large because I need to make inferences on the actual (and not observed) complementarities across issues and I would need to include in the calculations the potential number of generalists and specialists in each market. I aim to consider these extensions in the future and I include a short discussion of some empirical patterns for several issues in section 2.6.4.2.

Abstracting from the current application, the analysis underscores the potential of non-incentive integration theories to explain the effect of technology change in the way the firms use knowledge workers. Here I have provided a simple setting to study these effects but I have not provided empirical evidence on the way that changes in communication costs affect integration patterns. A natural sequel of this chapter would be an empirical study of an event in which the cost of transmitting firm-specific skills has been modified, and to test, whether this has any implications for the integration decision.

2.6 Appendix

2.6.1 Regulation in the OGA

The Honest Leadership and Open Government Act of 2007 is a law that amended parts of the LDA of 1995. To follow, there is a summary of the main changes that the bill introduced:

- Increase the *cooling off* period for Senators from one to two years before they can lobby Congress.
- Prohibits Cabinet Secretaries and other very senior executive personnel from lobbying the department or agency in which they worked for two years after they leave the position.
- Prohibits senior Senate staff and Senate officers from lobbying contacts with the entire Senate for one year, instead of just their former employing office.
- Prohibits senior House staff from lobbying their former office or Committee for one year after they leave House employment.
- Prohibits lobbyists from providing gifts or travel to Members of Congress.
- Prohibits Members and their staff from influencing hiring decisions of private organizations on the sole basis of partisan political gain (imprisonment for up to 15 years).
- Requires lobbyist disclosure filings to be filed twice as often.
- Increases civil penalty for knowing and willful violations of the Lobby Disclosure Act from $50,000 to $200,000 and imposes a criminal penalty of up to five years for knowing and corrupt failure to comply with the Act.
- Denies Congressional retirement benefits to Members of Congress who are convicted of bribery or other related crimes.
- Requires that candidates, other than those running for a seat in the House, pay the fair market value of airfare (charter rates) when using non-commercial jets to travel.
- Increases the penalty for Members of Congress, Senior Staff and Senior Executive officials for falsifying or failing to report financial disclosure forms from $10,000 to $50,000 and establishes criminal penalties of up to one year of imprisonment.

- **Requires the disclosure of businesses or organizations that contribute more than $5,000 per quarter and actively participate in lobbying activities by certain coalitions and associations.**[136]

- Requires that travel by members financed by outside groups be posted on a searchable, sortable and downloadable website by August 1, 2008.

- Prohibits Members from attending parties held in their honor at national party conventions if they have been sponsored by lobbyists.

- Prohibits Members from engaging in any agreements or negotiations about future employment until a successor has been selected.

- Requires that Members prohibit their staff from having any lobbying contact with the Member's spouse if such individual is a registered lobbyist or is employed or retained by a registered lobbyist to influence legislation.

- Revokes floor privileges and the use of the Members' exercise facilities and parking for former Senators, former Secretaries of the Senate, former Sergeants at Arms of the Senate and former Speakers of the House who are registered lobbyists.

[136] The cut-off before was 10.000 per semester. Therefore, there is not a change in the semester total amount cut-off.

2.6.2 Tables

Variable	**Before**	**After**	**% Change**
External Clients (E)	11207.3	11709	4.5
Internal Clients (I)	1952	2208	13.1
Total Clients $(E+I)$	13159.3	13917.6	5.8
Fraction VI $(I/(E+I))$	14.83	15.87	7
Estimated Coefficient $(\widehat{\lambda})$	1.1413	0.913	-20
Lower Limit (95%)	1.0949	0.8749	-20.1
Upper Limit (95%)	1.1908	0.9535	19.9
Cost (c)	1.0380	0.8237	-20.6
Lower Limit (95%)	0.9958	0.7893	-20.7
Upper Limit (95%)	1.0831	0.8602	-20.6
Industry-Knowledge (h)	0.4718	0.4716	-0.04
M	21585.84	22590.93	4.7
Knowledge Inequality $(\frac{z^{**}}{z^{*}})$	9.92	9.3	-6.2

Table 2.1: Input and Output of the Structural Estimation.

Note: The top panel shows the main inputs of the exercise and the bottom panel shows the outputs. External clients represent the total number of clients that never use in-house lobbyists. Internal clients represent the total number of clients using in-house lobbyists. The rate parameter of the distribution is calculated by Maximum Likelihood. The table reports the point estimate and the calculated value one standard deviation above and one standard deviation below the point estimate. With these estimates and using the system of equations, the table reports the estimated values for the cost of acquiring knowledge, the industry-specific knowledge, the total number of potential clients (M) and the knowledge inequality of the economy.

Contribution of each component in the change of VI			
	Demand	Supply	Industry-Knowledge
Estimated Coefficient	46.98	52.96	0.07
Lower Limit (95%)	44.9	54.99	0.11
Upper Limit (95%)	49.11	50.85	0.04

Table 2.2: Fraction of VI Changes Explained by Demand, Supply and Industry-specific Knowledge (Communication Costs) Components.

Note: The table reports the decomposition of each of these components on the total change of vertical integration of the economy. Each row denotes a different rate parameter of the distribution of the problems. The first row presents the decomposition when the rate parameter is the point estimate of the Maximum Likelihood estimation. The next rows reports the decomposition when the rate parameter is calculated one standard deviation above and one standard deviation below the point estimate.

	Fraction of V.I. Clients $(I/(E+I))$		*Leverage of the Economy* (n)		*Knowledge Inequality* $(\frac{z^{**}}{z^{*}})$	
	c_B= 1.03	c_A= 0.82	c_B= 1.03	c_A= 0.82	c_B= 1.03	c_A= 0.82
λ_B= 1.14	14.83	36.89	2.33	2.93	9.92	4.3
λ_A= 1.07	6.13	32.3	2.19	2.76	23.49	4.82

Table 2.3: Counterfactual Exercises.

Note: Each row denotes a different level in the parameter of the distribution of problems; each column represent different values of the cost of acquiring knowledge. Starting from left, each panel shows the counterfactual values for the fraction of integrated clients, the leverage of the economy and the knowledge inequality measure. As the model imposes the restriction that $\lambda > c$, I use λ_A= 1.075 for the rate parameter of the distribution after the OGA. This value represents 4 deviation standards above the estimated coefficient λ_A= 0.913.

		2008		**2009**		**2010**	
Semester		**I**	**II**	**I**	**II**	I	II
Fraction entrants	*Int.*	3.9	2.7	3.6	2.2	5.1	6.8
Spending <10K	*Ext.*	3.8	4.5	5	5.1	5.2	5.2
Mean Expenditure	*Int.*	167299.5	123559.1	122789.5	129382.1	252445.1	120182.8
	Ext.	30700.6	35181.8	31982.7	32546.4	32626.3	35839.6

Table 2.4: Characteristics of Entrant Clients After the OGA.

Note: The first two rows of the table report the fraction of clients that report lobbying expenditures less than 10.000 US dollars over the total number of entrant clients , discriminated by their use of in-house lobbyists. The last two rows present average lobbying expenditures for those clients, discriminated by their use of in-house lobbyists. Int means that the client is using in-house lobbyists whereas Ext means that the client is using external lobbyists. The table provides information on 6 semesters for the period 2008 to 2010. For the row denoted as semester, I means first semester (January-June) while II means the second semester (July to December).

2.6.3 Figures

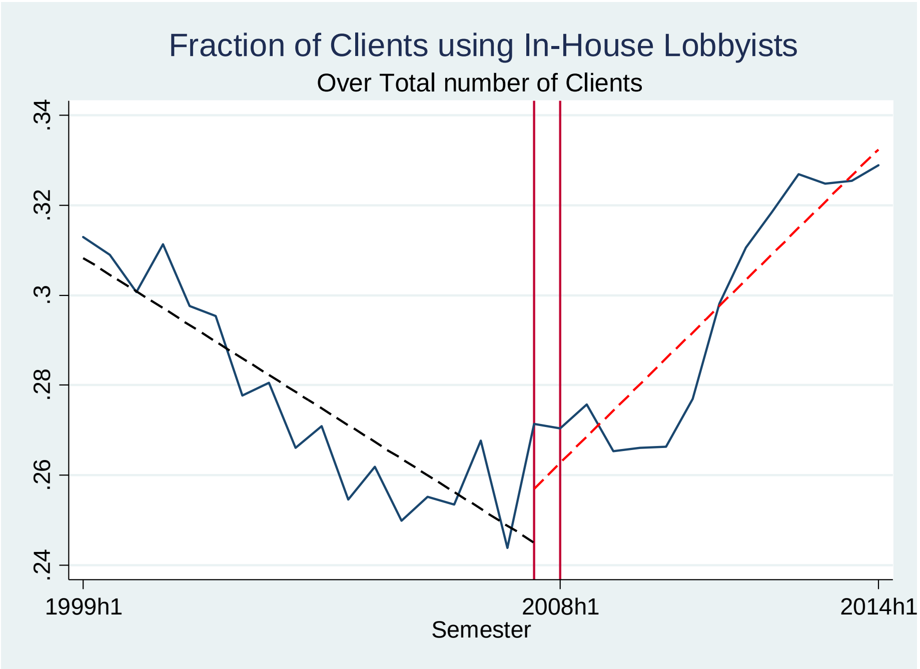

Figure 2.1: Time Series of the Fraction of Clients Using In-House Lobbyists. Note. The two vertical lines denote the second semester of 2007 and the first semester of 2008. An observation is a semester.

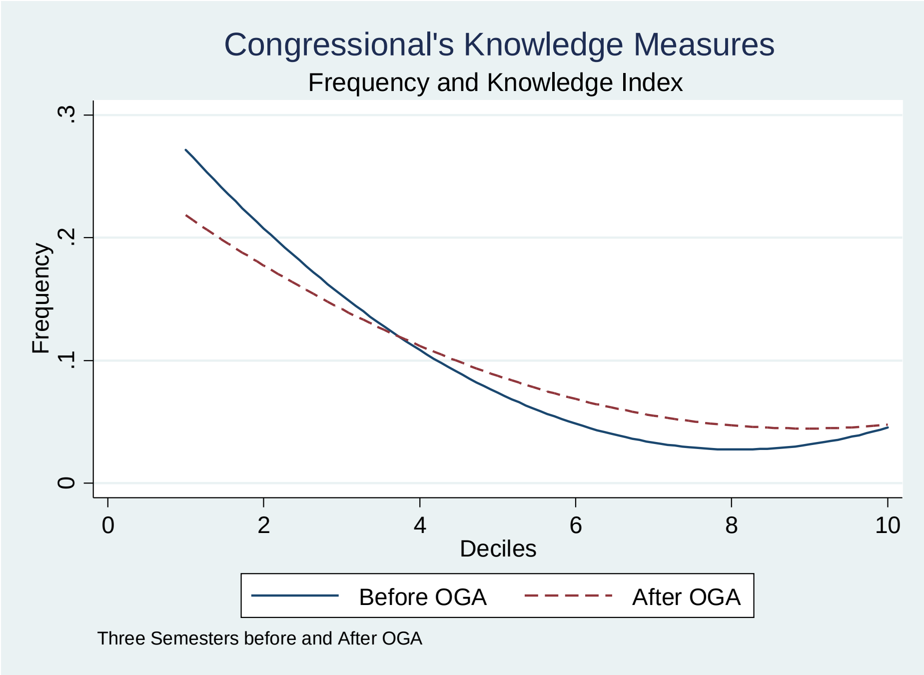

Figure 2.2: Differences in Congressional Knowledge Measures Before and After the OGA. Note: The x-axis represents the deciles of the index and the y-axis shows the normalized number of bills for each decile. The group *Before OGA* represents three semester before the OGA and the group *After OGA* represents three semesters since the OGA.

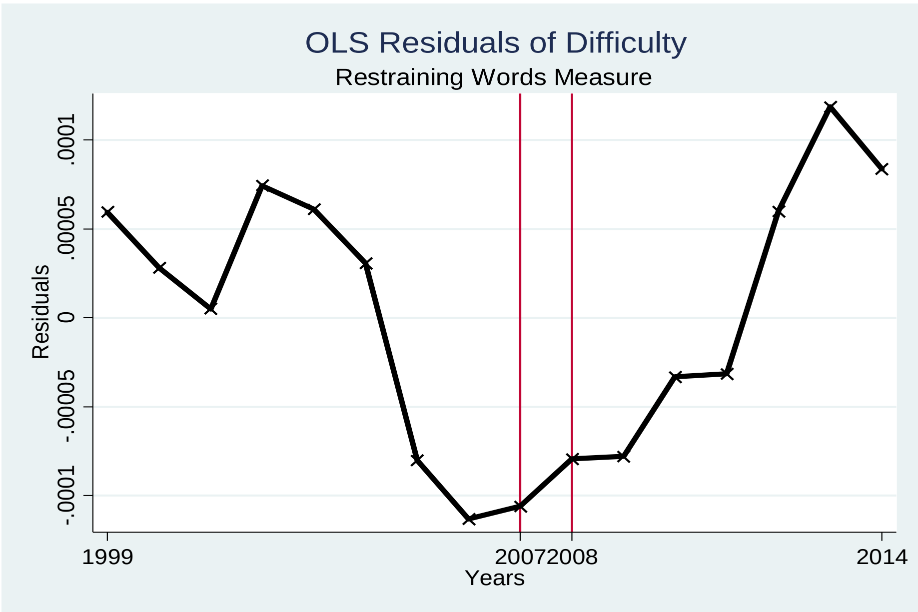

Figure 2.3: Plotted Residuals of the Economy-Wide Difficulty Measure Over Time. Note: Difficulty is measured with the ratio restraining words over regulating words. These calculations include all the industries in the economy. An observation is a year.

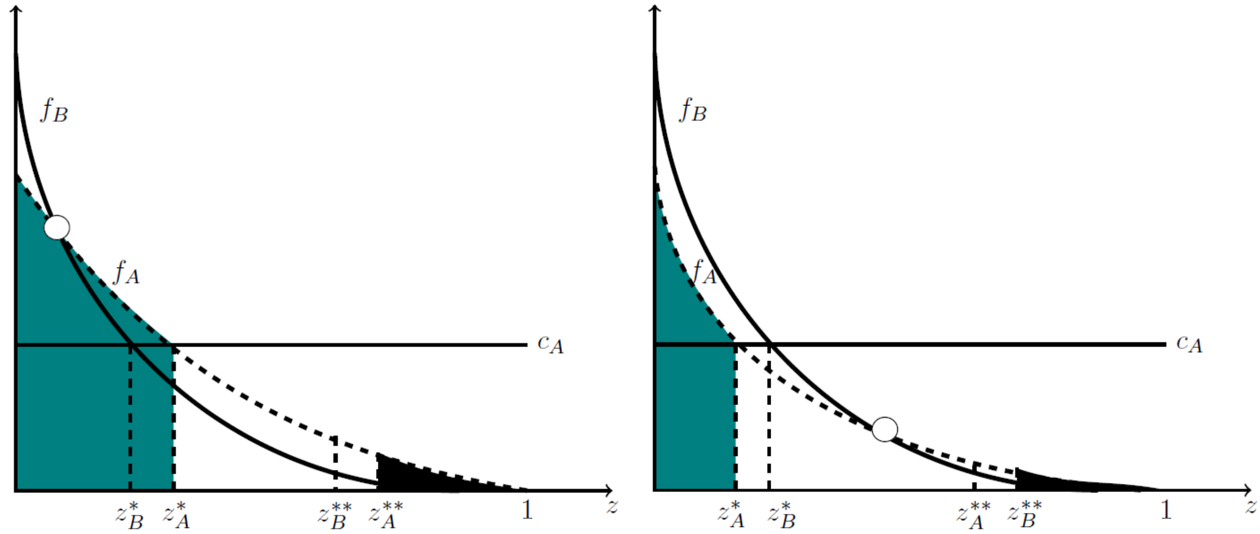

Figure 2.4: Theoretical Predictions of the Effect of a Change in the Frequency. Note: B denotes before the OGA and A denotes after the OGA. LHS: In this case, the densities cut above the cost of acquiring knowledge and the firm boundary shifts to the right. RHS: In this case, the densities cut below the cost of acquiring knowledge and the firm boundary shifts to the left.

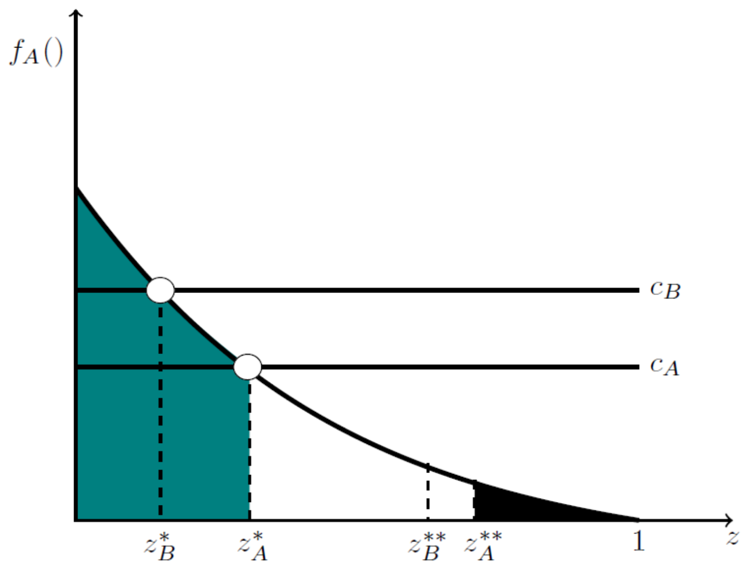

Figure 2.5: Theoretical Predictions of the Effect of a Change in the Costs. Note: B denotes before the OGA and A denotes after the OGA.

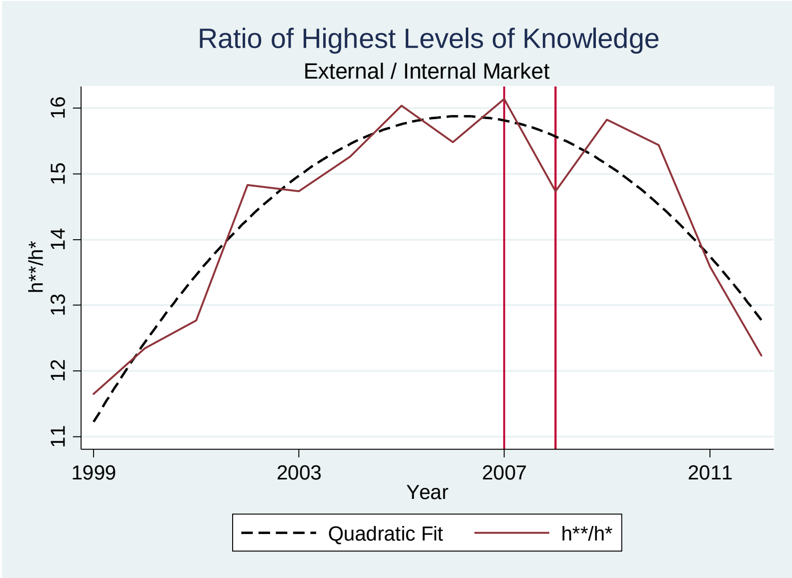

Figure 2.6: Time Series on the Knowledge Inequality. Note: The dashled line represents the quadratic fit whereas the other line shows the fitted line.

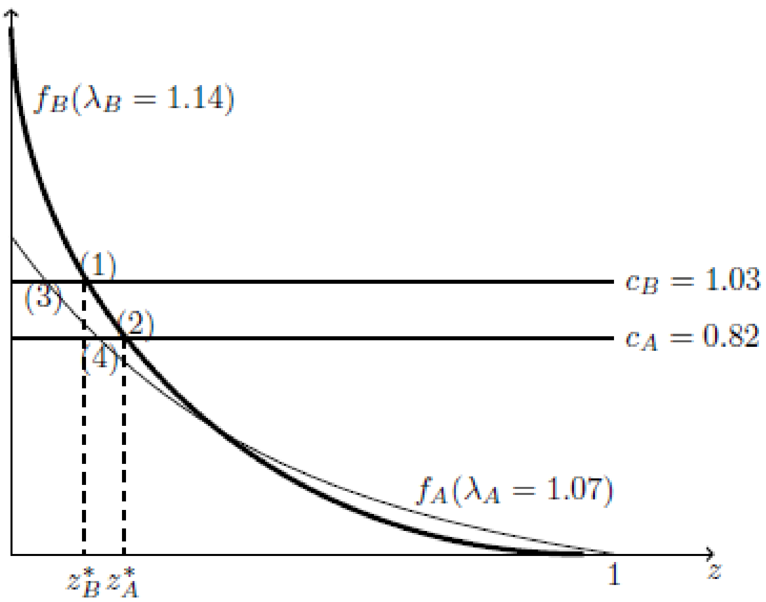

Figure 2.7: Graphical Explanation of the Counterfactual Exercises. Note: B denotes before the OGA and A denotes after the OGA. To simplify the presentation of this figure, I do not present the market boundary levels. However, the counterfactual exercises consider them. Point (1) represents the intersection that defines the firm boundary before the OGA and point (4) the intersection defining the boundary after the OGA. Point (2) holds fixed the density function of problems before the OGA but changes the cost of acquiring knowledge. Point (3) represents the intersection of the acquiring cost technology before the OGA with the density function of problems after the OGA.

2.6.4 Other Patterns observed in OGA

The aim of this section is to provide a clearer description of the main changes observed around the OGA. Graphical evidence that the industry became larger both in terms of clients and number of registered lobbyists is presented the Figure **2.A.1**. The first vertical line denotes the first semester of 2006 whereas the second vertical line shows the first semester of 2008. The main take away from this graph is that the number of clients and lobbyists are characterized by both an increasing trend before 2008 and a constant or decreasing trend after the OGA took effect. We can also observe a change in the slope for the clients line between 2006 to 2008. The growth rate was higher than ever before. Anecdotally, one can connect this change in the slope with the Jack Abramoff scandal.[137] Abramoff is a former American lobbyist who was at the center of an extensive corruption investigation where not only he but also White House officials, one congressman and nine other lobbyists and congressional staffers were convicted of corruption. Abramoff was convicted of fraud and conspiracy -trading expensive gifts, meals and sports trips in exchange for political favors- in the first semester of 2006. [138][139]

2.6.4.1 Transition Matrices

To understand the observed flows of lobbyists around the OGA, Table **2.A.1**, describes a transition matrix of lobbyists. Among the rows there are four different categories: Internal, external, mixed and new lobbyists. Among the columns there are four categories: the first three same categories of the rows and the exit lobbyists. Among each row-category there are three rows. The first one (i.e. #) shows the actual number of lobbyists, whereas the second (third) one shows the percentage of the number of lobbyists as a proportion of the total row (column). The table cuts the sample into two different periods: 3 years before 2008 and 3 years after 2007. Modifying the number of periods before and after provide qualitatively similar results.

In the first period there were about 19.000 lobbyists and in the second period there were 27.000. It seems natural to ask who are the lobbyists leaving and entering the market. Given that the threshold to be a (good) lobbyist increased, I expect to see that the entrant lobbyists has a better set of skills than the exit lobbyists.

[137] https://en.wikipedia.org/wiki/Jack_Abramoff_scandals

[138] It is also interesting to mention that according to several sources, this scandal was one of the main motivations for Congress to introduce, pass and sign the OGA.

[139] The change in the pattern for clients is not trivial to understand as there may be two effects in place: On the one hand, under higher scrutiny from the public, clients that lobby without reporting their activities will be more prone to start reporting lobbying actions. On the other hand, the whole lobbying activity has not only become more difficult but also more disreputable. The first effect will increase the number of clients we observe whereas the second effect will decrease it. The magnitude of these effects is an empirical matter. From this graph, it seems that the first effect has dominated for the period 2006-2008.

In order to test this hypothesis, I use the data based on *lobbyist.info*. This database contain information on previous working experience for the lobbyists. Table **2.A.2** classifies lobbyists in two categories. Lobbyists that leave the lobbying market within three years after the OGA took effect and lobbyists that entered the market for the same period. Changing the time span around the OGA does not change the main results. I focus on three variables: experience in the US Senate, experience as a political counsel, and experience as Chief staff of any politician. For each of these variables, I conduct t-statistic tests to see if there are statistically significant differences between the group of lobbyists exiting the market right after the OGA took effect and the group of lobbyists that entered the market. Among each previous job experience, I conduct this exercise for the lobbyists according to the market they worked on.

The bottom line of this table is that the exit lobbyists have a worse set of advocacy skills than their counterparts.[140] The lobbyists leaving the market have less political working experience.[141] For instance, 12% of the new lobbyists have working experience in the Senate whereas only 7% of the ones that exit had some experience. Similar patterns also emerge for the other two categories: political counsel and chief staff experience. Also interestingly, for the first two variables the differences in the quality can be explained only with the internal lobbyists. For instance, the fraction of internal lobbyists entrants with the Senate experience is about three times larger than the fraction of in-house lobbyists leaving the market with this type of experience. However, the difference for the case of external lobbyists is not statistically significant.[142] Results not presented here also show that internal and external exit advocates tend to have worse skills than the sample of lobbyists that remain in the market.

2.6.4.2 Results for different Issues

In this section, I conduct some empirical exercises to understand the variation of the data focusing on issue-year combinations. At this stage I have abstracted for the complementarity problem but I aim to include it in the future. In this section I proceed in two steps. I first provide empirical evidence there was a change in the lobbying difficulty for each topic when the OGA took effect. Then, I estimate the structural parameters using non-linear squares estimations.

Changes in the Lobbying Difficulty per Issue Most of the federal lobbying activity is destined to affect congressional bills. The policy agendas project (henceforth PAP) classifies the bills proposed in the US Congress in 20 major topics and

[140] Other variables do not turn to be significantly different for both group of lobbyists.

[141] Similar conclusions emerge when I change the three year period window.

[142] For sake of brevity I have decided to not include the numbers for the mixed types lobbyists. This type of lobbyists correspond to a small fraction of both set of lobbyists: entrants and exiters.

220 subtopics.[143] In this classification all the topics and sub-topics are consecutive and the code system is mutually exhaustive and hierarchical (i.e. only one topic for each bill title and every subtopic belongs to a major topic).[144] The Congressional Bill Project has a separate database providing bill-level information for more than 400.000 bills from 1947 to 2012 and use the PAP classification system.[145]

I propose to proxy the level of difficulty of a topic with the level of difficulty of passing bills related to that topic. This is to use better the variation of the data as only some bills are studied in Congressional committees. The idea is that topic A is more difficult than topic B if in average topic A has bills that are more difficult to pass than topic B. Although I can focus on the actual passing bills ratios per topics I prefer to construct an index by using six different variables using Principal Components Analysis[146]: Number of bills where the bill 1. was sponsored by a member of the majority. 2. passed the House debate 3. passed the Senate debate 4. Became law and the inverse of the fraction of bills that 5. received veto 6. were proposed by a first-time congressmen.[147].

I construct these six variables for each of 11 topics for each semester of the period 1999-2012.[148] The index is decreasing in the level of difficulty. That is, the larger the index the easier is to pass bills. If the bills are easy to pass, the lobbying activity is easier. Therefore, the lower the index, the more difficult the topic is.

Given that the first principal component is explaining a large amount of the variance and for the sake of parsimony, I will use this component as the difficulty index.[149]Figure **2.A.2** shows the observations of the first principal component discriminated by issue. For each issue, the figure also shows the mean value and one standard deviation above and below the mean value. Red circles denote observations since 2008 whereas grey circles denote observation before this date.

The figure shows several patterns. First, there is a lot of variation in the dispersion of the observations between topics. Budget and Appropriations and Trade have both large variance in the index whereas communication, taxes and transport

---

[143] The project offers a description about each of these classifications at: http://www.policyagendas.org/.

[144] Congressional Bills project (http://congressionalbills.org/codebooks.html) strives "... *for 90% interannotator reliability at the major topic level, and 80% at the subtopic level.*"

[145] http://congressionalbills.org/index.html.

[146] PCA reduces the dimensionality of the data by providing uncorrelated linear combinations of the variables that contain most of the variance of the intial data.

[147] I have used the inverse instead of the fraction of bills in the last two variables to make easier the interpretation of the index.

[148] The 11 topics I use are: Health, Tax, Finance, Budget and Appropriations, Defense, Energy, Education, Communications, Transport, Trade and other Issues.

[149] As the PCA analyzes the correlation matrix of the variables, each variable is standardized to have unit variance. Therefore, the total variance is the number of principal components (i.e. 6). The first principal component is explaining a great amount of the variance of the data: 72% (eigenvalue of the first component is 4.3 divided by the total variance). Given the first principal component is explaining a large amount of the variance and for the sake of parsimony, I will use this component as the difficulty index.

have low values. Second, Health and budget are the easiest topics to pass bills whereas tax and communications are the most difficult ones. Third, among the positive outlier values of the index (i.e. larger than 1.7) only one point is after the OGA. In addition, it can also be seen that the levels of difficulty after the OGA tend to be below the ones before the OGA.

Figure **2.A.3**. shows the evolution of the index for 10 topics (I exclude the *other* category) and linear predictions for both periods: after and before the OGA. The bottom line of this figure is the change in the pattern of the observation once the OGA took effect. From the first four plots is clear that the level of difficulty start increasing (recall that the index is decreasing in the level of difficulty) since the first semester of 2008. However, the last plot does not show a clear pattern in the change. The figure also show cycles in the levels of the index. The index tend to be larger in the first semester of each year and then in the second semester tend to decrease. To have a more robust empirical evidence of this change in the pattern, I also run fixed-effect regressions. The interested reader can see results in the following sub-section.

Overall this section has shown a change in the difficulty of lobbying activities per topic. However, there are two important shortcomings with this approach. First, for each bill there are two sides: one in favor and one against. The empirical exercise I conducted in this section is showing the difficulty has changed for clients in favor of bills. Of course, the total effect on the difficulty will depend on the distribution of clients in favor and against the bills. Second, not all the lobbying reports have information about the bills. Therefore, I am measuring the effect of the OGA for only a sample of the firms.

