Institut d'études politiques de Paris

École Doctorale de Sciences Po

Doctorat en Sciences économiques

Department of Economics

Thesis supervised by Émeric Henry,

Associate Professor at the Department of Economics, Sciences Po

Defended on July 2, 2019 by:

Arthur Guillouzouic

Local Public Goods and the Geography of Economic Activity

Jury

Philippe Aghion	Professor, Collège de France & Professor, London School of Economics
Johannes Boehm	Assistant Professor, Department of Economics, Sciences Po
Keith Head	Professor, Sauder School of Business, University of British Columbia (referee)
Émeric Henry	Associate Professor, Department of Economics, Sciences Po (supervisor)
Margaret Kyle	Professor, CERNA, MINES ParisTech (referee)

Note to the reader

The three chapters of this dissertation are self-contained research articles and can be read separately. They are preceded by an introduction which summarizes the research presented in this dissertation. The terms "paper" or "article" are used to refer to chapters. Chapters 1 and 3 are coauthored, which explains the use of the "we" pronoun.

Remerciements

Je remercie en premier lieu mon directeur de thèse, Émeric Henry, qui m'a apporté un soutien constant du début à la fin du doctorat. J'ai eu la chance d'être encadré par une personne toujours à l'écoute, disponible, investie, et en qui je pouvais avoir totale confiance. Émeric, je mesure à quel point cela a été une chance, tant pour l'apprentissage de la recherche qu'il m'a été possible de mener à ton contact que pour le cadre rassurant dans lequel j'ai pu vivre ces années de doctorat. Je suis heureux que nous ayons également eu l'occasion de travailler sur un projet commun, et j'espère que d'autres projets suivront.

Je remercie également les autres personnes avec qui j'ai eu la chance de coécrire, sur des papiers ayant constitué la base de deux des chapitres de cette thèse. Pierre, tout d'abord, avec qui nous nous sommes engagés il y a près de trois ans dans un projet qui s'est avéré être un travail de longue haleine. Le papier a grandement bénéficié de ta rigueur et de ta minutie. Joan then, I have learnt a lot working with you, in particular on how to integrate theoretical and empirical investigations and to have high standards on both sides.

I also wish to thank the members of the thesis committee. I am very glad and honored to have such great scholars read and comment on my research, and I am sure this will be extremely fruitful for me. Professeur Aghion, c'est évidemment un immense honneur pour moi que vous ayez accepté de siéger dans mon jury de thèse. Je suis enthousiasmé à l'idée de confronter votre vision à mes travaux, tout particulièrement sur les questions d'innovation. Professor Boehm, I am very grateful that you accepted to be part of this thesis committee. I received very useful advice from you on an early version of the second chapter of this dissertation, and I look forward to receiving new advice. Professor Head, I thank you sincerely for agreeing to be a referee for this dissertation. You already had a decisive impact on the first chapter: beyond the fact that you helped us settle on a title, your input and encouragements at a very early stage of this work were crucial for us to persevere. Professor Kyle, I am also deeply thankful that you accepted to referee my thesis. It is an opportunity for me to receive extremely useful feedback, and I hope

it will also lay the groundwork for future exchanges.

Plus qu'un manuscrit, cette thèse a aussi été un quotidien partagé quelques années avec un super groupe de doctorants. Souvent décrites comme une expérience de solitude, j'ai eu la chance immense de vivre mes années de thèse entouré d'amis. Je remercie Étienne, compagnon de pause et de RU d'une fiabilité à toute épreuve, Charles, patriote polyglotte, Pierre D., breton wanderer, Nicolò, qui m'a fait découvrir Venise et la précision dans la cuisine italienne, Tyler, the American bleausard, pour nos nombreuses sorties escalade, Lucas, pour notre prise de conscience écossaise, JB pour l'introduction au web-scrapping, Elisa, pour nos cafés littéraires, Jan, pour son énergie communicative, Ludo, mi-homme mi-Google scholar, Clément M., pour nos discussions montagnes et nos sorties multisports, Stefan, voisin idéal, pour l'horizon chinois. Je remercie également Alaïs, Aseem, Assia, Clément B., Dylan, Edgard, Edo, Guillaume, Jean-Benoit, Jean-Louis, Marcos, Odran, Paul, Riccardo, Sophie. Merci aussi aux équipes administratives de Sciences Po, et en particulier Claudine Lamaze, Pilar Calvo et Sandrine Le Goff, pour leur efficacité.

Je tiens également à remercier toute l'équipe de l'IPP, qui m'a permis d'assurer une transition idéale avec le doctorat. Je remercie bien entendu Antoine Bozio, dont la capacité à lier excellence dans la recherche académique et un éclairage rigoureux du débat public est une grande source d'inspiration pour moi. Je remercie également mes co-bureau Brice, Claire et Madhi, qui m'ont bien soutenu dans la dernière ligne droite, Clément pour ses conseils d'identification, et tous les autres membres de cette merveilleuse équipe pour l'excellente ambiance qui y règne.

La réalisation de cette thèse doit également beaucoup au soutien de ma soeur et de mes parents, qui sont formidables. Lola, ton environnement professionnel m'a bien souvent permis de relativiser les menus problèmes que je rencontrais. Bien au-delà de leur soutien, je dois aussi à mes parents l'estime que j'ai reçue d'eux pour cet absolu qu'est la production de connaissance. J'ai d'abord cru m'éloigner de la recherche avec pour projet d'être juriste, puis d'intégrer l'administration, mais j'y suis inexorablement revenu, et la décision de s'engager dans une thèse a finalement été assez évidente. La curiosité et l'ambition de bâtir des explications rigoureuses dont j'hérite n'y est certainement pas étrangère. Je suis infiniment chanceux de bénéficier d'un environnement familial qui valorise autant l'épanouissement intellectuel.

Je remercie enfin Caroline, qui partage ma vie, pour m'avoir aidé à remonter la pente dans les montagnes russes de la thèse, et pour m'avoir aidé à redescendre des montagnes kirghizes.

Contents

Note to the reader iii

Remerciements v

Contents vii

Introduction 1

1 The Percolation of Knowledge across Space 11

1.1 Introduction . . . 12

1.2 Data and Stylized Fact . . . 17

1.2.1 Data . . . 17

1.2.2 Stylized Fact: the Persistent Effect of Distance on Knowledge Flows . . . 21

1.3 Micro Evidence of Networked Knowledge Search . . . 23

1.3.1 Empirical strategy . . . 23

1.3.2 Results . . . 29

1.3.3 Robustness . . . 31

1.3.4 Spatial search of knowledge . . . 32

1.4 Theory: Network Origins of the Distance Effect . . . 35

1.5 Estimation of Aggregate Predictions . . . 38

1.5.1 Pareto distribution of innovator sizes . . . 38

1.5.2 Distance of citations as an increasing function of innovator size . . . 40

1.5.3 Time variations and their implications . . . 42

1.6 Conclusion . . . 44

Appendices 51

1.A Technical Appendix . . . 51

1.A.1 Description of the PATSTAT database . . . 51
1.A.2 Description of examiner citations . . . 53
1.A.3 Description of the variables used in Section 1.3. . . . 54
1.B Additional Elements on Gravity Estimates . . . 55
1.C Network search: Robustness . . . 56
1.D Additional aggregate results . . . 64
1.D.1 Additional tables . . . 64
1.D.2 Additional figures . . . 67
1.E Theory Appendix . . . 68
1.E.1 Proof of Proposition 1 . . . 68

