Sourcing of Expertise and the Boundaries of the Firm: The Case of Lobbyists*
(Job Market Paper)

Miguel Espinosa
London School of Economics
[Click here to download the latest version]

November 10, 2016

Abstract

This paper 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, using information at both the industry and bill levels, I validate the main theoretical predictions using client fixed-effects estimations. 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 it led to a disproportionate increase in the use of external lobbyists for the affected industry. Lastly, I argue that the 2007 Open Government Act modified both the distribution of problems that firms faced and the technology to acquire knowledge. Estimating the underlying parameters of the integration decision, I explain how these two changes modified the integration patterns of the industry.

*I am deeply indebted for the precise and insightful comments made by Jordi Blanes i Vidal, Luis Garicano, Gilat Levy, 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, Jim Snyder, 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 Introduction

Firms' dependence on knowledge workers such as lawyers and managers has increased remarkably in recent years.[1] 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.[2] 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.[3] 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 paper.

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 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 knowledge. Although outsourcing allows to access to a larger amount of issue-specific knowledge, it implies paying 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

[1] 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.

[2] 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.

[3] 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.

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.[4] 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 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 comprises the universe of US federal lobbying transactions in which I can differentiate not only 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.[5] To test the predictions from my theoretical framework, I enrich this database by including the industry of the firms, lobbyists' sociodemographic variables, and information about the committees studying the lobbied bills,

[4]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.

[5]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 good examples of papers with information on the supply side of the market, but a lack of information about the identities of the demand side of the market.

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 my 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.[6] 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 the theoretical variables is *RegData 2.2*, a database that uses text analysis and machine learning algorithms to create regulation measures at the industry-year level.[7] I use two measures from this database: the number of words related to the regulation of the industry and number of times that five strings -*measuring the difficulty of the regulation*- appear among these regulation words.

I conduct three sets of empirical exercises with the aforementioned data: Fixed Effects, Difference in Difference and Structural 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.

[6]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.

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

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.

In the final empirical exercise, 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 changed the lobbying difficulty not only for a set of firms -*as the BP spill did*- but for the whole economy. The Act closed lobbying channels and improved the technological access to information. This has two implications. First, 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. Second, using equilibrium conditions from my model to back out unknown parameters, I show that, consistent with the fact that the Act facilitated access to information, the cost of acquiring knowledge decreased. With the parameters at hand, 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 counterfactual exercises. For instance, I show that if the economy had faced only the *difficulty effect*, then the fraction of firms with in-house lobbyists would have decreased by 59%, whereas if it had faced only the *technological effect*, this fraction would have increased by 150%.

To tackle the concern on 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.[8] 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 paper, 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 that finds that low-skill workers earn more in the internal market, external service providers in PBS industries earn more than in-house workers.[9]

This paper enriches our understanding of the vertical integration decision in several ways. First, to the best of my knowledge, this is the only paper to simultaneously propose and test a vertical integration theory with knowledge workers that does not consider incentive-alignment is-

[8]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.

[9]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)

sues.[10] Second, using a 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 that complements it with information at the firm, lobbyist, industry, bill and committees levels.[11] 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 paper, I enrich the understanding of the scarcely studied firms-lobbyists relationship.[12] Furthermore, to the best of my knowledge, the effect of the OGA and 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 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, my paper finds support for some of the previous predictions but questions the general applicability of other results.[13] 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.[14] 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.

Another prediction from this literature is that the more difficult the transaction is, the more likely it is to be vertically integrated.[15] 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. My paper complements his insight by providing both theoretical and empirical support to this idea.[16]

---

[10] 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.

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

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

[13] This paper includes, in Section 7.1, an extensive discussion of the leading incentive-based vertical integration theories and their relationship with my theoretical framework.

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

[15] See, for instance, Monteverde and Teece (1982), Masten (1984), Anderson(1985) and Tadelis (2002).

[16] Strikingly, the issue of the frequency has not received much attention. For instance, two of the most compre-

The empirical context of this paper 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 machine and hiring someone in-house are totally different because the employee can always quit. I interpret my paper 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 a 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.[17] 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, my paper is the first 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.[18] The *knowledge-based-view* of the firm argues that knowledge is firms' most important resource or capability.[19] 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.[20] The economics literature has studied the integration decision concentrating on how to manage the misalignment of incentives. However, the

hensive and recent literature reviews of the subject, Lafontaine and Slade (2007) and Bresnahan and Levin (2012), neglected the topic.

[17]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.

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

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

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

knowledge-based-view of the firm, and my paper, point to the fact that knowledge plays a role in defining firm boundaries even in the absence of any misalignment of interests. In this paper, 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 paper 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 paper is divided into five sections. Section 2 presents my theoretical framework. In Section 3, I begin by presenting the data used and the institutional context. Then, I present the results of several fixed effect estimations and the relationships among generalists, specialists, firms' size and leverage. I end this section with the BP example. Section 4 starts by explaining the OGA and providing empirical evidence on its consequences. Then, I show the results of the structural estimations. Finally, in Section 5, I discuss the external validity of my theory with regard to other PBS occupations. I finish with a short discussion summarizing the main results from the paper and proposing further developments.

2 Theoretical Framework

This section is divided into two subsections. In the first part, I present the main setting in Section 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 2.2 and then I move to Section 2.3 where I present results for the case of two issues. The second part discusses the robustness of my theoretical framework in Section 2.4 and the relevant related literature in Section 2.5. I end up this Section by summarizing the main results of the model in Section 2.6. All the proofs can be found in Appendix 7.5.

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.[21] Clients receive exogenously one problem per unit of time spent in production for each issue: A and B.[22] 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 continuous 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

[21] 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.

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

agents know everything that the less knowledgeable ones do, and more. That is, knowledge is cumulative.[23] 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 in order to produce. 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.[24] 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 that this knowledge may differ across issues. A way to interpret h_i is with the average firm-specific knowledge of M clients.[25]

Any study on firm boundaries requires, at least implicitly, a definition of what a firm is. In this paper, 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

[23]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.

[24]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."

[25]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.

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.[26]

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.[27]

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 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 the paper, I assume that a matching technology allocates clients across markets by maximizing the total surplus of the economy. In section 2.4, I show that the main results obtained under 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.

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

[27] To simplify the presentation of the main setting, I will introduce the notation of the wages for generalists and specialists in Section 2.3.1.

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.

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.

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.

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 knowledge 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 $[F(z) - F(z^*)] - \frac{c}{n}z$.[28] 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 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. However, the beauty of the market

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

is that it allows several clients to share the acquiring knowledge 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.

Lemma 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 knowledge 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 my paper, 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 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 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 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.[29]This is straightforward and comes directly from applying the Envelope theorem. For the ex-post case, the main trade-off in the economy is given by the leverage and the

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

wage bill. When firm-specific knowledge is 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.

2.2.3 *Equilibrium*

Let N_I be the number of service providers in the internal 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:[30]

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

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,

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

the excess of gross earnings for external providers is proportional to the differences in knowledge levels and is given by $c\left[f^{-1}\left(ch(1-F(f^{-1}\left(c\right)))\right)-f^{-1}\left(c\right)\right]>0$.

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

$$VI=\int\limits_{0}^{f^{-1}(c)}f(x)dx/\int\limits_{0}^{f^{-1}\left(\frac{c}{n}\right)}f(x)dx$$

and using the fundamental theorem of calculus I get:

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

Theorem 1 *The fraction of vertically integrated clients increases with firm-specific knowledge h and can increase or decrease with the cost of acquiring issue-specific knowledge.*

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 relative sensitivity of these boundaries to the cost parameter depends on the specific shape of the distribution function of the problems.

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:[31]

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

and it is easy to see that $z^{**}>z^{*}$.[32] The levels of earnings in the economy are:

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

It is easy to see from these equations that external service providers earn more than in-house staff due to their additional issue-specific knowledge. 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:

$$N_{I}=M\frac{(\lambda-c)}{\lambda}\text{ and }N_{E}=Mhc^{2}\frac{(\lambda-ch)}{\lambda^{3}}$$

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

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

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

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

Once I have characterized the equilibrium objects, I can obtain some comparative static results. Table 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. As the leverage is defined by the time constraint, a decrease in the cost c it makes more likely that service providers face problems; therefore, each of them can work with fewer clients. Finally, the effect of the cost of acquiring knowledge on the fraction of clients with in-house providers is ambiguous.[33]

2.3 Problem with two Issues

2.3.1 *Internal Market: Firm Boundary*

The ex-ante objective function for the client is $\pi^S\left(z_A, z_B; c_A, c_B\right) = F_A\left(z_A\right) + F_B\left(z_B\right) - w_A^I z_A - w_B^I z_B$, 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:

---

[33] In concrete, $sign(\frac{\partial VI}{\partial c}) = sign(-\lambda^2 - c^2 h + 2ch\lambda^2)$. Intuitively, a decrease in c changes both the firm and market boundaries, and the total effect on vertical integration depends on which effect (firm or market boundary) moves faster. Here, the negative terms show the decrease in knowledge of the internal providers, and the positive term shows the decrease in knowledge of the external providers. Finally, the comparative static exercises for the parameter of the CDF, λ, show that $sign(\frac{\partial z^*}{\partial \lambda}) = sign(\ln(\frac{c}{\lambda}) + 1))$, $sign(\frac{\partial z^{**}}{\partial \lambda}) = sign(\ln(\frac{c^2 h}{\lambda^2}) + 2))$ 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.