Fixed Effect Regressions: Change in the difficulty by Issue To have a more robust empirical evidence of this change in the pattern, I also run the following fixed-effect regression:

$$I_{it} = \alpha_i + \gamma_t + \theta after_t + \varepsilon_t$$

In this regression, an observation is an issue-semester combination and I control for both issue and semester fixed effects. The variable I_{it} represents the difficulty level of the issue i at period t. I include issue dummies, α_i, to control for mean differences in the index across topics, and semester dummies γ_t, that control for index growth common to all issues. The variable $after_t$ takes a value of 1 for periods *after* the second semester of 2007 and 0 otherwise. I cluster standard errors at the issue level. Table **2.A.3** shows the results of the estimated coefficient θ for different combinations in the inclusion of the fixed effects. I include the calculated coefficients when I run the regression without fixed effects, with only one type of fixed effects and when I include both. I report the standard errors of the coefficients as well as the number of observations used, the overall R^2 and the Wald test.

The table shows a robust and consistent pattern. The level of difficulty of the issues increased once the OGA took effect. This result is robust to the combination of the fixed-effect controls we use. The standard deviation of the index is close to 2 points. Therefore, one way to interpret the result of this table is that the index become approximately one standard deviation more difficult after the OGA took effect.

2.6.5 Tables of the Appendix

Transition Matrix for Lobbyists

2005-2007 to 2008/2010						
		Internal	*External*	*Mixed*	*Exit*	*Total*
Internal	#	6854	43	5	4147	11049
	Row (%)	62	0.4	0	37.6	100
	Col(%)	56	0.5	6.7	63.2	
External	#	80	4726	12	3284	8102
	Row (%)	1	58.3	0.1	40.5	100
	Col(%)	0.6	60.4	16	50	
Mixed	#	6	11	12	44	73
	Row (%)	8.2	15	16.4	60.3	100
	Col(%)	0	0.1	16	0.7	
Entry	#	5398	3050	46		8494
	Row (%)	64	36	0.5		100
	Col(%)	44	40	61		
Total		12338	7830	75	7475	
Net Change		1289	-272	2		1019

Table 2.A.1: Transition Matrix for Lobbyists 2005-2010.

Note: An observation is a lobbyist. Among the rows there are four different categories: Internal, external, mixed and new lobbyists. Among the columns there are four categories: the first three same categories of the rows and the exit lobbyists. Among each row-category there are three rows. The first one (i.e. #) shows the actual number of lobbyists, whereas the second (third) one shows the percentage of the number of lobbyists as a proportion of the total row (column). The table cuts the sample into two different periods: 3 years before 2008 and 3 years after 2007. Modifying the number of periods before and after provide qualitatively similar results.

	Senate			*Political Counsel*			*Chief Staff*		
Sample	All	Int.	Ext.	All	Int.	Ext.	All	Int.	Ext.
Exit	7.2	3.6	11.2	5.35	2.22	7.93	2.78	0.92	5.0
Entrants	12.2	13.5	8.0	8.89	6.7	9.95	7.77	6.74	12.7
ttest	0.015	0.000	0.4205	0.049	0.000	0.606	0.000	0.000	0.012

Table 2.A.2: Entrant and Exit Revolving Doors.

Note: An observation is a lobbyist..Table A.2 classifies lobbyists in two categories. Lobbyists that leave the lobbying market within three years after the OGA took effect and lobbyists that entered the market for the same period. I focus on three variables: experience in the US Senate, experience as a political counsel, and experience as Chief staff of any politician. For each of these variables, I conduct t-statistic tests to see if there are statistically significant differences between the group of lobbyists exiting the market right after the OGA took effect and the group of lobbyists that entered the market. Among each previous job experience, I conduct this exercise for the lobbyists according to the market they worked on.

θ	-0.480***	-0.480***	-2.346***	-2.347***
Std Err.	0.148	0.148	0.479	0.480
Issues	*N*	*Y*	*N*	*Y*
Semesters	*N*	*N*	*Y*	*Y*
Obs	308	308	308	308
Overall R^2	0.012	0.650	0.102	0.740
Wald Test	10.49	549.75	107.04	770.44

Table 2.A.3.: Change in the Difficulty Before and After OGA.

Note: An observation is a semester. All the coefficients are significant at 1%. In this regression, an observation is an issue-semester combination and I control for both issue and semester fixed effects. The variable I_{it} represents the difficulty level of the issue i at period t. I include issue dummies, α_i, to control for mean differences in the index across indexes, and semester dummies γ_t, that control for index growth common to all issues. The variable $after_t$ takes a value of 1 for periods after the second semester of 2007 and 0 otherwise. I cluster standard errors at the issue level.

2.6.6 Figures of the Appendix

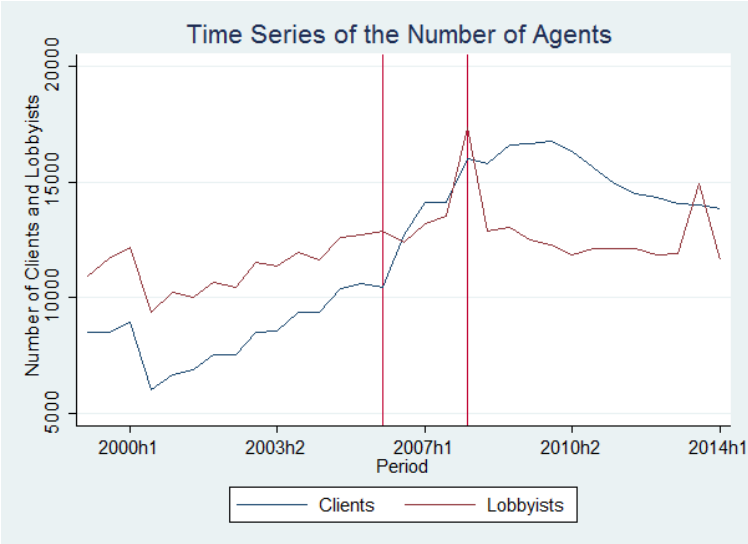

Figure 2.A.1: Time Series on the Total Number of Clients and Lobbyists.

Note: An observation is a semester.

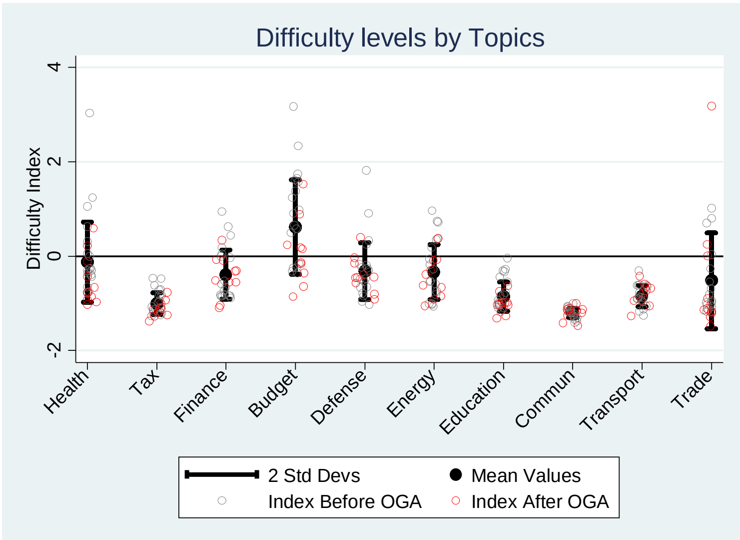

Figure 2.A.2: Values of the Difficulty Index for Each Topic-year Combination.

Note: The figure shows the values of the difficulty index discriminated by topic.

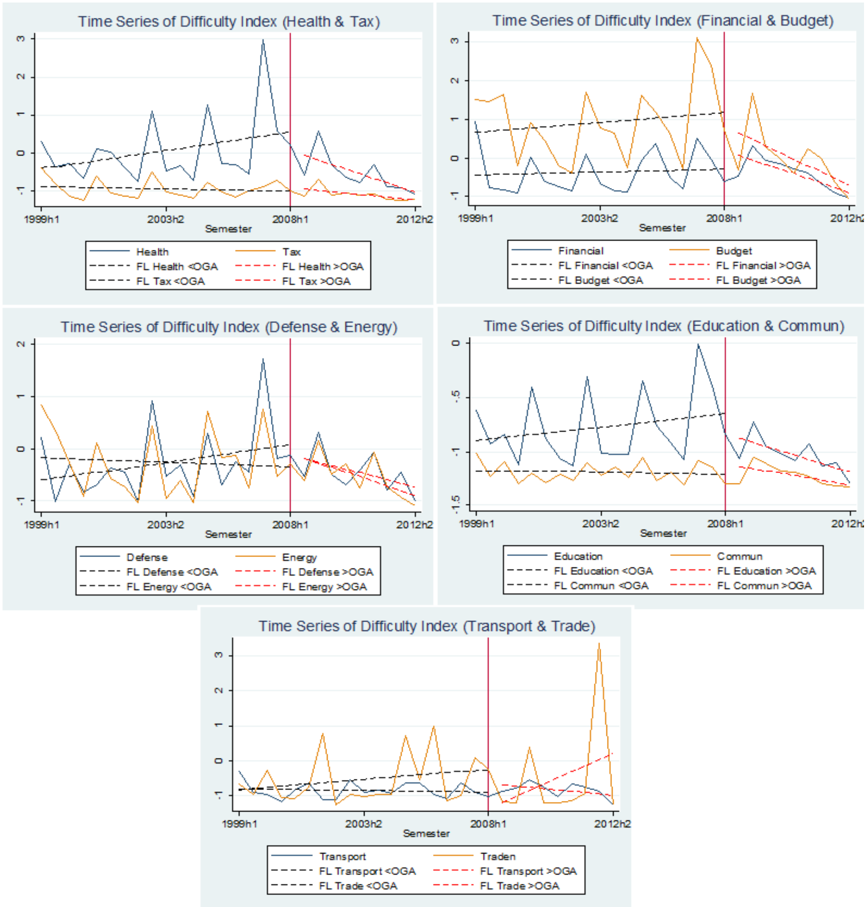

Figure 2.A.3: Time Series on the Difficulty Index for Each Topic.

Note: The figure shows the values of the difficulty index discriminated by topic. An observation is a semester.

Chapter 3

Market Concentration and Lobbying Expenditures[150]

Abstract

The collective action literature predicts that less-concentrated industries spend less than more concentrated industries on lobbying activities. This chapter presents a robust empirical fact that is at odds with this core result. To explain this fact, I include a neglected but, arguably, important dimension in the analysis: the level of excludability of the goods being lobbied. I present examples of excludable US political goals and, using new measures of excludability at the industry level, I show that less-concentrated industries tend to lobby more heavily for excludable goods. The central point is that neglecting the fact that different industries can lobby for goals that differ by their level of excludability can bias the estimates that link market concentration and group efforts. Then, I show that once one controls for the level of excludability in the industry-level lobbying goals, the standard collective action prediction is reestablished. I end this chapter by using national-level mergers as an exogenous source of changes in city market concentration. I show that, controlling for the level of excludability of their advocacy goals, firms that faced these shocks increased their lobbying expenditures disproportionately, providing validity to Olson's seminal prediction.

[150]I found the empirical fact presented in this chapter in 2014. All of the other material presented in this chapter is new. I am deeply indebted for the insightful conversations on this motivating result with Kevin Murphy. The following people provide insightful conversations to make the empirical fact more robust: Scott Ashworth, Marianne Bertrand, Jordi Blanes i Vidal, Tim Bresnahan, Andre Boik, Ethan Bueno de Mezquita, Austin Clemens, Alexia Delfino, Quoc-Anh Do, Liran Einav, Josh Feng, John de Figueiredo, Alfred Galichon, Luis Garicano, Alessandro Gavazza, Matthew Gentzkow, Brent Hickman, Ali Hortacsu, Brian Kelleher Richter, Steven Levitt, Leslie Marx, Michael Peters, Andrea Prat, Mar Reguant, Amit Seru, Jim Snyder, Jr., Paulo Somaini, John Sutton, Chad Syverson, Matt Tady, Richard Van Weelden, Glen Weyl, Hye Young You and Owen Zidar. I also acknowledge the comments and suggestions made by participants at Work in Progress seminars at UChicago and Stanford University. E-mail: m.espinosa@lse.ac.uk.

3.1 Introduction

There has been great interest in understanding the ways in which market structure affects economic and political outcomes. At least since Olson (1965), a particular branch of this literature has focused on individual and group efforts as the main choice outcome. The extensive theoretical collective action literature predicts that groups with fewer members exert higher levels of effort. A great number of empirical studies with very limited span (i.e., time period, number of industries and so on) have found that, in most cases this relationship does not appear significant, and for the cases that appear significant, the evidence is mixed. In this chapter, I present robust empirical evidence for a comprehensive set of industries and periods that is at odds with the theoretical prediction. I measure effort with lobbying expenditures and show that more-concentrated industries spend less on advocacy activities.

To present this puzzle, I use data first presented in Chapter 1, and I exploit the fact that I have been able to match the primary industry code for each of the interest groups that have lobbied at the federal level in the last 20 years. To complement this information, I use different data sources to extract diverse measures of market concentration, such as the total number of firms, concentration ratios, and the Herfindahl index. I show that this empirical puzzle seems robust to the time period chosen, the specific sample of industries used, the number of digits defining the industries, the ways in which we control for the size of the industry, and the way in which we measure market concentration.

To explain this puzzle, I include a new dimension to the collective action analysis: excludability of the goods for which firms lobby. I first present empirical evidence to support the fact that there are numerous examples of distinct political objects that have an important component of excludability. Examples include earmarks, tax breaks, duties giveaways, (legally intentional) fiscal loopholes, loans at favorable rates, price controls, and private bills, among others. To the best of my knowledge, there is no comprehensive dataset with information on all of these types of excludable goods. To advance in this direction, I focus on exploiting a new dataset I have constructed with information on earmarks. I take advantage of a transparency policy in which Congressmen have to report detailed information on earmarks, and I construct empirical measures that quantify the extent to which industries focus on the lobbying efforts for these political objects. I show that once one controls for this dimension, Olson's theoretical prediction is recovered.

To explain how the excludability of the goods being lobbied for can help to explain the empirical puzzle, I develop a theoretical framework in which firms face bills with different levels of excludability and choose whether or not to lobby and, if they decide to lobby, the magnitude of their lobbying expenditures. The key insight is that as the incentives to exert effort are increasing in the private benefit, excludable goods have larger associated lobbying expenditures. Then, I provide empirical evidence on less-concentrated industries lobbying more intensively (i.e., both at the extensive and intensive margins) than more concentrated industries for excludable goods. This implies that the observed positive relationship between the number of the firms and industry-level lobbying expenditures can be easily explained by the fact that different industries lobby for goods that differ by their level of excludability.

I end this chapter with an empirical exercise that allows me to explore the relationship between market structure and lobbying expenditures with exogenous shocks to industry-city concentration levels. The key idea is to explore how national-level mergers modify city-level lobbying expenditures for firms that are not involved in the merger but are affected differentially by changes in market structure. Mergers have a differential impact across cities, as the firms involved in the mergers have a presence only in a subset of these locations. Controlling for the level of excludability of the goods for which firms are lobbying, I find significant and robust evidence showing that firms and industries that face an increase in city market concentration (i.e., after a merger, there are fewer firms) increase their city-level lobbying expenditures.

This chapter makes three contributions. First, I use a newly compiled dataset to provide robust empirical evidence against the standard collective action prediction: more-concentrated industries tend to spend less money than less-concentrated industries on advocacy activities. Second, I provide a possible explanation for this empirical puzzle based on the excludability of the goods that are being lobbied for. I start by presenting some examples of excludable political objects that are lobbied in the US to show the empirical relevance of this dimension in the analysis. Then, I develop a simple theoretical framework that includes new theoretical predictions. Importantly, this model not only generalizes some of the most important models in the collective action literature, but, to the best of my knowledge, is also the first one to include the excludability dimension and to provide predictions on its effect on both the intensive and extensive margins of lobbying activities. Third, I compile a new dataset of city-level lobbying expenditures and study how these are affected when there are changes to the national-level market structure. To the best of my knowledge, there has been no other attempt to provide exogenous variation in market structure in order to measure its effect on city-level lobbying expenditures.

Related Literature

At least since the seminal work of Olson (1965) and Stigler (1971, 1974), academics have asked how special interest groups affect public policy.[151] Recently, there has been increasing academic interest in the specific way in which special interest groups affect public policy through lobbying.[152] However, to the best of

[151]For recent overviews on the topic Grossman and Helpman (2002a) and Ansolabehere, de Figueiredo, and Snyder (2003).

[152]See, for example: Richter et al. (2009), Blanes i Vidal et al. (2012), Bombardini and Trebbi

my knowledge, the focus on market concentration and lobbying expenditures has been neglected. Naturally, there are some papers measuring the relationship between special interest groups' characteristics and political influence.[153] But there are two main differences between these papers and my approach: 1. Some of them try to understand how firm or market characteristics are related to public policy outcomes. I focus only on lobbying expenditures. It makes more sense to focus on choice variables rather than on outcomes that depend on unknown interactions. 2. Among the papers focusing on these choice variables, all of them analyze the case of political campaign contributions (henceforth PAC) rather than lobbying measures. Empirically, lobbying expenditures are quantitatively more important than PAC expenditures, as they are, on average, six times bigger.

This chapter is divided into five sections. Section 3.2 presents the main empirical fact that motivates this chapter. In Section 3.3, I briefly present the theoretical framework that allows me to analyze the effect on the excludability of the benefits of the political goals. Here, I also present empirical evidence on the relevance of the excludability dimension by showing its relationship with market concentration, as well as its power to recover the standard collective action prediction. Section 3.4 takes a further step and presents some empirical evidence on the lobbying reports at the city level, showing that firms that face an increase in their city-industry market concentration increase their lobbying expenditures. Finally, in Section 3.5, I finish with a short discussion summarizing the main results and limitations of the chapter, and propose further developments.

3.2 Empirical Fact

In this section, I present the main empirical fact of this chapter: controlling for different measures of the size of the industry, less concentrated industries spend more on lobbying efforts than more concentrated industries.

3.2.1 Data Used

For this exercise, I use different data sources. On the lobbying expenditures side, I use the same data as in the two previous chapters and I refer the interested reader to seek the description of the data in Chapter 1. On the concentration measures, I use information from Orbis, Compustat and the Economic Census.[154]

(2012), Bertrand et al. (2014), Clemens (2014), Karam (2014), Kerr et al. (2014) and de Figueiredo and Richter (2014).

[153] See, for example: Pittman (1976, 1977, 1988), Salamon and Siegfried (1977), Mann and McCormick (1980), Esty and Caves (1982), Zardkoohi (1985), Andres (1985), Masters and Keim (1985), Munger (1988), Boies (1989), Snyder (1990), Grier, Munger and Roberts (1991, 1994), Mizruchi and Koenig (1988, 1991), Potters and Sloof (1996), Mitchell, Hansen, and Jepsen (1997), Hansen and Mitchell (2000), Schuler, Rehbein and Cramer (2002), Hansen, Mitchell and Drope (2004), Ansolabehere, Snyder and Ueda (2004), Ozer and Lee (2009), Macher, Mayo and Schiffer (2011), Vannoni (2013), Weymouth (2013).

[154] Orbis is a global database containing information on nearly 150 million companies worldwide, with an emphasis on private company information. In this section, I focus on US companies'

For each of these databases, I quantify market concentration with five different measures: total number of firms (henceforth n), three concentration ratios (henceforth C_4, C_8 and C_{20}) and the Herfindahl index (henceforth HHI).[155] I calculate these four measures using the Orbis data with the total number of employees for each firm for the year 2012. With Compustat, I calculate this information using data on number of employees for the period 1998 to 2014. Finally, I use the concentration measures that the Economic Census releases every five years. The Census gives these five concentration measures for all of the manufacturing industries. For non-manufacturing industries, the US Census Bureau only releases the first 4 concentration measures. For this data, I focus on the most recent available year: 2012. I obtain qualitatively similar results when I look at different years.

Each of these three data sources provide information for industries defined at the level four, three and two digits 2007 NAICS codes. Table 3.1 summarizes the data used showing the main data source, variables used, and the dates available. Tables 3.2 and 3.3 present descriptive statistics of the data used. Table 3.2 presents statistics for the cross section data (i.e., Orbis and Economic Census) and Table 3.3 presents information for the panel data, based on the Compustat database. As expected, the average number of firms grows as we use broader definitions of industries. Correspondingly, the HHI and the concentration ratios decrease monotonically as we move from four to two digits industries.

3.2.2 Main Result

In order to estimate the relationship between market concentation and lobbying expenditures, I conduct the following set of estimations. First, I use ORBIS and the Economics Census databases to run:

$$\frac{L_i}{X_i} = \alpha + \beta_1 MC_i + \varepsilon_i \qquad (1)$$

where L_i is the total lobbying expenditures for i-th industry and MC_i is one of the aforementioned concentration measures. To control for the size of the industry, I divide the total expenditure of the industry by proxies of industrial scale X_i such as total turnover, sales or employees. In this chapter, I present the results for the case in which X_i is measured by total sales. I obtain qualitatively similar results when I use different measures for industry size. In addition, with these databases I run:

information for the year 2012. I use compustat database annual information from January 1998 to January 2014 on active and inactive publicly held companies, comprising more than 28000 active and inactive firms in North-America. Finally, I use data for the Economic Census for 2012, 2007, 2002 and 1997. This data is classified and organized by the North American Industry Classification System (NAICS). For all of these databases I focus only on US firms.

[155] Let $S_j = \sum_{i=1}^{j} S_i$, where S_i is a size measure (sales, employees and so on) of firm i-th. Firms are organized by size, that is $S_i \geq S_{i+1}$. Then, $C_4 = \frac{S_4}{S_n}$ and C_8 and C_{20} are defined accordingly. Finally, $HHI = \sum_{i=1}^{n}(S_i/S_n)^2$.

$$\frac{l_{ji}}{x_{ji}} = \alpha + \beta_2 MC_i + \varepsilon_i \qquad (2)$$

where l_{ji} is the lobbying expenditure of j-th firm in i-th industry and x_{ji} is a measure of its size. Finally, I use Compustat to run:

$$\frac{L_{it}}{X_{it}} = \gamma_t + \beta_3 MC_{it} + \varepsilon_{it} \qquad (3)$$

where L_{it} is the lobbying expenditure of i-th industry in t-th year, X_{it} is as before, a measure of size, and γ_t represents time-fixed effects. I also include estimates at the firm level:

$$\frac{l_{jit}}{x_{jit}} = \gamma_i + \gamma_t + \beta_4 MC_{it} + \varepsilon_{it} \qquad (4)$$

where l_{jit} is the lobbying expenditure of j-th firm in i-th industry at t-th period and x_{jit} is a measure of its size. Table 3.4 to Table 3.7 report the main results of these exercises for the case that industries are defined with four and three digits code. I obtain qualitatively similar results when I define industries according to the two-digit NAICS classification. Table 3.4 shows the coefficient and standard errors for β_1 whereas Table 3.5 shows the results for β_2. Results for β_3 and β_4, are presented in Table 3.6 and Table 3.7, respectively. To facilitate the interpretation of the results, the independent variables are normalized by their standard deviation.

These tables show some robust patterns. First, looking at the significance of the coefficients, we see that all of the estimated coefficients are significant at least at 10% for equations (1) and (3), whereas *most* of the coefficients are significant for equations (2) and (4). Second, the four tables show two robust patterns. First, less concentrated industries spend more on lobbying. That is, controlling for the scale of the industry, an industry that is more concentrated (fewer firms, or larger levels of C_4, C_8, C_{20} or HHI) tend to have lower reported levels of lobbying expenditures than a less concentrated industry. Second, firms in more concentrated industries spend more on lobbying. That is, if we observe two firms with similar size, the firm in the more concentrated industry tends to spend more on lobbying than the firm in the less concentrated industry.

Orbis and Compustat also provide firm-level information on assets, total turnover and sales. In results not presented here, I construct the concentration measures using these alternative variables and I found the same patterns shown in Table 3.4 to Table 3.7. As an additional robustness check, I have also run the same estimations including other controls such as industry-level elasticities of substitution from Broda and Weinstein (2006) or measures of geographic and political concentration from Busch and Reinhardt (1999). These patterns are robust to the inclusion of these controls. In the next section, I propose a candidate explanation for this result.

3.3 Excludability of the Benefits

In this section, I argue that the introduction of the excludability of the benefits of political goals allows us to rationalize the empirical patterns observed in the previous section. This section is divided into three parts. I will first start explaining what I mean by excludability in the benefits and I will provide some examples. Then, I move to a simple theoretical model, that will allow me to have clear predictions on the relationship between market concentration, and firm and industry-level lobbying expenditures. I will explain how neglecting this dimension may bias the previous empirical results. I will end this section by providing empirical evidence on the relationship between market concentration and excludability of the benefits.

3.3.1 Types of Bills

The literature on lobbying has not emphasized a simple fact. Firms lobby for different types of bills. Although bills are complex, multidimensional objects, I think there are at least two dimensions that are of first order concern in my analysis: excludability and/or rivalry of the bills. Each combination of these two variables affects the benefits of the firms in a different way. As a consequence, the decision whether to lobby or not, how much money spend on it, and the effect of lobbying expenditures in the economy as a whole are very different.

3.3.1.1 Excludable Bills

Excludable bills are bills in which it is possible to prevent certain agents from enjoying the benefits to the bill. In particular, it is interesting to think that these excluded agents are non-lobbying units. Notice that there are two natural levels to think about this excludability problem: firm and industry levels. The first level assumes it is possible to prevent non-lobbying firms within a beneficiary industry to receive the benefits, while the second level assume that it is possible to exclude non-lobbying industries from receiving the benefits that a bill provides.[156]

Althought there are several examples of excludable bills, I want to briefly comment on only three: Earmarks, Tax breaks (or duties suspension) and Private bills.

Earmarks. A typical example of excludable bills are earmarks. An earmark is a legislative money provision that directs approved funds to be spent on specific projects, or that directs specific exemptions from taxes or mandated fees. There has been an extensive debate around these legislative objects. In fact, earmarks to for-profit institutions were banned in March 2010 by the House Appropriations Committee. Just to have a sense of how important these expenditures are, according to the Office of Management and Budget of the White House, the total amounts given by earmarks in Congress for the years 2008 to 2010 were 39.3, 39.2 and 37.8

[156]In addition, it may be difficult to explictly assess what a lobbying unit means as it may be necessary to include firms that have not only federally lobbied but that also are politically active in other dimensions (i.e., grassroot lobbying, PAC, SuperPACs, 527 groups and so on).

US billion dollars, respectively.[157]

Tax Breaks. The main difference between this category of excludable bills and earmarks is that the first directs benefits to specific units while the second provides benefits to a group of agents with certain defined characteristics. Tax breaks are mainly of two types: Transitory or permanent, although the first type is more common. In December 2014 the US Congress passed the bill H.R. 5771 with more than 50 tax breaks targeting specific groups and industries such as small companies, public transportation commuters, and teachers who spend their own money on classroom supplies.[158] Congressional estimates calculate that this bill is worth over 42 US billion dollars of federal budget.[159] Some recent examples of tax breaks for businesses in other legislative objects are tax credits for R&D, tax exemptions for financial companies with foreign profits, provisions to allow retailers to write off capital investments easily, and breaks for racehorse owners, film, TV, alternative fuel, and rum producers.

Private Bills. These bills are "designed to provide legal relief to specified persons or entities adversely affected by laws of general applicability." [160] Private laws apply only to the person named in the law and grant a benefit from the government to that person, not otherwise authorized by law. An alternative House of Representatives' definition is "A private bill is a bill for the relief of one or several specified persons, corporations, institutions, etc., and is distinguished from a public bill, which relates to public matters and deals with individuals only by classes."[161] No House rule defines what bills qualify as private, but most private bills have official titles stating them to be "for the benefit of" named individuals. Subjects of contemporary private bills include: Immigration and claims matters (is the most common subject), Patents and copyrights, Taxation (e.g., income tax liability, tariff exemptions), Public lands (e.g., sales, claims, exchanges, mineral leases) among others. Most of these bills have been targeted to specific individuals in recent years.

3.3.1.2 Rival Bills

Rival bills are bills whose consumption by one firm prevents simultaneous consumption by other firms. Typical examples of rival bills are law initiatives that give a fixed monetary subsidy to an industry. The more firms there are in the industry, the less each firm gets. Rival bills can also be thought of as bills that provide congestible

[157]For more information, see here. Earmarks have wrongly been associated as a synonym of Pork barrel legislation. Although they are similar in several dimensions, they are not exactly the same, as there is a more objective determination in the spending of an earmark, while pork-barrel spending tends to be more subjective. However, with the aim to know the order of magnitude of these expenditures, according to the organization Citizens Against Government Waste, the average pork barrel spending in the years 1998-2010 was of the order of 19.2 billion US dollars, which is about half of the size of the amount of the estimated earmarks.

[158]For more information, see Journal News.

[159]For more information, see H.R.5771 of 113th.

[160]For more information, see Oleszek et al. Congressional Procedures.

[161]For more information, see Hinds' Precedents of the House of Representatives.

goods. Typical examples of congestible bills are private or public parks.

Examples of non-rival bills are subsidies as a proportion of any observable scale of the firm or tax breaks. Although it is probable that the number of firms was a variable that legislators took into consideration when they set the percentage of the subsidy or the size of the tax break, the benefit received by one company does not affect the subsidy received by another firm.

3.3.1.3 Some Examples

Stigler (1971) thinks there are four policies that industries lobby for: 1. Direct Subsidy or Money. 2. Control Over New Entrants (by price policies or vertical integration). 3. Affecting Substitutes or Complements Goods. 4. Price Fixing. Although there may be several particularities about how these policies are defined, in general, policy one can be excludable within an industry whilst policies two to four are non-rival policies. To follow, I provide one example of each combination of rival/excludable bills.