2 Moving R&D Labs **71**
2.1 Introduction . . . 72
2.2 Motivating Facts . . . 75
2.2.1 Innovative Firms Relocations . . . 75
2.2.2 Localization of Innovation Networks . . . 79
2.2.3 Plant relocations and local connections . . . 81
2.2.4 Discussion . . . 84
2.3 Model . . . 84
2.3.1 Setup . . . 84
2.3.2 Stable networks . . . 87
2.3.3 Game with simultaneous relocations . . . 88
2.4 Sequential relocations and cascading effects . . . 92
2.4.1 Environment . . . 92
2.4.2 Cascades . . . 95
2.4.3 Long-run Equilibrium . . . 96
2.5 Conclusion . . . 98

Appendices **103**
2.A Data . . . 103
2.A.1 R&D employment . . . 103
2.A.2 Patent applicants - Firms matching . . . 103
2.B Additional Tables and Figures . . . 104

2.B.1 Desc. Stat on mobility . . . 104
2.C Proofs . . . 108

3 The Public Sector and the Distribution of Economic Activity 113
3.1 Introduction . . . 114
3.2 A baseline spatial equilibrium model with two sectors . . . 117
3.3 Data . . . 119
3.3.1 Employment Data . . . 119
3.3.2 Wages . . . 120
3.3.3 Local productivity estimates . . . 121
3.3.4 Rents and housing prices . . . 121
3.4 Empirical evidence . . . 122
3.4.1 Institutional setting and summary statistics . . . 122
3.4.2 Cross-sectional evidence . . . 123
3.4.3 Spillovers: the RGPP experiment . . . 127
3.5 A spatial equilibrium model of the public sector with nominal rigidities and spillovers 134
3.5.1 Model with spillovers . . . 135
3.5.2 Model with geographical rigidity of wages . . . 137
3.6 Conclusion . . . 139

Appendices 143
3.A Proofs . . . 143
3.B Additional figures and tables . . . 149
3.B.1 On cross-sectional evidence . . . 149
3.B.2 On the RGPP reform . . . 151

Résumé en français 163

List of figures 175

List of tables 179

Introduction

A striking fact in the geography of economic activity is its extremely unequal distribution. Some areas thrive and are attractive to firms, reaching extreme levels of human concentration. In contrast, other places offering similar locational advantages are sparsely populated or seem trapped in decline. This suggests that being located in a dense environment is beneficial to these firms, beyond the natural locational advantages places may have, as well as the congestion associated with density. The analysis of so-called agglomeration economies can be traced back to Marshall (1890), and much of the research conducted by urban economists has focused either on the measurement of their strength, or on the mechanisms underlying them. Duranton and Puga (2004) provide a typology of these forces. They distinguish between three main mechanisms underlying agglomeration: sharing externalities, referring to the ability to share facilities or suppliers; matching externalities, which mostly refer to the ability to provide better matches in thicker labor markets; and learning externalities, which refer to the enhanced possibility to take advantage of knowledge produced in one's direct environment. A different approach to these agglomeration forces can be to state that, keeping locational advantages constant, workers and firms locate wherever they can benefit from better public goods. Some are global (*e.g.* military defense, democratic institutions), others are local, meaning that they only affect the agents located in the same region, city or even block as where they are provided. This thesis takes interest in two mechanisms generating spatial heterogeneity in the provision of local public goods.

In the first case, the public good is technological knowledge. It is approached as a local public good because of the high transaction costs implied in its diffusion. Knowledge should in theory be a pure public good: it is both non-rival and non-excludable. The empirical study of knowledge diffusion is however marked by two important observations. First, although one might expect ideas to flow freely across space, they diffuse much better locally. Second, social and business networks strongly condition the diffusion of ideas. Introducing networks in the analysis

can reconcile the fact that ideas should be a pure public good with the empirical finding that they diffuse better at short distances. They provide a natural representation in which knowledge travels along links at no cost, but link formation is costly and this cost increases with distance as the probability of social interaction decreases. The fact that knowledge is a local public good is particularly decisive: innovation is known to be the main engine of economic growth (Aghion and Howitt, 1992; Aghion and Jaravel, 2015), and differences in technology adoption are responsible for a large share of cross-country income differences (Comin and Hobijn, 2010). While the above representation of knowledge diffusion is convenient, networks and spatial features often correlate so strongly that their separate measurement and identification is not possible (Topa and Zenou, 2015). This makes it difficult to establish causal associations between any outcome of interest and either space or network links, all the more so as networks are usually partly or fully unobserved by the econometrician. Yet, while the social and spatial dimensions may influence each other, it is important to know if what matters fundamentally is for agents to be located close to each other, or to be strongly connected. For instance, it determines whether policies trying to facilitate the diffusion of knowledge between innovators should focus on having these innovators move to the same area, or invest into joint research programs and exchanges between them.

The second mechanism studied in this dissertation implies heterogeneity in the availability of a local public good produced by the public sector. In many developed countries, the public sector constitutes a large part of total employment[1] and is in charge of providing broader services than mere governmental functions. The functions of police, justice, education and training, health, social assistance, transportation, or energy often depend partly or fully on the public sector. All the above functions can only be produced and consumed locally, but are often essential to workers and firms alike. Contrary to knowledge, they are subject to congestion when population in an area increases. Therefore, a heteregeneous provision of these local public goods should have deep consequences on the distribution of economic activity.

This dissertation studies how the above two mechanisms, fostering heterogeneity in local public goods, determine the spatial distribution of economic activity. It examines what causes such heterogeneity, the implications it has on the location choices of economic agents as well as the overall consequences on spatial economic patterns. The first two chapters of the dissertation focus on the interplay between innovation networks, the geography of knowledge flows, and the

1. Between 15% and 25% of total employment in most European countries. Source: OECD.Stats, Government at a glance, 2017.

location of innovative firms. The first chapter explores how the connections between innovators affect the geography of knowledge transfers. The second chapter then studies the extent to which the localisation of innovation networks determines the location choices of innovators. The third chapter focuses on a different problem, in which the local public good of interest is provided by the public sector. Because of institutional constraints (namely, centralized wage-setting), all areas do not benefit from the same level of public good. If the public sector generates positive spillovers on the private sector, this should distort the distribution of city sizes and productivities.

All three chapters adopt similar methods and tools and all combine theory and empirics. Empirically, they rely exclusively on large administrative databases to provide quantitative evidence of the phenomena at work. In particular, the first two chapters make extensive use of the Patstat database, which records all patent applications and many of their features for almost all patent offices in the world, often since their creation. Similarly, chapters two and three rely extensively on the comprehensive French linked employer-employee dataset DADS Postes, which is used to analyse both the location choices of innovative firms and the spatial distribution of the public sector. In the analysis, I structure and connect the empirical results with some elements of economic theory to understand their roots and implications.