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

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 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:[34]

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

and solving for t_A: [35]

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

Proposition 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 2 provides more intuition on this result. Plugging t_A^* into the objective function, I get:

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

[34] 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.

[35] 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.

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

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

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

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 specialists rather than one generalist; that is,

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

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 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 *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 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 in 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.

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:

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

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:

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

Lemma 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* $\widetilde{z}_B \in \left(z_I^B, Q_B \left(1 - \frac{n_A^g h_A}{n_B^g h_B} (1 - F_A(\widetilde{z}_A))\right) \right)$ *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.[36] The joint surplus for the external service provider and her clients is:

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

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:

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

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

[36] 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.

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

Lemma 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 spend all their time on one issue, they reach 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 7 *For a given issue, external providers always acquire more knowledge than any type of internal providers.*

This results is similar to Lemma 1, and implies that external service providers have higher earnings.

Table 2 shows all the possible combinations of vertical integration and external contracting that clients can face. Figure 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.

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:

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

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

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 $w^I \left[t_A^* z_i^* + \left(1 - t_A^*\right) z_i^*\right] - c_A z_i^* - c_B z_i^*$, 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 $\left(c_A + c_B\right) z_i^*$ 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 8 *The level of vertical integration in the industry is decreasing in the firm-specific knowledge levels h_i and h_j.*

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 problems solve their problems in-house, whereas clients with harder problems ask for help in the external market.

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 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 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.[37] 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

[37] 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 $\frac{c}{n_i^2} > 0$.

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 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 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 $f(z_i^{**}) = \frac{c}{n_i}$.

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.

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 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 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.[38]

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 $M_1 + M_2F(z^*)$ 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 $M_1F(z^*)$. With ex-ante homogeneous service providers and infinite supply, wages compensate for the cost of acquiring knowledge (i.e., $w = c$). The prediction 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.

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

[38] 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 $\lambda > ch$.

$F(z^*)^2$ 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 *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 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.*

2.5 Discussion

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. My paper is different because I focus on the vertical integration problem and not on the 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 their models allow the possibility of not having problems in one issue.

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 patterns of knowledge levels across markets hold for both cases: ex-ante homogeneous and heterogeneous service providers.

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 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.

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 level 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 decreasing in firm-specific knowledge, and it can be either increasing or decreasing in both the cost of acquiring knowledge c and 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.

3 Empirics

This section is divided into four subsections. In Section 3.1, I present the data and the institutional context. Section 3.2 presents the fixed effect estimation results at both the client and the transaction levels. Then, in Section 3.3, I provide some empirical evidence for other results derived from the theoretical section. Finally, in Section 3.4, I use the BP oil spill event to provide causal evidence on the effect of difficulty on vertical integration decisions. Additional description of the data, results and other tests can be found in the Online Appendix **A**.

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.[39] 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.[40]

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.[41] Each report is classified as in-house or external. 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.

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.[42] After introduction, the bill is referred to the appropriate committee or committees, based on the committees' jurisdiction, which is defined by congressional rules.[43] 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.[44]

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

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

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

[41] Although, lobbying reports contain information on expenditures, I have not considered this variable in most of my analysis, as in-house reports contain information for both lobbyists work and office space.

[42] 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.

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

[44] The Senate has established this figure since 1937.

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.[45]

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.[46] 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 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.[47] 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 lobby 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.[48] RegData classifies industries at the NAICS four-digit code levels for the period 1999 to 2014.[49] For each industry-year combination, I use two variables from this database: 1) the number of words related to the regulation of the industry;[50] and 2) the number of restriction

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

[46] For more information refer to Appendix **A**.

[47] More details can be found in Appendix **A**.

[48] 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.

[49] 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.

[50] Proxying regulation with these types of measures is not completely new. Coffey et al. (2012) and Mulligan

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 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 provide some statistical evidence on the existence of communication costs in Appendix **A**.

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).

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

Figure 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.[51] 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.[52] Both of these

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.

[51] 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.

[52] 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 the

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.

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 Appendix **A**, I show that the total number of regulating words is significantly correlated with both the displacement and trade literature measures.

3.2 Main predictions

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

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

Appendix **A**, 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.

cofound the vertical integration decision.[53] 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 product of the financial crisis, changes in market uncertainty or lobbying regulation.[54] 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. In this case, I cluster the standard errors at the industry level. 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.[55]

The key independent variable in these estimations, X, represents frequency, difficulty or industry-specific knowledge. For the frequency variable, I present here only the results on the committee-based knowledge requirement measures, but the interested reader can see the results for the other frequency variable in Appendix **A**. 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).[56]

3.2.2 Some Descriptive Statistics

Table 3 shows some examples of the least and most four-digit industries by the fraction of restraining words over the total number of words and the 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 total level of regulating words. Both columns use 2014 data. Similar rankings are obtained for different time periods.

Table 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.

[53]For examples of the relevance of size of the client, see Section 3.3.2 of this paper; for union status, see Abraham and Taylor (1996); and on geographical variation, see Chinitz (1961) and Autor (2003).

[54]For the effect of the Crisis on integration patterns, see Knudsen and Foss (2014); for economy-wide related time varying patterns, see Abraham and Taylor (1996); and for the effect of policy changes, see Section 4.1.1 of this paper.

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

[56]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.

3.2.3 Frequency

Table 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 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%.

3.2.4 Difficulty

Table 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 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 vertical integration measure by an amount between 2.2 and 2.7 (1.4 and 2.1) percentage points. Similarly, an additional standard deviation in the difficulty proxy (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%).

3.2.5 Industry-Specific Knowledge

Table 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.[57] 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

[57] 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.

a decrease in the probability of integration of three to six percentage points and 2% to 4% on the fraction of internal reports.

3.3 Other Predictions

In this section, I focus on three empirical patterns accounted for 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.

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.[58] The categories I use in this section are as follows: 1) internal lobbyists (56.1%), who have **always** lobbied as in-house lobbyists; and 2) external lobbyists (43.9%), who are intermediaries that have **never** lobbied as in-house lobbyists.[59] 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.[60]

Table 9 shows the main descriptive statistics of this exercise.[61] 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

[58] 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.

[59] 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.

[60] 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 three periods with the following allocation: first period: 20, 40, 20, 20; second period: 40, 20, 20, 20; and third period: 20, 20, 40, 20. 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.

[61] 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.

classified as generalists, 78% are in-house lobbyists, while 22% are external lobbyists. Table 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.[62] 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 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 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.[63]

A possible concern from these results is that a lobbyist can erroneously be classified as a specialist if she has worked only few times. I include in Appendix **A** descriptive statistics, controlling for the tenure of the lobbyists, showing that the patterns highlighted here between vertical integration and specialization are robust.

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.[64] Grossman, Helpman, and Szeidl (2006) theoretically predict that 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 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 on 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

[62] 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.

[63] 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.*"

[64] 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).

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 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.[65] I use two variables to proxy for the size of the client: sales and employees. Table 10 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 7 shows the firm size distribution by type of client.[66] 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 *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 3.1.1. To 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. For the sake of space, I do not present the results

[65] 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.

[66] 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.

of these estimations here, but the interested reader can see the results in Appendix **A**. 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, I expect to find that larger vertically integrated clients will have a larger fraction of internal specialists than smaller internal clients will.[67] Figure 8 is a scatter plot between the HHI of lobbyists and size of the clients, providing evidence from the previous intuition. Internal specialists tend to be with the largest clients. The generalists working internally tend to be hired by smaller internal clients.

3.3.3 Earnings, leverage and knowledge

Lemma 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. The interested reader can see descriptive statistics of federal government experience discriminated by in-house and external lobbyists in Appendix **A**.

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.[68] 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.[69]

Figure 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 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

[67] 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.

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

[69] 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.

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 9 provides evidence on the positive relationship between 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 9.

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 quasiexperiment to explore the validity of this prediction.

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.[70]

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 examples: the reactions of Congress, the federal Government and (potential) voters.

Congress. The oil industry has a strong and long-standing relationship with Republican policy makers.[71] 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.[72] 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.[73] Some of these

[70] 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).

[71] 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.

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

[73] 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

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.[74] 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;[75] 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,[76] and Secretary of the Interior Ken Salazar suspended deepwater offshore drilling on the Outer Continental Shelf for a period of six months.[77]

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 had an important effect on both prices and volume of sales. They interpret this finding as a shift in the 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.[78] 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.[79]

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.

S.598 of 2013 proposed to prohibit royalty incentives for deepwater drilling.

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

[75] Ibid.

[76] Johnston, Nicholas and Nichols, Hans (1 May 2010)."Obama Says New Oil Leases Must Have More Safeguards." Bloomberg.