Non Rival/ Excludable: *Schrimp Importation Financing Fairness Act. (H.R. 155 of the 108-th Congress).*[162] This bill prohibits the Secretary of Commerce from imposing any new restrictive regulations on the domestic shrimping industry within the area that is under the jurisdiction of the Gulf of Mexico Fishery Management Council. Besides that, this bill eliminates the financial help available that the United States provide to other countries exporting shrimp to the country. This is excludable at the industry level and non rival within the beneficiary industry. In this case, the bill is restricting the competition of production of Shrimp in the United States. So, only that industry and only the firms producing schrimps in that geographic region are actually benefited (excludable) and the fact that one schrimp producing firm in that region receives the benefit, does not imply that other firms with similar characteristics receive less from that benefit (non-rival).[163]

Non Rival/Non Excludable: *A bill to designate Taiwan as a visa waiver program country. (S.1545 of the 112-th Congress).* These legislative proposals are by definition public goods: they are neither rival nor excludable. The bill aims to avoid Taiwan citizens to require a visa to legally enter the United States. The

[162]Another example is the Middle Class and Small Business Tax Relief Act of 2012 H.R.6262 of 112 th Congress. The bill makes provisions of a 2003's act that reduce the tax rate on dividend and capital gains income for taxpayers whose incomes do not exceed the base amount permanent (i.e., $200,000 for individual taxpayers and $250,000 for married couples filing jointly). Increases to 20% the tax rate on dividend and capital gains income for taxpayers whose incomes are above the base amount. Of course, this bill exclude people from their income but for every person who is in the aceptable income range, has a tax relief. Another example is the bill S.1808 of the 99th Congress. The bill amends the 1954's IRC to exempt from taxation trusts which acquire and manage real property for certain exempt organizations. More information about this bill can be consulted here. Details of thie Schrimp bill can be seen here.

[163]There are basically 4 regions of Shrimp production New England, South Atlantic, Gulf and Pacific. The Gulf contains about 90% of the total production. More information about this industry can be consulted here. Information of this bill can be found here.

beneficiaries of the bill are Taiwan citizens. The bill does not exclude the waiving privilege to a certain subset of Taiwan citizens so it is non-excludable. As the benefit that one Taiwanese person receives does not decrease the future benefit of another Taiwan citizen, the bill is also not rival.[164]

Rival/ Excludable: *Transportation Infrastructure Grants and Economic Reinvestment Act. (S.942 of 112-th Congress).* This bill directs the Secretary of Transportation (DOT) to establish a national infrastructure investment program with a finite budget to provide competitive grants (of \$10-\$500 million), secured loans, and loan guarantees to a state, local government, or transit agency for eligible transportation projects. This bill benefits all the firms in the construction of transportation infrastructure industry, so it is excluding by industry. This bill is also clearly rival not only because the allocation of grants is competitive, but the total amount of the programme is bounded.[165]

Rival/ Non Excludable: *Harriet Tubman National Historical Parks Act. (S.247 of 112-th Congress).* This bill establishes that in some parts of Dorchester and Maryland a historical Park. This park does not charge any fees, so the park is a non-excludable good. However, as with other common goods, it is congestible. In particular, the park has some small historical sites, that get crowded easily.[166]

I acknowledge that the key requirement, to be able to classify bills within these four categories, is to distiniguish the good offered. For instance, in the second example, I assumed that the good offered was visa waiving to Taiwanese people. However, if I would have considered that the good offered was visa waiving rights, then although the good is non rival it is excludable, because it is only providing the rights to Taiwanese citizens. The aim of this subsection was to provide simple examples in which the levels of excludability and rivalry of the bills differ. The specific way in which these levels are defined is not relevant for the rest of the chapter.

3.3.2 Theoretical Framework

To understand better the difference in the incentives for lobbying units among these combinations, let us consider a simple model. Let l_i be the lobbying expenditure of the i-th firm and let G be the benefit of the bill. Let $c(l_i)$ be an increasing function representing the cost of lobbying l_i dollars and $\widetilde{G}$ the benefit of the bill when the i-th firm does not lobby.[167] Finally, let n be the number of firms in the industry

[164] The details of this bill can be consulted here.

[165] Information of this bill can be found here.

[166] Information of this bill can be found here.

[167] At this stage the reader should recognize that this is quite general. For instance, if there are budget constraints, firms can borrow money to lobby, then the cost to a lobbying firm is just $(1 + r)l_i$ where r is the interest rate. If the lobbying units are measured in personal effort or time, or even income but the capital markets are imperfect, then it may make sense to think that additional units of effort are increasingly costly, making the cost function a convex one. If the convexity is explained by the imperfections of the credit market, the distorsion will be higher for

and n_l is the number of firms lobbying. The number of firms lobbying is defined as $n_l : \left\{ n | \frac{G}{n_l} - c(l_i) \geq 0 > \frac{G}{n_l+1} - c(l_i) \right\}$.

A firm will lobby if the net benefit of doing so is larger than the net benefit of not lobbying. Table 3.8 shows the net benefit of lobbying and the net benefit of not lobbying according to the level of rivalry and excludability of the bill being lobbied. The first row differs from the second in that the benefit of the bill is spread out to other firms in the case the firm is lobbying, while the first column predicts a zero payoff for non-lobbying firms. Intuitively, when goods are rival, firms have a smaller incentive to lobby because they do not receive fully (proportionally) the benefit of the good. On the other hand, non-excludable goods provide bigger incentive to free-ride than excludable goods. Notice that if the good is excludable, firms will lobby as long as $G > n_l c(l_i)$, while if the good is non-excludable, firms will lobby if: [168]

$$G - \widetilde{G} > nc(l_i) \qquad (*)$$

3.3.2.1 Collective Action Problem

The typical story assumes that the benefits of lobbying are non-excludable goods. That is the goods are either rival (common goods) or non rival (Public Goods). For most firms, especially the small ones, the costs of lobbying will outweigh the marginal benefits of lobbying. So, as market concentration tends to be an increasing function of the average size of the firm in the industry, Olson suggests that concentrated industries are more likely to overcome this problem. To be more specific, Olson's prediction is that as n is higher in less-concentrated industries, there is less lobbying in these industries because equation $(*)$ tend to not hold. There are two parts to this explanation and each of them is changing one side of inequality $(*)$.

The first explanation is that the RHS of inequality $(*)$ increases as n increases. The second explanation is that more concentrated industries tend to have fewer firms, so we expect to see that the actions of each of these firms has a stronger impact on the others firms in the industry. As this impact is measured by $G - \widetilde{G}$, we expect to see that this difference is an increasing function in the level of concentration in the industry. Then, less concentrated industries have a lower LHS of $(*)$. The prediction, then, is that more concentrated industries will lobby more. If you also add to this intuition the fact that firms in more concentrated industries tend to be richer (so they have more money available to spend) then the stylized fact exposed in the previous section becomes a puzzle.

---

the poorer (less concentrated industries). Then, the cost for the less concentrated industries firms will rise more steeply with expenditure. In this chapter, I abstract from these issues.

[168] That is, for the case of excludable goods, if the firms lobby for the case in which the good is rival (i.e., $G > n_l c(l_i)$) they will also do it when the good is not rival (i.e., $G > c(l_i)$). For the case, the good is not excludable, if firms lobbying when the good is rival, they will do it when the good is not rival because $\frac{G-\widetilde{G}}{n} > c(l_i) \rightarrow G - \widetilde{G} > c(l_i)$ for a fixed $n > 0$ and l_i.

3.3.2.2 A model: *There ain't no such thing as a free lunch*

The typical model used to understand the collective action problem does not include the excludability dimension in the analysis. In this sub-section, I propose a way to do it. All of the proofs can be found in the Appendix of this chapter.

Setting Let $\pi(L)$ be the probability of passing the bill, where $L = \sum_{j=1}^{n} l_j$ is the total lobbying expenditure of the industry and assume the benefit of the bill depends on the total lobbying expenditures, that is $G(L)$. I assume $\frac{\partial \pi}{\partial l_i}, \frac{\partial G}{\partial l_i} > 0 \; \forall i : 1, ..., n$. As before, let n be the number of firms in the industry. This will be the measure of market concentration. The benefit that each firm receives depends on the divisibility of the benefit. For instance, if $G(L)$ is purely congestible good, the firm will receive $\frac{G(L)}{n}$; whereas, if $G(L)$ is non congestible at all, the firm will receive $G(L)$. Then, a convenient way to represent these extremes as well as intermediate cases where the benefit is partially divisible is $\frac{G(L)}{n^{\beta}}$ where $\beta \in [0, 1]$ is an exogenous constant that represent how rival the bill is. If $\beta = 1$ the bill is purely rival and if $\beta = 0$ the bill is non-rival.[169]

To include heterogeneity in a simple way, let us assume that the expected benefit for the i-th firm is simply $\frac{\pi(L)G(L)}{n^{\beta}} S_i(n)$, where S_i is a variable that measures the scale of the firm, such as the sales level or the number of employees that depends on the total number of firms in the industry.[170] In the Appendix, I discuss how we can relate this variable with the firm size distribution. Finally and inspired by Mitra (1999) and Kerr et al (2014), let $c_i(l_i)$ be the cost function depending on the i-th firm. This last assumption allows firms to differ by organizational abilities, relationships of the CEO's of the firm with lobbyists and senators, previous political activities (inertia in lobbying), frictions in the lobbying entry etc.

The excludability of the bill can be due to several factors. Here I focus on the case of empirical relevance, that is, when bills are excludable, only lobbying firms receive the bill's benefits. Let us denote n_θ as the number of firms sharing the benefit of the bill. Two extreme cases seem relevant. If the bill is excludable, then n_θ is equal to the number of firms lobbying, n_l; whereas n_θ is equal to the total number of firms in the industry, n if the bill is perfectly non-excludable. Then, a representation of this term is n_θ as the convex combination of the total number of firms in the industry and the number of firms lobbying, that is $n_\theta = \theta n + (1 - \theta) n_l$ where $\theta \in [0, 1]$ is a dummy variable equals to one if the bill is perfectly non-excludable and zero if it is perfectly excludable. When the bill is neither perfectly excludable not perfectly non-excludable $\theta \in (0, 1)$, and therefore, $n_\theta \in (n_l, n)$. Notice that $\sum_{j=1}^{n} l_j = \sum_{j=1}^{n_l} l_j$ as $l_j = 0$ for non-lobbying firms.

[169] Borcherding and Deacon (1972) and Bergstrom and Goodman (1973) were the first papers to use a similar specification to represent the divisibility problem.

[170] Hillman et al (2004) has an excellent review of the variables employed in empirical studies to explain political participation of firms. He concludes that the firm's size is the most common and perhaps the best proxy of the scale of the firm to explain political participation.

The objective function for the i-th firm is:

$$\frac{\pi(\sum_{j=1}^{n} l_j)G(\sum_{j=1}^{n} l_j)}{n_\theta^\beta} S_i\left(n\right) - c_i\left(l_i\right)$$

Notice that if the bill is excludable then the expected benefit is only shared by the lobbying firms, however, if the bill is not excludable the benefit is shared among all the firms in the industry.

The timing of the game is as follows. First, there is a random draw of the bill and its characteristics, β and θ. This information as well as n and the functions π, G, S_i and c_i are public knowledge. The model has two stages. In the first stage of the game, firms decide whether or not to lobby. This stage is related to the *extensive margin* of the lobbying game and can be understood as a stage where we analyze the probability of lobbying. From this stage we know the identities of the firms willing to lobby and the number of these firms (i.e., n_l).

In the second stage, the firms that have decided to lobby in the first stage make a choice about the level of the lobbying expenditures by maximizing their objective function. Each firm takes the lobbying expenditures of other firms as given. This stage is related to the *intensive margin* of the lobbying game and it characterizes the equilibrium amount of lobbying expenditures for each of the lobbying firms. The problem is solved by starting from the second stage. An equilibrium in this game is the number of firms lobbying n_l and a n_l-dimensional vector of firm-level lobbying expenditures.

3.3.2.3 Main Results

I start with the extensive margin.[171]

Theorem 3.1 *Firms are more likely to lobby when they are bigger, when they can be affected more by the bill, when the firm can affect more the expected benefit of the bill, if the lobbying costs are smaller or when the bills are more excludable and/or less congestible.*

The positive relationship between size and lobbying participation has been empirically explored in several papers, for instance in Bombardini (2008), Macher et al. (2011), Hill et al. (2013) and Kerr et al. (2014). Notice this is a simple way to see the Olson (1965)'s famous "*exploitation of the large by the small.*" Hill et al. (2013) provide evidence that firms lobby most actively if there are greater potential payoffs from favorable policy and regulations. That is, it will be more likely to see firms lobbying, when either the benefit of the bill or the probability of passing the

[171]Let $\sum_{j=1,j\neq i}^{n} l_i = L_{-i}$ be the total lobbying expenditure of all firms except firm i-th. Taking $n_l - 1$ firms lobbying, let firm i-th decide whether to lobby or not. Then, firm i-th will lobby if $\frac{[\pi(L)G(L)-\theta\pi(L_{-i})G(L_{-i})]S_i}{n_\theta^\beta} > c_i\left(l_i\right)$. Notice that as far as $\pi(L_{-i})G(L_{-i}) > 0$ lobbying for the non-excludable good imply lobbying for the excludable good, *ceteris paribus*.

bill are larger.[172] To the best of my knowledge, the predictions about excludability and rivalry are new.[173]

Now I provide a result on the effect of market structure on the extensive margin decision. Let $\in_b = \frac{b_i'(n)n}{b_i(n)}$ with $b_i(n) = [\pi(L)G(L) - \theta\pi(L_{-i})G(L_{-i})] S_i$, then:

Lemma 3.1 $\frac{\partial Pr(l_i>0)}{\partial n} = \in_b - \beta\frac{\partial n_\theta}{\partial n}$

This result simply says that if the expected benefit is sufficiently sensitive to the number of firms (i.e., large $\in_b$), less concentrated industries have a larger probability than more concentrated industries of having firms lobbying. I include a discussion of this lemma in section 3.6.3.2. Now, I start by presenting two simple results on the intensive margin decision. The interested reader can see more theoretical results not presented here, in the Appendix of this chapter. I start by presenting a result that links firm- and industry-level lobbying expenditures with market concentration. First, I assume that the lobbying expenditures are symmetric, that is $l_i(n) = l(n)$, $\forall i : 1, ..., n$.

Theorem 3.2 $sign(\frac{\partial L}{\partial n}) = sign(1 - \in_{l,n})$ *where* $\in_{l,n} = -\frac{nl'(n)}{l(n)}$.

This result states the relationship between the elasticity of firm- and industry levels lobbying expenditures.[174] Notice that the average lobbying expenditure is increasing (decreasing) if $l'(n) > (<)\frac{l(n)}{n}$. There are three relevant regions. Two in which the industry-level lobbying expenditure is increasing in the number of firms and one in which is decreasing. When the firm-level lobbying function is increasing then the industry-level function is always increasing. However, when the firm-level

[172]The prediction on the lobbying costs is intuitive but to the best of my knowledge it has not been empirically explored.

[173]A discussion with David Baron has called my attention to the following not obvious effect. First, notice that the equation defining the extensive margin decision also works in a case where we analyze the lobbying decision with respect to the status quo benefits. Second, and more importantly, firms will be more likely to lobby when the bill negatively affects them than in other cases. To see these facts, let $G_{sq} \geq 0$ be the benefits that the firm has in the status quo. Then, the firm can decide whether to lobby or not. If it lobbies it will get G_l, while if it does not lobby, it will get G_{nl}. Then, the firm will lobby if $G_l - G_{sq} \geq G_{nl} - G_{sq}$. This is the same condition stated above. To see the second result, notice that the incentives to lobby may be different if the firm can increase its profits or if the firm tries to avoid a decrease in its benefits. To understand this, note that if $G_{nl} < G_{sq}$, the RHS of the extensive margin inequality is negative, so as far as $G_l - G_{sq} \geq 0$, the firm will always decide to lobby. However, if $G_{nl} > G_{sq}$, $G_l - G_{sq} \geq 0$ it is not a sufficient decision to make firms lobby. It will also depend on $G_l - G_{nl} \geq 0$. This simple idea shows that **it is more likely to see firms lobbying when they are going to be adversely affected by the bill than when they can be positively affected.**

[174]If $l'(n) \geq 0$, $\in_{l,n} < 0$ and then $\frac{\partial L}{\partial n} \geq 0$. On the other hand, if $l'(n) < 0$, $\in_{l,n} > 0$, then we need the elasticity to be inelastic to have $\frac{\partial L}{\partial n} > 0$ and elastic to have the reverse case. Inelastic would mean the average lobbying expenditure is larger than the marginal lobbying expenditure (i.e., $\frac{l(n)}{n} > -l'(n)$). Notice also that if there is no relationship between the firm-level lobbying expenditure and the number of firms $l'(n) = 0$, then the firm-level lobbying elasticity is zero $\in_{l,n} = 0$ which implies that $\frac{\partial L}{\partial n} > 0$. This is the pure effect of having a large number of firms in the industry. Finally, notice that $\frac{\partial L}{\partial n} = 0$ if and only if $\in_{l,n} = 1$. The elasticity reachs this value when $l'(n) < 0$ and the marginal lobbying expenditure is exactly equal to the average expenditure.

lobbying function is decreasing the elasticity of the firm-level function (i.e., $\in_{l,n}$) will determine whether the industry-level lobbying function is increasing in the number of firms or not.

Figure 3.1. shows these three regions. In the first region, $l'(n), L'(n) \geq 0$. Region 2 represents the case in which $l'(n) < 0$ but the marginal expenditure is larger than the mean expenditure, therefore, $\in_{l,n} > 1$ and as a consequence $L'(n) < 0$.[175] Finally, the last region represents the case in which $l'(n) < 0$ and $L'(n) \geq 0$. This last region is the only region representing the empirical results of Section 3.2. Less concentrated industries spend more than more concentrated industries, and firms in industries with more firms spend less than firms in industries with fewer firms. The take away here is that with symmetric lobbying expenditure functions, the only way to rationalize the empirical fact from Section 3.2 is when $\in_{l,n} < 1$, that is when firm-level lobbying expenditures are inelastic to the number of firms in the industry.

The following result states the relationship between the elasticities when I allow for heterogeneity in lobbying expenditures. A thorough discussion on the differences between symmetric and asymmetric lobbying expenditure functions can be found in the Appendix of this chapter. Let $w_i = \frac{l_i}{L}$ be the proportion of the i-th lobbying expenditure to the total industry lobbying expenditure, $\in_{l_i,n}$ is the elasticity of the lobbying expenditure of the i-th firm to the number of firms of the industry and $\in_{l_i,L}$ is the elasticity of the firm-level lobbying expenditure of the i-th firm to the industry-level lobbying expenditure, then

Theorem 3.3 $sign(\frac{\partial L(n)}{\partial n}) = sign(\sum_{i=1}^{n} w_i \in_{l_i,n}) sign(1 - \sum_{i=1}^{n} w_i \in_{l_i,L})$

This result simply states that the total industry lobbying expenditure elasticity depends on all the individual elasticities of both the number of firms and the total lobbying expenditure weighted by their relative size on the lobbying expenditure. One implication of this equation is that even if most of the companies tend to decrease their lobbying expenditures as the number of firms in the industry increases $\frac{\partial l_i}{\partial n} < 0$ and these firms are very sensitive to the total lobbying expenditure (i.e., $\frac{\partial l_i}{\partial L} > 1$), if there are just a few big companies in the industry for which $\frac{\partial l_i}{\partial n} > 0$ and $\frac{\partial l_i}{\partial L} < 1$, the total industry elasticity may be positive.

Now I state, two results for the intensive margin. Assume there is a single lobbyist whose income is L, and who is going to choose a signal s to send to the politicians she wants. Then, in the same spirit of Bernheim and Whinston (1986) the lobbyist is maximizing the total lobbying expenditures received $L(s)$.[176] In equilibrium, total lobbying expenditures depend on the number of firms lobbying $L(n)$. To follow, I focus on characterizing the symmetric equilibrium schedule lobbying functions.

[175] If $l(n)$ is a concave function in n, then $l'(n)(n' - n) \geq l(n') - l(n)$ for all $n\prime, n$. If $n\prime = 0$, then $l'(n) \geq \frac{l(n)}{n}$.

[176] The reader should notice that the extension to the case where the lobbyist care about other variables such as her reputation is straightforward.

Theorem 3.4 *The lobbying expenditure (truthful schedule) $l(n)$ of one firm is given by* $\max(c^{-1}(-ng'(n), 0)$.

The lobbying expenditure of one firm compensates for the reduction in the gross welfare of the other lobbying firms brought about by the decision of the firm to lobby. Now, I present the key result of this section.

Lemma 3.2 $\frac{\partial l(n)}{\partial \theta}, \frac{\partial L(n)}{\partial \theta} < 0$.

This result states that the lobbying expenditures both at the firm and industry level are larger when they lobby for bills that are more excludable. That is, holding everything constant, more excludable bills have larger associated lobbying expenditures.

Discussion of the Model This model provides a generalization of some of the most important papers in the literature. In section 3.6.3.1, I explain how my model generalizes Olson (1965), Stigler (1974), Esteban and Ray (2001) and Pecorino and Temimi (2008). The most important take away here is that none of these papers consider excludable goods at all.

I am not the first one to study the relationship between excludability and collective action. Chamberlin (1974), among other authors, points out that Olson's proposition of an inverse relationship between effective collective action and group size hinges on the assumption that the collective good is purely non-excludable. However, these authors do not provide a theoretical framework to understand its dynamics.

The typical collective action story does not emphasize that the size of the bill can be affected and completely neglects that the probability of passing the bill can be manipulated. [177] If a firm does not lobby, it incurs in two costs: 1. The probability that the bill is passed is reduced, so the expected gain is also reduced. 2. In certain cases, the size of the benefits of the bills can be affected by lobbying. So, the size of these benefits is also reduced, decreasing again the expected gain.[178] **Rides are not free**. When firms do not lobby, firms change their expected benefits, that is not only the size of the pie but the probability of seeing the pie. To the best of my knowledge, my model is the first one including not only the excludability dimension but the fact that *rides are not for free.*

Main Implication To understand better why the excludability of the bill can explain the main empirical fact presented in the previous section, let's consider Figure 3.2. This graph represents the relationship between industry-lobbying expenditures and number of firms in the industry for two different levels of excludability of the

[177] For more information on the limitations of the collective action theory, see Wilkerson, Smith and Stramp (2015)

[178] According to Stigler (1974) this idea was realized by Harold Demsetz.

bill. The right (left)-hand decreasing function represents the case of lobbying expenditures for excludable (non-excludable) goods.

The graph represents a situation in which less concentrated industries have lower expenditures, just as Olson's main prediction. For the sake of argument, imagine we start in point A. This point corresponds to the combination (n, L) (i.e., an industry with n firms spending L dollars). Now, let's imagine we see a different market structure, that is a horizontal move from n to $n^* > n$. This change can be due to a change in the market structure of the industry that was located at point A or simply a different industry with more firms. The relationship between lobbying expenditures and the number of firms in the industry will imply that we have to reach point B, which is (n^*, L'). However, as excludable bills have a higher lobbying expenditure for all levels of market concentrations, as we stated in the last result of the previos sub-section, we can change the market structure and be in a point like C, that is (n^*, L''). This can happen for instance, if the change in the market structure may have also changed the type of bills the industry is interested in lobbying for. This point has the same market structure but it has a higher lobbying expenditure $L'' > L'$.

Then, the main implication of this analysis is that if different market structures lobby for legislative objects with different excludability levels, the main empirical finding of the previous section can be explained by the simple fact, that less concentrated industries tend to lobby for more excludable goods. The aim of the next sub-section will be to provide empirical evidence in favor of this hypothesis.

3.3.3 Empirical Evidence

In this section, I provide empirical evidence that firms in less-concentrated industries tend to lobby more heavily for excludable goods. In order to conduct this exercise, I use information on earmarks. Earmarks are appropriations of the federal budget that members of Congress give to firms or organizations. Deciding on the recipient necessarily means excluding other potential firms or organizations.

I base my empirical exercises on information publicly released by Taxpayers for Common Sense (henceforth *TCS*). This database contains the recipients of earmarks for the period 2008 to 2010. TCS started creating this database using information released by members of Congress, during a time when a transparency policy was in place. Under this policy, politicians had to release the description and amount of money committed to a project, as well as the main requester of the earmark. TCS expanded this database by including earmarks that were not reported by any member but that meet the definition of an earmark.

I have webscrapped the industry of the main recipients of the earmarks and have matched the earmark information with market concentration measures from the Economic Census. Table 3.9 shows descriptive evidence on the relationship between industry market concentration and earmarks recipients. Here, I measure market concentration with the total number of firms in the industry. The main results of this section are robust to alternative ways of measuring market concentration. To facilitate the interpretation of the table, I have standarized the levels of concentration by quintiles.[179] Larger quintiles represent less-concentrated industries. The main takeaway from this table is the sharp concentration of recipients among larger quantiles. For each year of the database, at least 63% of the recipients are in the two largest quintiles. This is more than expected, as these two quantiles represent only 40% of the total number of industries. It is also remarkable that one out of three recipients is in the largest quintile.

I have also conducted these cross-tabulations for the sample of recipients that were *also* lobbying in the years before the earmarks were awarded. The conclusions are robust. Most of the recipients that were lobbying belong to industries with larger number of firms.

An alternative way to confirm the extent to which industries differ by their interest in earmarks is to use information about all of the bills that are lobbied and to see the extent to which the bills with earmarks are a large or small fraction of the total number of bills. Unfortunately, TCS does not report the exact bill that allocated the earmark money. To overcome this limitation, I have webscrapped all of the information about earmark declarations.[180]

For the period 2008 to 2010, policy makers had to report detailed information about the earmark. This exercise, contrary to TCS' information, provides me with information about the specific bill that introduced the earmark. The limitation is that I include only earmarks that have been declared by congressmen, and, contrary to TCS' information, I would not be able to see the universe of those.

The main results of this exercise are given in Table 3.10. This table is divided into two panels. Panel *A* shows the fraction of bills with earmarks lobbied over the total number of bills lobbied for each quintile and year. For instance, the table shows that among all of the bills lobbied in the less-concentrated industries (i.e., Quintile 5), in 2008, 55% of these legislative proposals had an earmark attached. Panel *B* shows the fraction of expenditures on bills that had any earmark over the total lobbying expenditure. The table shows that about 60% of the lobbying expenditures by the less-concentrated industries were targeted to bills that contain earmarks. Panel *A* of this table shows that less-concentrated industries lobby more for bills with earmarks than for other types of bills. The percentage for the largest quintile ranges from 55% to 62%. In contrast, this percentage is not larger than 24% for the first three quintiles. That is, more-concentrated industries do not spend

[179] To make this standarization, I calculate the quintiles on the total number of firms for all of the four-digits code industries reported by the Economic Census. For these analyses, I use 2007 data and no other year, for two reasons. First, it is more accurate, given that the earmark data are from one to three years after the Census information; and, second, not all of the industries have information for 2012, but all of them have them information for 2007.

[180] For more details on the main source of the earmark declarations, see **here**.

more than one fourth of their advocacy efforts on bills with earmarks. Panel B shows similar patterns. The least-concentrated industries always devote more than 52% of their expenditures to bills with earmarks, whereas this percentage is never greater than 24% for the case of the first three quintiles. The main takeaway from Table 3.10 is that, if we measure the level of excludability in lobbying activities with either the number of bills with earmarks or the money spent on those bills, less-concentrated industries tend to disproportionately exert more effort than more concentrated industries on excludable bills.

3.3.3.1 Omitted Bias

All of these results suggest that the main empirical fact of Section 3.2 can be explained using the omitted variable bias logic. To better understand the problem, imagine that we want to estimate the lobbying expenditures equation as follows:

$$L_i = \beta_1 MC_i + \beta_5 Ex_i + \varepsilon_i$$

where L_i are the lobbying expenditures of the i-th industry; MC_i is the level of market concentration of the industry; and Ex_i is the level of excludability of the goods for which industry i is lobbying. In Section 3.2, we ran the estimation above without considering a variable for the level of excludability. However, as the theoretical section has shown, the level of excludability of the goods explains heterogeneity in the lobbying expenditures. Therefore, we may be estimating:

$$\widehat{\beta_1} = \beta_1 + \beta_5 \gamma,$$

where γ measures the correlation between MC_i and Ex_i. The evidence given in Tables 3.9 and 3.10 suggests that $\gamma < 0$, whereas the theoretical section argues for $\beta_5 > 0$; therefore, we have that:

$$\widehat{\beta_1} < \beta_1$$

To investigate this possibility, I focus on market concentration data for four-digit industries from the Economic Census of 2007, as well as on lobbying expenditures for 2007. To measure the excludability dimension, I use 2008 earmarks data from TCS.[181] I propose a measure of excludability based on congressional committees, in the same spirit of Chapter 1. For each committee, I calculate the fraction of bills that receive an earmark. Then, to have a measure at the industry level, I simply calculate the average (weighted by the number of bills) of these fractions over all of the committees that study bills for that industry.

Table 3.11 shows the results of this exercise. Here, I present the estimates of the coefficient when I include only market concentration measures (n, C_4, C_8, C_{20}

[181] Unfortunately, this seems the most plausible estimation at the industry level. Orbis data are from 2012, and the most recent year with earmark information is 2010.

and HHI) and when I include both the concentration measures and the measure of excludability. When I include only the concentration measures, I obtain results that are qualitatively and significantly similar to the ones shown in Section 3.2 -in particular, Table 3.4.[182] Less-concentrated industries tend to spend more on lobbying efforts. However, the inclusion of the excludability dimension changes the sign and significance of the market concentration effect. First, the results for the market concentration measures are less significant. For instance, measures C_8 and C_{20} are not significant, while measures n and HHI move from being significant at 95% level to 90% significance level. Second, the sign of the coefficient is completely reversed, implying that, controlling for the level of excludability, more-concentrated industries spend more on lobbying, just as in the standard Olson's prediction.