Chapter 1 The first chapter of this dissertation is entitled "The Percolation of Knowledge across Space" and is based on a joint paper with Pierre Cotterlaz. It aims at explaining how innovators learn about each others' ideas in spite of frictional knowledge diffusion, and how these learning dynamics may help decipher the peculiar geography of knowledge flows.

This work is rooted in a vast body of literature showing that knowledge flows exhibit a strong spatial bias. This literature can be traced back to Jaffe et al. (1993), who use patent citations to show that an inventor is more likely to cite patents developed around her than similar innovations developed further away. While this finding led to a fierce debate (Thompson and Fox-Kean, 2005; Henderson et al., 2005), many papers confirmed it both at the very micro level (Thompson, 2006; Murata et al., 2014) and on international or interregional citation flows (Bottazzi and Peri, 2003; Peri, 2005; Li, 2014). The mechanism driving this role of distance still needs to be formally established. While it seemed intuitively linked to the lower probability of social interactions at greater distances, the difficulty to identify social interactions separately from geography in the data was not overcome until recently. Several papers focus on a specific

measure of social proximity and show that its inclusion in the analysis reduces the measured effect of distance (Singh, 2005; Breschi and Lissoni, 2009; Agrawal et al., 2008; Kerr, 2008). Focusing on citations by researchers in mathematics, Head et al. (2018) show more formally that social ties inherited from one's career determine which articles they cite, what in turn explains a large part of the effect of distance on citations.

Building on the above work, this chapter studies the dynamics of link formation between innovators, and how they relate to the aggregate effect of distance on knowledge flows. It intends to characterize the evolution of firms' links with other innovators over their life-cycle. To achieve this, it studies empirically the influence of what an innovator knows about at a given point in time on what she is likely to learn about in the future. In practice, we use patent citations as the "paper trail" of their knowledge set, and compare citations made over time to track its evolution. We initialize it in an early year and define as contacts the other innovators that a studied firm already knows about. We then compare the subsequent citations to a set of relevant citations which could be expected if knowledge diffusion was frictionless. To isolate such set for each patent, we rely on patent citations added by the experts in charge of the examination of patent applications in patent offices, who are in charge of making the reference list comprehensive. We show that firms learn about new knowledge step by step, through the contacts of their own contacts. This fact is used to build a model following Chaney (2018), in which the dynamics of link formation have implications on the effect of distance faced by firms depending on their age. Small firms are born with a set of knowledge originating from firms located at rather short distances, and then expand their set of knowledge over their life-cycle. This initial spatial feature of where a firm's contacts are located can in turn explain why such a persistent effect of distance is found on knowledge flows. It provides two stark aggregate predictions: these predictions fit the data exceptionally well, and deliver two stylized facts. First, the distribution of innovators sizes follows a Zipf law. Second, there exists a log-linear relationship between a firm's size and the average (squared) distance of its citations.

This result can provide a tentative answer to a puzzle: why has the aggregate effect of distance remained constant since the 1980s, in a period when ICTs developed massively? Tools such as Google Patents, which made any patent immediately searchable, should indeed have made innovation diffusion much less frictional. This chapter provides some explanations: while the effect of distance has not decreased, the link between a firm's size and the distance at which it cites has. A possible solution to the puzzle coming out of our analysis therefore is that the

effect of distance on each firm taken individually has indeed decreased, but was compensated by an increase in the share of small firms, facing stronger effects of distance, relative to large ones.

Chapter 2 The second chapter, entitled "Moving R&D Labs", takes as given the local knowledge networks that the previous chapter has analysed. It seeks to answer, both empirically and theoretically, how firms internalize the existence of such local networks, and the unequal spillovers they can expect to receive from them, to make their location choices. The analysis is motivated by the need to link two established facts on the geography of innovation: the fact that knowledge diffuses better at short distances, and the fact that innovation activities are geographically more concentrated than economic activity in general (Audretsch and Feldman, 1996; Buzard et al., 2017). While knowledge spillovers provide an additional reason to cluster in space, few papers have focused on the location choices of innovators in particular. This chapter studies relocation choices, spillovers firms abandon when they move and spillovers they expect from the new location.

The understanding of this phenomenon is very relevant to the design of public policies. Indeed, based on the example of the flourishing Silicon Valley cluster, most developed countries have tried to foster the creation of similar clusters through billions of dollars in subsidies or tax credits.[2] These policies have been (sometimes harshly) criticized by economists both because of effects that were often perceived as weak compared to the induced public investments, and because of misspecified theoretical foundations (Glaeser and Gottlieb, 2008; Duranton, 2011). This motivates a deeper understanding of the link between firms' location and relocation choices, and the knowledge spillovers they receive from their direct environment.

The first part of the chapter describes this mobility phenomenon in the data. I use the DADS Postes data, which contain exhaustive linked employer-employee information in France, and characterize innovative firms as the ones employing workers with a position explicitely labelled as R&D. I then study their propensity to move their R&D workers and the distance at which they move. Two salient facts emerge from the analysis: innovative firms are more likely to relocate than other firms employing high-skilled workers, and they move further away. I then construct a measure of how valuable the place that the firm joins and the one it leaves are for its innovation. I match these firms data with patent data, and depict the centrality of moving firms in their local knowledge network. I find that having better connections in their original

2. As an example, the iconic "Pôles de compétitivité" policy in France has induced around 2.5 million euros of public expenditures between 2009 and 2014 (European Commission, 2016).

location makes firms less likely to relocate, and that movers tend to join locations where they can expect to benefit from larger spillovers.

The second part of the chapter builds on the above facts and formalizes the problem faced by R&D firms striving to optimize the knowledge spillovers they receive. To tie their links to other agents with their location decision, I build a theoretical model relying on tools from the "games on networks" literature (Jackson and Zenou, 2015). I abstract from traditional industrial organization features relative to the production function and competition faced by the firm: this is key to emphasize the main factors influencing joint relocation decisions while keeping the model tractable. In this model, firms are born in a location with a set of links towards some of their geographical neighbors, from which they receive positive knowledge spillovers. Within a location, firms differ in their centrality, a measure of how well connected they are to other firms in their location. These firms have limited information on the level of spillovers they could receive in the other location. I first study stability, understood as the fact that no firm would be better off in the other location. I show that it depends simply on the difference between the lowest centrality in a location and the average centrality in the other location. I then study a two-stage game, in which firms may first want to relocate if this gives them higher expected payoffs, and then choose their optimal R&D effort. I characterize equilibria if firms made their relocation choices simultaneously as the set of stable situations, and describe further an equilibrium which is both Pareto-optimal and welfare-maximizing. I then provide an analysis of a process in which movements take place sequentially: it shows under which conditions cascades of movements can be expected.

Chapter 3 The third chapter of this dissertation is entitled "The public sector and the distribution of economic activity", and is part of a joint project with Emeric Henry and Joan Monras. It intends to explain the extent to which the location of civil servants who produce local public goods consumed by workers and firms determines the spatial distribution of the private sector. Using comprehensive data on both public and private employment in France, we show that two distinctive features make the public sector stand out compared to the standard spatial equilibrium model.