[77] https://www.doi.gov/news/pressreleases/Interior-Issues-Directive-to-Guide-Safe-Six-Month-Moratorium-on-Deepwater-Drilling.

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

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

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.[80] This change in the structure 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 codes above. Examples of those are codes 3251 (Basic Chemical Manufacturing), 324191 (Petroleum Lubricating Oil and Grease Manufacturing) and 324199 (All Other Petroleum and Coal Products Manufacturing).[81] For this exercise, I focus on quarterly reports, which are mandatory since the OGA. The period of analysis runs from the first quarter of 2008 to the second quarter of 2014. Figure 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 presented in Appendix **A**. 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 Figure 12, in which the fraction of internal reports began to increase after the OGA took effect for all 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 11 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.[82] 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 sections of Figure 11 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

[80] www.nytimes.com/diminished-bp-in-deepwater-horizon-settlement.html

[81] These exclusions leave me with 52 clients in the treated group and 218 in the control group. I use the classification from *Open Secrets* to assign clients to the Oil and Gas industry. 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.

[82] Unfortunately, there are not enough quarter-level observations to construct a meaningful time series with this level of aggregation.

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 as of 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. Second, 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 reports originated and 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 the 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 11 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 vertical integration patterns, 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.

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 use of in-house staff. A decrease in the use of in-house lobbyists would be inconsistent with the time constraint explanation. In Appendix **A**, I provide 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 including either the total number of reports or the total expenditures of the client as an additional control variable. The main results presented in the text 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 to work 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 Appendix **A**, I provide evidence that the equilibrium payment in the internal market decreased, supporting the idea that the change is explained by demand channels.

A second concern is that there may have been negative spillovers to other firms in the oil and gas industry. Although Figure 10 and 11 do not seem to show any change for the control group around the event, in Appendix **A**, I include additional econometric estimations using different control groups and I show that the main results presented in this section hold.[83] A third concern is that the variation in the interaction variable is not at the client but the industry level. To control for this type of autocorrelation, in Appendix **A**, 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. (2016), I aggregate the data at the client-period level (before and after).

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 Appendix **A**, 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.

4 The Effect of the Open Government Act

In this section, I use the Open Government Act (henceforth OGA) as an additional quasi-experiment to test my theoretical framework. I proceed in two steps. In Section 4.1, I explain the main changes introduced by the OGA and how the Act changed the vertical integration patterns in the industry. Then, in Section 4.2, I provide maximum likelihood estimations results and using my theoretical framework, I explain how and why the OGA affected the vertical integration pat-

[83] 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 internal reports.

terns of the economy. I end up this section by presenting the main results of the counterfactual exercises and a brief discussion of the robustness in my estimations.

4.1 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.[84] Remarkably, the Act did not include a single regulation discriminating between internal and external 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.[85] 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.[86] Unfortunately, the survey did not ask lobbyists why or how the difficulty of their activities increased.

4.1.1 Consequences and Interpretation of the OGA

Although the OGA requires quarterly reports, in Figure 12, 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 12 shows the time-series patterns of the fraction of clients using in-house lobbyists.[87] 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 main takeaway from this figure is the change in the pattern around 2007. After the

[84]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 Appendix **A**.

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

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

[87]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.

OGA, the fraction of in-house clients increased from 25% to almost 34%.[88]

Change in the Density of Problems Politicians need knowledge to make informed decisions.[89]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 13 shows the density functions before and after the OGA using the congressional knowledge measures introduced in Section 3.1.1. 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 13 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.[90] Similar conclusions emerge when I estimate these density functions for different periods around the enactment of the OGA.[91]

To provide additional empirical evidence on the increase in difficulty around the OGA, Figure 14 shows the estimated residuals obtained from running the ratio of restraining words over the total number of words on the years considered.[92] The figure shows a remarkable similarity with Figure 12. 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 knowledge.

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 15 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 16 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

[88]One way to understand the decreasing trend before the OGA 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 lobbyists. Graphical evidence of the increase in the market is given in Appendix **A**.

[89]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.

[90]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.

[91]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.

[92]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.

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 contrained, 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 know to what extent each of these components, density change and cost change help us understanding the total change in the 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.

4.1.2 Parametric Solution

In the case of only one issue and $z_B \sim Exp\left(\lambda_B\right)$, 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 $z_B^{**} = -\frac{1}{\lambda_B}\ln(\frac{c_B^2 h_B}{\lambda_B^2})$. The earnings for the in-house lobbyists are $c_B z_B^*$ and for the external lobbyists $n_B\left(\frac{c_B}{n_B}\right)z_B^{**}$. Clearly, external lobbyists earn more as they acquire larger amount of knowledge in equilibrium.

Remark 1 *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.*

The following equations characterize the equilibrium in the economy: total number of internal clients I_B, total number of external clients E_B, and the time constraint of the external lobbyists:

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

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

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

The fraction of vertically integrated clients is given by $VI_B = \frac{\lambda_B(\lambda_B - c_B)}{\lambda_B^2 - c_B^2 h_B}$. This fraction is decreasing in the communication costs h_B and it can be increasing or decreasing in λ_B and c_B.[93]

[93] 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

4.2 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.[94][95] However, there are four unknowns: λ_B, c_B, h_B and M_B.[96] 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 proxy $z_{i,B}$ with the fraction of witnesses with high knowledge in that congressional hearing. Assuming independent and identically distributed observations, the likelihood function is:

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

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:[97]

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

I follow this methodology for both the periods before and after the OGA. Table 12 shows the input variables used, the estimated coefficients $\widehat{\lambda}_B$ and the estimated unknowns.[98] 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 the external and internal clients, the latter variable increases by about three times as much as the former one. The estimated coefficient for the parameter of the distribution before the OGA 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 about the increase in the difficulty of their lobbying activities.

number of words regulating all the lobbying industries, the main results do not change.

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

[95] 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.

[96] 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.

[97] 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 $I_B(n_B - 1) > 0$. Therefore, as far as $n_B > 1$, $\lambda_B > c_B$, $\lambda_B > c_B h_B$ and $M_B, c_B, h_B > 0$.

[98] 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 $h_B = \frac{\lambda_B}{n_B c_B} = \frac{1}{n_B K}$.

The estimated cost of acquiring knowledge c decreases by 20%. While the coefficient for industry-specific knowledge and the total number of (potential) clients M also change, this does not seem large. I include an estimate of what I call the knowledge inequality of the economy, $\frac{z_B^{**}}{z_B^*} = \ln(\frac{c_B^2 h_B}{\lambda_B^2}) / \ln(\frac{c_B}{\lambda_B})$. In my theoretical framework, 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 knowledge and earnings levels between the market and firm boundaries. Table 12 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%.

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 as follows. 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 $dx_t \approx x_A - x_B$.[99] Table 13 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.[100] This result is confirmed when I use the RegData to proxy for industry-specific knowledge. Table 13 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 provide here. The Act closed lobbying channels and decreased the cost of acquiring knowledge. 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 13 and theoretically this is translated in a change of the density of the problems. The Act also decreased the cost of acquiring knowledge as a new technology to acquire relevant information to the 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.

---

[99] 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 $\frac{\partial VI_B}{\partial h_B} = \frac{\lambda_B(\lambda_B - c_B)c_B^2}{\left(\lambda_B^2 - c_B^2 h_B\right)^2}$.

[100] 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.

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.

4.2.1 Counterfactual Exercises

To simplify the burden of the paper, I call the change in the distribution, the *difficulty effect* and the change in the estimated cost of acquiring knowledge, 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 17 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 or the density function 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 their demand due to the 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.[101]

Table 14 shows the main results of these counterfactual exercises. The *technology effect* in-

[101] This fraction is equal to $\frac{c_A}{\lambda_A^2}\left(\lambda_A - c_A h_A\right) - \frac{c_B}{\lambda_B^2}\left(\lambda_B - c_B h_B\right)$, 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.

creases 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 learning cost 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 14 shows that the leverage increases in 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, decreases the number of clients that external service providers have.

Robustness in the estimations One may wonder to what extent the main results from these exercises depend on the assumption on 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.[102] 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.[103]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 knowledge. The idea is summarized as follows. The fraction of vertically integrated clients before the OGA is given by $\frac{F_B\left(z_B^*\right)}{F_B\left(z_B^{**}\right)} =$

[102]The density function of the Pareto distribution is strictly decreasing as assumed in the theoretical framework.

[103]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 difficulty lobbying activity of the economy.

$\frac{F_B\left(f_B^{-1}(c_B)\right)}{F_B\left(f_B^{-1}\left(\frac{c_B}{n_B}\right)\right)}$. 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 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. More details can be found in Appendix **A**.

These two previous exercises leave two concerns to be addressed. All of my estimations consider the case of one layer in the external market, and use equilibrium conditions for the case of only one issue. Extending this methodology to the case of multiple layers and one issue presents an important challenge to the way that I define the layers. When the fraction of problems unsolved is very low, all of the action will be among the changes in the firm and the last layer boundaries, and, as a consequence, the calculations for the counterfactual exercises will not be more difficult. However, a robust definition of a layer is not a trivial task. For the case of two issues and one layer, the burden of the estimations is large because I need to make inferences on the actual (and not observed) complementarities and consider the relative number of generalists and specialists in each market. I aim to consider these extensions in the future.