These results support the idea of the omitted variable bias. In Section 3.2, the estimated coefficient is about -0.01, whereas it is about 0.01 in Table 3.11. This is explained by the fact that the -0.02 difference is equal to the coefficient of the excludability on market concentration $\widehat{\beta}_5$ =0.04 and the correlation between market concentration and the excludability measure, which is about -0.5.

3.4 Identification

As a final empirical exercise in this chapter, I investigate if the negative correlation I found between market concentration and lobbying expenditures in Section 3.2 can be interpreted causally. One way to investigate this possible causation effect is by using *Difference in Difference analysis.* The idea is to explore a national-level exogenous shock to the market structure of an industry that affects city markets differently. Its effect is manifested by changes in lobbying expenditures. Standard econometric estimations measuring the relationship between market concentration and advocacy efforts suffer from reverse-causality problems. In this setting, this concern is implausible since it is difficult to argue that city-level lobbying expenditures can change the national-level market structure for a certain industry. This is the case as city-level lobbying expenditures have very concrete goals that target changes in the status quo of *local* as opposed to *national markets.*

3.4.1 Data Used

There are two main sources of data: reports of lobbying expenditures at the city level and, a mergers database.

3.4.1.1 Lobbying Expenditures

I have collected city-level lobbying expenditures for firms in different industries. I focus on the nine biggest cities that have lobbying disclosure requirements allowing me to identify the client, lobbyist, lobbying expenditure, main issue for discussion

[182]The main difference between these estimates and the ones in Table 3.4 is that there, I use data for the Economic Census of 2012, whereas in this table, I use data for the 2007 census.

and time of the lobbying activity.[183] There are other cities in the US with disclosure requirements, but some do not make the information publicly available, are too small (i.e., scarce advocacy activities), or do not require lobbyists to register relevant information such as the specific client or the amount of lobbying expenditures.

Table 3.12 presents the main features of this database. The collected data shows large geographical dispersion, as I have cities in states as distant as California, Washington, New York and Florida. The cities make available lobbying reports starting from different years. Some have been reporting data for long periods, whereas others have started more recently. For instance, there have been data available on New York City since 1998, while for Chicago, the data have been available only since 2012.

In five of these cities, lobbyists have to make these reports every quarter; in three cities, they have to make reports every time that they start a contract; and only in San Francisco, lobbyists have to make these reports monthly. For comparability, I focus on time variation at the quarter level.[184]

3.4.1.2 Mergers Database

I use the Thomson Reuters SDC Platinum database.[185] This database provides detailed information on worldwide mergers, acquisitions and alliances. The database contains both US and non-US companies' targets. This database is updated daily and uses, as its source, Securities and Exchange Commission filings with their international counterparts, over 200 news sources in English and other languages, trade publications, and others sources. The database contains all public and private corporate transactions involving at least 5% of the ownership of a company, where the transaction was valued at $1 million or more or where the value of the transaction was undisclosed. The initial database contains 823,337 observations of mergers and acquisitions of more than 220 countries for the period 1995 to 2016. The database reports information about the location and industry of both target and acquiring companies. Besides information on the target and acquirer, each observation contains a small summary of the history of the merger, information on the status of the operation, total shares acquired (if the merge was successfull otherwise the number of shares promised to be transferred), the value of the transaction, and other financial information such as the forecast value of the companies, as well as sales and revenues after taxes.

This database also provides the month of the announcement of the merger, as well as the month in which the merger was effective. The status of the merger has

[183]These nine cities are Chicago, Dallas, Los Angeles, Miami, NYC, Philadelphia, San Diego, San Francisco and Seattle.

[184]In the case of contracts with a duration longer than a quarter, I assume that the expenditure is divided uniformly across months, and I allocate the corresponding fraction to each quarter in which the contract was available. For instance, if there is a contract from January to April for $400, I will register $300 for the first quarter and $100 for the second quarter.

[185]For more information, see Thomson Reuters.

three different categories: Category 1 (77%) is when there was a successfull merger. Among the observations in this category, 91% contain a merger in which 100% of the targeted company was acquired. For the baseline estimations, I will focus on this sub-sample. Category 2 (3%) represents unsuccessful mergers -that is, circumstances in which the parties intended to merge, but it did not happen. Finally, Category 3 (20%) groups all the other cases that have neither been successful nor withdrawn.

Table 3.13 shows statistics descriptive of this database classified for three different levels of industry aggregation: two-, three- and four-digit SIC codes. The first part of the table shows the mean and standard deviation (between and within) for the number of mergers by category. The bottom part of the table contains information on the successful mergers. For this part, I focus on only three variables: total value of the mergers at the industry level; average percentage of shares acquired; and average value of the transaction. This last variable is simply the total value of all the mergers divided by the total number of successful mergers, for each industry.

3.4.2 Estimation

With this database, I am interested in understanding how changes in market concentration brought about by national-level mergers and acquisitions affect both firm- and industry-level lobbying expenditures. In order to investigate this relationship, I estimate:

$$l_{jict} = \gamma_{ic} + \gamma_{it} + \gamma_{ct} + \beta_6 D_{ict} + \varepsilon_{jict}$$

where j index firms, i industries, c cities and t quarters. The model provides full nonparametric control for industry-specific city effects (γ_{ic}), time-varying industry effects (γ_{it}) and city-specific time effects that are common across industries (γ_{ct}). The variable l_{jict} is the firm-level lobbying expenditure. The coefficient β_6 measures the relationship of interest. D_{ict} takes the value of 1 for mergers and acquisitions that affected industry i in city c at period t. Here, if $D_{ict} = 1, D_{ics}$ is also 1 for all of the periods $s > t$. The variable takes the value of 0 otherwise.

A second estimation of interest is to run the above equation with a dependent variable at the industry level. That is, L_{ict}. As in Section 3.2, this is simply the sum of all the firm-level expenditures of industry i. For these exercises, and with the aim of obtaining cleaner results, I consider only the lobbying expenditures of the firms that were not involved in the merger. These estimations cluster the standard errors at the industry-city level.

3.4.2.1 Selection

In the following exercises, I focus on observations where both the target and acquiring firms are based in the United States and belong to the same two-digit industry, and I examine them only for the years 2012 to 2016. Furthermore, I restrict the set of mergers to the ones in which there is a transfer of 100% of the shares, and the total value of the transfer is above 95% of the population of transfer values.[186]

A possible empirical exercise is to classify all the city-level political goals in a dimension of excludability and create an empirical measure of the intensity of the excludability goals at the local market level. This seems a challenging exercise as there is not a clear way to classify some lobbying objects. In this chapter, I focus only on the sample of lobbying reports in which there is a clear excludable component in the lobbying goal. Examples of those are earmarks, tax breaks, duties giveaways, (legally intentional) fiscal loopholes, loans at favorable rates, price controls, and private bills, among others. In exercises not shown here, I have conducted the same estimations as presented below for the non-excludable goals sample of lobbying activities and I found qualitatively similar results.

Tables 3.14 and 3.15 present information on the 15 unique mergers that conform to all of the criteria above. Table 3.14 provides information on the specific names of the companies in the merger, as well as the industries to which the companies belong. The information on the industry is shown with both the two-digit SIC code and the specific name of the industrial sector. Table 3.15 shows the specific month and quarter in which the merger took effect. The quarter with more mergers is the last quarter of 2014, while the quarter with fewer mergers is the second quarter of 2014. The first quarter of 2014 and the second and third quarters of 2013 each have three mergers. This table also provides the specific value of the transaction, which ranges from 550 to 9.700 million dollars. Finally, the table provides information on the total number of cities affected by each merger. A city is affected if both of the firms involved in the merger were conducting lobbying activities in that city for at least one year prior the merger. This number ranges from three to five cities. On average, a merger affects four cities.

Table 3.16 presents descriptive information on the city-level lobbying expenditures. This table presents the total number of firms and the two-digit industrial codes for each city. The number of industries ranges from 61 in San Francisco to 78 in New York City. The average number of firms actively lobbying in a given quarter ranges from 387 in San Diego to 621 in New York City. Finally, the table presents the average firm-level lobbying expenditure per quarter for each of the nine selected cities. San Diego and Dallas have the lowest average firm-level lobbying expenditures, while Chicago and New York City have the highest.

3.4.3 Results

Table 3.17 presents the main results of the above estimations. The first two columns of this table present the estimated coefficient β_6. The first column presents the results for the case in which the unit of observation is firm-level lobbying expenditures, whereas the second column presents the results for the case in which we consider

[186]Percentile 95% corresponds to US 2300 million dollars.

total industry lobbying expenditures.

For both dependent variables, the coefficient is positive and significant, meaning that both firms and industries that faced an increase in their market concentration (through a merger that affected their *local market* -industry*city-) increased their average lobbying expenditures. The average firm-level lobbying expenditure is about 6200, while the point estimate is about 690. This represents an increase of about 11%. For the case in which the dependent variable is the industry-lobbying expenditures, this increase is about 10%, as the point estimate is about 9900 and the average value industry-level lobbying expenditures are 102000.

To get a sense of these magnitudes, note that, on average, each local market (an industry-city combination) has seven firms. Therefore, a merger, which decreases the number of firms by one, represents a decrease in the total average number of firms by about 14%=1/7. Putting these estimates together, we see that a decrease in the number of firms by 14% increases the firm and industry-lobbying expenditures by about 11%. That is, the relationship between lobbying expenditures and market concentration is inelastic.

3.4.3.1 Robustness

The previous estimation is the baseline estimation in this section. This estimation is robust to different changes in the selection of the sample. In Table 3.17, I also present the results with different samples of mergers. For all of these cases, I focus on mergers in which both the target and acquiring firms are companies based on the United States and only for years 2012 to 2016. I conduct several exercises. First, I consider the largest mergers for companies that belong to different industries, and I focus on analyzing the effect of the firms that belong to the same industry as the *acquiring* firm. That is, the point estimates present the effect of lobbying expenditures for firms that belong to the same local market of the acquiring firm. The intuition is that local markets face a change in the market structure not by means of a change in the number of firms, but by a change in the size of one of the companies in this market. There are 37 mergers with those criteria. Columns 3 and 4 of Table 3.17 present the results of these exercises, showing that the coefficients for both firm and industry estimations are positive and with a similar magnitude as before, but not significant. This means that there is no statistical evidence to imply that companies increase their lobbying expenditures when one of the companies in the same local market has acquired a company in another industry.

An alternative exercise is to focus on firms that belong to the same industry but to change the definition of a merger. In the baseline estimation, we consider mergers in which there was a 100% transfer of shares and that belong to transaction values of at least the 95th percentile of the total distribution of values. Columns 5 and 6 of Table 3.17 present the results when I also include mergers in which there was at least an 80% transfer of the shares. In columns 7 and 8, I present the results when I include transfer values of at least the 90th percentile of the distribution of values. For both of these exercises, the coefficient is still positive and significant. This implies that the selection of mergers in terms of percentage of transfer of shares and value of the transaction do not have an impact on the results in this section. In results not presented here, I show that I obtain qualitatively the same results, when I use different thresholds for the two aforementioned variables.[187]

For all the previous estimations, I estimated the impact of a merger on *other* firms belonging to the same local market of the firm in the merger. Columns 9 and 10 of Table 3.17 show the results when I also include the merged firms in the calculations. The results are still significant and positive. However, the point estimates are smaller than the effects found in the baseline estimations. Overall, Table 3.17 present evidence on the positive robust effect of mergers on city-level lobbying expenditures.

3.4.3.2 Parallel Assumption

An important assumption in the difference-in-difference methodology is the parallel trend assumption before the shocks. In order to test this assumption, I conduct a modified version of the baseline estimation by changing $\beta_6 D_{ict}$ by $\sum_{\tau=t-3}^{t+3} \beta_\tau D_{ic\tau}$. In this case, $D_{ic\tau} = 1$ if $t = \tau$ and 0 otherwise, and $D_{ict} = 1$ in the quarter in which the i-th industry at city c faced a merger. Table 3.18 presents the results of this exercise. As in the previous tables, I present the results for both firm and industry-level lobbying expenditures. The table shows that the coefficient estimates for the periods before the merger are positive but not significant, meaning that there are no statistical differences between the treated and control groups before the market concentration shocks. On the other hand, the coefficients for the period of the merger and the following two quarters are significant, implying a differential effect on lobbying expenditures for local markets that faced a merger. For the case of the firm-level estimations, the effect of the merger remains up to the second quarter after the merger. However, for the case of the industry-level estimations, this effect persists for three quarters afterwards. Overall, the main message from Table 3.18 is that the assumption of parallel trends holds and the effect of market concentration on lobbying expenditures has an impact in the same quarter that the merger occurs, as well as half a year after.

3.4.3.3 Placebo Tests

I end this section by conducting alternative estimations. First, I consider the case in which the observed mergers occur at a date different from the one observed. For

[187] I have also conducted estimations in which I consider mergers in which there is a transfer of at least 70, 75, 85, 90 percent of the shares, and I obtain the same results. These results are also robust when I try different percentiles for the distribution of transfer values, such as 75, 80, 85. Finally, one can also try mergers defined by combinations of these two thresholds, and the main results still hold.

each existing merger, I change the actual date of the merger to a random quarter. This randomly selected quarter can be before or after the actual event took place. Second, I fix the dates in which the mergers occur, but I allocate two different companies randomly selected to be in the merger. That is, I hold the change in the market structure for each local market (the same change in each industry for each city) constant, but I change the identity of the merged firms. These two exercises aim to distinguish whether or not there are some events explaining the effects on the lobbying expenditures due to some unobservable characteristics of the industry that received the shock or the date on which the event occurred, respectively.

Third, I consider mergers with the same characteristics of the selected sample of the baseline estimation (i.e., at least 95th percentile of the transaction value, only 100% shares transferred, etc.) in which a merger was announced but did not happen. Finally, for the mergers of the baseline sample, I change the dates on which the event occurred. I consider the date of the initial announcement instead of the day on which the merger took place, as in the baseline estimation. On average, the announcement happened two quarters before the merger took place.

Table 3.19 show the results of these exercises. For all of these estimations, the effect of the merger is not statistically different from 0. The only coefficient from these estimations with some statistical significance is for the fourth exercise, in which I measure the effect of a merger (with its initial announcement date, instead the date in which the merger took place) on industry-level lobbying expenditures. An interpretation of this result is that the industry as a whole started reacting to an announced merger as soon as it was announced. However, the coefficient represents a smaller effect than the estimates in which the date of the merger is the actual effect date (when the merger took place instead of when it was announced). Overall, this implies that there is an industry-level reaction once the merger is announced but that the reaction to it is smaller than the reaction at the time that the merger takes place. A possible way to explain this result is that there are some circumstances in which the intention to merge is withdrawn once it has been announced, and, therefore, there is no need to adjust for advocacy strategies once the announcement takes place.

The results collected in this section provide empirical evidence on the effect of merger on city-level lobbying expenditures. Firms and industries that face an increase in their market concentration increase their lobbying expenditures. This result is robust to different ways of defining the sample of mergers, as well as to different placebo tests.

3.5 Final Discussion

The aim of this chapter has been to understand the way in which the excludability of the goods for which interest groups lobby can have an impact on the way that market structure is related to lobbying expenditures.

The central point is that even if this dimension has been largely neglected by the literature, it is important because it allows us to understand why we observe that less-concentrated industries tend to spend more on lobbying than more-concentrated industries. Mancur Olson was the first scholar to establish a well-structured link between the number of individuals in the group and individual efforts. His main claim is that unless the number of individuals in a group is quite small, self-interested individuals will face the free-riding problem and, therefore, will not overcome the collective-action problem. My contribution to this debate is that as groups can have goals that differ by their excludability level, we can empirically observe that Olson's prediction does not hold. I show that once one controls for a measure of the excludability in effort goals, we reestablish his prediction. The main takeaway is, therefore, that future studies should keep in mind, in a general context, the main characteristics of the groups' goals and, in particular, the level of excludability of these goals.

Although I believe that this chapter has progressed our understanding of the collective-action problem, there are some weaknesses that should be addressed in future work. First, the theoretical framework developed here relates lobbying efforts to several underlying variables. Here, I focused on the level of excludability of the lobbying goals, as there were no clear ways to construct empirical measures for other variables. Future work could explore how changes in the rivalry of these goods or how the perceived effect of the lobbying expenditures can explain the variance in these efforts.

Second, in Section 3.3, I measure the level of excludability of the goals of the groups using earmark data. Although I think that this is an appropriate measure for excludable goods, it is an incomplete one, and, therefore, it should be interpreted as a lower bound of the total excludability measures we may observe. In the future, scholars may consider measuring other types of excludable goods or extending the earmark time period.[188] Finally, in Section 3.4, I focused on only nine cities, as they were among the few asking detailed information about advocacy activities. The transparency rules regulating advocacy activities at both the city and state levels are changing rapidly. I expect, in the near future, to see a larger set of cities and states with detailed information of their lobbying activities. A natural sequel to this chapter would be to extend this work to more cities and to explore the possibility of better understanding the lobbying world with the use of state-level advocacy activities.

Abstracting from the current application, the analysis underscores the potential

[188]For the first part, I refer the interested reader to some early attempts to categorize tax exemptions. For instance, the literature in Accounting and Law have studied some tax exemptions in a broad category called rifle-shot rules (Hanna (2006)). For the second part, I refer the reader to the possibility of using a promising but incomplete text data analysis on earmarks detection at DSSG.

of including excludability in the collective-action analysis. Here, I have provided a simple setting to study its effects in the lobbying context. A natural continuation of this chapter would be to see the extent to which this dimension contributes to our understanding of collective action in other economic and political contexts.

3.6 Appendix

3.6.1 Tables

	Data Source		
Variable Used	**Orbis**	**Compustat**	**Economic Census**
n	Yes	Yes	Yes
C_4, C_8 and C_{20}	Yes	only C_4, C_8	Yes
HHI	Yes	Yes	only Manufacturing
Years Available	2012	1998-2014	1997, 2002, 2007 and 2012
Definition Industry	SIC 2, 3, 4 digits	SIC 2, 3, 4 digits	SIC 2, 3, 4 digits
Variables Used	Turnover, Employees	Assets, Employees, Revenue, Sales	Sales, Employees

Table 3.1: Variables Used by Data Source.

Note: For each of these databases, I measure marker concentration with five different measures: total number of firms (henceforth n), 3 concentration ratios (henceforth C_4, C_8 and C_{20}) and Herfindahl index (henceforth HHI). I calculate these measures using the Orbis information with both total turnover and total number of employees for each firm for the year 2012. For Compustat, I calculate this information using data on total assets, number of employees, total revenue and total sales for the period 1998 to 2014. Finally, I use the concentration measures that the Economic Census release every five years (1997 to 2012). The Census give these concentration measures for all of the manufacturing industries. For non-manufacturing industries, the US Census Bureau only releases the first 4 concentration measures.

		Orbis			Economic Census		
		4 Digits	**3 Digits**	**2 Digits**	**4 Digits**	**3 Digits**	**2 Digits**
n	*Mean*	15844.05	48641.24	194565	35509.37	115860.2	405371.8
	Stdv	39029.91	98731.47	193772.7	55164.97	163462	284203.7
C_4	*Mean*	8.20	6.46	3.23	24.00	22.94	8.04
	Stdv	16.26	12.63	10.65	19.48	20.31	6.52
C_8	*Mean*	8.37	6.60	3.31	31.45	30.30	12.47
	Stdv	16.25	12.62	10.64	22.80	22.40	8.78
C_{20}	*Mean*	8.45	6.68	3.38	40.96	39.41	20.8
	Stdv	16.23	12.61	10.63	24.84	24.39	13.89
HHI	*Mean*	85.08	67.17	34.25	401.64	178.65	NA
	Stdv	162.22	12.60	10.62	389.40	188.50	NA
	Industries	307	100	25	258	74	12

Table 3.2: Descriptive Statistics for Orbis and Economics Census.

Note: The table reports mean and standard deviation for concentration measures n, C_4, C_8 and C_{20} and HHI. All the statistics are calculated for year 2012 except for the industries Utilities (2007 NAICS code 22) and Administrative and support and waste management and remediation services (2007 NAICS code 22) for the Economic Census. For these two industries, 2007 is the most recent data available. For the Economic census, n is calculated with the total number of establishments whereas it is the total number of firms for Orbis. For Orbis, Concentation measures $C4$, $C8$, $C20$ and HHI are calculated using data on the number of employees per firm. Similar results emerge when I calculate these estimations with total revenue. For the Economic Census, the concentration measures $C4$, $C8$ and $C20$ are calculated using information on Sales, receipts, or revenue. Finally, for the Census, HHI is only calculated for manufacturing industries (codes 31-33) for the 50 largest companies for three and four 2007 NAICS digits for the year 2007. There are 21 (85) 3 (4) digits 2007 NAICS code industries among codes 31-33. The Census do not report information for Mining (21), Construction (23) and Management of Companies and Enterprises (55).

			4 Digits		3 Digits		2 Digits	
			Mean	**Stdv**	**Mean**	**Stdv**	**Mean**	**Stdv**
	n	*Total*	20.04	56.60	31.84	96.90	121.98	233.60
		Bet/With		51.6/21.4		90.3/32.4		222.3/68.3
Sales	HHI	*Total*	0.45	0.28	0.39	0.27	0.21	0.24
		Bet/With		0.26/0.13		0.26/0.12		0.27/0.07
	C_4	*Total*	0.87	0.17	0.83	0.19	0.63	0.23
		Bet/With		0.15/0.08		0.17/0.07		0.23/0.06
	C_8	*Total*	0.94	0.12	0.92	0.13	0.77	0.2
		Bet/With		0.11/0.06		0.12/0.05		0.19/0.05
Emp	HHI	*Total*	0.46	0.27	0.40	0.28	0.22	0.24
		Bet/With		0.25/0.14		0.25/0.13		0.25/0.07
	C_4	*Total*	0.87	0.19	0.83	0.2	0.62	0.24
		Bet/With		0.16/0.10		0.18/0.08		0.24/0.07
	C_8	*Total*	0.94	0.15	0.92	0.15	0.76	0.22
		Bet/With		0.12/0.09		0.14/0.07		0.21/0.06
	Observations		6370	440/12	4011	275/14	1047	72/14

Table 3.3: Descriptive Statistics for Compustat.

Note: Initially I have 127715 observations, comprending 14 years (1998-2014) and 16581 companies. The panel is unbalanced because some companies enter or exit during this period. The statistics are classified by the number of digits for which an industry is defined. Within each industry classification, the first column represents the mean value, while the second column presents the standard deviation. The variables of interest are number of firms n, HHI, C_4 and C_8. For each firm, I present the mean for the sample and the standard deviation between (across industries) and within (across years). After the statistics representing the number of firms, there are two panels. The first panel is based on the sales of the firms, while the second panel is based on the number of employees of the firm. The last row contains information on the total number of observations, industries and years used in the analysis.

Dependent Variable: $\frac{L_i}{X_i}$ [Estimates of β_1]

4 Digits

n	0.012**					0.015**				
	(0.008)					(0.010)				
C_4		-0.012**					-0.013**			
		(0.009)					(0.010)			
C_8			-0.013**					-0.013**		
			(0.008)					(0.008)		
C_{20}				-0.010**					-0.010**	
				(0.005)					(0.005)	
HHI					-0.017**					-0.016**
					(0.010)					(0.010)
Obs			307					258		
Database			Orbis					Economic Census		
Adj R^2	0.478	0.456	0.421	0.433	0.542	0.490	0.479	0.446	0.471	0.569

3 Digits

n	0.020*					0.024**				
	(0.015)					(0.013)				
C_4		-0.019**					-0.021**			
		(0.011)					(0.012)			
C_8			-0.020*					-0.023*		
			(0.016)					(0.018)		
C_{20}				-0.012**					-0.015**	
				(0.007)					(0.009)	
HHI					-0.019*					-0.024**
					(0.014)					(0.013)
Obs			100					74		
Database			Orbis					Economic Census		
Adj R^2	0.419	0.386	0.356	0.375	0.462	0.428	0.391	0.347	0.368	0.471

Table 3.4: Estimates of Equation (1).

Note: The table shows the estimated coefficients, standard errors, total number of observations and adjusted R^2 from equation (1). The table has two different panels. The top panel presents the results for the case that the industry is defined with four-digits whereas the bottom panel presents the results for the case the industry is defined with three-digits. In these estimations, an observation is an industry. The left-side part of the table uses data from Orbis, whereas the right side part of the table uses data from the US Economic Census. The dependent variable is the same for these 20 econometric estimations. Each pair of rows present results for different ways to measure market concentration. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors in parenthesis are clustered at the industry level.

Dependent Variable: $\frac{l_{ji}}{x_{ji}}$ **[Estimates of** β_2**]**

					4 Digits					
n	-0.024*					-0.023				
	(0.019)					(0.021)				
C_4		0.022*					0.020*			
		(0.017)					(0.015)			
C_8			0.023*					0.025*		
			(0.016)					(0.017)		
C_{20}				0.015*					0.018	
				(0.010)					(0.017)	
HHI					0.025*					0.023*
					(0.019)					(0.018)
Obs			43125					29568		
Database			Orbis					Economic Census		
Adj R^2	0.258	0.243	0.221	0.213	0.262	0.218	0.213	0.191	0.201	0.222
					3 Digits					
n	-0.024					-0.025*				
	(0.022)					(0.019)				
C_4		0.020*					0.021*			
		(0.015)					(0.016)			
C_8			0.028*					0.028*		
			(0.019)					(0.019)		
C_{20}				0.019*					0.019	
				(0.014)					(0.018)	
HHI					0.023*					0.024*
					(0.018)					(0.019)
Obs			43125					29568		
Database			Orbis					Economic Census		
Adj R^2	0.208	0.193	0.178	0.186	0.214	0.198	0.173	0.168	0.172	0.208

Table 3.5: Estimates of Equation (2).

Note: The table shows the estimated coefficients, standard errors, total number of observations and adjusted R^2 from equation (2). The table has two different panels. The top panel presents the results for the case that the industry is defined with four-digits whereas the bottom panel present the results for the case the industry is defined with three-digits. In these estimations, an observation is a firm. The left-side part of the table uses data from Orbis, whereas the right side part of the table uses data from the US Economic Census. The dependent variable is the same for these 20 econometric estimations. Each pair of rows present results for different ways to measure market concentration. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors in parenthesis are clustered at the industry level.

Dependent Variable: $\frac{L_{it}}{X_{it}}$ [Estimates of β_3]					
			4 Digits		
n	0.025**				
	(0.014)				
C_4		-0.021**			
		(0.012)			
C_8			-0.029**		
			(0.015)		
C_{20}				-0.020**	
				(0.010)	
HHI					-0.025*
					(0.017)
Obs			4908		
Database			*Compustat*		
Adj R^2	0.912	0.823	0.901	0.899	0.877
Industry FE	X	X	X	X	X
Year FE	X	X	X	X	X
			3 Digits		
n	0.027**				
	(0.015)				
C_4		-0.022*			
		(0.015)			
C_8			-0.029**		
			(0.016)		
C_{20}				-0.020**	
				(0.010)	
HHI					-0.026*
					(0.019)
Obs			1589		
Database			*Compustat*		
Adj R^2	0.859	0.786	0.876	0.875	0.841
Industry FE	X	X	X	X	X
Year FE	X	X	X	X	X

Table 3.6: Estimates of Equation (3).

Note: The table shows the estimated coefficients, standard errors, total number of observations and adjusted R^2 from equation (3). The table has two different panels. The top panel presents the results for the case that the industry is defined with four-digits whereas the bottom panel present the results for the case the industry is defined with three-digits. In these estimations, an observation is an industry*year. These estimations are based on the Compustat database. The dependent variable is the same for these 10 econometric estimations. Each pair of rows present results for different ways to measure market concentration. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors in parenthesis are clustered at the industry level. All of the results include the estimation of both industry and year fixed-effects.

Dependent Variable: $\frac{l_{it}}{x_{it}}$ [Estimates of β_4]					
			4 Digits		
n	-0.022*				
	(0.012)				
C_4		0.020*			
		(0.014)			
C_8			0.020		
			(0.018)		
C_{20}				0.018*	
				(0.009)	
HHI					0.024*
					(0.018)
Obs			436789		
Database			Compustat		
Adj R^2	0.732	0.765	0.702	0.657	0.621
Industry FE	X	X	X	X	X
Year FE	X	X	X	X	X
			3 Digits		
n	-0.024*				
	(0.013)				
C_4		0.026			
		(0.024)			
C_8			0.024		
			(0.021)		
C_{20}				0.019*	
				(0.011)	
HHI					0.029*
					(0.021)
Obs			436789		
Database			Compustat		
Adj R^2	0.729	0.741	0.76	0.729	0.718
Industry FE	X	X	X	X	X
Year FE	X	X	X	X	X

Table 3.7: Estimates of Equation (4).

Note: The table shows the estimated coefficients, standard errors, total number of observations and adjusted R^2 from equation (4). The table has two different panels. The top panel presents the results for the case that the industry is defined with four-digits whereas the bottom panel present the results for the case the industry is defined with three-digits. In these estimations, an observation is a firm*year. These estimations are based on the Compustat database. The dependent variable is the same for these 10 econometric estimations. Each pair of rows present results for different ways to measure market concentration. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors in parenthesis are clustered at the industry level. All of the results include the estimation of both industry and year fixed-effects.

	Excludable		Non Excludable	
	Lobbying	Non-Lobbying	Lobbying	Non-Lobbying
Rival	$\frac{G}{n_l} - c(l_i)$	0	$\frac{G}{n} - c(l_i)$	$\frac{\widetilde{G}}{n}$
	Lobbying	Non-Lobbying	Lobbying	Non-Lobbying
No-Rival	$G - c(l_i)$	0	$G - c(l_i)$	$\widetilde{G}$

Table 3.8: Net Benefit of Lobbying by Excludability/Rivalry of the bill.