The recent crisis of the "yellow vests" in France has brought back at the center of the political discussion the issue of an equal access of territories to basic public services. This resonates with the polarizing political question of whether and where some public jobs should be cut. Yet, the

evidence of the effects such jobs have on economic activity at the local level is recent and does not yet provide a clear answer to whether they crowd in or out private sector activity. Part of this undeterminacy lies in the fact that most of the exploitable quasi-experiments relocate civil servants who really produce a global public good, such as government workers or militaries (Becker et al., 2013; Jofre-Monseny et al., 2019; aus dem Moore and Spitz-Oener, 2012), rather than local public goods supplied by schools, hospitals or police stations. The work presented in this chapter tries to formalize and illustrate the effects one can expect from local public goods provided by public sector workers, in the particular case of centralized wage-setting in this sector. It shows how this deviates from the baseline spatial equilibrium framework with two sectors (Rosen, 1974; Roback, 1982).

The chapter first highlights that the central setting of wages, with limited leeway to adjust them to local prices, implies different wages across sectors within a city. Following the standard result that mobile private sector workers must be equally well in every city as they get compensated for amenity differences implies that public sector workers' utilities are not equalized across places. In the French context, this is illustrated by many well-known facts, for instance the stark age sorting of teachers across areas, with young workers starting their careers in less desirable places before being able to move away. Because a higher productivity of the private sector puts pressure on the housing market for all workers in a city, civil servants are better off in places with lower rents, thus lower productivity (keeping amenities constant). The analysis shows that this might rationalize part of the imbalances between the respective numbers of workers in the public and in the private sector across the French territory.

The second important factor through which the location of civil servants influences local economic activity is the spillovers they produce on the private sector. We use large public job cuts which occured during the period 2008 to 2012 and affected areas differently to highlight the implied losses in private employment in subsequent years. Consistently with the literature, we find that places suffering larger losses also lost employment in the non-tradable sector, which is easily explained by the drop in local demand for non-tradable goods and is well documented in the literature (Moretti, 2010; Faggio and Overman, 2014; Faggio, 2019). In contrast with existing literature however, we find that places affected more strongly by the policy also suffer large losses in the tradable sector, which can only be rationalized through the existence of positive spillovers from the public sector to the tradable sector.

The above findings are then integrated in a spatial equilibrium model with two sectors,

where public sector workers are sent in a city at random, have a fixed wage, and generate positive spillovers on the other sector. The model shows that the fixed-wage structure tends to increase the gap in labor market size between low and high productivity cities. This is the case because the wage-constrained sector can spend less in housing than if wages were flexible, which puts less pressure on the housing sector in the more productive cities, in turn leading to even starker differences of size between cities with different productivities. Spillovers have the opposite effect. A higher productivity in the private sector attracts private sector workers at the expense of public workers and congests the public good produced by the public sector. Therefore, the existence of spillovers affecting positively the private sector lowers the gap in size between the most and the least productive cities.

References

Aghion, P., and P. Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60 (2): 323–51.

Aghion, P., and X. Jaravel. 2015. "Knowledge Spillovers, Innovation and Growth." *Economic Journal,* no. 583: 533–573.

Agrawal, A., D. Kapur, and J. McHale. 2008. "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics* 64, no. 2 (September): 258–269.

Audretsch, D. B., and M. P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (3): 630–640.

Becker, S., S. Heblich, and D. Sturm. 2013. *The impact of public employment: evidence from Bonn.* Technical report.

Bottazzi, L., and G. Peri. 2003. "Innovation and spillovers in regions: Evidence from European patent data." *European economic review* 47 (4): 687–710.

Breschi, S., and F. Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of Economic Geography:* lbp008.

Buzard, K., G. A. Carlino, R. M. Hunt, J. K. Carr, and T. E. Smith. 2017. "The agglomeration of American R&D labs." *Journal of Urban Economics* 101:14–26.

Chaney, T. 2018. "The Gravity Equation in International Trade: An Explanation." *Journal of Political Economy* 126 (1): 150–177.

Comin, D., and B. Hobijn. 2010. "An Exploration of Technology Diffusion." *American Economic Review* 100 (5): 2031–59.

Duranton, G. 2011. "California Dreamin': The Feeble Case for Cluster Policies." *Review of Economic Analysis* 3 (1): 3–45.

Duranton, G., and D. Puga. 2004. "Micro-foundations of urban agglomeration economies." Chap. 48 in *Handbook of Regional and Urban Economics,* 1st ed., edited by J. V. Henderson and J. F. Thisse, 4:2063–2117. Elsevier.

European Commission. 2016. *Smart Guide to Cluster Policy.* Technical report.

Faggio, G. 2019. "Relocation of public sector workers: evaluating a place-based policy." *Journal of Urban Economics,* no. forthcoming.

Faggio, G., and H. Overman. 2014. "The effect of public sector employment on local labour markets." Spatial Dimensions of Labor Markets, *Journal of Urban Economics* 79:91–107.

Glaeser, E. L., and J. D. Gottlieb. 2008. "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity* 2008 (1): 155–253.

Head, K., Y. A. Li, and A. Minondo. 2018. "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics." *Review of Economics and Statistics,* no. forthcoming.

Henderson, R., A. Jaffe, and M. Trajtenberg. 2005. "Patent citations and the geography of knowledge spillovers: A reassessment: Comment." *American Economic Review* 95 (1): 461–464.

Jackson, M. O., and Y. Zenou. 2015. "Games on networks." In *Handbook of game theory with economic applications,* 4:95–163. Elsevier.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *the Quarterly journal of Economics:* 577–598.

Jofre-Monseny, J., J. I. Silva, and J. Vázquez-Grenno. 2019. "Local labor market effects of public employment." *Regional Science and Urban Economics,* no. forthcoming.

Kerr, W. R. 2008. "Ethnic scientific communities and international technology diffusion." *The Review of Economics and Statistics* 90 (3): 518–537.

Li, Y. A. 2014. "Borders and distance in knowledge spillovers: Dying over time or dying with age?—Evidence from patent citations." *European Economic Review* 71:152–172.

Marshall, A. 1890. *Principles of economics.* Digireads. com Publishing.

Aus dem Moore, J. P., and A. Spitz-Oener. 2012. *Bye bye, GI-The impact of the US military drawdown on local german labor markets.* Technical report. Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Moretti, E. 2010. "Local Multipliers." *American Economic Review P&P* 100, no. 2 (May): 373–77.

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura. 2014. "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *The Review of Economics and Statistics* 96 (5): 967–985.

Peri, G. 2005. "Determinants of knowledge flows and their effect on innovation." *Review of Economics and Statistics* 87 (2): 308–322.

Roback, J. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6):1257–1278.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82:34–55.

Singh, J. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5): 756–770.

Thompson, P. 2006. "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations." *The Review of Economics and Statistics* 88 (2): 383–388.

Thompson, P., and M. Fox-Kean. 2005. "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review* 95 (1): 450–460.

Topa, G., and Y. Zenou. 2015. "Neighborhood and network effects." In *Handbook of regional and urban economics,* 5:561–624. Elsevier.