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 or by other regulations of the OGA not related to knowledge-requirement changes. Appendix 7.2 presents evidence that economic shocks do not have an effect on lobbying integration patterns. In Appendix 7.3, I argue that other changes brought about by the OGA cannot fully explain the aforementioned patterns.

5 Final Discussion

This section is divided into two parts. In Section 5.1, I give preliminary evidence on the external validity of my results for other PBS industries. In Section 5.2, I briefly summarize this paper and propose ways to extend it.

5.1 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 (OES) program.[104]This program from the Bureau of Labor Statistics produces employment and wage esti-

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

mates annually for over 800 occupations in the United States, sampling over 200,000 non-farm business establishments every semester.[105] In this section, I focus on national occupational estimates for specific industries for the period 2002 to 2014.[106]

5.1.1 Data

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. I focus on five different occupations: lawyers, managers, IT personnel, accountants and lobbyists. Table 15 summarizes the input used for this exercise.[107] 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 workers with occupational 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 is to use four-digits NAICS codes (5418), which, unfortunately, include other occupations, such as advertising services.

5.1.2 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 18 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 in both the internal and external markets.[108]

---

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

[106] 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, 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.

[107] 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.

[108] In units of one million employees.

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, as explained in Section 4.1.1. 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, supporting the argument on the lack of effect of the financial crisis on vertical integration patterns with knowledge workers, Figure 18 does not show any change in the integration levels around 2008 for all the occupations, excluding lobbyists. Third, around the time that the OGA was approved, 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. The patterns for lobbyists, seen in Figure 18, are consistent with empirical patterns shown in Figure 12.[109]

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 16 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.[110] 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 differentials or demand-side rent differentials. Appendix **A** presents the same patterns when I consider annual instead of hourly earnings.[111]

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.[112] For instance, code 11-1021:

[109] In Appendix **A**, I show that the total employment of lobbyists changed around 2008. In this year the total employment of lobbyists reached its highest point and since then, the total employment of lobbyists decreased.

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

[111] 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, high-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.

[112] 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*

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. [113]

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."[114]

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 17.[115] Table 18 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 19 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 9. I take this piece of evidence as a first step towards the development of a comprehensive vertical integration theory of knowledge workers.

5.2 Concluding Remarks

The aim of this paper 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 to conduct 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.

---

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).

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

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

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

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. Finally, I exploit the effect that the OGA had in the industry's integration patterns. I show that the Act changed the knowledge intensity of the lobbying activities and that this, in turn, modified the type of problems that firms faced. As the easy lobbying activities became less common and as the acquiring cost technology became cheaper -*due to the technological shift*-, the fraction of clients using in-house lobbyists increased.

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 paper 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 paper, I have focused on understanding the causes rather than the consequences of vertical integration.[116] 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 interest 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?

[116] 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).

6 References

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

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.

Ansolabehere, Stephen; Snyder, James M. and Micky Tripathi (2002): "Are PAC Contributions and Lobbying Linked?," *Business and Politics* 4 (2): 1-26.

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.

Berlingieri, Giuseppe (2014): "Outsourcing and the rise in services," *CEP Discussion Paper No. 1199.*

___________ (2015): "Managing Export Complexity: the Role of Service Outsourcing," manuscript.

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.

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.

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.

Caliendo, Lorenzo and Esteban Rossi-Hansberg (2012) "The Impact of Trade on Organization and Productivity," *Quarterly Journal of Economics*, 127 (3): 1393–1467.

Carrington, William J. (1993): "Wage Losses for Displaced Workers: Is It Really the Firm That Matters?" *Journal of Human Resources*, 28 (3): 435–62.

Chinitz, Benjamin (1961): "Contrasts in Agglomeration: New York and Pittsburgh," *American Economic Review, PandP*, 51(2): 279-289.

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.

De Figueiredo, John M.and Brian S. Silverman (2006): "Academic earmarks and the returns to lobbying." *Journal of Law and Economics*, 49 (2): 597–625.

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.

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.

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.

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.

Hirsch, Alexander V. and Pablo Montagnes (2015): “The Lobbyist’s Dilemma: Gatekeeping and the Profit Motive,” *Working Paper*. Accesed at: http://people.hss.caltech.edu/~avhirsch/HirschMon-LobbyistsDilemma-Wallis.pdf

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.* Accessed at: https://hyeyoungyou.files.wordpress.com/2015/08/matchmaker1.pdf

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.

Masten, Scott (1984): "The Organization of Production: Evidence from the Aerospace Industry." *Journal of Law and Economics*, 27 (2): 403–17.

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 Science*, (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.

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.

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.

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.

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.

Simon, Herbert A. (1951): "A Formal Theory of the Employment Relationship." *Econometrica,* (19): 293–305.

Stigler, George (1951): "The Division of Labor is Limited by the Extent of the Market," *Journal of Political Economy*, (59): 185-193.

Sullivan, John (2007): "How Our Laws Are Made," The Library of Congress. Link at: https://www.congress.gov/resources/display/content/How+Our+Laws+Are+Made+Learn+About+the+Legislative+Process

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.

Wernerfelt B. (1984): "A Resource-Based View of the Firm," *Strategic Management Journal*, 5(2): 171-180

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.

Zabojnik, J (2002): "A Theory of Trade Secrets in Firms," *International Economic Review*, (43): 831–855.

7 Appendix

7.1 Discussion of the Model and Related Literature

The integration decision literature has focused on the problem of incentives. In this paper, 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 paper 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 my paper. 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).[117] 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.[118]

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 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

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

[118] 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.

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.[119] 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.[120] 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.

Although my paper 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. My paper 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. To the best of my knowledge this is the first paper in the literature to put these two concepts together.

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.[121]

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-

[119] 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.

[120] 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.

[121] 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.

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.[122]

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

[122]Notice for instance, (anna) the lobbying reports neither provide information on the contacted policy makers nor the specific message transmitted.

7.2 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 the vertical integration patterns. First, according to figure 18 there are not changes in integration patterns for other PBS occupations outside 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) predicts 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 12.

To make a step further 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.[123] **2**. I construct shocks for each industry in each year using the annual industry-specific nominal Gross Domestic Product (GDP) as the primary business cycle indicator.[124] To calculate the shocks I apply the Hodrick-Prescott filter using a smoothing parameter of 6.25.[125] 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 industry n demand shocks 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 industry n-th at period t-th. A negative and significant η_z suggests that a positive shock in the demand faced by firms from industry j-th decreases the level of vertical integration of industry j-th. 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

[123] 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 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.

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

[125] 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).

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. The interested reader can find these results in Appendix **A**.[126]

7.3 Other Changes brought by the OGA

7.3.1 Change in Mandatory reports

Before the OGA, firms have 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 dollar per quarter.

If a firm spends any value 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 the vertical integration in the industry. In Appendix **A**, 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 than 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 strongly reject this alternative.

7.3.2 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 the 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 Figuereido and Kim (2004) argue that firms tend to use in-house lobbyists for firm-specific topics prone to sensitive-information leakage.[127] 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

[126] 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, my interpretation on the non-significant effect of the Crisis over the lobbying integration patterns is rationalized if the latter effect is compensated with the former one.

[127] 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.

arguments supported by personal interviews with lobbyists that leads to think 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 two entities that benefits 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 the 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 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, numerous anecdotal evidence show that sensitive information leakage tend to be more common among employees or former ones rather than external service providers.[128] 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 increases, the expected cost of hiding increases and therefore it will be more likely that the clients start reporting. Given that 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.[129] They argue that is unlikely that clients will avoid lobbying just for this 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.

To sum up, the change in the reporting incentives may have changed the number of observed internal clients after the OGA. However, given the arguments provided here, it is difficult to argue

[128] 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.

[129] 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.

that the change in the incentives to report can explain the observed big shift in the integration patterns.

7.4 Tables

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

Table 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 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 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 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 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 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 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 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 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.

		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 10: 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.

	Fraction of Reports				*Fraction of 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 11: 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.

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 12: 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 13: 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 14: 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.

Occupation	**Occupational Code**	**NAICS**
Lawyers	23 (excluding 23-1023)	5411
Managers	11	5511
IT Personnel	15-10[130]	5415
Accountants	13-2011	5412
Lobbyists	27-3031	5418[131]

Table 15: 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.

[130] 15-11 after 2010. The definition for this occupation only change the occupational code but not the description of activities.

[131] 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[132]
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 16: 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 17: 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.

		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 18: 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.

[132] 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*
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 19: 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.