Note: The table represents the benefits of goods according to the combination of excludability and rivalry of the bills that the firms lobby. The first row differs from the second in that the benefit of the bill is spread out to other firms in the case the firm is lobbying, while the first column predicts a zero payoff for non-lobbying firms. The number of firms lobbying can be defined as $n_l : \left\{ n | \frac{G}{n_l} - c(l_i) \geq 0 > \frac{G}{n_l+1} - c(l_i) \right\}$.

% of Recipients	2008	2009	2010
Quintiles			
1	6.48	6.37	4.81
2	9.87	6.7	8.15
3	21.2	22.3	17.86
4	24.95	29.65	31.17
5	37.5	34.98	38.01
Total	100	100	100
Total Recipients Firms	1531	1543	4502

Table 3.9: Relationship Between Market Concentration and Earmarks Recipients.

Note: The table shows descriptive evidence on the relationship between industry market concentration and earmarks recipients. Here I measure market concentration with the total number of firms of the industry. The main results of this section are robust to alternative variables measuring market concentration. I have standarized the levels of concentration by quintiles. To make this standarization, I calculate the quintiles on the total number of firms for all of the 4-digits code industries that report the Economic Census. Larger quantiles represent less concentrated industries. For these analyses, I use 2007 data for two reasons. First, it is more accurate, given that the earmarks data is three years after the Census information and second, not all of the industries have information for 2012 but all of them have them for 2007. The table provides information for each year of data available (2008-2010). Each column adds to 100.

Panel A			
% of Bills Lobbied with Earmarks			
Quintiles	**2008**	**2009**	**2010**
1	17.61	20.12	21.34
2	18.91	19.59	24.67
3	24.57	20.15	22.75
4	47.58	48.5	40.17
5	54.55	61.26	58.7
Panel B			
% of Expenditures in Bills with Earmarks			
Quintiles	**2008**	**2009**	**2010**
1	23.59	22.63	11.90
2	19.05	11.14	17.44
3	21.05	17.52	16.47
4	54.17	50.63	41.97
5	60.72	53.51	51.26

Table 3.10: Bills and Expenditure with Earmarks by Quintiles.

Note: The table is divided in two panels. Panel A shows the fraction of bills with earmarks lobbied over the total number of bills lobbied for each quintile and year. Panel B shows the fraction of expenditure spend on bills that had any earmark over the total lobbying expenditure.

Dependent Variable: $\frac{L_i}{X_i}$ [Estimates of β_1]

	4 Digits									
n	0.014**	-0.012*								
	(0.008)	(0.008)								
Ex		0.04**								
		(0.022)								
C_4			-0.016*	0.011*						
			(0.011)	(0.009)						
Ex				0.035*						
				(0.026)						
C_8					-0.014	0.012				
					(0.012)	(0.011)				
Ex						0.041**				
						(0.022)				
C_{20}							-0.009**	0.009		
							(0.005)	(0.009)		
Ex								0.038*		
								(0.026)		
HHI									-0.019**	0.013*
									(0.011)	(0.009)
Ex										0.042**
										(0.024)
Obs	258									
Database	*Economic Census*									
Adj R^2	0.486	0.589	0.478	0.592	0.458	0.610	0.465	0.584	0.489	0.568

Table 3.11: Effect of the Inclusion of Excludability Measures.

Note: The table shows the estimated coefficients, standard errors, total number of observations and adjusted R^2 from estimating the effect of market concentration and excludability measures on lobbying expenditures. The table presents the results for the case that the industry is defined with four-digits. In these estimations, an observation is an industry. The table uses data from the 2007 US Economic Census. The dependent variable is the same for these 20 econometric estimations. Each pair of rows present results for different ways to measure market concentration. Each pair of columns use the same concentration measure. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors in parenthesis are clustered at the industry level.

City	State	Beginning period	Timing Reports	Source
Chicago	Illinois	2012	Quarter	*Web*
Dallas	Texas	2010	Quarter	*Web*
Los Angeles	California	2003	Quarter	*Web*
Miami	Florida	2006	Contracts	*Web*
New York City	New York	1998	Contracts	*Web*
Philadelphia	Pennsylvania	2012	Quarter	*Web*
San Diego	California	2011	Quarter	*Web*
San Francisco	California	2010	Monthly	*Web*
Seattle	Washington	2008	Contracts	*Web*

Table 3.12: A subset of cities with lobbying disclosure requirements.

Note: The table presents the main information used in Section 3.4. The dataset has nine cities. Each row represents a different city. The second column represent the state of the city. The third column present the first year for which there is available information for the city. The fourth column presents the periodicity in which lobbyists have to report advocacy activities. The last column presents the link in which the main information has been extracted.

		4 Digits		3 Digits		2 Digits	
Category #		*Mean*	*Stdv*	*Mean*	*Stdv*	*Mean*	*Stdv*
1	*Total*	2.85	12.07	8.03	38.37	46.34	114.88
	Bet/With		11.47/3.75		37.54/8.14		111.91/26.57
2	*Total*	0.08	0.50	0.22	1.28	1.32	3.73
	Bet/With		0.34/0.37		0.94/0.86		2.76/2.53
3	*Total*	1.09	4.64	2.67	9.93	13.88	29.47
	Bet/With		3.94/2.46		8.73/4.75		25.73/14.6
Obs.		14416	901	6240	390	1280	80
Tot. Value	*Total*	1414.68	6094.46	2267.13	8811.48	7280.4	19538.73
	Bet/With		2522.82/4527.22		4740.95/6085.51		13652.04/ 13282.7
Obs.		3781	677	2749	329	1016	73
% Shares	*Total*	93.37	17.40	93.63	15.13	94.57	8.70
	Bet/With		12.27/14.39		10.08/12.59		4.87/7.26
Obs.		6229	867	3827	383	1135	79
Ave. Value	*Total*	163.96	570.03	153.94	489.07	163.21	466.89
	Bet/With		394.49/ 467.21		259.32/410.51		316.20/359.60
Obs.		3780	677	2748	329	1016	73

Table 3.13: Descriptive Statistics for SDC Platinum.

Note: This table presents mean value, number of observation as well as standard deviation for between and within groups. Because all the variables contain information for 16 years, the rows labelled observations represent the total number of observations and the number representing the between categories that correspond to the number of industries considered. The data is grouped under three different definitions of what is an industry according to the level of digits: four, three and two digits. The first part of the table present statistics for three different categories of mergers: Category 1 is when there was a successfull merger, Category 2 is an unsucessfull merger, where the parties had an intention to merge but it did not happen and Category 3 groups all the other merger attempts that has neither been sucessfull nor withdrawn. The second part of the table present descriptive statistics for three variables (across all of the categories for a given industry): Total value of the mergers at industry level, average percentage of shares acquired and average value of the transaction. This last variable is simply the total value of all the mergers divided by the total number of successfull mergers, for each industry.

Merger	Company 1	Company 2	SIC	Industry
1	Longview Timber LLC	Weyerhaeuser Co	8	Forestry
2	Consolidation Coal Co	Murray Energy Corp	12	Coal Mining
3	Apache Corp-Shelf Assets	Fieldwood Energy LLC	13	Oil & Gas Extraction
4	Onyx Pharmaceuticals Inc	Amgen Inc	28	Chemical & Allied Prod.
5	BP PLC-Carson Refinery,ARCO	Tesoro Refining & Marketing Co	29	Petroleum & Coal Prod.
6	LSI Corp	Avago Technologies Ltd	36	Electronic & Other Elect. Equip.
7	US Airways Group Inc	AMR Corp	45	Transportation by Air
8	T-Mobile USA Inc	MetroPCS Communications Inc	48	Communications
9	NV Energy Inc	MidAmerican Energy Holdings Co	49	Electric, Gas, & Sanitary Services
10	Harris Teeter Supermarkets Inc	Kroger Co	54	Food Stores
11	NYSE Euronext	IntercontinentalExchange Inc	62	Security & Communication Brokers
12	Coventry Health Care Inc	Aetna Inc	63	Insurance Carriers
13	Cole Real Estate Investments	American Rlty Capital Ppty Inc	67	Holding & Other Investment Offices
14	ExactTarget Inc	Salesforce.com Inc	73	Business Services
15	Health Management Assoc Inc	Community Health Systems Inc	80	Health Services

Table 3.14: Basic Information of 15 Selected Mergers.

Note: The table provides information on 15 selected mergers. The table presents the name of the companies involved in the mergers in columns two and three. The table presents information on the specific 2-digits SIC code as well as the industrial sector name.

Merger	**Effective Date**	**Quarter**	**Value Transaction (Mill)**	# Cities
1	7/2013	3-2013	2650	3
2	12/2013	4-2013	3462	3
3	9/2013	3-2013	5250	3
4	10/2013	4-2013	9692.50	4
5	6/2013	2-2013	2425	4
6	5/2014	2-2014	6685.51	4
7	12/2013	4-2013	3080	3
8	5/2013	2-2013	3689.93	4
9	12/2013	4-2013	553.96	4
10	1/2014	1-2014	2543.53	5
11	11/2013	4-2013	8052.28	4
12	5/2013	2-2013	5695.60	3
13	2/2014	1-2014	6952.60	4
14	7/2013	3-2013	2640.79	3
15	1/2014	1-2014	7547.68	5

Table 3.15: Complementary Information of 15 Selected Mergers.

Note: The table provides information on 15 selected mergers. In columns two and three, the table presents month, quarter and year in which the merger took place. The table also provides information on the total value of the transaction and the number of cities affected by the merger. A city is affected if both of the firms involved in the merger were conducting lobbying activities in that city for at least one year prior the merger.

City	2-D SIC Codes	Firms	Lobbying Expend./Quarter
Chicago	72	560	8657.23
Dallas	65	389	4857.21
Los Angeles	71	589	6578.12
Miami	68	435	4956.01
New York City	78	621	9124.43
Philadelphia	64	417	6867.87
San Diego	65	387	4678.91
San Francisco	61	423	5678.76
Seattle	69	565	5245.77

Table 3.16: City-level Lobbying Expenditures.

Note: The table provides information of number of industries, firms and average lobbying expenditure per quarter for each of the nine cities selected for this exercise. The second column represents the average number of two-digits SIC codes across all of the quarters. The third column presents the average number of active lobbying firms across all of the quarters. The last column presents the average lobbying expenditure per active firm across quarters.

	Firm Level	Industry Level	Firm Level	Industry Level	Firm Level	Industry Level	Firm Level	Industry Level	Firm Level	Industry Level
β_6	692.69*	9876.65**	729.65	10112.36	717.82*	9765.25*	743.56*	9986.23**	612.71*	9771.48*
St. Err.	477.71	5129.13	629.01	8793.36	535.69	6259.78	442.60	5043.55	422.55	6386.58
Obs	72891	12608	82080	13320	85680	13500	83880	13860	76500	12492
Adj R^2	0.895	0.921	0.845	0.896	0.876	0.902	0.891	0.911	0.868	0.915
Avg. Dependent	6156.2	101568.1	6892.5	98695.7	6456.2	99568.3	6789.6	102365.9	6473.2	112571.2
Time Period	2012-2016		2012-2016		2012-2016		2012-2016		2012-2016	
Total Mergers	15	15	37	37	34	34	31	31	15	15
Fixed Effects										
γ_{ic}	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
γ_{it}	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
γ_{ct}	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sample	Baseline		Industry Acquiring		$>$ 80% Shares		$>$ 90% Transaction Value		Including Merged firms	

Table 3.17: The Table Presents the Results of the Triple-dif Estimations.

Note: The table presents the coefficient estimates and standard errors for the coefficient of interest. The table provides information on the number of observations used, adjusted R^2, average value of the dependent variable, time period and combination of fixed-effects used. An observation in the columns marked as firm-level estimations is a firm*city*quarter. An observation in the industry-level estimations is an industry*city*quarter. Columns named Firm-Level mean that the dependent variable is the firm-level lobbying expenditure whereas columns named Industry-Level mean that the dependent variable is the lobbying expenditures of the industry affected by the merger. All of the estimation comprised the time period 2012-2016. The table presents the results for both firm and industry level lobbying expenditures for five different samples. The first sample named "Baseline", includes the results of the sample explained in the text. The pair of columns marked "Industry Acquiring" measures the effect for the case there is a merger between companies that belong to different industries. The point estimates presents the effect of lobbying expenditures for firms that belong to the same local market of the acquiring firm. The sample denoted as "$>$ 80% Shares" incudes all of the mergers in which there was a transfer of shares of at least 80%. The sample denoted as "$>$ 90% Transaction Value" includes all of the mergers in which the transaction value of the merger was at least as large as the percentile 90 of the whole distribution of transaction values. The last pair of columns include the lobbying expenditures of all of firms in the local market including the merged firms. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors are clustered at the industry-city level.

	Firm Level	**Industry Level**
β_{t-3}	87.96	1562.85
St. Err.	76.09	1467.46
β_{t-2}	68.95	1325.66
St. Err.	59.65	1244.75
β_{t-1}	156.44	1601.55
St. Err.	135.33	1503.80
β_{t}	215.65*	4156.98**
St. Err.	148.11	1908.62
β_{t+1}	245.68**	3256.89*
St. Err.	123.58	2049.65
β_{t+2}	289.57**	2899.45**
St. Err.	145.66	1331.24
β_{t+3}	105.66	658.96*
St. Err.	91.40	414.70
Obs	72891	12608
Adj R^2	0.895	0.921
Average Dependent	6156.23	101568.1
Time Period	2012-2016	
Total Mergers	15	15
Fixed Effects		
γ_{ic}	Y	Y
γ_{it}	Y	Y
γ_{ct}	Y	Y
Sample	Baseline	

Table 3.18: Leads and Lags Estimations.

Note: The table presents the coefficient estimates and standard errors for the Leads and Lags exercise. The table provides information on the number of observations used, adjusted R^2, average value of the dependent variable, time period and combination of fixed effects used. An observation in the columns marked as firm-level estimations is a firm*city*quarter. An observation in the industry-level estimations is an industry*city*quarter. Columns named Firm-Level mean that the dependent variable is the firm-level lobbying expenditure whereas columns named Industry-Level mean that the dependent variable is the lobbying expenditures of the industry affected by the merger. All of the estimation comprised the time period 2012-2016. The table presents the coefficient estimates for three periods before and after the merger occur as well as the estimate for the quarter in which the merger occurred. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors are clustered at the industry-city level.

	Firm Level	**Industry Level**	**Firm Level**	**Industry Level**	**Firm Level**	**Industry Level**	**Firm Level**	**Industry Level**
β_6	12.36	5.65	11.65	10.98	8.96	13.46	456.78	1568.21*
St. Err.	13.01	6.31	12.80	13.29	8.53	12.02	415.25	1120.15
Obs	72891	12608	72889	12604	76680	12781	72891	12608
Adj R^2	0.695	0.621	0.658	0.635	0.678	0.627	0.758	0.810
Average Dependent	6156.23	101568.1	6125.98	103998.8	5956.78	98765.32	6156.23	101568.1
Time Period	2012-2016		2012-2016		2012-2016		2012-2016	
Total Mergers	15	15	15	15	19	19	15	15
Fixed Effects								
γ_{ic}	Y	Y	Y	Y	Y	Y	Y	Y
γ_{it}	Y	Y	Y	Y	Y	Y	Y	Y
γ_{ct}	Y	Y	Y	Y	Y	Y	Y	Y
Sample	Random Dates		Random Companies		Withdrawn Mergers		Announcement Time	

Table 3.19: Placebo Tests.

Note: The table presents the coefficient estimates and standard errors for the coefficient of interest. The table provides information on the number of observations used, adjusted R^2, average value of the dependent variable, time period and combination of fixed effects used. An observation in the columns marked as firm-level estimations is a firm*city*quarter. An observation in the industry-level estimations is an industry*city*quarter. Columns named Firm-Level mean that the dependent variable is the firm-level lobbying expenditure whereas columns named Industry-Level mean that the dependent variable is the lobbying expenditures of the industry affected by the merger. All of the estimation comprised the time period 2012-2016. The table presents the results for both firm and industry level lobbying expenditures for five different samples. The first sample named “Random Dates”, includes the mergers of the baseline sample with randomly allocated dates for these mergers. The second sample titled “Random Companies”, allocate randomly companies not involved in the merger as the primary actors of the merger. The third sample named “Withdrawn Mergers” includes mergers that were announced as a merger, that fulfil the characteristics used in the baseline mergers used in the baseline, but that did not happen. Finally, the last sample, study the mergers of the baseline but instead of considering the date in which the merger took place, it considers the date in which the merger was announced. (*) means significance at 10%, whereas (**) and (***) stand for significance at 5% and 1%, respectively. Obs is the number of observations used in order to run the regression. Standard errors are clustered at the industry-city level.

3.6.2 Figures

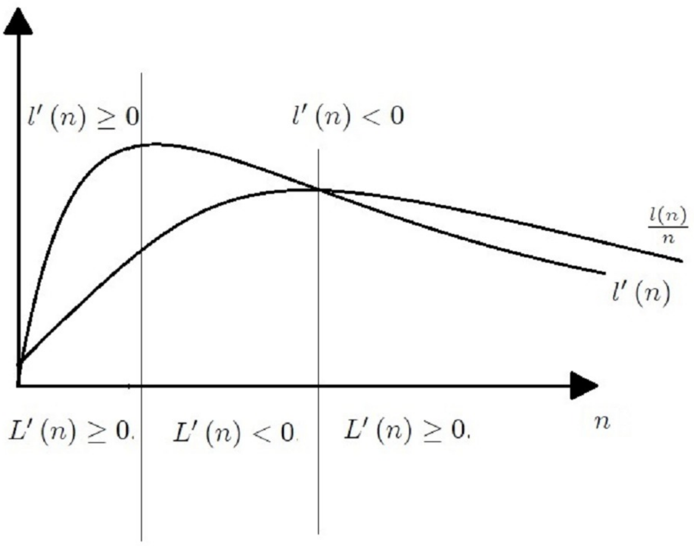

Figure 3.1:
Three Relevant Regions when firm-level lobbying functions are symmetric.

Note: The figure represents three relevant regions. Two in which the industry-level lobbying expenditure is increasing in the number of firms and one in which is decreasing. When the firm-level lobbying function is increasing then the industry-level function is always increasing. However, when the firm-level lobbying function is decreasing the elasticity of the firm-level function will determine whether the industry-level lobbying function is increasing in the number of firms or not. The third Region is the only region representing the empirical results in the previous section.

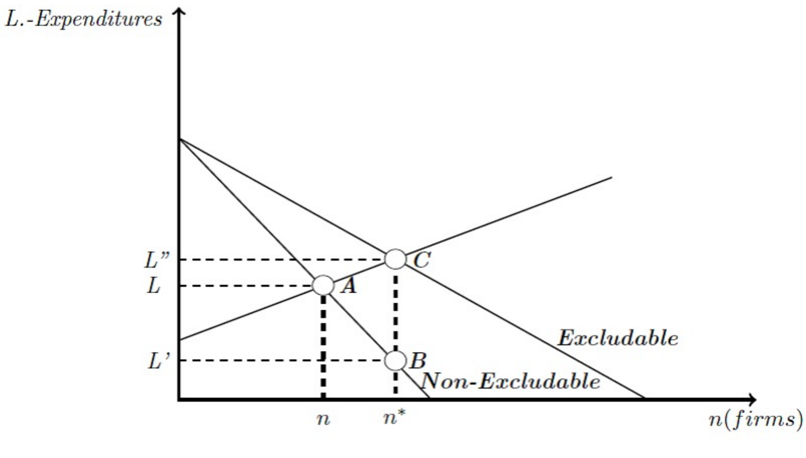

Figure 3.2:
The effect of Excludability in Lobbying Expenditures.

Note: The figure represents two industry level lobbying expenditures as a function of the number of firms in the industry. Right decreasing curve represents the case in which the industry lobbies for excludable goods, whereas Left decreasing curve represents the case for non-excludable goods. The figure shows the possibility in which two different market structures lobby for different types of bills and therefore, we observe an increasing pattern between lobbying expenditures and number of firms in the industry.

3.6.3 Theoretical Appendix

3.6.3.1 This Model is General

I think this simple model may be able to rationalize my data and provide a generalization to some of the most important papers in the literature. To follow I explain how my model generalizes Olson (1965), Stigler (1974), Esteban and Ray (2001) and Pecorino and Temimi (2008). The most important take away here is that these papers do not consider exludable goods at all.

Olson (1965): Olson's objective function is $V_i - c\left(l_i\right)$, where α_g is the size of the group, F_i is the fraction of the individual gain, V_g is the group value, T is the rate of level at which the collective good is obtained and V_i is the gain to the individual i-th. In his setting, $F_i = \frac{V_i}{V_g}$ and $\alpha_g T = V_g$.

You get Olson (1965)'s setting if $\pi(\sum_{j=1}^{n} l_j) = \beta = 1$, $\alpha_g = n$, $c\left(l_i\right) = c_i\left(l_i\right)$, $V_g = G$, $T = \frac{G}{\alpha_g}$ and $V_i = \frac{GS_i}{n}$. The two main results in Olson are: **1**. Individual i will contribute if $F_i > \frac{c}{V_g}$, the gain to an individual from seeing that the good is provided will exceed the cost. As the only associated cost is the lobbying cost, it seems this setting is exactly for excludable goods. **2**. In the optimum, the rate of gain to the group $(\frac{dV_g}{dl_i})$ must exceed the rate of increase in cost $\left(\frac{dc(l_i)}{dl_i}\right)$ by the same multiple that the group gain exceeds the gain to the individual concerned $(\frac{1}{F_i} = \frac{V_g}{V_i})$, that is $\frac{dV_g}{dl_i} / \frac{dc(l_i)}{dl_i} = \frac{V_g}{V_i}$. $\frac{dG(l_i)}{dl_i} / \frac{dc(l_i)}{dl_i} = \frac{n}{S_i}$.

Notice a very important difference is that the payoff is not a function of group effort, that is $G(l_i) = G$.

Stigler (1974): The objective function in Stigler is $G(m, e) - e(m)$, where $G(m, e)$ is the gain to the individual if collective action is taken, m is the number of individuals joining the coalition and $e(m)$ is the expenditure per individual who joins.

You get Stigler (1974) setting if $S_i = 1 = \pi(\sum_{j=1}^{n} l_j)$, $c_i(l_i) = l_i$, $e(m) = l_i(n)$ and $G_i(m, e) = \frac{G\left(\sum_{j=1}^{n} l_j\right)}{m^{\beta}}$.

The main difference between his setting and mine is that while I am talking here about the probability of passing a bill, Stigler is talking about the probability of a given individual joining the collective action.

Esteban & Ray (2001): Their objective function is $\frac{A_i}{A}\left[\lambda P + (1-\lambda)\frac{M}{N_i}\right] - v\left(a_i\right)$, where A_i is the total effort contributed by group i, N_i is the size of the group i, $A = \sum_{i=1}^{N} A_i$, $A_i = \sum_{i=1}^{N_i} a_i$, $\frac{A_i}{A}$ is the probability of sucess of group i, P is the public component of the good, M is the private component, and $v\left(a_i\right)$ is the cost function. In intuitive terms, P represents the non rival part whilst M represents the rival part.

You get Esteban & Ray's (2001) setting if $S_i = 1$, $\pi = \frac{A_i}{A}$[189], $c_i\left(l_i\right) = v\left(a_i\right)$ and finally $\frac{G(\sum_{j=1}^{n} l_j)}{n^{\beta}} = \lambda P + (1-\lambda)\frac{M}{N_i}$ with $\beta = 1 - \lambda$.

[189] In my case, this probability will be $\frac{l_i}{L}$, which means that only one group will get the good.

Chamberlin (1974) explains that we may expect $\frac{\partial L}{\partial n} < 0$ when the good is rival but $\frac{\partial L}{\partial n} > 0$ when the good is fully public. Esteban & Ray's main result is that $\frac{\partial L}{\partial n} > 0$ for both perfectly private and perfectly public goods if the cost function increase sufficiently fast. In other words $sign\left(\frac{l_i \frac{\partial^2 c}{\partial l_i^2}}{\frac{\partial c}{\partial l_i}} - 1\right) = sign\left(\frac{\partial L}{\partial n}\right)$. Although I can model the competition between groups, the way I am doing it simplifies the analysis without leaving aside the main objective of capturing the essential mechanisms in mind. Another difference between their model and mine is that their payoff in the event a group's policy is chosen is not a function of group effort. This setting is by assumption perfectly excludable at the group level, and non-excludable within a group.

Pecorino & Temimi (2008): Their objective function is $\frac{G(S_{-i}+s_i)}{n^\beta} - v(s_i)$. From the main objective function, we can get Pecorino & Temimi (2008) if we set $\pi(\sum_{j=1}^{n} l_j)S_i = 1$ and $c_i(l_i) = v(a_i)$. The main results from this paper are robust to the presence of small fixed costs of participation (*i.e.* $\lim_{s\to 0} v(s) = c$) in the case of a non-rival good, but not in the case of a fully rival good. This is true, regardless of the value of the elasticity $l_i \frac{\partial^2 c}{\partial l_i^2} \frac{1}{\frac{\partial c}{\partial l_i}}$. The intuition is that if the good is rival, the per person benefit approaches to 0 as the group grows large.

This shows that my simple setting is a generalization of these 4 models. My main contribution to these 4 models are: **1**. None of those contain the excludability/Rivalry dimension. I have shown in the body of this chapter that there are several bills with these combinations, so it seems important to include these differences. Furthermore, all these models analyze exclusively the case where the goods are non-excludables. None of those discuss the excludable goods' case. **2**. The last three models implictly assume symmetric solutions by imposing homogeneity in the individuals. I try to go a step further as in Olson (1965) to allow any degree of heterogeneity. Although there may be several ways to do that I use a scale parameter S_i multiplying the expected benefit. **3.** I do not restrict my analysis to particular functional forms such as full certainty $\pi(\sum_{j=1}^{n} l_j) = \beta = 1$ in Olson (1965), linear cost functions $c(l_i) = l_i$ in Stigler (1974), $c(l_i) = c_i(l_i)$ exactly equal cost function, particular probability function $\pi = \frac{l_i}{L}$ in Esteban and Ray (2001) or $\pi(\sum_{j=1}^{n} l_j)S_i = 1$ in Pecorino and Temimi (2008).

3.6.3.2 Discussion on Lemma

Let $\Delta_i = \pi(L)G(L) - \theta\pi(L_{-i})G(L_{-i})$ be the net effect of the lobbying expenditure of firm i-th on the expected benefit. Then, it is more likely to see firms lobbying if they can exert a greater change in the expected benefit. The change in the expected benefit only matters when the good is non perfectly excludable (i.e., $\theta > 0$), otherwise, this effect is equal to the expected effect of the bill. Most of the discussion of Olson (1965) is focused here. This factor can be understood as the perceived effectiveness of the contribution. Then, the fewer firms an industry has, the more likely each firm thinks it is important for the industry, then, the larger the incentive to contribute. For instance, if the group is large is very likely that each individual cannot exert such a big change in $\pi(L)G(L)$ (i.e., as $L \to L_{-i}$), so $\Delta_i(n)$ will tend to be decreasing in n. To better understand this term lets add $\theta\pi(L)G(L_{-i}) - \theta\pi(L)G(L_{-i})$ to Δ_i. Then, we get:

$$\pi(L)G(L) - \theta\pi(L_{-i})G(L_{-i}) + \theta\pi(L)G(L_{-i}) - \theta\pi(L)G(L_{-i})$$

factoring common terms:

$$\pi(L)\left(G(L) - \theta G(L_{-i})\right) + \left(\pi(L) - \pi(L_{-i})\right)\theta G(L_{-i})$$

and redefining terms we get

$$\Delta_i = \Delta_G \pi(L) + \Delta_\pi \theta G(L_{-i})$$

where $\Delta_G = G(L) - \theta G(L_{-i})$ and $\Delta_\pi = \pi(L) - \pi(L_{-i})$. Notice that $\Delta_i = \pi(L)G(L)$ if $\theta = 0$, and $\frac{\partial \Delta_i}{\partial \theta} = -\pi(L_{-i})G(L_{-i}) < 0$. Then, the net effect of the participation of the firm i-th can be decomposed into two parts: The expected change due to the change in the size of the good Δ_G for a given probability of passing the bill if the firm i-th participates $\pi(L)$ and the expected change due to the change in the probability of passing the bill Δ_π for a given size if the firm i-th does not participate $G(L_{-i})$. So the total effect is a combination of the change in the size of the good weighted by the probability and the change in the probability weighted by the size of the good.[190] Now I am interested to have predictions on the effect of changes in both the number of firms in the industry and the number of firms lobbying.

First notice that Δ_i depends on the total lobbying expenditures L, and those will depend on equilibrium on the total number of firms of the industry n as well as the total number of firms lobbying n_l. That is, $\Delta_i(n, n_l)$. To keep things tractable, I assume that the cost of lobbying does neither depend on the number of firms of the industry nor the number of firms lobbying but the reader should realize that this extension is straightforward. Now, in order to know how this condition changes as we change the number of firms n, and taking into account that n_l is an endogenous object, $n_\theta = \theta n + (1-\theta) n_l(n)$ we can take derivatives of the equation defining the extensive margin with respect to n to get:

$$\frac{\left[\Delta'_{i,n}(n, n_l) S_i(n) + \Delta_i(n, n_l) S'_i(n)\right] n^\beta - \beta\left(\theta + (1-\theta)\frac{\partial n_l}{\partial n}\right)\Delta_i(n, n_l) S_i(n) n_\theta^{\beta-1}}{n_\theta^{2\beta}}$$

[190]Notice that if I would have added $\pi(L_{-i})G(L) - \pi(L_{-i})G(L)$ to Δ_i I would have got $\Delta_i = \Delta_G \pi(L_{-i}) + \Delta_\pi G(L)$ which has a similar interpretation.

where $\Delta'_{i,n}(n, n_l)$ represents the total derivative with respect to the number of firms in the industry $\frac{\partial \Delta_i(n,n_l)}{\partial n} = \frac{\partial \Delta_i(n,n_l)}{\partial n} + \frac{\partial \Delta_i(n,n_l)}{\partial n_l}\frac{\partial n_l}{\partial n}$.