Chapter 1

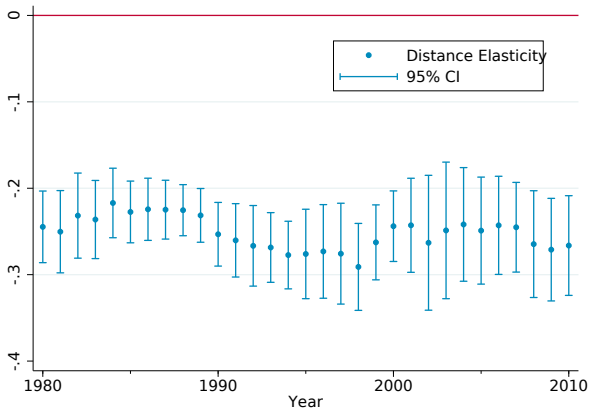
The Percolation of Knowledge across Space

This chapter is a joint work with Pierre Cotterlaz.

1.1 Introduction

Despite considerable improvements in information and communication technologies in the past three decades, geographical distance remains a serious hindrance to knowledge diffusion. As Figure 1.1 shows, the elasticity of international patent citation flows with respect to geographical distance hovers around -0.3, meaning that a 10% increase in the distance between two countries is associated with a 3% decrease in citations between them. This is all the more surprising as most conventional distance-related costs, such as transport costs or tariffs, do not apply to ideas. While it would be reasonable to expect a decrease in its magnitude since the introduction of search technologies such as Google Patents, it seems on the contrary to have remained very stable since the 1980s.

Figure 1.1: Elasticity of international patent citation flows with respect to distance, over time.



Note: The above figure displays estimates and confidence intervals obtained on the distance coefficient in structural gravity estimations - equation (1.1) -. Section 1.2 provides more information on the specification and the data used to perform these estimations. Self-citations and intranational citations are excluded.

Attempting to solve the above puzzle, several papers have argued that social links play an important role in knowledge diffusion:[1] they typically find that controlling for social distance decreases the effect attributed to spatial distance. This paper seeks to go one step further. It starts by describing precisely how links form: it documents a phenomenon called "triadic closure", in which firms disproportionately form links with firms two steps away from them (i.e. contacts of contacts). We show that this determines how knowledge flows between innovators, and then formally tie it to the distance effect observed in aggregate. Both our micro and our

1. See for instance Singh, 2005; Kerr, 2008; Agrawal et al., 2008; Breschi and Lissoni, 2009.

aggregate findings uncover large differences according to firms' size in the extent to which they rely on their network, and thus are affected by distance.

We use past patent citations to build network links in a flexible way, and implement a novel identification strategy using citations added by examiners as counterfactual citations to show the influence of the network on the probability that a link forms. It provides evidence that firms are more aware of knowledge originating from firms they are linked with (their contacts), and are prone to linking with the contacts of their contacts. This is reminiscent of the physics phenomenon of percolation, in which knowledge is approached as a fluid making its way from an inventor to another along network paths. We then use this finding as a key ingredient of a model able to explain the aggregate effect of distance, and show that the data fit these predictions. Thus, our paper provides evidence that the spatial clustering of social and business ties is the root of the distance effect observed in aggregate. More specifically, it highlights the dynamics of these ties' formation. It shows that small firms tend to rely substantially more on their existing links, while in aggregate, large firms tend to cite further away. These facts are tied together by a model in which firms grow over time as their network spreads step by step, implying that firms are less and less affected by distance as their size/age increases, simply because of the time they have had to expand their network.

Having an acute understanding of the forces underlying the imperfect dissemination of knowledge is of prime importance. Innovation and technology diffusion are essential for growth as well as convergence patterns between countries (Nelson and Phelps, 1966; Aghion and Jaravel, 2015; Akcigit et al., 2018). Indeed, a small group of high income countries achieves a disproportionate share of technological knowledge production,[2] while productivity growth in other countries depends considerably on knowledge flows from those few highly innovative economies which are likely to condition technology adoption.[3] While the focus of the current paper is on international knowledge flows, the extreme spatial concentration of innovation means this reasoning can be extended at smaller scales such as regions or urban areas.

This paper starts by documenting important facts on the network formation process between innovators. Our test for diffusion along the network links relies on the use of examiner-added citations to build a counterfactual for what innovators would cite if they knew every relevant

2. For instance, in 2011, roughly 80% of triadic patent families (patents applied for in USPTO, EPO and JPO) were achieved by applicants residing in only 5 countries (Japan, the US, Germany, France, Korea). Source: http://stats.oecd.org.

3. Along these lines, Comin and Hobijn (2010) estimate that cross-country variation in the timing of technology adoption accounts for 25% of per capita income differences.

patent. When applying for a patent, innovators are required to give a list of all the patents on which their invention builds. This list is completed by experts from the patent office, who make on average 60% of all citations. Our main identifying assumptions for the effect of the network are that in a frictionless world, firms would cite all relevant patents, and that the union of references added by applicants and by examiners is a good proxy for the set of relevant patents. Therefore, examiner citations help forming an ideal group of counterfactual citations: they are relevant to the patented invention and are observably similar to applicant citations, but were not known by it (otherwise it would have cited them). We form links using past patent citations by applicants, which indicate that they knew about this innovation. By looking at whether, among our group of relevant references, patents from linked firms are found disproportionately often in applicant-added citations, we can identify the effect of the network of innovators on the use of knowledge.

Our estimates show that firms are 30% more likely to cite a patent belonging to one of their contacts than if it were from outside their network. This is however strongly heterogeneous, since firms not belonging to the 1% largest are almost three times more likely to cite patents if they belong to contacts of them than if they do not. Moreover, percolation really operates since this effect expands beyond direct links: we also find that the citation of a patent is 70% more likely when this patent had previously been cited by at least one of the firm's contacts than when it was unknown from its contacts. Our estimates are robust to the introduction of a range of control variables, as well as to a variety of robustness tests.

We then undertake to bridge the above finding with the aggregate distance effect. The fact that distance negatively affects bilateral flows of goods has been widely studied in trade economics (Head and Mayer, 2014), through the so-called gravity equations. Interestingly, recent developments in trade gravity models provide insights on the determinants of such spatial frictions for knowledge flows, even though the nature of the object they apply to is different in many aspects (it is often assumed to be non-rival and non-excludable, in contrast with traded goods). Abstracting from trade costs, Chaney (2018) builds a dynamic model of network formation with search for international trade partners through the network, adapting the established idea of triadic closure in the social networks literature, *i.e.* the disproportionately high likelihood to make friends with friends of friends (Jackson and Rogers, 2007). The model describes an economy in which firms get knowledge from contacts located further and further away as they grow older, but in which a constant growth rate in the number of firms generates a large popula-

tion of new and small firms relative to old and large ones. This model generates predictions which connect directly to the distance feature of gravity equations. Our hypothesis is that an analogous phenomenon takes place for knowledge flows: firms initially access knowledge from spatially clustered contacts, and sequentially obtain new sources of spillovers through their existing contacts. In order to adapt the theoretical framework to the search of knowledge sources, we slightly modify Chaney (2018)'s model to allow for the possibility of "spatial search",[4] which we model as the possibility for firms to find new partners in the places where they already have a contact.