7.5 Proofs

Lemma 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 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 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 *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 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 *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 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 *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 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 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 7 *For a given issue, external providers always acquire more knowledge than any type of internal providers.*

Proof. It follows from the text. ■

Lemma 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 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 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 *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 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 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. ■

7.6 Figures

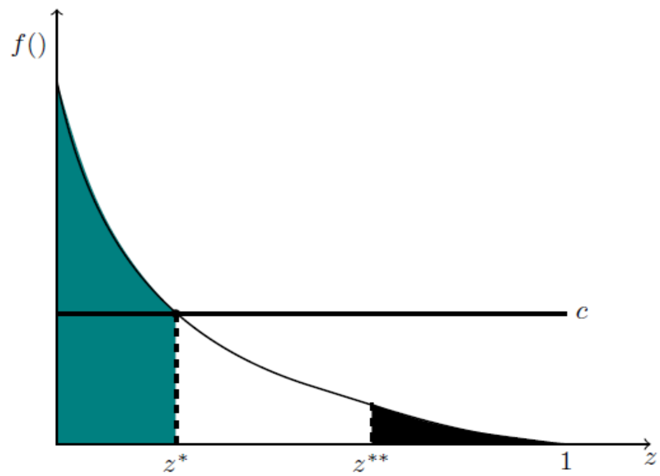

Figure 1: Solution with One Issue. Note: The left shaded (green) area represents the activities for which the marginal benefit of conducting the activities in-house is larger than or equal to the marginal cost. The first cut-off represents the firm-boundary, whereas the second cut-off denotes the market 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.

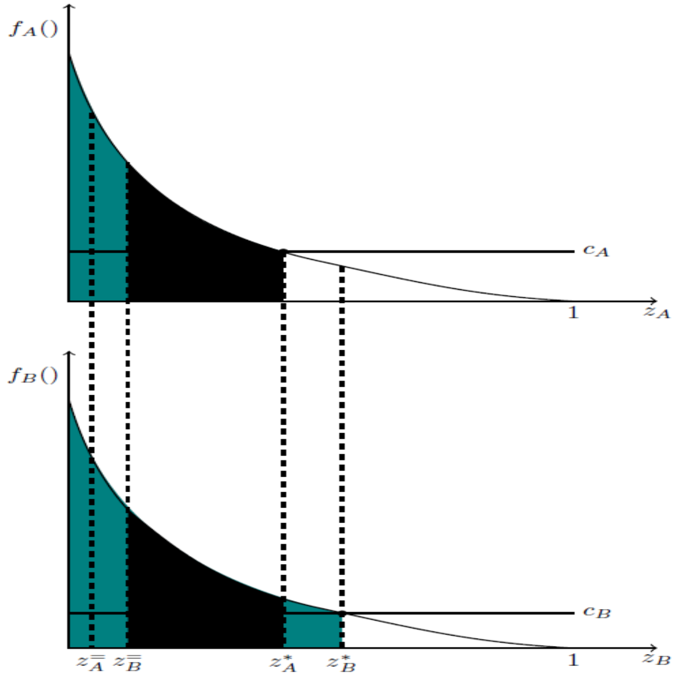

Figure 2: Solution with Two Issues and Generalists in the Internal Market. Note: The dark (black) area represents the levels of knowledge that internal generalists acquire.

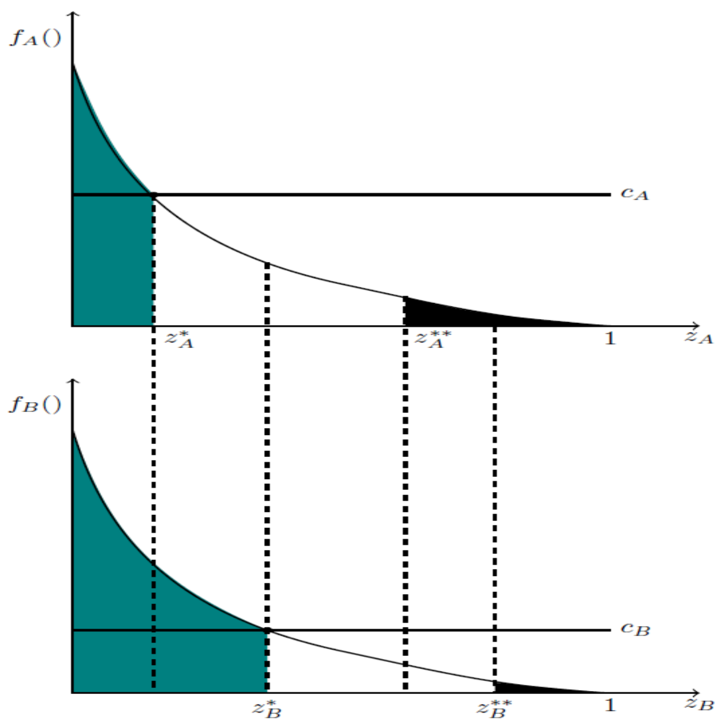

Figure 3: Solution with Two Issues. Note: This figure shows 5 out the 6 possibilities discussed in table 2. The case where firms use in-house staff for some issue and leave the problem unsolved for the other issue can be easily included for a low enough cost of acquiring knowledge in issue B.

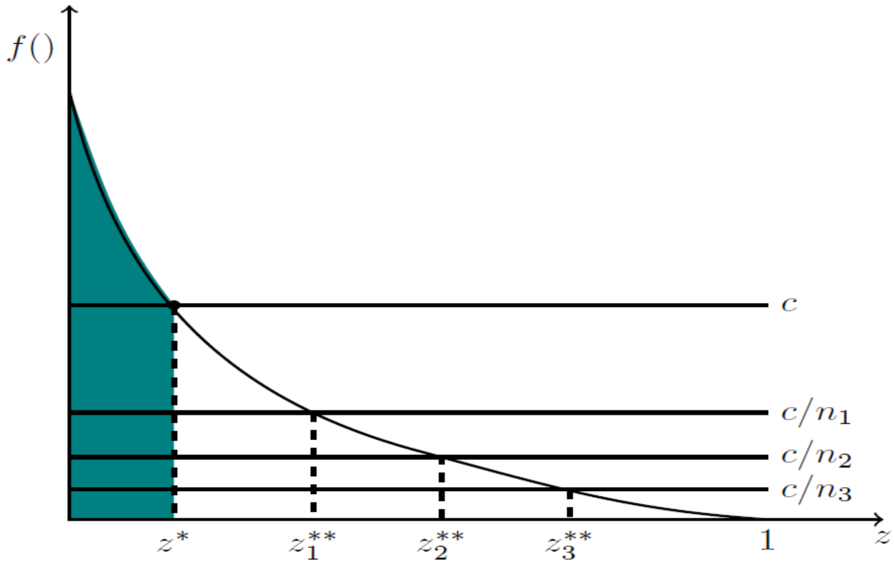

Figure 4: Solution to the Problem with Multiple Layers in the External Market: *The Level of Specialization of the External Providers is Limited by the Extent of the Market.* Note: The figure represents the case with three layers in the external market. The first cut-off represents the firm boundary. Each cut-off in the external market denotes a different layer, implying a different level of knowledge and leverage.

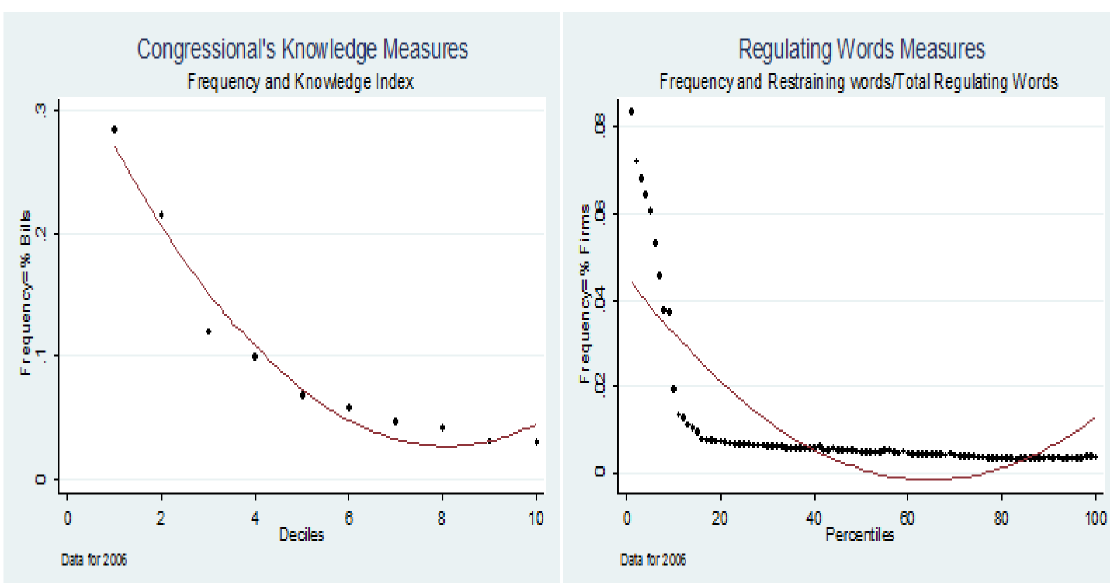

Figure 5: Knowledge Proxied by Congressional's and Regulating Words Measures. Note: LHS: Frequency function of the committee-knowledge requirements' index. The x-axis shows the deciles of the index and the y-axis shows the normalized (fraction over the total) number of bills for each decile. RHS. Frequency 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.