Then, this condition is positive if $\frac{\left[\Delta'_{i,n}S_i + \Delta_i S'_i\right]}{\Delta_i S_i} > \frac{\beta\left(\theta + (1-\theta)\frac{\partial n_l}{\partial n}\right)}{n_\theta}$. Now, let $b_i(n) = \Delta_i S_i$ and let $\in_b = \frac{b'_i(n)n}{b_i(n)}$ be the elasticiy of $b_i(n)$ with respect to the number of firms, then our condition is summarized as:

$$\frac{\partial Pr(l_i > 0)}{\partial n} = \in_b - \beta\left(\theta + (1-\theta)\frac{\partial n_l}{\partial n}\right) = \in_b - \beta\frac{\partial n_\theta}{\partial n}$$

That is, if the expected benefit is sufficiently sensitive to the number of firms , then more firms in the industry imply a higher probability of seeing these firms lobbying.[191] Some Remarks about equation$\frac{\partial Pr(l_i>0)}{\partial n}$:

1. The sign of $\Delta'_{i,n}$ does not determine the sign of $\frac{\partial Pr(l_i>0)}{\partial n}$ but the interaction of the elasticity of $\Delta_i(n)$ and the elasticity of the scale of the firm. If we denote $\in_{\Delta_i} = \frac{\Delta'_{i,n}n}{\Delta_i}$ and $\in_{S_i} = \frac{S'_i(n)n}{S_i(n)}$ then the equation defining $\frac{\partial Pr(l_i>0)}{\partial n}$ is equivalent to $\in_{\Delta_i} + \in_{S_i} - \beta\frac{\partial n_\theta}{\partial n}$.

2. Notice that if $\Delta'_{i,n}, S'_i(n) < 0 < \frac{\partial n_\theta}{\partial n}$, then less concentrated industries will be less likely to lobby. However, if the first two derivatives are positive, I cannot say something about the effect of the number of firms in the probability.

3. Let $\in_E = \frac{E'(n)n}{E(n)}$ be the elasticity of the expected benefit $E(L) = \pi(L)G(L)$. Then, the $sign(\frac{\partial Pr(l_i>0)}{\partial n})$ can be summarized in Table A.1.

3.6.3.3 Other Results on the Intensive Margin

Asumming positive lobbying expenditures and taking derivatives on the equilibrium truthful schedule with respect to the number of firms we get:

$$sign(l'(n)) = sign(-c^{-1\prime}(\cdot)\left[g'(n) + ng''(n)\right])$$

Then, the sign of the lobbying expenditure of a given firm will depend on the concavity the expected benefit as well as both the slopes of the expected benefit and cost functions. As we know that the cost function is increasing by assumption, the sign of $sign(l'(n))$ is simplified to be

$$-sign(\left[g'(n) + ng''(n)\right])$$

Now, lets inspect each part of this equation. First the derivative of the total expected benefit $g'(n)$ is equal to $\frac{\left[\frac{\partial E(n)S(n)}{\partial n}\right]n^\beta - E(n)S(n)\beta n^{\beta-1}}{n^{2\beta}}$ which means that

[191]Notice that in order to assure unicity of the equilibrium in the number of firms let's suppose that $\beta\frac{\partial n_\theta}{\partial n} \geq 0$ .Then, we simply need to show that either 1. If $b'_i(n) > 0$, then $b_i(0) \leq \beta\frac{\partial n_\theta}{\partial n} \leq b_i(N)$ or 2. If $b'_i(n) < 0$, then $b_i(N) \leq \beta\frac{\partial n_\theta}{\partial n} \leq b_i(0)$. I still need to get more information about the $b'_i(n)$ to be able to guarantee the unicity. We can also work out some static comparative on the parameters of the function on $b_i(n)$.

$$sign(g'(n)) = sign(\in_E + \in_{S_i} - \beta)$$

Some tedious algebra shows that $g''(n) = \in_E \left(\gamma_E - \frac{1}{n}[\in_E - 1]\right) + \in_{S_i} \left(\gamma_{S_i} - \frac{1}{n}[\in_{S_i} - 1]\right)$ where $\gamma_x = \frac{X''(n)}{X'(n)}$. Finally

the term $g'(n) + ng''(n)$ is equal to:

$$\in_E + \in_{S_i} - \beta + \in_E (n\gamma_E - [\in_E - 1]) + \in_{S_i} (n\gamma_{S_i} - [\in_{S_i} - 1])$$

and simplifying we get $\in_E (n\gamma_E - \in_E + 2) + \in_{S_i} (n\gamma_{S_i} - \in_{S_i} + 2) - \beta$. Which can be summarized as:

$$sign(l'(n)) = -sign(\in_E \alpha_E + \in_{S_i} \alpha_{S_i} - \beta)$$

where $\alpha_X = 2 + n\gamma_X - \in_X$.[192] Notice here that the sign of the lobbying expenditure depends in a complicated way on both the concavity of the expected benefit and scale function as well as the number of firms and the rivalry of the bill.

Elasticities at the firm and and Industry level This subsection tries to clarify the main differences between the possible combinations of asymmetric/symmetric firm-level expenditures and endogenous/exogenous number of lobbying firms. First, notice that in general $L = \int_{j=1}^{n_l} l_j dj$ but if the firm lobbying expenditures are symmetric, $L = n_l l(n_l, n, \theta, \beta)$. If the number of firms lobbying is also endogenous, then in equilibrium n_l also depends on (n, θ, β). Let's see the four possible combinations between the asymmetric assumption on the firm-level expenditures and endogeneity in the number of lobbying firms.

- l symmetric and n_l exogenous.

In this case $L = n_l l(n_l, n, \theta, \beta)$. Then, taking derivatives with respect to the number of firms in the industry we get $\frac{\partial L}{\partial n} = n_l \frac{\partial l(\cdot)}{\partial n}$ which is the same as

$$\varepsilon_{L,n} = \varepsilon_{l,n}$$

The term $\varepsilon_{L,n}$ ($\varepsilon_{l,n}$) represents the elasticity of the industry (firm) lobbying expenditures to the number of firms of the industry. Notice that in general $\frac{\partial L}{\partial \gamma} = n_l \frac{\partial l(\cdot)}{\partial \gamma}$ with $\gamma = \theta, \beta$. So, $sign(\frac{\partial L}{\partial \gamma}) = sign\left(\frac{\partial l(\cdot)}{\partial \gamma}\right)$ or $\varepsilon_{L,\gamma} = \varepsilon_{l,\gamma}$. In Section 3.2, I have empirically shown that $\varepsilon_{L,n} > 0 > \varepsilon_{l,n}$, then the assumptions used under this scheme (i.e., l symmetric and n_l exogenous) are empirically ruled out at least for the case of the analysis on the effect of a change in the number of firms in the industry (i.e., n).

[192] If for instance, I assume a Pareto firm size distribution with exponent $\frac{1}{\gamma}$ (i.e., $S(n) = \frac{A}{n^{\gamma}}$) then $\in_{S_i} \alpha_{S_i} = -\gamma$.

- l symmetric and n_l endogenous.

In this case, $L = n_l(n,\theta,\beta)\, l(n_l(\cdot), n,\theta,\beta)$. Then, taking derivatives with respect to the number of firms in the industry we get $\frac{\partial L}{\partial n} = \frac{\partial n_l}{\partial n} l(n) + n_l \left[\frac{\partial l(\cdot)}{\partial n} + \frac{\partial l(\cdot)}{\partial n_l}\frac{\partial n_l}{\partial n}\right]$. Multiplying both sides by $\frac{n}{L}$, we get $\varepsilon_{L,n} = \frac{\partial n_l}{\partial n} l(n)\frac{n}{L} + n_l\left[\frac{\partial l(\cdot)}{\partial n}\frac{n}{L} + \frac{\partial l(\cdot)}{\partial n_l}\frac{\partial n_l}{\partial n}\frac{n}{L}\right]$ which is the same as

$$\varepsilon_{L,n} = \underbrace{\varepsilon_{n_l,n}\left(1+\varepsilon_{l,n_l}\right)}_{IE} + \underbrace{\varepsilon_{l,n}}_{DE}$$

where $\varepsilon_{n_l,n}$ is the elasticity of the number of lobbying firms with respect to the number of firms in the industry, and ε_{l,n_l} is the elasticity of the firm-level lobbying expenditures with respect to the number of lobbying firms. This equation says that in order to understand the effect of the market structure in the industry lobbying expenditures, there are two main effects: indirect and direct effect. The latter is simply the relationship between firm-level lobbying expenditures and market structure whereas the indirect effect shows the effect of the market structure on the number of lobbying firms. This indirect effect is composed of two parts: $\varepsilon_{n_l,n}$ and $\varepsilon_{n_l,n}\varepsilon_{l,n_l}$. The first part shows how the market structure directly affects the number of lobbying firms. The second term shows how the firm-level lobbying expenditures are affected by the market structure through the effect that the market structure has on the number of lobbying firms. Therefore, the direct effect measures all the effects that explain a firm-level expenditure not related to the externalities effects of the other firm-level expenditures. Notice, the case studied here is not ruled out by the empirical evidence as far as $\frac{\partial n_l}{\partial n} > 0$. Also notice that for $\gamma = \theta, \beta$, the assumptions above imply

$$\varepsilon_{L,\gamma} = \varepsilon_{n_l,\gamma} + \varepsilon_{l,\gamma}$$

That is, the total effect of the excludability (rivalry) of the bill is decomposed in two effects: 1. How the excludability (rivalry) of the bill affects the number of lobbying firms and 2. How the excludability (rivalry) level affects the firm-level lobbying expenditure.

My problem with this combination of assumptions is that if all firms are equal, then either all of them lobby $n_l = n$ or none of them $n_l = 0$. That is, if all the firms are equal to each other, it is difficult to understand why $n_l \in (0,n)$ in some industries as empirical evidence shows. If $n_l = 0$, then total lobbying expenditures are equal to zero (i.e., $L = 0$) for all levels of excludability and rivalry of the bills as well as all market concentration levels.

However, for the case where all firms lobby in the industry the condition relating elasticities at the firm and industry levels can be summarized as:

$$sign(\varepsilon_{L,n}) = sign(1 - \in_{l,n})$$

which is the case explored in the text.

- l asymmetric and n_l exogenous.

Notice that $L = \int_{j=1}^{n_l} l_j(n_l, n, \theta, \beta)\, dj = \int_{j=1}^{n} l_j(n_l, n, \theta, \beta)\, dj$ as $l_j = 0$ in the complement of n_l. The objective function for the firm i-th is $\frac{\pi(L)G(L)}{n^\beta} S_i - c_i(l_i)$. This equation can be written in a more concise way as $B(l_i, L) - c_i(l_i)$. Notice that $\frac{\partial B(l_i,L)}{\partial l_i} \geq 0$ and if $\frac{\partial B(l_i,L)}{\partial L} > (<) 0$ there are positive (negative) externalities. In the lobbying context there are positive externalities, because larger lobbying expenditures not only increase the size of the bill but also raise the probability of the bill being passed. Then, taking the first order condition of the objective function with respect to the lobbying expenditure implies $\frac{\partial B(l_i,L)}{\partial l_i} - \frac{\partial c_i(l_i)}{\partial l_i} = 0$ which is equal to

$$\frac{\partial B(\cdot)}{\partial l_i} + \frac{\partial B(\cdot)}{\partial L}\frac{\partial L}{\partial l_i} = \frac{\partial c_i(l_i)}{\partial l_i}$$

Notice that $\frac{\partial B(\cdot)}{\partial l_i}, \frac{\partial B(\cdot)}{\partial L}, \frac{\partial L}{\partial l_i}, \frac{\partial c_i(l_i)}{\partial l_i} > 0$. Then, inverting the equation we get the demand of political manipulation or simply the lobbying expenditure function $l_i = d_i(n, L)$. In equilibrium $\int_{i=1}^{n} l_i di = \int_{i=1}^{n} d_i(n, L) di = L$ then taking derivatives with respect to the number of firms n we get:

$$\frac{\partial L}{\partial n} = \int_{i=1}^{n} \left(\frac{\partial d_i}{\partial n} + \frac{\partial d_i}{\partial L}\frac{\partial L}{\partial n} \right) di$$

Then, $\frac{\partial L}{\partial n} = \frac{\int_{i=1}^{n} \frac{\partial d_i}{\partial n} di}{1 - \int_{i=1}^{n} \frac{\partial d_i}{\partial L} di}$ which implies:

$$\varepsilon_{L,n} = \frac{\int_{i=1}^{n} w_i \varepsilon_{l_i,n} di}{1 - \int_{i=1}^{n} w_i \varepsilon_{l_i,L} di}$$

where $w_i = \frac{l_i}{L}$ is the proportion of the i-th lobbying expenditure to the total industry lobbying expenditure, $\varepsilon_{l_i,n}$ is the elasticity of the lobbying expenditure of the i-th firm to the number of firms of the industry and $\varepsilon_{l_i,L}$ is the elasticity of the firm-level lobbying expenditure of the i-th firm to the industry-level lobbying industry. Then, the total industry lobbying expenditure elasticity depends on all the individual elasticities of both the number of firms and the total lobbying expenditure weighted by their relative size on the lobbying expenditure. One implication of this equation is that even if most of the companies tend to decrease their lobbying expenditures as the number of firms in the industry increases $\frac{\partial l_i}{\partial n} < 0$ and these firms are very sensitive to the total lobbying expenditure $\frac{\partial l_i}{\partial L} > 1$, if there are just a few big companies in the industry for which $\frac{\partial l_i}{\partial n} > 0$ and $\frac{\partial l_i}{\partial L} < 1$, the total industry elasticity may be positive. If we define $\varepsilon_{l,n} = \int_{i=1}^{n} w_i \varepsilon_{l_i,n} di$ and $\varepsilon_{l,L} = \int_{i=1}^{n} w_i \varepsilon_{l_i,L} di$, we conclude that

$$-sign(\varepsilon_{L,n}) = sign(\varepsilon_{l,L} - 1) \cdot sign(\varepsilon_{l,n})$$

The empirical finding shown in section 2 of this draft is $-sign(\varepsilon_{L,n}) < 0$, which means that the signs of the RHS should be different (i.e., $sign(\varepsilon_{l,L} - 1) = -sign(\varepsilon_{l,n})$). That means that if the weighted demand elasticity of the lobbying expenditure is inelastic (elastic) (i.e., $\varepsilon_{l,L} < (>) 1$), then the weighted demand elasticity of the number of firms should be positive (negative). A possible way to interpret our finding of a positive correlation between firm-level expenditure and number of firms is that on average $\varepsilon_{l,n} > 0$. If that is the case, then, we can conclude that $\varepsilon_{l,L} < 1$. That is, on average the firm-level expenditures do not respond *strongly* to the industry expenditures.[193]

- l asymmetric and n_l endogenous.

Similar to before but allowing the term n_l to be endogenous, we get
$L = \int_{j=1}^{n_l} d_j \left(L, n, n_l\left(n\right)\right) dj$, then $l_i = d_i(n, L, n_l\left(n\right))$.
In equilibrium $\int_{i=1}^{n} l_i di = \int_{i=1}^{n} d_i(n, L, n_l\left(n\right)) di = L$, then taking derivatives with respect to the number of firms n we get:

$$\frac{\partial L}{\partial n} = \int_{i=1}^{n} \left(\frac{\partial d_i}{\partial n} + \frac{\partial d_i}{\partial n_l\left(n\right)} \frac{\partial n_l\left(n\right)}{\partial n} + \frac{\partial d_i}{\partial L} \frac{\partial L}{\partial n} \right) di$$

Then, $\frac{\partial L}{\partial n} = \frac{\int_{i=1}^{n} \left(\frac{\partial d_i}{\partial n} + \frac{\partial d_i}{\partial n_l(n)} \frac{\partial n_l(n)}{\partial n} \right) di}{1 - \int_{i=1}^{n} \frac{\partial d_i}{\partial L(n)} di}$ which implies:

$$\varepsilon_{L,n} = \frac{\int_{i=1}^{n} w_i \left(\varepsilon_{l_i,n} + \varepsilon_{l_i,n_l} \varepsilon_{n_l,n} \right) di}{1 - \int_{i=1}^{n} w_i \varepsilon_{l_i,L} di}$$

If we define $\varepsilon_{l,n} = \int_{i=1}^{n} w_i \varepsilon_{l_i,n}$ and $\varepsilon_{l,L} = \int_{i=1}^{n} w_i \varepsilon_{l_i,L}$, and $\varepsilon_{l,n_l,n} = -\int_{i=1}^{n} w_i \varepsilon_{l_i,n_l} \varepsilon_{n_l,n}$ we conclude that:

$$-sign(\varepsilon_{L,n}) = sign(\varepsilon_{l,L} - 1) \cdot sign(\varepsilon_{l,n} - \varepsilon_{l,n_l,n})$$

The empirical finding shown in Section 3.2 is $\varepsilon_{L,n} > 0$ and we can argue that $\varepsilon_{l,n} < 0$. Then, if $\varepsilon_{l,n} - \varepsilon_{l,n_l,n} < 0$, $\varepsilon_{l,L} < 1$ and if $\varepsilon_{l,n} - \varepsilon_{l,n_l,n} > 0$, $\varepsilon_{l,L} > 1$.

[193] This is obviously a simplification of a more serious interpretation because the elasticies are weighted.

To sum up, Table A.2 shows the main results of this sub-section.[194]

According to this table the main difference between the first and second column is the indirect effect of the market structure on the number of lobbying firms, whereas the main difference between the first and second row is the externalities effect that appears in the denominator of the equation.

In intuitive terms, the relationship between industry lobbying expenditures and market structure (i.e., $\varepsilon_{L,n}$) depends on two key factors:

1. The level of firm-level lobbying expenditure (i.e., l_i) and how this function depends on the market concentration (i.e $\varepsilon_{l_i,n}$),the number of firms lobbying (i.e., ε_{l_i,n_l}) and the other firms' lobbying expenditures ($\varepsilon_{l_i,L}$).

2. The number of lobbying firms (i.e., n_l) and how this number is related to the market structure (i.e., $\varepsilon_{n_l,n}$).

Only the fourth case when the firm-level lobbying expenditures are asymmetric and n_l is endogenous captures all the effects mentioned above.

Impact ot the Rivalry of the Bill Notice that according to the equilibrium found above the firm-level lobbying expenditure is $l\left(n\right)=c^{-1}\left(-ng'\left(n\right)\right)$. Then, in order to know the effect of the rivalry of the bill in the lobbying expenditure, we can take derivatives with respect to β and get $\frac{\partial l}{\partial \beta}=\frac{\partial c^{-1}(\cdot)}{\partial(\cdot)}\frac{\partial(\cdot)}{\partial g'(n)}\frac{\partial g'(n)}{\partial\beta}$. As the cost function is increasing in the lobbying expenditure,we can conclude that the sign of $sign(\frac{\partial l}{\partial \beta})$ is equal to $sign(-\frac{\partial g'(n)}{\partial \beta})$. Then, as $g'\left(n\right)=\left[\frac{\partial E(n)S(n)}{\partial n}\right]\frac{1}{n^{\beta}}-\frac{E(n)S(n)\beta}{n^{\beta+1}}$ we can take derivatives with respect to β to get:

$$\frac{E\left(n\right)S\left(n\right)\left[\ln n\beta-1\right]}{n^{\beta+1}}-\left[\frac{\partial E\left(n\right)S(n)}{\partial n}\right]\frac{\ln n}{n^{\beta}}$$

which is simplified to

$$\ln n\left[\beta-\in_E-\in_{S_i}\right]-\frac{1}{n^{\beta}}$$

which shows how the expected benefit and the scale of the firm elasticities matter to understand the impact of the rivalry of the firm in the lobbying expenditures. To understand better the relationship between industry-level lobbying expenditures and rivalry of the bill, let's analyze the following figures.

Figure A.1 represents three regions. Region I is the region where $L'\left(n\right)>0$ because $l'\left(n\right)>0$. Region II, has both decreasing firm and industry level lobbying

[194] For the case of $\gamma=\theta,\beta$, following the order of this section, these elasticities are as follows: $\varepsilon_{l,\gamma}$, $\varepsilon_{n_l,\gamma}+\varepsilon_{l,\gamma}$, $\frac{\varepsilon_{l,\gamma}}{1-\varepsilon_{l,L}}$ and $\frac{\varepsilon_{l,\gamma}-\varepsilon_{l,n_l,\gamma}}{1-\varepsilon_{l,L}}$. Notice that all but one are very similar to the elasticities of the market structure. The only difference is when the firm-level lobbying expenditure is symmetric and the number of firms is endogenous. This is due to the total number of lobbying firms n_l appearing in the equation defining the industry-level lobbying expenditures.

expenditures $l'(n), L'(n) < 0$. Finally, region III $|l'(n)| < \frac{l(n)}{n}$ which implies that $L'(n) > 0$.

This graph is obviously not general as I have assumed that the average lobbying expenditure is decreasing in the the rivalry of the firm. However, given the equilibrium found we can make some remarks on the behavior of the firm-level lobbying function on the level of rivalry of the firm. In this section I am concerned about the existence of region 3. First, notice that if $g'(n) > 0$, the the lobbying function is 0. That means, that our lobbying function cannot be negative. Second, notice that $g'(n) = \left[\frac{\partial E(n)S(n)}{\partial n}\right]\frac{1}{n^{\beta}} - \frac{E(n)S(n)\beta}{n^{\beta+1}}$.

Then, when the bill is not rival this function is $\left[\frac{\partial E(n)S(n)}{\partial n}\right]$, which means that if $\in_E + \in_{S_i} < 0$, $\frac{l(n)}{n} > 0$, while $\in_E + \in_{S_i} > 0$, $\frac{l(n)}{n} = 0$. Notice that if the bill is purely rival $l(n) \propto -[\in_E + \in_{S_i} - 1]$. Then, either $l(n) = 0$ if $\in_E + \in_{S_i} > 1$ or $l(n) > 0$ if $\in_E + \in_{S_i} < 1$.

If the average lobbying expenditure is decreasing there will always be points where the (negative) marginal lobbying expenditure is lower than the average lobbying expenditure. Also, notice there cannot be horizontal lines representing the relationship between average lobbying expenditure and rivalry of the firm. Also notice that if $l(\beta = 0) > 0$ or $l(\beta = 0) = 0$, but $\frac{\partial l}{\partial \beta} > 1$ there is no any problem as there will always be point to the left of the 45 degrees line below the average lobbying expenditure.

Figure A.2 represents two relevant increasing curves. Both curves are characterized by $l(\beta) = 0$ for some range of values in the domain of β. Both curves start at the right of the point (0,0) but one has a higher slope than the other. The curve at the left allows the existence of region 3 while the second curve is the only case where region 3 does not exist. Increasing curves happen if we have $\frac{1}{n^{\beta}} + \ln n\left[\in_E + \in_{S_i} - \beta\right] > 0$. Notice that if $\in_E + \in_{S_i} - \beta > 0$, $g'(n) > 0$ and therefore $l(\beta) = 0$ for all β. On the other hand, if $\in_E + \in_{S_i} - \beta < 0$, $g'(n) < 0$ and therefore $l(\beta) > 0$. This is the only relevant case where region III may not exist. In this case we also need that $\frac{1}{n^{\beta}} > -\ln n\left[\in_E + \in_{S_i} - \beta\right]$. There is nothing that guarantees us that this is impossible.

Finally, notice that as $l(n) = \max(c^{-1}(-ng'(n), 0)$ which is equal to $g'(n) = \left[\frac{\partial E(n)S(n)}{\partial n}\right]\frac{1}{n^{\beta}} - \frac{E(n)S(n)\beta}{n^{\beta+1}}$, then $sign(\frac{\partial l(n)}{\partial L}) = -sign(\frac{\partial g'(n)}{\partial L})$. Then notice that $sign(\frac{\partial g'(n)}{\partial L})$ is equal to the sign of $[\frac{\partial ES}{\partial n \partial L}]\frac{1}{n^{\beta}} - \frac{\partial ES}{\partial L}\frac{\beta}{n^{\beta+1}}$ which is:

$$sign\left(\left[\frac{\partial ES}{\partial n \partial L}\right] - \frac{\partial ES}{\partial L}\frac{\beta}{n}\right)$$

and defining $ES_L = \frac{\partial ES}{\partial L}$ and applying Young's theorem we get:

$$sign\left(\left[\frac{\partial ES_L}{\partial n}\right] - \frac{ES_L \beta}{n}\right) = sign\left(\left[\frac{\partial ES_L}{\partial n}\frac{n}{ES_L}\right] - \beta\right)$$

$$sign(\frac{\partial l\left(n\right)}{\partial L}) = -sign(\in_{ES_L,n} - \beta)$$

This means that if the expected benefit is submodular in the total lobbying expenditures and the number of firms, then there are positive externalities, while if these benefits are supermodular in L and n, there are negative externalities.

An alternative and more intuitive way to see the previous result is the following. Let's denote $F = \frac{S_i}{n^\beta}\left[\frac{\partial\pi(L)G(L)}{\partial l_i}\right] - \frac{\partial c_i(l_i)}{\partial l_i} = 0$. Notice that this equation is continuous and $\frac{\partial F}{\partial l} \neq 0$, then l may be expressed as a function of n in some domain of F. That is, there exists a function over the domain such that $l = l(n)$. Then, $l'(n) = -\frac{\frac{\partial F}{\partial n}}{\frac{\partial F}{\partial l}}$ equals to:

$$-\left[\frac{-\beta\frac{S_i}{n^{\beta+1}}\left(\pi\left(L\right)G\left(L\right)\right)'}{\frac{S_i}{n^\beta}\left(\pi\left(L\right)G\left(L\right)\right)'' - c_i\left(l_i\right)''}\right]$$

Notice that the sign of this derivative is represented by the following function:

$$sign(l'(n)) = -1_{sign((\pi(L)G(L))'')=sign(c_i(l_i)'')}\left\{sign\left(\frac{S_i}{n^\beta} - \frac{c_i\left(l_i\right)''}{\left(\pi\left(L\right)G\left(L\right)\right)''}\right)\right\}$$

where $1_{sign((\pi(L)G(L))'')=sign(c_i(l_i)'')}$ is an indicator function that is equal to $sign\left(\frac{S_i}{n^\beta} - \frac{c_i(l_i)''}{(\pi(L)G(L))''}\right)$ if the the expected benefit and the cost function are both concave or both convex. In the first case, only one of these functions is concave and the other is not, then $l'(n) < 0$.

If we are interested to understand the effect of the rivalry of the bill in the total lobbying expenditures, we know that $\frac{\partial L}{\partial l}\frac{\partial l}{\partial \beta}$, then using the same intuition as above we know that $\frac{\partial l}{\partial \beta} = -\frac{\frac{\partial F}{\partial \beta}}{\frac{\partial F}{\partial l}}$ equals to:

$$-\left[\frac{-\frac{S_i}{n^\beta}\left(\pi\left(L\right)G\left(L\right)\right)'\ln(n)}{\frac{S_i}{n^\beta}\left(\pi\left(L\right)G\left(L\right)\right)'' - c_i\left(l_i\right)''}\right]$$

Then, $sign(\frac{\partial L}{\partial \beta}) = sign(\frac{S_i}{n^\beta}\left(\pi\left(L\right)G\left(L\right)\right)'' - c_i\left(l_i\right)'')$. A common assumption will be that the expected benefit is concave and the cost function is convex, then, in this case, industries lobbying for non-congestible bills will spend more on lobbying than industries lobbying for fully rival bills.

The results from this section summarize the importance of the concavity assumption in both functions $\pi\left(L\right)G\left(L\right)$ and $c_i\left(l_i\right)$.

3.6.3.4 Other Results of the Model

In the previous section we were able to establish that $sign(\frac{\partial Pr(l_i>0)}{\partial n}) = sign(\in_{\Delta_i} + \in_{S_i} - \beta)$. I have been able to make some empirical inference to support the idea that $\in_{S_i} < 0$ but I have still remained silent about the possible sign of $\in_{\Delta_i}$. In order to advance in this direction notice that:

$$\frac{\partial \Delta_i \left(L\left(n\right), L_{-i}\left(n\right)\right)}{\partial n} = \frac{\partial \Delta_i \left(L\left(n\right)\right)}{\partial L\left(n\right)} \frac{\partial L\left(n\right)}{\partial n} + \theta \frac{\partial \Delta_i \left(L_{-i}\left(n\right)\right)}{\partial L_{-i}\left(n\right)} \frac{\partial L_{-i}\left(n\right)}{\partial n}$$

Notice that by assumption and definition $\frac{\partial \Delta_i(L(n))}{\partial L(n)} > 0 > \frac{\partial \Delta_i(L_{-i}(n))}{\partial L_{-i}(n)}$. Then, for the sake of simplicity we can assume that $sign(\frac{\partial L(n)}{\partial n}) = sign(\frac{\partial L_{-i}(n)}{\partial n}))$. Then,

$$sign(\frac{\partial \Delta_i \left(L\left(n\right), L_{-i}\left(n\right)\right)}{\partial n}) = sign(\underbrace{\underbrace{\frac{\partial \Delta_i \left(L\left(n\right)\right)}{\partial L\left(n\right)}}_{+} + \theta \underbrace{\frac{\partial \Delta_i \left(L_{-i}\left(n\right)\right)}{\partial L_{-i}\left(n\right)}}_{-}}_{Part\ I}) sign(\underbrace{\frac{\partial L\left(n\right)}{\partial n}}_{Part\ II})$$

Notice that if the bill is perfectly excludable, the effect of n on the industry-level lobbying expenditures uniquely determines the effect of the market concentration in the level of Δ_i.[195] However, if the bill is not perfectly excludable, then we need to not only see the sign of $\frac{\partial L(n)}{\partial n}$ but also the relative size of $\theta \frac{\partial \Delta_i(L_{-i}(n))}{\partial L_{-i}(n)}$ with respect to $\frac{\partial \Delta_i(L(n))}{\partial L(n)}$. Finally if $\frac{\partial \Delta_i(L(n))}{\partial L(n)} = \theta \frac{\partial \Delta_i(L_{-i}(n))}{\partial L_{-i}(n)}$, then $\frac{\partial \Delta_i(n)}{\partial n} = 0$ and it is true that industries with more firms will tend to lobby less (i.e., $\frac{\partial Pr(l_i > 0)}{\partial n} < 0$).