Finally, we bring to the data the two key theoretical predictions of the network formation model, which are sufficient to explain the observed negative distance elasticity. Firstly, the size distribution of innovators should be Pareto. Secondly, an increasing power function should link the average squared distance at which firms cite to their size. We find that these features hold remarkably well in the data. On top of being sufficient conditions to generate a constant negative distance elasticity, these two predictions of the model are interesting stylized facts in themselves. Indeed, we show that, beyond being well-described by a Pareto distribution, the size distribution of innovators actually enters the class of economic objects following a Zipf law. Similarly, the systematic relationship between an innovator's size and the distance at which it is able to access to knowledge is a novel finding, which we find to hold very well in a variety of settings, both in cross-section and over time.

An important takeaway of this paper therefore is that small firms are the main contributors to the aggregate effect of distance. Innovators start off relying on knowledge produced by contacts located close to them, and get links with innovators located further away as they grow through network search. We find that while the overall effect of distance remained constant over time, the relationship between size and distance of citations weakened in our period of study: we show that this was caused by small innovators accessing more distant knowledge. Although this should have implied a decrease in the overall effect of distance, it seems to have been offset by an increased share of small innovators versus large ones.

This paper relates to several other strands of the literature. Micro evidence of spatial frictions in the diffusion of knowledge were first brought out in Jaffe et al. (1993), comparing the colocation rates of realized vs non realized citations, and was later discussed and refined by Thompson and Fox-Kean (2005) and more recently by Murata et al. (2014). Thompson (2006) and Alcácer and

4. Note that Chaney (2014) studies the network formation at the individual level and allows for an analogous type of spatial search.

Gittelman (2006) contributed to this literature by using citations added by examiners to set forth the local bias of applicants in their citations. We use the same tool in our identification strategy, this time neutralizing the network bias of applicants rather than their spatial bias. The other main approach has used aggregate bilateral flows between geographical units and measured whether these aggregated flows were affected by geographical variables (mostly administrative borders and distance). This approach was pioneered by Maurseth and Verspagen (2002), and later used by Peri (2005) and Li (2014). These papers also found a decay in the probability of a patent citation with distance. Additionally to being less intense, knowledge spillovers between remote locations also take longer to occur, as evidenced by Griffith et al. (2011), who showed that there exists a home bias in the speed of citation, meaning that domestic institutions are quicker to cite domestic patents than foreign institutions, a finding later confirmed by Li (2014).

The effect of social networks on the diffusion of technological and scientific knowledge was first studied using specific types of links. Singh (2005) studied interpersonal links through coinvention within patents (which our analysis largely excludes by removing applicants' self-citations) and found that controlling for ties diminishes greatly the effect of geographical variables on the probability of a citation. Similarly, Breschi and Lissoni (2009) found that controlling for mobility of skilled workers between firms reduced the effect of distance. Agrawal et al. (2008) and Kerr (2008) proxied social proximity with ethnicity as revealed from names and found it increased the probability of citation. Head et al. (2018) studied citations between research articles in mathematics, and controlled for social ties in a more elaborate way, building connections based on past acquaintances (working in the same institution, being one's PhD supervisor, etc.), and reached a similar conclusion. In the same vein, Iaria et al. (2018) found that, by disrupting encounters and exchanges between scientists of both sides of the conflict, WWI greatly reduced international knowledge flows, while Catalini et al. (2018) showed that the opening of a low-cost airline increased collaboration between scientists at both ends, implying that travel costs were an important friction to knowledge diffusion. Hypothesizing that social interactions between adopters and non-adopters of a technology are at the root of technology adoption, Comin et al. (2012) studied how a set of important technologies diffused in space, exploring an hypothesis relying on traveling routes and social interactions.

In contrast with the above strand of the literature, we do not restrict our attention to a particular type of links. This flexible approach is allowed by the fact that we use past patent citations to construct the network of innovators: rather than constructing links based for instance

on R&D collaborations, we initialize the contacts of an applicant using the citations made in a given year. We ask how likely it is that knowledge will flow again along a link, either through the citation of another of the contact's patents, or through the citation of a patent previously cited by the contact. This provides an asymmetric measure of links which is general enough to encompass many of the usual explanations of the localization of knowledge spillovers: citations could capture links as diverse as formal R&D collaboration agreements, linkages with geographical neighbours (*e.g.* inside clusters), inter-firm mobility of engineers, input-output linkages, acquaintances from college between inventors, etc. While we lack information on the nature of these links, such generality is a major advantage if one wants to explain phenomena observed in aggregate.

The remainder of the paper is organized as follows. The next section describes the data and replicates the stylized fact that distance negatively affects the intensity of international knowledge flows. Section 1.3 provides micro evidence of knowledge percolation from actual link formation between contacts, while section 1.4 builds a theoretical framework linking dynamic network formation with the effect of distance. Finally, section 1.5 empirically shows that the model's aggregate predictions hold on patent data.

1.2 Data and Stylized Fact

1.2.1 Data

Patent Citations The standard approach in the literature to track knowledge flows has been the use of patent citations: when applying for a patent, the applicant is required to cite the relevant prior art on which its invention builds. Therefore, the widespread assumption made by this literature is that a patent citation reflects a knowledge transfer from the cited patent to the citing patent. Patent citations are an imperfect proxy of knowledge flows. Reasons include the fact that many patents are valueless, that citation rules vary across offices, that citations can be handled by lawyers rather than inventors, include some strategic considerations (Lampe, 2012), or that inventions are rarely patented in some industries. In a nutshell, assimilating patent citations to knowledge flows could both introduce many citations having led to no knowledge transfer at all and miss knowledge transfers which did not lead to a citation. In particular, a long-standing criticism towards the use of patent citations as a proxy for knowledge flows has been the potential bias induced by the presence of examiner-added citations among the

citations. Indeed, although patent applicants are compelled to cite prior art on which their invention relies in their initial application, this list is often incomplete, and examiners working for patent offices are in charge of making it comprehensive. Applicant added citations are a priori the only citations which may reflect knowledge transfers. Moreover, surveying patent applicants at the USPTO, Jaffe et al. (2000) found that a sizable share of citations did lead to a knowledge transfer. In the same spirit, Duguet and MacGarvie (2005) surveyed French applicants at the EPO and found that citations indeed correlate with ways for inventors to learn about new knowledge such as R&D collaboration and technology licensing.

Fortunately, our database (PATSTAT, Fall 2016 edition) includes, for patent applications made in the early 2000s onward, a variable indicating whether the citation was added by the applicant itself or by the examiner during prosecution time. This piece of information was made available by the USPTO in 2001 and in 1978 for the EPO. Consequently, it becomes widely available in the database for patents applied for in the years 2000s (as shown in Figure 1.A.1 in the Appendix). In the population from which we draw our samples (USPTO patents posterior to 2000), each patent has on average 5 applicant-added citations, and 12 examiner-added citations (see Figure 1.A.3). This means that, on average, out of 17 relevant patents, applicants have actually cited 5 of them. An interesting fact is that the sequentiality of the citation procedures (applicant then examiner) does not seem to make overlapping citations impossible: indeed, out of the 73 million citations made within the USPTO from 2000 on, 13% of citations are made by the examiner even though the applicant had already made them, and this share rises to 20% when only the 47 million citations from patents with at least one applicant-added citation are considered.