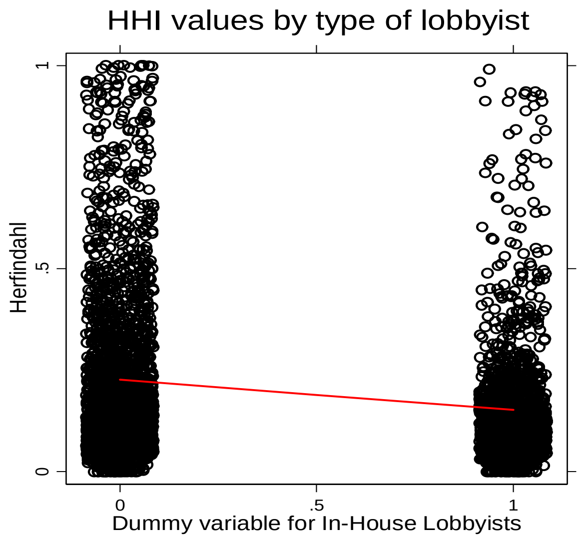

Figure 6: Patterns of HHI Across Markets. Note: The value 1 corresponds to internal lobbyists and 0 corresponds to external advocates. The line shows the linear prediction of the HHI on the dummy of internal lobbyists. The fitted regression is $HHI_i = 0.22 - 0.07 \cdot D_i + \varepsilon$ where $D_i = 1$ 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.

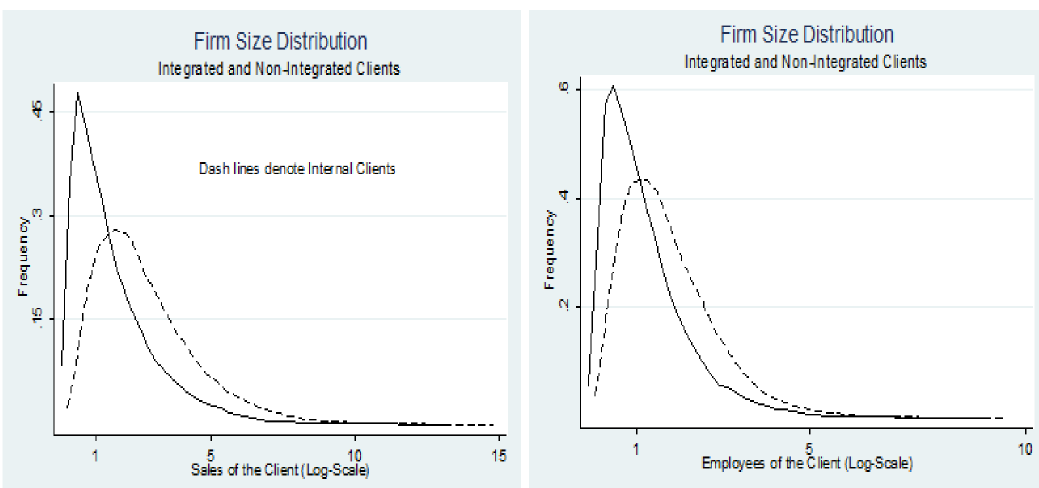

Figure 7: Relationship Between Firm Size and Status of Integration. Note: LHS: Firm size distribution by type of vertical integration relationship when the size is proxied by the log of the sales of the client. RHS: Firm size distribution by type of vertical integration relationship when the size is proxied by the log of the number of employees of the client.

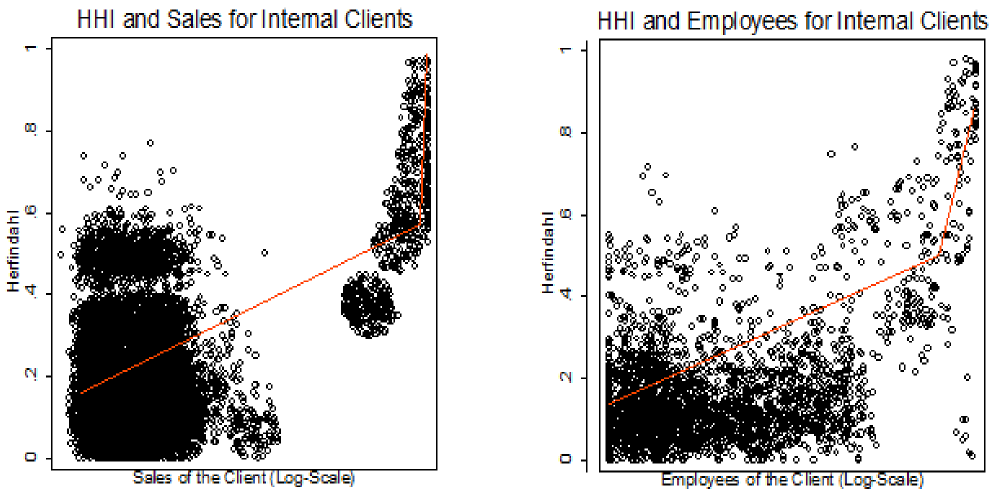

Figure 8: Relationship Between Firm Size of Internal Clients and HHI. Note: LHS: Scatter plot of the Herfindahl index and the log of the sales of internal clients. RHS: Scatter plot of the Herfindahl index and the log of the number of employees of internal clients. I restrict this figure to the sample of lobbyists with at least 5 years of lobbying activity. Similar patterns emerge for alternative time spans. An observation is a lobbyist.

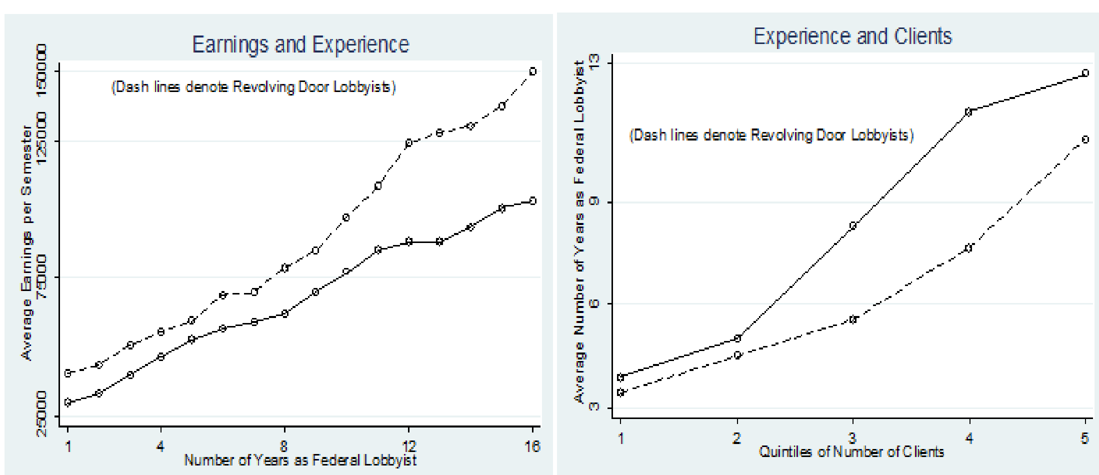

Figure 9: Earnings, Experience and Leverage. Note: LHS: Relationship between earnings (y-axis) and experience, as proxied by the number of years as a Federal Lobbyist (x-axis). RHS: Relationship between experience and number of clients. The x-axis reports the quantiles of the distribution of the number 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.

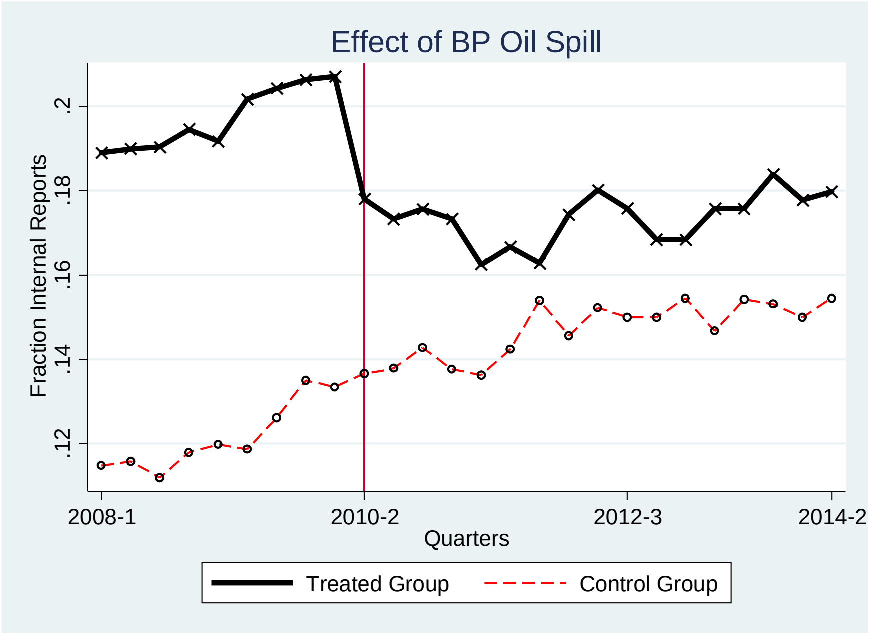

Figure 10: BP Oil Spill and Fraction of Internal Reports for Firms in the Oil Industry. Note. The treated group includes all the firms conducting lobbying activities that belong to 2007 NAICS codes 211111, excluding BP. The control group includes all the firms that belong to the Oil and Gas industry, excluding the codes above. An observation is a group- quarter.