To analize this condition, I will separate the analysis between two parts: Part I which talks about the relative size of the derivatives of Δ_i with respect to the total lobbying expenditures and Part II which is concerned about the effect of the number of firms in the total lobbying expenditures.

Part I. Although we know that $\frac{\partial \Delta_i(L(n))}{\partial L(n)} > 0$, it is possible that the slope of the function Δ_i in the point $L_{-i}(n)$ is larger than the slope in the point $L(n)$. This for example would happen if the curve is sufficiently concave such that the function has decreasing marginal benefit. The next result shows that if the expected benefit function is concave in the lobbying expenditure, this function is actually concave in the total lobbying expenditure.

Lemma A.1. If the expected benefit function is concave in the firm-level lobbying expenditure, then it is concave in the industry-level lobbying expenditure.

Proof. We know that $\frac{\partial \pi(L)G(L)}{\partial l_i} = \frac{\partial \pi(L)G(L)}{\partial L} \frac{\partial L}{\partial l_i}$, then $\frac{\partial^2 \pi(L)G(L)}{\partial l_i^2} = \left(\frac{\partial^2 \pi(L)G(L)}{\partial L^2}\right) \frac{\partial L}{\partial l_i} \frac{\partial L}{\partial l_i} + \frac{\partial \pi(L)G(L)}{\partial L} \frac{\partial^2 L}{\partial l_i^2}$. As $L = \sum_{i=1}^{n} l_i$, then $\frac{\partial L}{\partial l_i} = 1$ and $\frac{\partial^2 L}{\partial l_i^2} = 0$. Then, $\frac{\partial^2 \pi(L)G(L)}{\partial l_i^2} = \left(\frac{\partial^2 \pi(L)G(L)}{\partial L^2}\right)$. If $\pi(L)G(L)$ is a concave function in L, then$[\pi(L)G(L)]'(L_{-i} - L) \geq \pi(L_{-i})G(L_{-i}) - \pi(L)G(L)$ for all L_{-i}, L. If $L_{-i} = 0$, then $[\pi(L)G(L)]' \geq \frac{\pi(L)G(L)}{L}$
.

Unfortunately, this would mean that we are not be able to say too much about the sign of the first part. To understand better the problem, let me explain the problem with the Figure A.3. The figure shows the expected benefit as a function of the industry-level lobbying expenditure. There are three points of interest: A, B and

[195] As in Section 3.2 we showed that $\frac{\partial L(n)}{\partial n} > 0$, we can conclude that if $\theta = 0$, $\frac{\partial \Delta_i(n)}{\partial n} > 0$.

C. Let's denote L_x as the lobbying expenditure at the point $x = A, B$ and C. The slope of these points are represented by $\frac{\partial \pi(L_x)G(L_x)}{\partial L_x}$ and notice that $\frac{\partial \pi(L_B)G(L_B)}{\partial L_B} > \frac{\partial \pi(L_A)G(L_A)}{\partial L_A} > \frac{\partial \pi(L_C)G(L_C)}{\partial L_C} \geq 0$. Notice that the assumption of $\frac{\partial \pi(L)}{\partial l_i}, \frac{\partial G(L)}{\partial l_i} > 0$, rules out the cases where $\frac{\partial \pi(L_x)G(L_x)}{\partial L_x} < 0$.

Let's assume that $L_{-i} = L_A$. That is, the point A corresponds to the level of industry expenditure without the lobbying expenditure of firm i-th. Then, firm i-th lobby the amount $l_i > 0$. Then, the new point will be L. As we know that $L - L_{-i} > 0$, there are three relevant cases where L may be located: In the boundary of A, in B or in C.

Case I. $L - L_{-i} = \varepsilon$. That is, the lobbying expenditure of firm i-th is relatively small to the preceding total lobbying industry (i.e., $\frac{l_i}{L_{-i}} \to 0$). In this case the new level of industry-level lobbying expenditure is around a vecinity of the previous point. This implies that $\frac{\partial \pi(L_{-i})G(L_{-i})}{\partial L_{-i}} \approx \frac{\partial \pi(L)G(L)}{\partial L}$, which makes $\frac{\partial \Delta_i(n)}{\partial n} \to 0$. Therefore $\frac{\partial Pr(l_i > 0)}{\partial n} < 0$. The same conclusion also applies if we start from point B and C.

Case II. $L = L_B$. In this case, the new point has a higher slope than the previous point $\frac{\partial \pi(L)G(L)}{\partial L} > \frac{\partial \pi(L_{-i})G(L_{-i})}{\partial L_{-i}}$. Then, if $\frac{\partial \pi(L)G(L)}{\partial L} > \theta \frac{\partial \pi(L_{-i})G(L_{-i})}{\partial L_{-i}}$ we know that $\frac{\partial \Delta_i(n)}{\partial n} > 0$. If point B is sufficiently steep, then $\in_{\Delta_i}$ can be sufficiently high to make $\frac{\partial Pr(l_i > 0)}{\partial n} > 0$. However, this case does not imply that we will always get a positive relation between the probability of firms lobbying and the concentration of the industry as we can have $\frac{\partial \pi(L)G(L)}{\partial L} > \frac{\partial \pi(L_{-i})G(L_{-i})}{\partial L_{-i}}$ with $\frac{\frac{\partial \pi(L)G(L)}{\partial L}}{\frac{\partial \pi(L_{-i})G(L_{-i})}{\partial L_{-i}}} < 0$.

Case III. $L = L_C$. In this case, the new point has a smaller slope than the previous point, so $\frac{\partial \pi(L)G(L)}{\partial L} < \frac{\partial \pi(L_{-i})G(L_{-i})}{\partial L_{-i}}$. If point C is sufficiently steep (i.e., $\theta < \frac{\frac{\partial \pi(L_{-i})G(L_{-i})}{\partial L_{-i}}}{\frac{\partial \pi(L)G(L)}{\partial L}}$), then $\frac{\partial \Delta_i(n)}{\partial n}$ is negative and as a consequence $\frac{\partial Pr(l_i > 0)}{\partial n} < 0$. The same conclusion also applies if we start from point B and we move to a point C. Case III represents a case where the lobbying expenditure of a firm is relatively high compared to the already existing lobbying expenditures, that is $\frac{l_i}{L_{-i}} >> 0$, while case II can be considered as an intermediate firm case.

To sum up, if we start from any point and a new firm contributes small lobbying expenses, then the probability of lobbying from a new firm decreases as there are more firms in the market. However, the entry of a firm willing to contribute medium or big lobbying expenditure (case II or case III) can increase or decrease the probability of lobbying as the market structure changes. In these two last cases, this probability will depend on the relative size of the rivalry of the bill θ.

Part II This part is included in the text and corresponds to the discussion of $sign(\frac{\partial L(n)}{\partial n})$.

Comparing Industries and Entry (Exit) of Firms This section is answering two questions. First, if we are in an equilibrium what happens if there is entry or exit of firms. Second, how can we compare the equilibrium expenditures among different industries when they only differ by the number of firms.

The lobbying firms choose the level of lobbying expenditure when the marginal cost equals the marginal benefit. That is $\frac{S_i}{n^\beta}\left[\frac{\partial \pi(L)}{\partial l_i}G\left(L\right)+\frac{\partial G(L)}{\partial l_i}\pi\left(L\right)\right]=\frac{\partial c_i(l_i)}{\partial l_i}$ (assuming an interior solution, otherwise $\frac{S_i}{n^\beta}\left[\frac{\partial \pi(L)}{\partial l_i}G\left(L\right)+\frac{\partial G(L)}{\partial l_i}\pi\left(L\right)\right]-\frac{\partial c_i(l_i)}{\partial l_i}\leq 0$, $l_i \geq 0$). Then, marginal benefit equals marginal cost can be represented as:

$$mb_i=\frac{\partial \pi(L)G(L)}{\partial l_i}\frac{S_i}{n^\beta}=\frac{\partial c_i\left(l_i\right)}{\partial l_i}=mc_i$$

Now, let's take derivatives of both marginal benefit and cost with respect to the number of firms. Let's first analyze the case of the marginal benefit $mb_i(n)$. Taking derivatives with respect to n, we get:

$$\frac{\partial mb_i}{\partial n}=\frac{\partial \pi(L)G(L)}{\partial l_i \partial n}\frac{S_i}{n^\beta}+\frac{\partial \pi(L)G(L)}{\partial l_i}\left(\frac{S_i' n^\beta-\beta S_i n^{\beta-1}}{n^{2\beta}}\right)$$

which is equal to

$$\frac{\partial \pi(L)G(L)}{\partial l_i \partial n}\frac{S_i}{n^\beta}+\frac{\partial \pi(L)G(L)}{\partial l_i}\left[\frac{S_i'}{n^\beta}-\frac{\beta S_i}{n^{\beta+1}}\right]$$

and dividing by $\frac{\frac{\partial \pi(L)G(L)}{\partial l_i}S_i}{n^\beta}>0$ we get:

$$\frac{\frac{\partial \pi(L)G(L)}{\partial l_i \partial n}}{\frac{\partial \pi(L)G(L)}{\partial l_i}}+\left[\frac{S_i'}{S_i}-\frac{\beta}{n}\right]$$

Then, multiplying by n we get

$$sign(\frac{\partial mb_i}{\partial n})=sign(\in_{l_i,n}+\in_{S_i}-\beta)$$

where $\in_{l_i,n}=\frac{\partial \pi(L)G(L)}{\partial l_i \partial n}\frac{n}{\frac{\partial \pi(L)G(L)}{\partial l_i}}$ is the elasticity of the marginal expected benefit with respect to the number of firms.

Now, to see what happens with the marginal cost $mc_i(n)$, let's take derivatives with respect to the number of firms:

$$\frac{\partial mc_i}{\partial n}=\frac{\partial c_i\left(l_i\right)}{\partial l_i \partial n}$$

Then, if we are initially in an equilibrium and there is entry (exit) of firms, the lobbying expenditures of the old (remaining) firms will raise or fall according to both the signs of $\frac{\partial mb_i}{\partial n}$ and $\frac{\partial mc_i}{\partial n}$. In particular, there are 6 possible cases. The first two cases are when the signs of both derivatives are equal and the four remaining cases where the signs of the derivative of the marginal benefit and marginal cost with respect to the number of firms differ:

Case I. In this case, both derivatives are non-negative, so the new equilibrium lobbying expenditure is located to the right (or the same level) of the initial equilibrium. That is, as $\frac{\partial mb_i}{\partial n}, \frac{\partial mc_i}{\partial n}\geq 0$ the sign of $l'\left(n\right)$ is non-negative.

Case II. In this case, both derivatives are non-positive, so the new equilibrium lobbying expenditure is located to the left (or the same level) of the initial equilibrium. That is, as$\frac{\partial mb_i}{\partial n}, \frac{\partial mc_i}{\partial n} \leq 0$ the sign of $l'(n)$ is non-positive.

Case III. In this case, the marginal benefit is non-negative but the marginal cost is non-positive. That is $\frac{\partial mb_i}{\partial n} \geq 0 \geq \frac{\partial mc_i}{\partial n}$. Then, there are two relevant cases, one where the marginal cost is more sensitive to the entry (exit) of firms than the marginal benefit and other where the marginal cost is less sensitive. Case IIIa represents the situation where the marginal benefit is more sensitive than the marginal cost: $\frac{\partial mb_i}{\partial n} \geq \left|\frac{\partial mc_i}{\partial n}\right|$. Therefore, in this case the firm-level lobbying expenditure is non-decreasing $l'(n) \geq 0$. The other case is Case IIIb and happens when the marginal cost is more sensitive than the marginal benefit $\frac{\partial mb_i}{\partial n} \leq \left|\frac{\partial mc_i}{\partial n}\right|$. In this case, the lobbying expenditure is lower (higher) when there is entry (exit) of firms $l'(n) \leq 0$.

Case IV. This case can be analized similarly to case three. In this case $\frac{\partial mc_i}{\partial n} \geq 0 \geq \frac{\partial mb_i}{\partial n}$. Then, there are two possible cases, one where the marginal benefit is more sensitive than the marginal cost and other in which the opposite happens. The first case is Case IVa and is represented by the condition $\left|\frac{\partial mb_i}{\partial n}\right| \geq \frac{\partial mc_i}{\partial n}$. This inequality implies that $l'(n) \leq 0$. Finally, we have Case IVb where $\left|\frac{\partial mb_i}{\partial n}\right| \leq \frac{\partial mc_i}{\partial n}$ and then the lobbying expenditure is not to the left of the initial equilibrium $l'(n) \geq 0$. A more concise way to summarize these results is the following. If $sign\left(\frac{\partial mb_i}{\partial n}\right) = sign\left(\frac{\partial mc_i}{\partial n}\right)$, then

$$sign\left(l'(n)\right) = sign\left(\frac{\partial mc_i}{\partial n}\right)$$

On the other hand, if $sign\left(\frac{\partial mb_i}{\partial n}\right) = -sign\left(\frac{\partial mc_i}{\partial n}\right)$, then

$$sign\left(l'(n)\right) = sign\left(\left|\frac{\partial mb_i}{\partial n}\right| - \left|\frac{\partial mc_i}{\partial n}\right|\right)$$

That is, if the signs are equal, the supermodularity of the cost function in the lobbying expenditures and the number of firms, determines if more firms imply more lobbying expenditures or not. If the signs are not equal, then the relative sensitivity of the marginal benefit with respect to the marginal cost determine the sign of $l'(n)$.

Lets assume that the cost function is convex in l_i and the $\pi(L)G(L)$ is concave in l_i. This assumption implies that mb_i is decreasing in l_i and mc_i is increasing in l_i. Figure A.4 represents the problem. Let's assume we start in the point where $mc_i(n) = mb_i(n)$. Then, assume there is entry of firms in the industry. The firm-level lobbying expenditure will increase if there are more firms in cases I, IIIa, and IVb. A necessary condition for the first two regions to happen is that the expected benefit is positive, while a necessary condition to see cases I and IVb happening is that the lobbying cost function is supermodular in the level of the lobbying expenditure and the number of firms. Therefore, a necessary condition to observe $l'(n) \geq 0$ is that either the expected benefit $\pi(L)G(L)$ or the cost function is supermodular in the lobbying expenditure and the number of firms.

The function $E(L,n) = \pi(L,n)G(L,n)$ is supermodular in L and n if when $L > L_{-i}$, $E(L,n) - E(L_{-i},n)$ is non decreasing in n. If E is supermodular in (L, n), then the incremental gain to choosing to have a larger lobbying expenditure is greater when the number of firms is higher. This definition is equivalent to $n > n\prime$ implies $E(L,n) - E(L,n\prime)$ non decreasing in L.[196]

Firm Size distribution We can assume different distribution functions for the scale of the firm. The notation assumes that firms are organized. In the case of the arithmetic progression, firms are organized in ascending order (that is, S_1 is the smallest firm) whereas the last four columns are organized in descending order (that is, S_1 is the largest firm of the industry). S_n should be understood as the size of the n-th firm (the largest for the arithmetic progression and the smallest one for the firm rank specifications). There are least three variables where the assumption of the relationship between the scale firm and the number of firms may have an impact. First, how the scale parameter of these specification change the probability of the n-th firm lobbying $\frac{\partial Pr(l_n>0)}{\partial \delta}$. Second, how the assumption in each of these specifications affect the elasticity of the scale of the firm with respect to the number of firms for the n-th firms (i.e., $\in_{S_n}$) . This elasticity in turn, would affect not only $\frac{\partial Pr(l_n>0)}{\partial n}$ but $sign(l'(n))$ and $sign(\frac{\partial mb_n}{\partial n})$ as it was specified above. Finally, the functional form of these specification affect α_{S_n} which in turn affect $sign(l'(n))$. Table A.3. shows a summary of the impact on the assumed distribution. Notice that each of these five specifications allow both even and uneven distributions. The advantadge of the arithmetic progression besides its simplicity is that the HHI can be written in terms of only one firm. HHI can also be written for each of these specifications as a function of the number of firms. As sometimes these equations are long I have decided to leave it out. The advantadge of the last four specification is that the parameters of the model can be statistically estimated using existing data. If S_i is distributed pareto the probability of seeing firm i-th lobbying will be

$$\Pr(firm\ i-th\ lobbying) = \left\{ \begin{array}{ccc} \frac{S_m n^\beta}{c_i(l_i)[\pi(L)G(L)-\theta\pi(L_{-i})G(L_{-i})]} & \text{if} & \frac{c_i(l_i)[\pi(L)G(L)-\theta\pi(L_{-i})G(L_{-i})]}{n^\beta} \geq S_m \\ 1 & \text{if} & \text{otherwise} \end{array} \right\}$$

This simple step allows me to estimate by GMM or Maximum likelihood β or θ when I assume parametric forms for the cost and the expected benefit.

[196] Applying Topki's theorem, we know that if E is supermodular in (L; n), then $L(n)$ is non decreasing in n. Notice that if $E(L,n)$ is supermodular, then the entire objective function is also supermodular, as the cost part does not depend on the (L,n). $E()$satisfies the single crossing condition if for all $L > L'$ and $n > n\prime, E(L,n) - E(L',n) \geq 0$ implies $E(L,n') - E(L',n') \geq 0$. If E is single crossing in (L,n), then $L^*(n) = \arg\max_x E(L,n)$ is nondecreasing in n. Moreover, if $L^*(n)$ is nondecreasing in n for all choice sets, then $E(L,n)$ is single-crossing in (L,n).

Lobbying associations The difference between private and social benefits could be seen by contrasting the marginal benefits of a lobbying association (joining all the firms in the industry) and a typical firm. The marginal benefits differ by the effect of a change in total lobbying expenditure on a firm's marginal benefit.

The total benefit of the industry-association is $B_A = \sum_{i=1}^{n} B_i(l_i, L)$. Taking derivatives with respect to L

$$\frac{\partial B_A}{\partial L} = \sum_{i=1}^{n} \frac{\partial B_i}{\partial L} + \sum_{i=1}^{n} \frac{\partial B_i}{\partial l_i} \frac{\partial l_i}{\partial L}$$

Then substracting from both sides the marginal private benefit we know that $\frac{\partial B_A}{\partial L} - \frac{\partial B_i}{\partial l_i}$ is equal to:

$$\sum_{i=1}^{n} \frac{\partial B_i}{\partial L} + \sum_{i=1}^{n} \frac{\partial B_i}{\partial l_i} \left(\frac{\partial l_i}{\partial L} - 1 \right)$$

3.6.4 Proofs

Theorem 3.1 *Firms are more likely to lobby when they are bigger, when they can be affected more by the bill, when the firm can affect more the expected benefit of the bill, if the lobbying costs are smaller or when the bills is more excludable and less congestible.*

Proof. Follows from the text. ■

Lemma 3.1 $\frac{\partial Pr(l_i>0)}{\partial n} = \in_b - \beta\frac{\partial n_\theta}{\partial n}$.

Proof. See section 3.6.3.2. ■

Theorem 3.2 $sign(\frac{\partial L}{\partial n}) = sign(1 - \in_{l,n})$ *where* $\in_{l,n} = -\frac{nl'(n)}{l(n)}$.

Proof. As the total lobbying expenditure is $L(n) = \sum_{j=1}^{n} l(n) = nl(n)$. Taking derivatives with respect to the number of firms we get $\frac{\partial L}{\partial n} = l(n) + nl'(n)$.[197]Then, this condition can be summarized as $sign(\frac{\partial L}{\partial n}) = sign(1 - \in_{l,n})$. ■

Theorem 3.3 $sign(\frac{\partial L(n)}{\partial n}) = sign(\sum\limits_{i=1}^{n} w_i \in_{l_i,n}) sign(1 - \sum\limits_{i=1}^{n} w_i \in_{l_i,L})$.

Proof. The objective function for the firm i-th is$\frac{\pi(L)G(L)}{n^\beta}S_i - c_i(l_i)$. This equation can be written in a more concisely way as follows, $B(l_i, L) - c_i(l_i)$. Notice that $\frac{\partial B(l_i,L)}{\partial l_i} \geq 0$ and if $\frac{\partial B(l_i,L)}{\partial L} > (<)0$ there are positive (negative) externalities. In the lobbying context there are positive externalities, because larger lobbying expenditures not only increase the size of the bill but also changes the probability of the bill being accepted. Then, taking the First Order Condition of the objective function with respect to the lobbying expenditure implies $\frac{\partial B(l_i,L)}{\partial l_i} - \frac{\partial c_i(l_i)}{\partial l_i} = 0$ which is equal to $\frac{\partial B(\cdot)}{\partial l_i} + \frac{\partial B(\cdot)}{\partial L}\frac{\partial L}{\partial l_i} = \frac{\partial c_i(l_i)}{\partial l_i} = mc_{l_i}$. Notice that $\frac{\partial B(\cdot)}{\partial l_i}, \frac{\partial B(\cdot)}{\partial L}, \frac{\partial L}{\partial l_i}, \frac{\partial c_i(l_i)}{\partial l_i} > 0$. Then, inverting the equation we get the demand of political manipulation or simply the lobbying expenditure function $l_i = d_i(mc_{l_i}, n, L)$. In equilibrium $\sum\limits_{i=1}^{n} l_i = \sum\limits_{i=1}^{n} d_i(mc_{l_i}, n, L) = L$ then taking derivatives with respect to the number of firms n we get $\frac{\partial L(n)}{\partial n} = \sum\limits_{i=1}^{n}\left(\frac{\partial d_i}{\partial mc_{l_i}}\frac{\partial mc_{l_i}}{\partial n} + \frac{\partial d_i}{\partial n} + \frac{\partial d_i}{\partial L(n)}\frac{\partial L(n)}{\partial n}\right)$. Now let's define $\frac{\partial d_i}{\partial n} + \frac{\partial d_i}{\partial mc_{l_i}}\frac{\partial mc_{l_i}}{\partial n} = \frac{\partial \widetilde{d_i}}{\partial n}$. Then, $\frac{\partial L}{\partial n} = \frac{\sum\limits_{i=1}^{n}\frac{\partial \widetilde{d_i}}{\partial n}}{1 - \sum\limits_{i=1}^{n}\frac{\partial d_i}{\partial L(n)}}$ which implies $\in_{L,n} = \frac{\in_{\widetilde{d}_i,n}}{1 - \in_{d_i,L}}$, where $\in_{L,n} = \frac{\partial L}{\partial n}\frac{n}{L}$, $\in_{\widetilde{d}_i,n} = \sum\limits_{i=1}^{n}\frac{\widetilde{d}_i}{L}\left(\frac{\partial \widetilde{d}_i}{\partial n}\frac{n}{\widetilde{d}_i}\right)$ and $\in_{d_i,L} = \sum\limits_{i=1}^{n}\frac{d_i}{L}\left(\frac{\partial d_i}{\partial L}\frac{L}{d_i}\right)$. Then, we conclude that $-sign(\in_{L,n}) = sign(\in_{d_i,L} - 1) \cdot sign(\in_{\widetilde{d}_i,n})$. The empirical finding showed at the beginning of this chapter is $-sign(\in_{L,n}) < 0$, which means that $sign(\in_{d_i,L} - 1) = -sign(\in_{\widetilde{d}_i,n}))$. That means that if the weighted demand elasticity of the lobbying expenditure is inelastic

[197]The reader should realize that $l'(n)$ includes both the indirect effect of n through the marginal cost and the direct effect.

(elastic) (i.e., $\in_{d_i,L}< (>)1$), then the weighted demand elasticity of the number of firms should be positive (negative). To understand better this result notice that $d_i = l_i$ and $\frac{\partial l_i}{\partial n} = \frac{\partial \widetilde{d}_i}{\partial n}$. Then, $sign(\frac{\partial L(n)}{\partial n}) = sign(\sum_{i=1}^{n} w_i \in_{l_i,n}) sign(1 - \sum_{i=1}^{n} w_i \in_{l_i,L})$, where $w_i = \frac{l_i}{L}$ is the proportion of the i-th lobbying expenditure to the total industry lobbying expenditure, $\in_{l_i,n}$ is the elasticity of the lobbying expenditure of the i-th firm to the number of firms of the industry and $\in_{l_i,L}$ is the elasticity of the firm-level lobbying expenditure of the i-th firm to the industry-level lobbying industry. ■

Theorem 3.4 *The lobbying expenditure (thruthful schedule) of one firm is given by* $l\,(n) = \max(c^{-1}\,(-ng'\,(n)\,, 0)$.

Proof. With n firms are lobbying and truthful contributions, the cost associated to the lobbying expenditure of one of these lobbying firms is $c(l\,(n)) = \frac{\pi(L)G(L)S(n)}{n^{\beta}} - a\,(n) = g\,(n) - a\,(n)$, where $a\,(n)$ is the payoff anchor of the contribution schedule of a lobbying firm. To find the equilibrium with n firms lobbying, let's assume that a measure $\triangle n$ of firms decide not to lobby. Then, the change of welfare for the lobbyist is $(n - \triangle n)\,[g\,(n - \triangle n) - a\,(n)]$. In equilibrium the lobbyist gets $(n)\,[g\,(n) - a\,(n)]$. Then, $a\,(n)$ has to be such that $(n - \triangle n)\,[g\,(n - \triangle n) - a\,(n)] = n\,[g\,(n) - a\,(n)]$ which is equal to: $\triangle n a\,(n) = n\,[g\,(n) - g\,(n - \triangle n)] + \triangle n g\,(n - \triangle n)$. Then, solving for the payoff anchor: $a\,(n) = \frac{n[g(n) - g(n - \triangle n)]}{\triangle n} + g\,(n - \triangle n)$. Now taking the limit when $\triangle n \to 0$, we get $a\,(n) = ng'\,(n) + g\,(n)$. By assumption, the cost function is increasing, so applying the inverse function theorem we know that the lobbying expenditure is $l\,(n) = \max(c^{-1}\,(-ng'\,(n)\,, 0)$. ■

Lemma 3.2 $\frac{\partial l(n)}{\partial \theta}, \frac{\partial L(n)}{\partial \theta} < 0$.

Proof. Follows from noticing that $\frac{\partial g(n)}{\partial \theta} > 0$ and that the relationship between θ and $l\,(n)$ is maintained when the firm-level expenditure is multiplied by a positive constant. ■

3.6.5 Tables of the Appendix

	Excludable	**Non Excludable**
Rival	$sign(\in_E + \in_{S_i} - \frac{\partial n_l}{\partial n})$	$sign(\in_{\Delta_i} + \in_{S_i} - 1)$
No-Rival	$sign(\in_E + \in_{S_i})$	$sign(\in_{\Delta_i} + \in_{S_i})$

Table 3.A.1. Sufficient Conditions Defining the Sign of $\frac{\partial Pr(l_i>0)}{\partial n}$

$\varepsilon_{L,n}$		n_l	
		Exogenous	*Endogenous*
l	*Symmetric*	$\varepsilon_{l,n}$	$\varepsilon_{n_l,n}(1+\varepsilon_{l,n_l})+\varepsilon_{l,n}$
l_i	*Asymmetric*	$\frac{\varepsilon_{l,n}}{1-\varepsilon_{l,L}}$	$\frac{\varepsilon_{l,n}-\varepsilon_{l,n_l,n}}{1-\varepsilon_{l,L}}$

Table 3.A.2. Relationship Between Industry and Firm Levels elasticicities.

	Arithmetic	As a function of the Firm Rank			
	Progression	**Linear**	**Power**	**Exponential**	**Quadratic**
Functional Form of S_n	$S_1 + (n-1)\,\delta$	$A + \frac{\delta}{n}$	$\frac{A}{n^{\delta}}$	$\frac{A}{e^{\delta n}}$	$A + \delta n + Cn^2$
$\frac{\partial Pr(l_n > 0)}{\partial \delta}$	$(n-1)$	$\frac{1}{n}$	$-\frac{A}{n^{\delta}} \ln n$	$-\frac{An}{e^{\delta n}}$	n
$\in_{S_n}$	$\frac{\delta n}{S_1 + (n-1)\delta}$	$-\frac{\delta}{An+\delta}$	$-\delta$	$-\delta n$	$\frac{\delta n + 2Cn^2}{A + \delta n + Cn^2}$
α_{S_n}	$\frac{2\left[S_1 + \left(\frac{n}{2} - 1\right)\delta\right]}{S_1 + (n-1)\delta}$	$\frac{\delta}{An+\delta}$	1	2	$\frac{\sum_{j=0}^{3} n^j \eta^j}{\sum_{j=0}^{3} n^j \mu^j}$ 198

Table 3.A.3: Firm Size Distribution.

Note: All the columns assume $A, \delta > 0$ except the quadratic form. All the specifications have $\frac{\partial Pr(l_n > 0)}{\partial \delta} > 0$, except the exponential. All the elasticities are negative exept the one using the arithmetic progression. α_{S_n} is positive for the arithmetic progression if there are at least two firms, and if there is only one we need that $S_1 > \frac{\delta}{2}$. α_{S_n} is positive for all the specification of the firm rank except the quadratic one, where is not possible to conclude the sign of this coefficient.

3.6.6 Figures of the Appendix

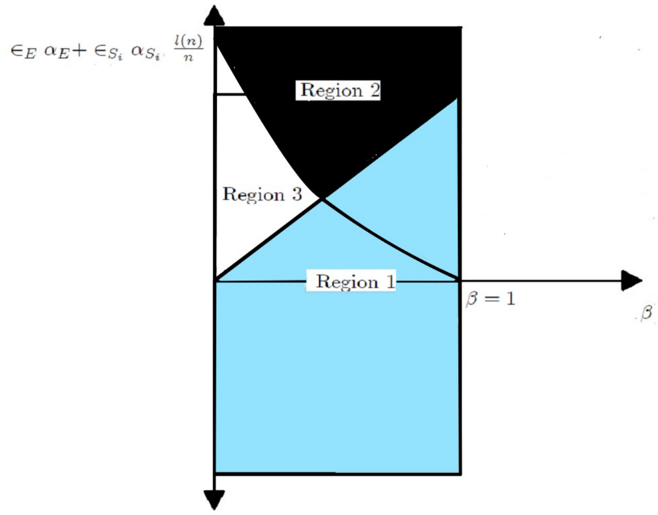

Figure 3.A.1: Three regions denoting the relationship between elasticities, average lobbying expenditure and rivalry of the firm.