The citation procedure is the following. At the time of the application, patent assignees are asked to cite the relevant prior art, which helps judge the patentability of the invention, and notably its novelty relative to the existing technological background. The exact nature of this requirement varies slightly across offices: for instance, applicants at the USPTO have the obligation (called "duty of candor") to do so for the patent to be enforceable once granted, while the requirement is softer at the EPO.[5] The application is then assigned to an office examiner in the relevant group called art unit. To assess novelty of each of the claims that the patent contains, the office examiner looks for relevant prior art and typically produces a comprehensive search report which has to be thorough and exhaustive, making use of the variety of tools at her

5. Yet, as Akers (2000) explains, applicants at the EPO have incentives to cite the relevant patents when they file their application.

disposal.[6] It contains the documents that she considers to be relevant prior art and patents with potentially overlapping claims. Based on this exhaustive search, the examiner adds references to the patent.

The information on whether citations come from the applicant is unfortunately not available for all patents. Notably, at the USPTO, it appears on patents published after 2001, thanks to a change in citations reporting.

Patent Applicants Patent applications distinguish between the people who actually developed the claimed invention (called the inventors) and those who will obtain the legal rights over the invention if the application is successful (equivalently called the applicants or the assignees throughout this paper). Notably, inventors are usually employees of the institution which obtains the rights over the invention. Therefore, inventors are always private individuals, while a vast majority of assignees are firms. Since our focus is on firms, we determine the country of a patent through the country of its assignee. However, for large firms, the country indicated on the patent may correspond to the location of the headquarters, instead of the location where the innovation process actually took place. In this case, using the country of the inventors would give a more accurate information on the place where research was conducted. Thus, we also present results obtained using the inventors to determine the patent's country as a robustness check. Finally, 11% of the applications have several assignees, potentially based in different countries. In such case, we consider the application to be located in the country that appears most frequently among the assignees (the mode), and if there is no mode, we assign randomly one of the assignees' countries to the patent.

Patents do not include unique firm identifiers, therefore the allocation of a patent to a firm can be made only through the assignee name indicated on the patent. A common issue is that the applicant's name may be different even for patents belonging to the same firm due to spelling mistakes, spelling variations, and national units of large companies. Therefore, some algorithms were developed to harmonize applicant names. PATSTAT contains several name harmonizations, of which we use the Patstat Standardized Name (PSN) applicant identifier.[7]

6. "Upon creation of a European search report [...], a pre-search algorithm generating a list of documents to be inspected by the examiner is triggered.[...] The examiner should start the search process by formulating a search strategy, i.e. a plan consisting of a series of search statements expressing the subject of the search, resulting in sections of the documentation to be consulted for the search." (EPO, 2016)

7. Provided by ECOOM `https://www.ecoom.be/en/EEE-PPAT` it is automated and is particularly accurate for the largest patentees, which is crucial when estimating a size distribution. Moreover, it is available for assignees at all offices represented in PATSTAT, while the HAN harmonization conducted by the OECD is mostly for the EPO.

Note also that along with name harmonization, PATSTAT contains information on the type (firm, university, etc.) of each applicant. Unless specified otherwise, we keep only the applicants that are signaled as firms in this harmonization. As a robustness check (shown in subsection 1.3.3), we also go a step further and conduct our analysis after matching names with the firm database Orbis, to check the consistency of the identifier and run robustness checks at the group level.

The information on the country of the assignee is only available for about half of the patents. Nevertheless, there is a simple way to improve this figure by making use of the name harmonization work performed by PATSTAT. Suppose the country is missing for a patent, but is available for another patent granted to the same assignee: we consider that the country of the former patent is also the one of the latter patent. Thanks to this method, we infer geographic information for about one third of the patents, which leaves us with only few patents without country information, as illustrated in Figure 1.A.2.

Contacts Our definition of contacts, which section 1.3 will use extensively, is the following. The set of contacts f of a given firm is defined as all the assignees of patents truly cited (ie cited by the applicant) for the first time in an initial year. Unless specified otherwise, contacts are initialized on citations made in year 2000, for the coverage reason mentioned above. For this measure to remain an acceptable proxy of an existing link between two applicants, we exclude citations towards very large applicants (*i.e.* applicants belonging to the top 1% of the size distribution, where size is measured as the total number of patent applications in the database): our assumption is that industry leaders are too widely visible for a citation towards them to be meaningful, and for differences of informational frictions between examiners and applicants to be exploitable. Additionally, it makes the construction of the database considerably lighter (since all the citations made by all their patent applications would have to be constructed).

Building distance 2 links (contacts of contacts) is a heavy procedure. To alleviate the analysis conducted in section 1.3, we randomly select a third of all firms which would enter our analysis, which amounts to more than 7,000 firms applying for 650,000 patents and citing more than 10 million patents.

1.2.2 Stylized Fact: the Persistent Effect of Distance on Knowledge Flows

We test for the existence of spatial frictions in the diffusion of knowledge by studying the sensitivity of the flows of outward patent citations to distance. Our aim is to determine whether, after accounting for countries' heterogeneity in size and technology levels, distance still affects the intensity of knowledge flows between two countries. This can be done using so-called gravity equations, a very standard and widely used specification in international economics. The citation flow from country i to country n (Y_{ni}) is the product of an origin specific component, O_i, a destination specific component, D_n, and a bilateral resistance term, related to the geographical distance between the countries (d_{ni}) and to unobserved factors (η_{ni}).

$$Y_{ni} = D_n \, O_i \, {d_{ni}}^{\zeta} \, \eta_{ni} \tag{1.1}$$

This equation can be estimated through OLS or through Poisson Pseudo Maximum Likelihood (PPML). All the country-specific elements, which make a location more likely to cite or be cited, are accounted for by a set of origin and destination fixed effects. Most notably, the fixed effects account for the "knowledge stock" of a country, without imposing any assumption on the functional form of this stock, but also for the propensity to patent and the propensity to cite. Data on the geographical distance between countries comes from the CEPII GeoDist.[8] There are several ways to compute such bilateral distances. The distance between the most populated city of each country is our baseline measure of distance, but we additionally report results obtained with a "weighted distance" between the main cities of each country provided in the above-cited database in the Appendix (Figure 1.A.7). We exclude citations of patents having the same applicant or the same inventor as the citing patent ("self-citations"[9]) from our analysis. These citations could lead to a spurious distance effect because they would increase the "home bias", and do not imply proper knowledge flows.

The first exercise we conduct is to estimate the distance elasticity on bilateral citations flows aggregated from 1980 to 2010. As Table 1.1 shows, distance significantly and strongly affects citation flows between countries. The first column of the table indicates the distance elasticity estimated on the complete sample of citations using OLS, while column 2 shows the same estimation using only citations added by the applicants. Columns 3 and 4 show the

8. See Mayer and Zignago, 2011, http://www.cepii.fr/cepii/fr/bdd_modele/presentation.asp?id=6.

9. We consider an outward citation to be a "self-citation" as soon as the cited and the citing patent have at least one common applicant or inventor.

corresponding estimates using PPML regressions.