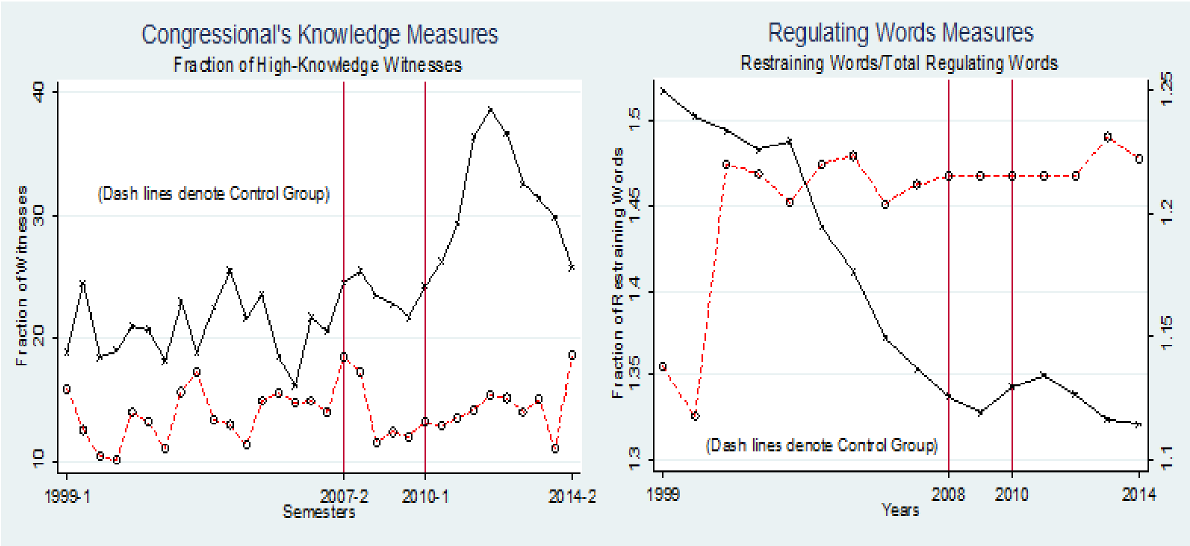

Figure 11: Effect of the BP Oil Spill on Knowledge Measures. Note: The LHS figure shows the knowledge measure proxied by the fraction of witnesses. To construct this measure, for each semester I considered all the bills that each of these firms were lobbying. To get a group-level measure, I have taken 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.

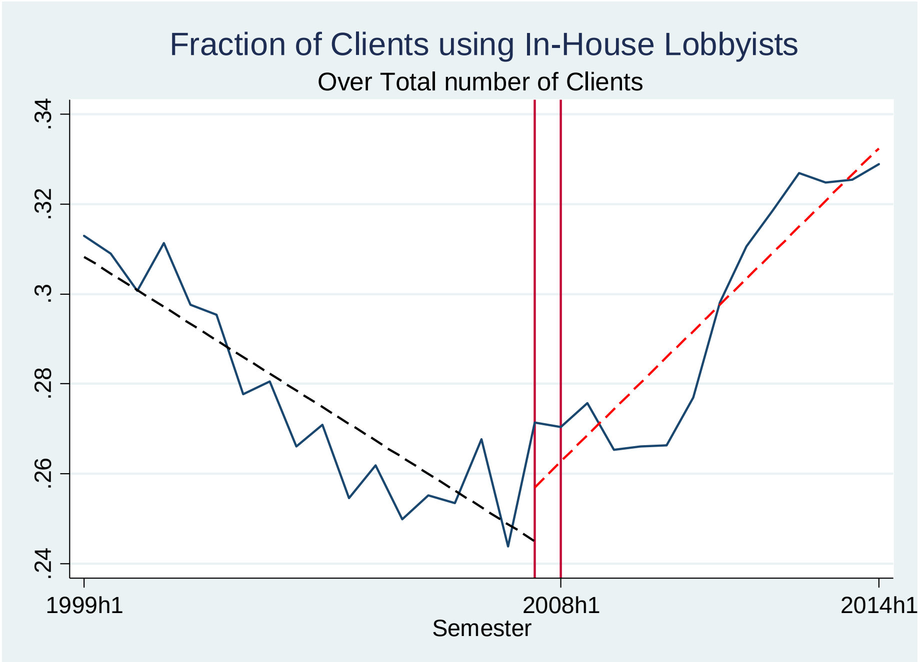

Figure 12: Time Series of the Fraction of Clients Using In-House Lobbyists. Note. The two vertical lines denote the second semester of 2007 and the first semester of 2008. An observation is a semester.

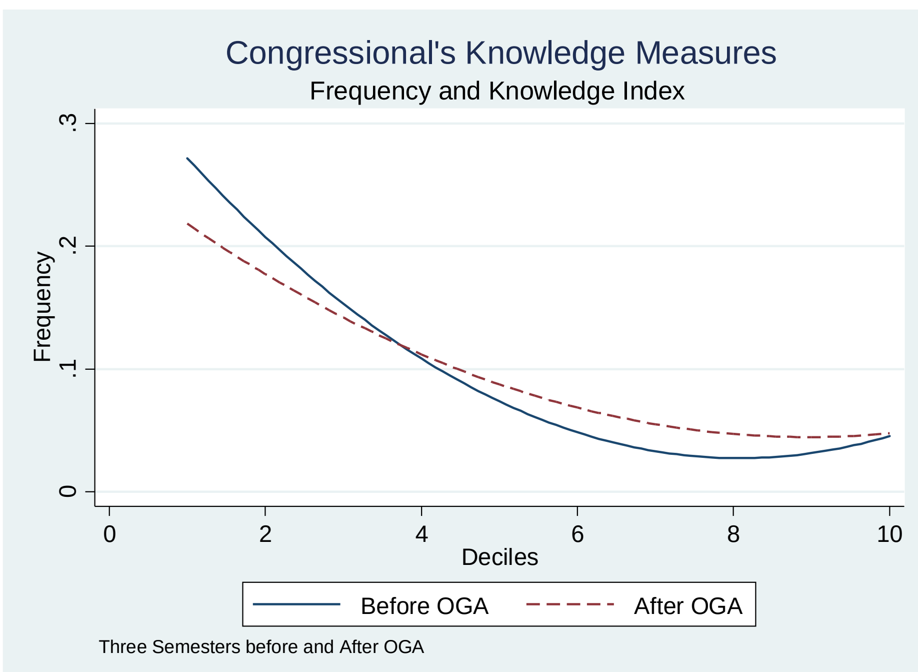

Figure 13: Differences in Congressional Knowledge Measures Before and After the OGA. Note: The x-axis represents the deciles of the index and the y-axis shows the normalized number of bills for each decile. The group *Before OGA* represents three semester before the OGA and the group *After OGA* represents three semesters since the OGA.

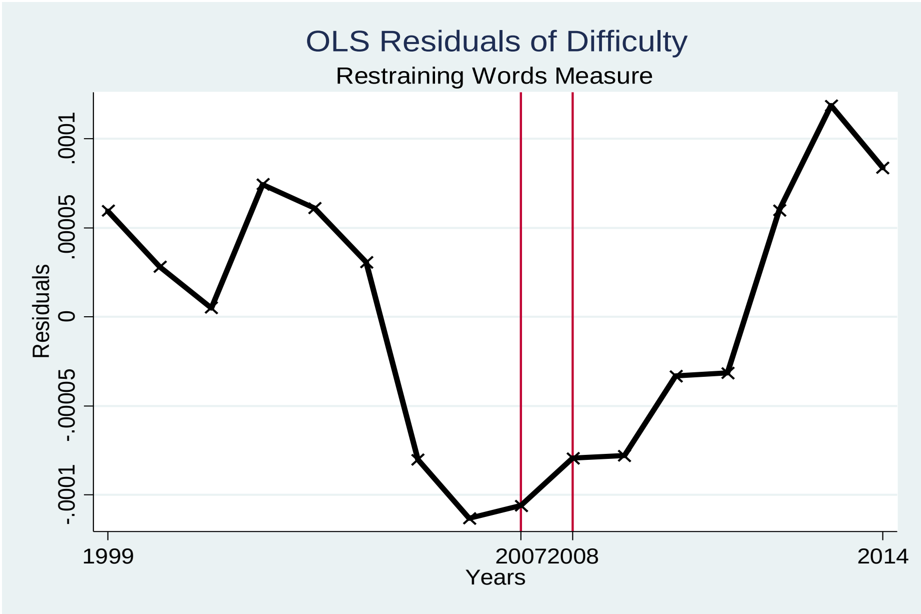

Figure 14: Plotted Residuals of the Economy-Wide Difficulty Measure Over Time. Note: Difficulty is measured with the ratio restraining words over regulating words. These calculations include all the industries in the economy. An observation is a year.