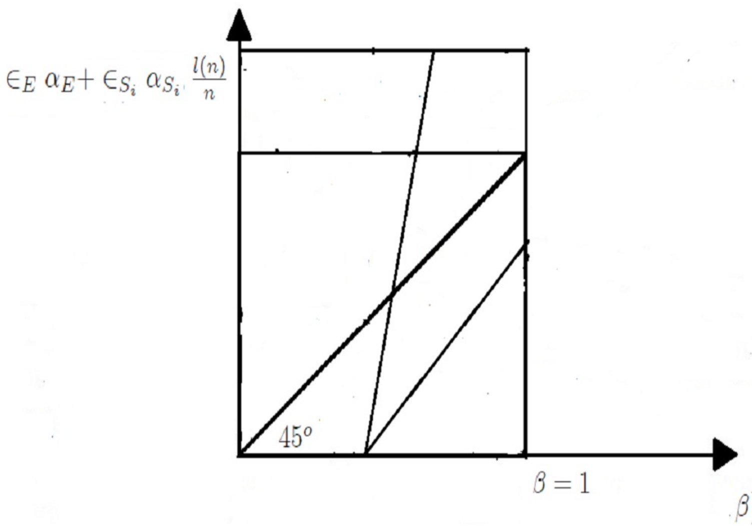

Figure 3.A.2: Relationship between elasticities, average lobbying expenditure and rivalry of the firm with.two relevant increasing curves

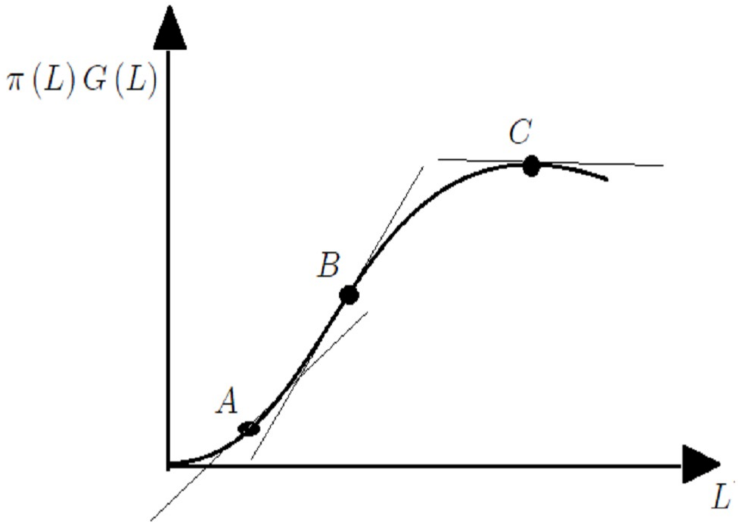

Figure 3.A.3: Three relevant cases where the slope of the expected

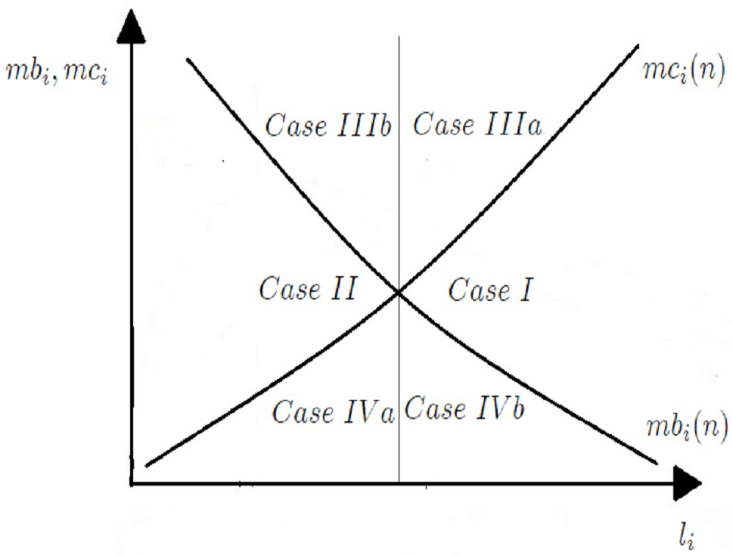

Figure 3.A.4: Six Relevant regions when there is entry (exit) of firms

Bibliography

Abraham, Katharine (1988): "Restructuring the Employment Relationship: The Growth of Market-Mediated Work Arrangements," In Robert A. Hart, Ed., Employment, Unemployment and Hours of Work, London: George Allen and Unwin, 288–311.

________________ and Susan K. Taylor(1996): "Firms' Use of Outside Contractors: Theory and Evidence," *Journal of Labor Economics*, 14 (3): 394-424.

Abramovsky, Laura, and Rachel Griffith (2006): "Outsourcing and Offshoring of Business Services: How Important is ICT?," *Journal of the European Economic Association*, 4 (2-3): 594-601.150

Acemoglu, Daron, Aghion Philippe, Clare Lelarge, John Van Reenen and Fabrizio Zilibotti (2007): "Technology, information and the decentralization of the firm," *Quarterly Journal of Economics* 122 (4):1759–1799.

Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti (2010): "Vertical Integration and Technology: Theory and Evidence," *Journal of the European Economic Association*, 8: 989–1033.

Al-Ubaydli, Omar and Patrick A. McLaughlin (2015): "RegData: A Numerical Database on Industry-Specific Regulations for All U.S. Industries and Federal Regulations, 1997-2012," *Regulation and Governance.* DOI: 10.1111/rego.12107

Anderson, Erin and David Schmittlein (1984): "Integration of the sales force: an empirical examination," *RAND Journal of Economics* (15): 385–395.

Anderson, Erin (1985): "The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis," *Marketing Science*, (4) 3: 234-254.

Andres, Gary (1985):"Business Involvement in Campaign Finance: Factors Influencing the Decision to Form a Corporate PAC," *PS: Political Science and Politics,* 18 (2): 215–219.

Ansolabehere, Stephen; James M. Snyder, Jr. and Micky Tripathi (2002): "Are PAC Contributions and Lobbying Linked?," *Business and Politics,* 4 (2): 1-26.

Ansolabehere, Stephen; John M. de Figueiredo, and James M. Snyder, Jr. (2003): "Why Is There So Little Money in US Politics? " *Journal of Economic Perspectives,* 17 (1): 105-30.

Ansolabehere, Stephen, James M. Snyder, Jr. and Michiko Ueda (2004): "Did Firms Profit from Soft Money? " *Election Law Journal,* 3 (2): 193-198.

Antras, Pol and Elhanan Helpman(2004): "Global Sourcing," *Journal of Political Economy*, 112 (3): 552-580.

Arrow, Kenneth (1974): "The Limits of Organization," *Norton.*

Atalay, Enghin; Hortaçsu, Ali and Chad Syverson (2014): "Vertical Integration and Input Flows," *American Economic Review*, 104 (4): 1120-48.

Autor, David H. (2003): "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing," *Journal of Labor*

Economics, (21) 1:1-42.

Azoulay, Pierre (2004): “Capturing Knowledge within and across Firm Boundaries: Evidence from Clinical Development,” *American Economic Review*, 94 (5): 1591-1612.

Baccara, Mariagiovanna and Ronny Razin (2007): “Bargaining on New Ideas: Rent Distribution and Stability of Innovative Firms,” *Journal of European Economic Association*, 5 (6): 1095-1129.

Baccara, Mariagiovanna and Heski Bar-Isaac (2008): “How to Organize Crime,” *Review of Economic Studies*, 75 (4): 1039- 1067.

Baker, George; Gibbons, Robert and Kevin J. Murphy (1994): “Subjective Performance Measures in Optimal Incentive Contracts,” *Quarterly Journal of Economics*,109 (4): 1125–56.

______________________________ (2002): “Relational Contracts and the Theory of the Firm,” *Quarterly Journal of Economics*, 117 (1): 39–84.

Barney J. (1991): “Firm Resources and Sustained Competitive Advantage,” *Journal of Management,*17 (1): 99-120.

Barrage, Lint; Eric Chyn, and Justine Hastings (2014): “Advertising, Reputation, and Environmental Stewardship: Evidence from the BP Oil Spill,” *NBER Working Paper No. 19838.*

Bartel Ann; Lach, Saul and Nachum Sicherman (2014): “Technological Change and the Make-or-Buy Decision,” *The Journal of Law, Economics and Organization, (*30) 1: 165-192.

Becker, Gary S. (1964): “Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.” New York: National Bureau of Economic Research.

_______ and Murphy, Kevin M. (1992): “The division of labor, coordination costs, and knowledge,” *Quarterly Journal of Economics* 107(4): 1137-1160.

Bergstrom, T. C., and R. P. Goodman (1973): “Private Demand for Public Goods,” *American Economic Review, (*63): 280–296.

Berlingieri, Giuseppe (2014): “Outsourcing and the rise in services,” *CEP Discussion Paper No. 1199.*

_________ (2015): “Managing Export Complexity: the Role of Service Outsourcing,” manuscript.

Bernheim, B.D., Whinston, M.D., (1986): “Menu Auctions, resource allocation, and economic influence,” *Quarterly Journal of Economics,* 101 (1): 1–31.

Bertrand, Marianne; Bombardini, Matilde and Francesco Trebbi (2014): “Is It Whom You Know or What You Know? An Empirical Assessment of the Lobbying Process,” *American Economic Review*, 104 (12): 3885-3920.

Blanes i Vidal, Jordi; Draca, Mirko and Christian Fons-Rosen (2012): “Revolving Door Lobbyists,” *American Economic Review*, 102 (7): 3731-48.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen (2014): "The Distinct Effects of Information Technology and Communication Technology on Firm Organization," *Management Science*, 60 (12): 2859–2885.

Boies, John L. (1989): "Money, Business, and the State: Material Interests, Fortune 500 Corporations, and the Size of Political Action Committees," *American Sociological Review,* 54 :821–33.

Bolton, Patrick, and Mathias Dewatripont(1994): "The Firm as a Communication Network." *Quarterly Journal of Economics*, 109: 809-839.

Bombardini, Matilde (2008): "Firm Heterogeneity and Lobby Participation," *Journal of International Economics,* 75 (2): 329–48.

_____________ and Francesco Trebbi (2012): "Competition and Political Organization: Together or Alone in Lobbying for Trade Policy?," J*ournal of International Economics,* 87 (1):18–26.

Borcherding, T. E. and R. T. Deacon (1972): "The demand for services of non-federal governments," *American Economic Review, (* 62):891–901.

Bresnahan, Timothy and Jonathan Levin (2012): "Vertical Integration and Market Structure," in R. Gibbons and D.J. Roberts, ed. *Handbook of Organizational Economics*, Princeton University Press.

Broda, Christian and David E. Weinstein (2006): "Globalization And The Gains From Variety," *Quarterly Journal of Economics*, (121) 2: 541-585.

Busch, Marc L. and Eric Reinhardt (1999): "Industrial location and voter participation in Europe," *American Journal of Political Science,* 43 (4): 1028-1050.

Caliendo, Lorenzo and Esteban Rossi-Hansberg (2012) "The Impact of Trade on Organization and Productivity," *Quarterly Journal of Economics*, 127 (3): 1393–1467.

Caroli Eve and John Van Reenen (2001): "Skill-Biased Organizational Change: Evidence from A Panel of British and French Establishments," *Quarterly Journal of Economics,* 116 (4):1448–1492.

Carrington, William J. (1993): "Wage Losses for Displaced Workers: Is It Really the Firm That Matters?" *Journal of Human Resources*, 28 (3): 435–62.

Chamberlin, John R (1974): "Provision of Collective Goods as a Function of Group Size," *American Political Science Review, (* 68): 707-16.

Chinitz, Benjamin (1961): "Contrasts in Agglomeration: New York and Pittsburgh," *American Economic Review, PandP*, 51(2): 279-289.

Clemens Austin (2014): "All Politics is Local, but Lobbying is Federal and Local: the validity of LDA data," *Business and Politics*, 16 (2): 267–289.

Coase, Ronald (1937): "The Nature of the Firm," *Economica*, 4 (16): 386-405.

Couch, Kenneth A. and Dana W. Placzek (2010): "Earnings Losses of Displaced Workers Revisited," *American Economic Review*, 100 (1): 572-589

Crémer, Jacques and Garicano, Luis and Andrea Prat (2007) : "Language and the theory of the firm," *Quarterly Journal of Economics*, 122 (1): 373-407.

De Figueiredo, John M. and James J. King (2004): "When do Firms Hire Lobbyists? The Organization of Lobbying at the Federal Communications Commission." *Industrial and Corporate Change* 13 (6): 883-900.

____________ and Brian S. Silverman (2006): "Academic earmarks and the returns to lobbying." *Journal of Law and Economics*, 49 (2): 597–625.

____________ and Brian Ritcher (2014): "Advancing the Empirical Research on Lobbying," *Annual Review of Political Science*, (17): 163-185.

Demsetz, H., (1988): "Theory of the firm revisited," *Journal of Law Economics and Organization,* (4): 141–162.

Demski, Joel; Lewis, Tracy; Yao, Dennis and Huseyin Yildrim (1999): "Practices for Managing Information Flows within Organizations," *Journal of Law, Economics and Organizations*, (15): 107–131.

Drutman, Lee (2010): "The Business of America is Lobbying: The Expansion of Corporate Political Activity and the Future of American Pluralism," *PhD Dissertation*, University of California, Berkeley.

Dube, Arindrajit, and Ethan Kaplan (2007): "The Labor Boundaries of the Firm: Employment and Independent Contracting," *manuscript.*

__________________ (2010): "Does Outsourcing Reduce Wages in the Low-Wage Service Occupations? Evidence from Janitors and Guards," *Industrial and Labor Relations Review*, 63 (2): 287–306.

Eisenhardt, K. M. and Martin, J. A. (2000): "Dynamic capabilities: what are they?," *Strategic Management Journal*, (21): 1105–1121.

Esteban, Joan and Debraj Ray (2001): "Collective Action and the Group Size Paradox," *American Political Science Review,* (95): 663–672.

Esty, D.C., and Caves, R.E. (1983): "Market structure and political influence: New data on political expenditures, activity, and success," *Economic Inquiry,* 21 (2): 4-38.

Foss, N. J. (1996): "More Critical Comments on Knowledge-Based Theories of the Firm", *Organization Science*, 7 (5): 519–523.

Galdon-Sanchez, Jose E., Gil, Ricard and Alberto Bayo-Moriones (2015): "Outsourcing of peripheral services: Evidence from Spanish manufacturing plant-level data," *European Economic Review* (78): 328–344.

Garicano, Luis (2000): "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*,108 (5): 874–904.

________ and Esteban Rossi-Hansberg (2006): "Organization and inequality in a knowledge economy. *Quarterly Journal of Economics*, 121:1383–435.

________ and Thomas Hubbard (2007): "Managerial Leverage is Limited by the Extent of the Market: Hierarchies, Specialization and the Utilization of Lawyers' Human Capital", *Journal of Law and Economics*, 50 (1): 1-45.

____________________ (2009): "Specialization, Firms, and Markets: The Division of Labor Within and Between Law Firms", *Journal of Law, Economics and Organization*, 25 (2): 339-371.

Grant, R.M. (1996): "Toward a Knowledge-Based Theory of the Firm," *Strategic Management Journal* (17): 109–122.

_________ (2002): "The Knowledge-Based View of the Firm", In C. W. Choo & N. Bontis (Eds.), *The Strategic Management of Intellectual Capital and Organizational Knowledge*: 133-148. Oxford: Oxford University Press.

Gibbons, Robert (2005): "Four Formal(izable) Theories of the Firm," *Journal of Economic Behavior and Organization* (58): 202-47.

Girma, Sourafel and Holger Görg (2004): "Outsourcing, Foreign Ownership, and Productivity: Evidence from UK Establishment-level Data," *Review of International Economics*, (12): 817–832.

Goldschmidt, Deborah and Johannes F Schmieder (2015): "The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure," *NBER Working Paper* No. 21366.

Grier, K. B., M. C. Munger, and B. E. Roberts (1991): "The industrial organization of corporate political participation," *Southern Economic Journal,* 57: 727-38.

Grossman, Sanford and Oliver Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy,* (94): 691-719.

Grossman, Gene and Elhanan Helpman (2002a): "Integration Versus Outsourcing In Industry Equilibrium," *Quarterly Journal of Economics*, 117 (1): 85-120.

____________________ (2002b): "Special Interest Politics," *MIT Press Books*, The MIT Press.

____________________ (2005): "Outsourcing in a Global Economy," *Review of Economic Studies*, 72 (1): 135-159.

____________________and Adam Szeidl (2006): "Optimal Integration Strategies for the Multinational Firm," *Journal of International Economics*, 70 (1): 216-238.

Guadalupe, Maria (2003): "The Hidden Costs of Fixed Term Contracts: The Impact on Work Accidents," *Labor Economics*, 10 (3): 339-357.

Hanna, Christopher (2006): The Magic in the Tax Legislative Process, *Southern Methodists University Law Review*, (59): 649-688.

Hansen, Wendy L. and Neil J. Mitchell (2000): "Disaggregating and Explaining Corporate Political Activity: Domestic and Foreign Corporations in National Politics," *American Political Science Review,* 94 (4): 891–903.

............................, Mitchell, N., Drope, J., (2005): "The logic of Private and Collective Action," *American Journal of Political Science,* 49 (1): 150–167.

Hart, Oliver and John H. Moore (1990): "Property Rights and the Nature of the

Firm," *Journal of Political Economy*, (98):1119-58.

Hart, Oliver (1995): "Firms, Contracts, and Financial Structure," New York: Oxford University Press, USA.

Hayek, Friedrich A. (1937): "Economics and knowledge," *Economica*, 4 (13): 33–54.

Heitshusen, Valerie (2015): "Senate Committee Hearings: Arranging Witnesses," Congressional Research Service, 7-5700.

Helper, Susan (1991): "Strategy and Irreversibility in Supplier Relations: The Case of the U. S. Automobile Industry," *Business History Review*, LXV: 781–824.

Hill, Matthew D., G. Wayne Kelly, G. Brandon Lockhart and Robert A. Van Ness (2013): "Determinants and Effects of Corporate Lobbying," 42 (4): 931–957.

Hillman, A.L.,Gerald D. Keim and Douglas Schuler (2004): "Corporate Political Activity: A Review and Research Agenda," *Journal of Management,* 30 (6): 837-857.

Hirsch, Alexander V. and Pablo Montagnes (2015): "The Lobbyist's Dilemma: Gatekeeping and the Profit Motive," *Working Paper.*

Holmstrom, Bengt, and Paul Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization,* (7): 24-52.

_____________________ (1994): "The firm as an incentive system," *American Economic Review,* (84): 972–991.

Holmstrom, Bengt, and John Roberts (1998): "The Boundaries of the Firm Revisited," *Journal of Economic Perspectives*,12(4): 73-94.

Holmstrom, Bengt (1999): "The Firm as a Subeconomy," *Journal of Law, Economics, and Organization*, 15 (1): 74-102.

Hortaçsu, Ali, and Chad Syverson (2007): "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices," *Journal of Political Economy*, 115 (2): 250–301.

Huckman, Robert S and Gary P. Pisano (2006): "The Firm Specificity of Individual Performance: Evidence from Cardiac Surgery," *Management Science,* 52 (4):473-488.

Jabbour, Liza (2013): "Market thickness, sunk costs, productivity, and the outsourcing decision: an empirical analysis of manufacturing firms in France," *Canadian Journal of Economics*, 46 (1): 103-134.

Jacobson, Louis, LaLonde, Robert and Daniel Sullivan (1993): "Earnings Losses of Displaced Workers." *American Economic Review,* 83 (4): 685-709.

Joskow, Paul (1988): "Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence," *Journal of Law, Economics and Organization*, 4 (1): 95-117.

Kahn, Shulamit (2000): "The Bottom-Line Impact of Temporary Work on Companies' Profitability and Productivity," In Nonstandard Work: The Nature and Challenges of Emerging Employment Arrangements, edited by Francoise Carre, Marriane A. Ferber, Lonnie Golden, and Stephen A. Herzenberg. New York: Industrial Relations Research Association.

Kang, Karam (2016): "Policy Influence and Private Returns from Lobbying in the Energy Sector," *The Review of Economic Studies,* 83 (1): 269-305.

Kang, Karam and Hye Young You (2016): "Lobbyists as Matchmakers in the Market for Access," *Working Paper.*

Kerr, William R.; William F. Lincoln, and Prachi Mishra (2014): "The Dynamics of Firm Lobbying," *American Economic Journal: Economic Policy,* 6 (4): 343–379.

Klein, Benjamin; Robert G. Crawford and Armen A. Alchian (1978): "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *The Journal of Law and Economics*, 21(2): 297-326.

Klein, Benjamin (1988): "Vertical Integration as Organized Ownership: The Fisher Body-General Motors Relationship Revisited," *Journal of Law, Economics and Organization*, (4):199-213.

Knudsen, Eirik and Kirsten Foss (2014): "The Effect of Recessions on Firms' Boundaries," *Industrial and Corporate Change*, (9) 9:1-28.

Kogut, Bruce, and Udo Zander (1992): "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology," *Organization Science,* 3(3): 383–397.

___________________________(2000): "The Network as Knowledge: Generative Rules and the Emergence of Structure," *Strategic Management Journal,* (21): 405–425.

Lafontaine, Francine and Margaret Slade (2007): "Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*, 45 (3): 629-685.

Macher, Jeffrey T. John W. Mayo, and Mirjam Schiffer (2011): "The Influence of Firms on Government," *The B.E. Journal of Economic Analysis & Policy*, (Contributions)(11) 1: 1-27.

Mann, H. M., & McCormick, K. (1980): "Firm attributes and the propensity to influence the political system," In J. J. Siegfried (Ed.), *The economics of firm size, market structure and social performance*: 300-313. Washington, DC: Federal Trade Commission Bureau of Economics.

Masten, Scott (1984): "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics*, 27 (2): 403–17.

Masters, M. F., and G. D. Keim (1985): "Determinants of PAC participation among large corporations," *Journal of Politics,* 47: 1158-7.

Mitchell, Neil J., Wendy L. Hansen, and Eric M. Jepsen (1997): "The Determinants of Domestic and Foreign Corporate Political Activity," *Journal of Politics,* 59 (4): 1096–1113.

Mitra, D.,(1999): "Endogenous lobby formation and endogenous protection: a long-run model of trade policy determination," *American Economic Review,* 89 (5): 1116–1134.

Mizruchi, Mark S., and Thomas Koenig (1988): "Economic Concentration and Corporate Political Behavior: A Cross- Industry Comparison," *Social Science Research,* 17: 287–305.

.... (1991): "Size, Concentration, and Corporate Networks: Determinants of Business Collective Action," *Social Science Quarterly,* 72 (2): 299–313.

Monteverde, Kirk and David Teece (1982): "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *Bell Journal of Economics*, (13): 206-213.

Monteverde, Kirk (1995): "Technical Dialog as an Incentive for Vertical Integration," *Management Scienc*e, (41): 1624-1638.

Moreira, Sara (2016): "Firm Dynamics, Persistent Effects of Entry Conditions, and Business Cycles", Ph.D Dissertation, University of Chicago.

Mulligan, Casey and Andrei Shleifer (2005): "The Extent of the Market and the Supply of Regulation," *Quarterly Journal of Economics*, (120): 1445–73.

Munger, Michael C. (1988): "On the Political Participation of the Firm in the Electoral Process: An Update," *Public Choice* 56 (3): 295–298.

Murphy, Kevin M.(1986): "Specialization and Human Capital," Ph.D Dissertation, University of Chicago.

Neal, Derek (1995): "Industry-Specific Human Capital: Evidence from Displaced Workers," *Journal of Labor Economics*, 13 (4): 653–77.

Nickerson, J. and Zenger, T. (2004): "A Knowledge-based Theory of the Firm: the Problem-Solving Perspective," *Organization Science*, 15(6): 617-632.

Nunn, Nathan (2007): "Relationship-Specificity, Incomplete Contracts and the Pattern of Trade," *Quarterly Journal of Economics.* 122 (2): 569-600.

Olson, Mancur (1971) [1965]: "The Logic of Collective Action: Public Goods and the Theory of Groups," (Revised edition ed.). *Harvard University Press.*

Ozer, M. and S.H. Lee (2009): "When do firms prefer individual action to collective action in the pursuit of corporate political strategy? A new perspective on industry concentration," *Business and Politics,* 11 (1): 1-21.

Parent, Daniel (2000): "Industry-Specific Capital and the Wage Profile: Evidence from the National Longitudinal Survey of Youth and the Panel Study of Income Dynamics," *Journal of Labor Economics*, 18 (2), 306-323.

Pecorino, P., & Temimi, A. (2008): "The Group-size Paradox Revisited," *Journal of Public Economic Theory*, (10): 785–799.

Penrose, E. (1959): "The Theory of the Growth of the Firm," Oxford: Basil Blackwell Publisher.

Peteraf M. A. (1993): "The Cornerstones of Competitive Advantage: A Resource-

Based View," *Strategic Management Journal*, 14(3): 179-191.

Pittman, R. (1976): "The Effects of Industry Concentration and Regulation on Contributions in Three 1972 US Senate Campaigns," *Public Choice*, (23): 71-80.

.... (1977): "Market Structure and Campaign Contributions," *Public Choice,* 31(1): 37–51.

.... (1988): "Rent-seeking and market structure: Comment," *Public Choice (*58): 173-185.

Potters, J., Sloof, R., (1996): "Interest groups: a survey of empirical models that try to assess their influence," *European Journal of Political Economy,* (12): 403–442.

Rajan, Raghuram and Luigi Zingales (2001): "The Firm as a Dedicated Hierarchy: A Theory of the Origin and Growth of Firms," *Quarterly Journal of Economics*, (116): 805–852.

Rauch, James E. (1999): "Networks versus Markets in International Trade," *Journal of International Economics*, XLVIII: 7–35.

Ravn, Morten, and Uhlig, Harald (2002): "On adjusting the hodrick-prescott filter for the frequency of observations," *Review of Economics and Statistics* 84 (2): 371–376.

Richter Brian K., Samphantharak Krislert, Timmons Jeffrey F. (2009): "Lobbying and taxes," *American Journal of Political Science*, 53 (4): 893–909.

Rosen, Sherwin (1983): "Specialization and Human Capital," *Journal of Labor Economics,* (1):43–49.

Sachs, Richard (1999): "Hearings in the House of Representatives: A Guide for Preparation and Conduct," Congressional Research Service, 96-623.

Salamon, Lester M., and John J. Siegfried (1977): "Economic Power and Political Influence: The Impact of Industry Structure on Public Policy." *American Political Science Review,* 71 (3): 1026–43.

Schuler, Douglas A., Kathleen Rehbein, and Roxy D. Cramer (2002): "Pursuing Strategic Advantage Through Political Means: A Multivariate Approach," *Academy of Management Journal,* 45 (4):659–72.

Simon, Herbert A. (1951): "A Formal Theory of the Employment Relationship," *Econometrica,* (19): 293–305.

Snyder, James M., (1990): "Campaign Contributions as Investments: The U.S. House of Representatives 1980-1986," *Journal of Political Economy,* 98 (6): 1195-1227.

Stigler, George (1951): "The Division of Labor is Limited by the Extent of the Market," *Journal of Political Economy*, (59): 185-193.

Stigler, George (1971): "The Theory of Economic Regulation," *Bell Journal of Economics*, 2 (1) : 3-21.

..........................(1974): "Free Riders and Collective Action: An Appendix to Theories of Economic Regulation," *Bell Journal of Economics*, 5 (2) : 359-365.

Sullivan, John (2007): "How Our Laws Are Made," The Library of Congress. Link at: https://www.congress.gov

Tadelis, Steven (2002): "Complexity, Flexibility, and the Make-or-Buy Decision," *American Economic Review*, 92 (2): 433-437.

__________ and Oliver Williamson (2012): "Transaction Cost Economies," in R. Gibbons and D.J. Roberts, ed. *Handbook of Organizational Economics*, Princeton University Press.

Teece DJ, Pisano G, Shuen A. (1997): "Dynamic Capabilities and Strategic Management," *Strategic Manage- ment Journal,* 18 (7): 509-5.

Vannoni, Matia (2013): "The Determinants of Direct Corporate Lobbying in the EU: A Multi-dimensional Proxy of Corporate Lobbying," *Interest Groups & Advocacy, (*2): 71–90.

Wernerfelt B. (1984): "A Resource-Based View of the Firm," *Strategic Management Journal*, 5(2): 171-180.

Weymouth, Stephen (2012): "Firm Lobbying and Influence in Developing Countries: A Multilevel Approach," *Business and Politics,* 14 (4): 1-26.

Williamson, Oliver (1971): "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, (61): 112-23.

____________ (1973): " Markets and hierarchies: some elementary considerations," *American Economic Review,* (63): 316–325.

____________(1975): "Markets and Hierarchies: Analysis and Antitrust Implications," *New York: Free Press.*

____________ (1979): "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 22 (2): 233-261.

___________(1985): "The Economic Institutions of Capitalism," *New York: Free Press.*

___________(1991): "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives," *Administrative Science Quarterly*, 36 (2): 269-296.

Wilkerson John, Smith, D. and Stramp, N. (2015): "Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach," *American Journal of Political Science*, 59 (4): 943–956.

Zabojnik, J (2002): "A Theory of Trade Secrets in Firms," *International Economic Review*, (43): 831–855.

Zardkoohi, Asghar (1985): "On the Political Participation of the Firm in the Electoral Process," *Southern Economic Journal* 51 (3): 804–817.