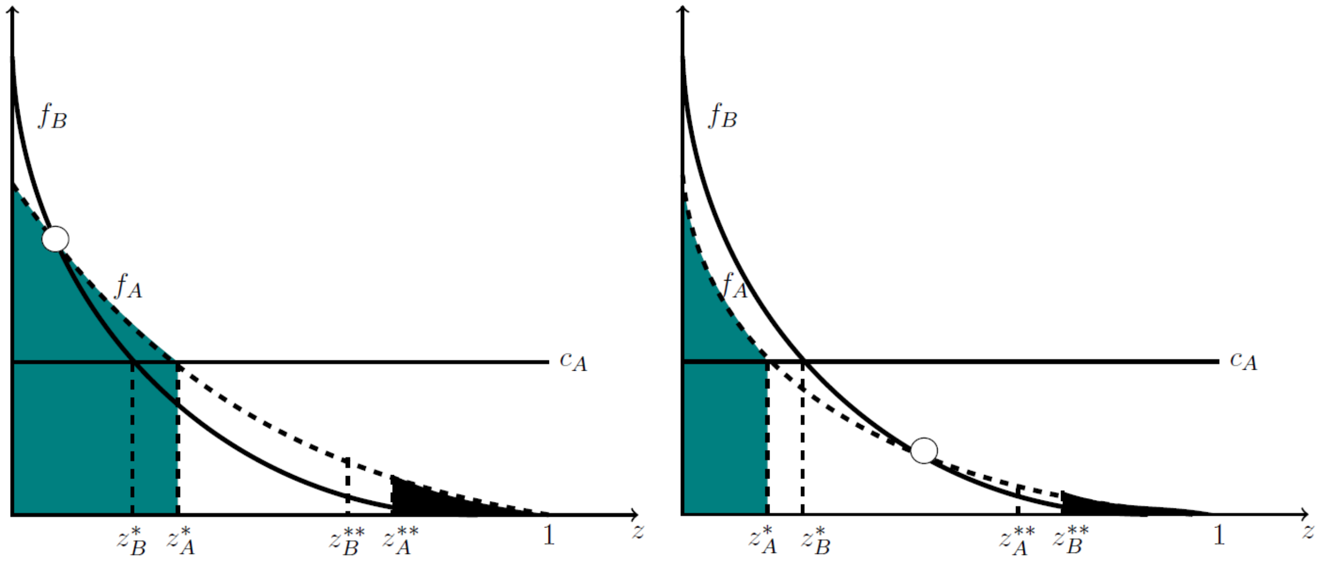

Figure 15: Theoretical Predictions of the Effect of a Change in the Frequency. Note: B denotes before the OGA and A denotes after the OGA. LHS: In this case, the densities cut above the cost of acquiring knowledge and the firm boundary shifts to the right. RHS: In this case, the densities cut below the cost of acquiring knowledge and the firm boundary shifts to the left.

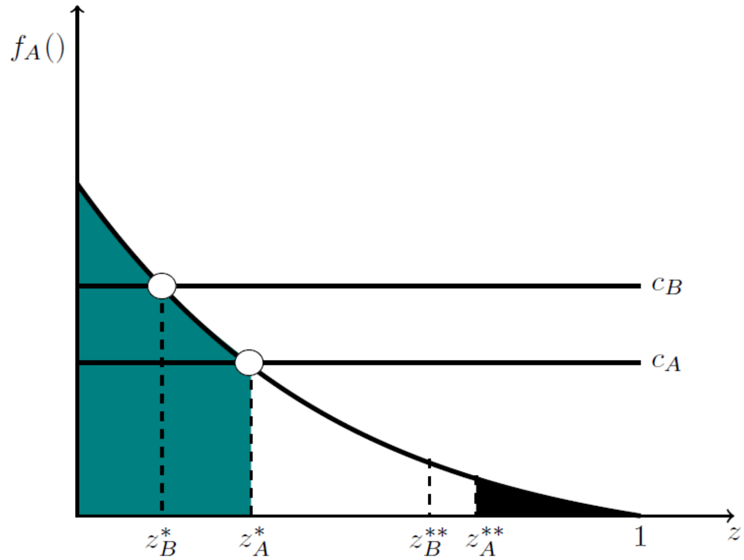

Figure 16: Theoretical Predictions of the Effect of a Change in the Costs. Note: B denotes before the OGA and A denotes after the OGA.

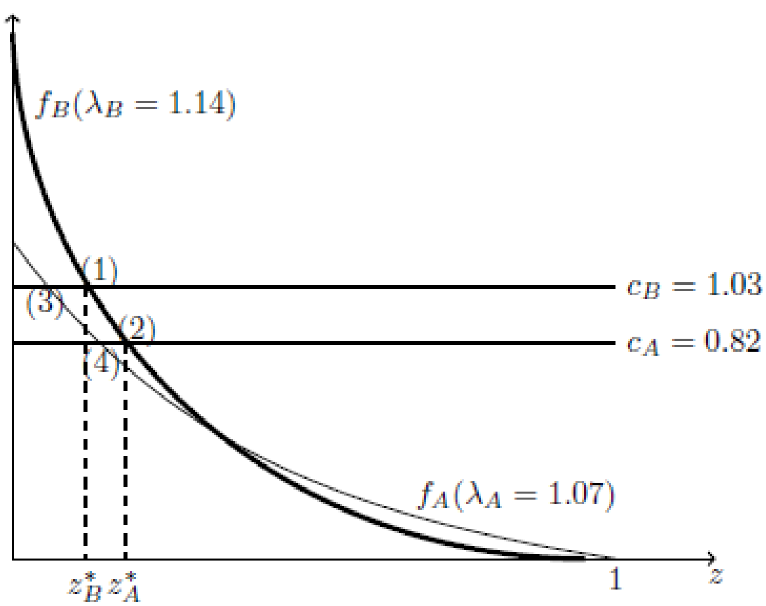

Figure 17: Graphical Explanation of the Counterfactual Exercises. Note: B denotes before the OGA and A denotes after the OGA. To simplify the presentation of this figure, I do not present the market boundary levels. However, the counterfactual exercises consider them. Point (1) represents the intersection 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.

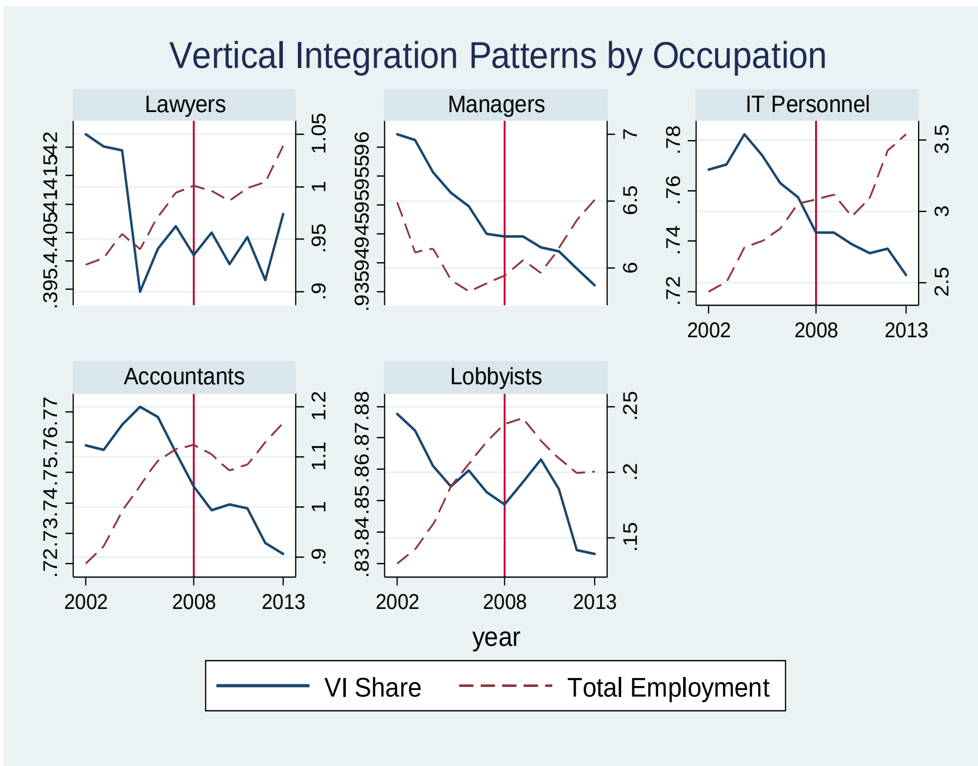

Figure 18: Time Series on Total Employment and Fraction of Integrated Clients for Five Knowledge-Intensive Occupations. Note. An observation is a year.