The second exercise consists in running a series of yearly cross section estimations. This permits to shed light on how the spatial decay of knowledge flows evolved over time. In order to ensure that the set of dyads used in the PPML estimation does not vary over time, we balance our database by ensuring that each potential pair of country is present at every point in time, potentially with a zero citation flow. The results of these estimations were provided in the introduction (Figure 1.1) for the PPML estimates. The distance elasticity hovers around -0.3 and is remarkably stable over time. The OLS estimates provide a similar picture (see Figure 1.A.5 in the Appendix).

The negative effect of distance on the intensity of international knowledge flows is a very robust finding. In particular, it holds when the sample is disaggregated between the three main patent offices (EPO, JPO and USPTO), and between wide technological sectors (sections of the International Patent Classification, hereafter IPC). We also estimate equation (1.1) using a different distance measure, and considering an alternative way to determine the country of each patent. In all cases, the distance elasticity of citation flows remains clearly negative (see Figures 1.A.6 and 1.A.7 in the Appendix for further explanations and results).

Table 1.1: Estimates of the distance elasticity of citation flows (ζ).

	(1)	(2)	(3)	(4)
	Cit. flow			
Distance	-0.375^a	-0.356^a	-0.297^a	-0.281^a
	(0.0349)	(0.0372)	(0.0301)	(0.0377)
Orig. and dest. FE	Yes	Yes	Yes	Yes
Estimation	OLS	OLS	PPML	PPML
Sample	All cit.	AA cit.	All cit.	AA cit.
Nb of dyads	7166	4667	36485	28863

Note: Distance elasticity estimated using equation (1.1). Distance is measured as the geodesic distance between the main city of each country. The country of each patent is determined based on its applicants. Self-citations and intranational citations are excluded. No dyadic control variables are included. Columns (1) and (2): s.e. clustered by origin and destination country. Columns (3) and (4): robust s.e. Significance levels: $^a : p < 0.01$; $^b : p < 0.05$; $^c : p < 0.1$

1.3 Micro Evidence of Networked Knowledge Search

This section aims at highlighting how the network is formed between inventors. This analysis requires information on the network of innovators. As described in section 1.2.1, we recover this network from past knowledge flows. Based on this definition of the network, we ask two questions:

1. Is an innovator more likely to cite patents of one of its contacts (than a similar patent owned by an applicant it is not linked to)?

2. Is an innovator more likely to cite a patent known by at least one of its contacts (than a similar patent unknown from its contacts)?

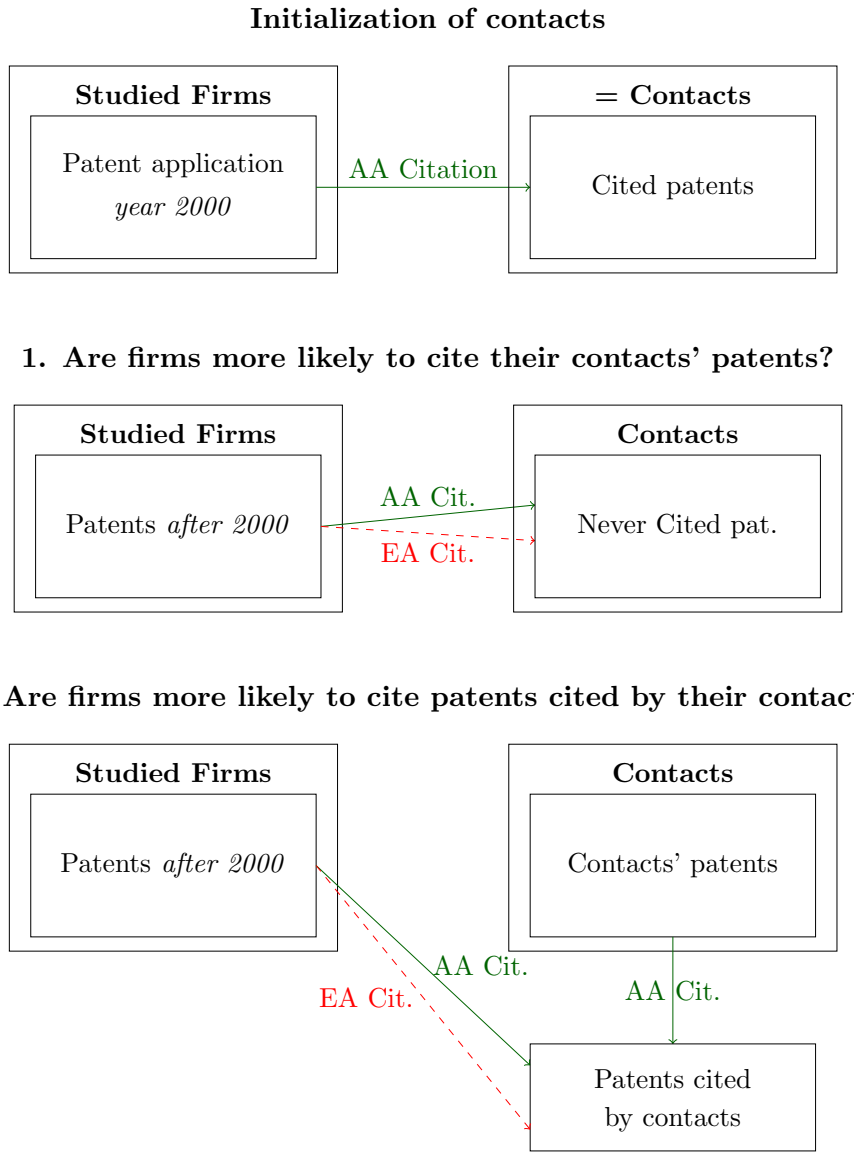
The first test aims at providing evidence of the role of networks in the circulation of knowledge, as well as asserting the quality of our proxy for contacts. The second one unveils a network formation process, by looking at the existence of triadic closure, *i.e.* links being formed between an innovator and a contact of one of its contacts.

These two tests are depicted graphically in Figure 1.2, as well as through the following example. Consider patent a, which was applied for by firm A, with priority year 2000. Among others, this patent cited another patent b_1 applied for by applicant B. Our first test assesses whether in its subsequent patent applications, firm A is more likely to cite patents of its contact B (i.e. patents $b_2, \ldots, b_n$) than similar control patents. In an earlier application, B had cited patent c owned by applicant C. Our second test investigates whether firm A is also more likely to cite patents previously cited by applicant B such as patent c.

1.3.1 Empirical strategy

Identification For each patented invention, there are millions of patents that the applicant could potentially cite, which makes it computationally unfeasible to consider the complete set of potential choices. In other words, the patents which are relevant to an applicant's invention but that she does not know about are unobserved. Therefore, we need to proxy it and restrain the set of potential alternatives to a set of patents with characteristics such that they had a high probability of being cited. To achieve this, we argue that the patents added by the patent office during the examination process constitute a credible set of potential yet non-realized citations. They are similar to the patents actually cited both in terms of observable characteristics (we explore this point below and are able to control for it), which most of the literature has focused

Figure 1.2: Design of the tests



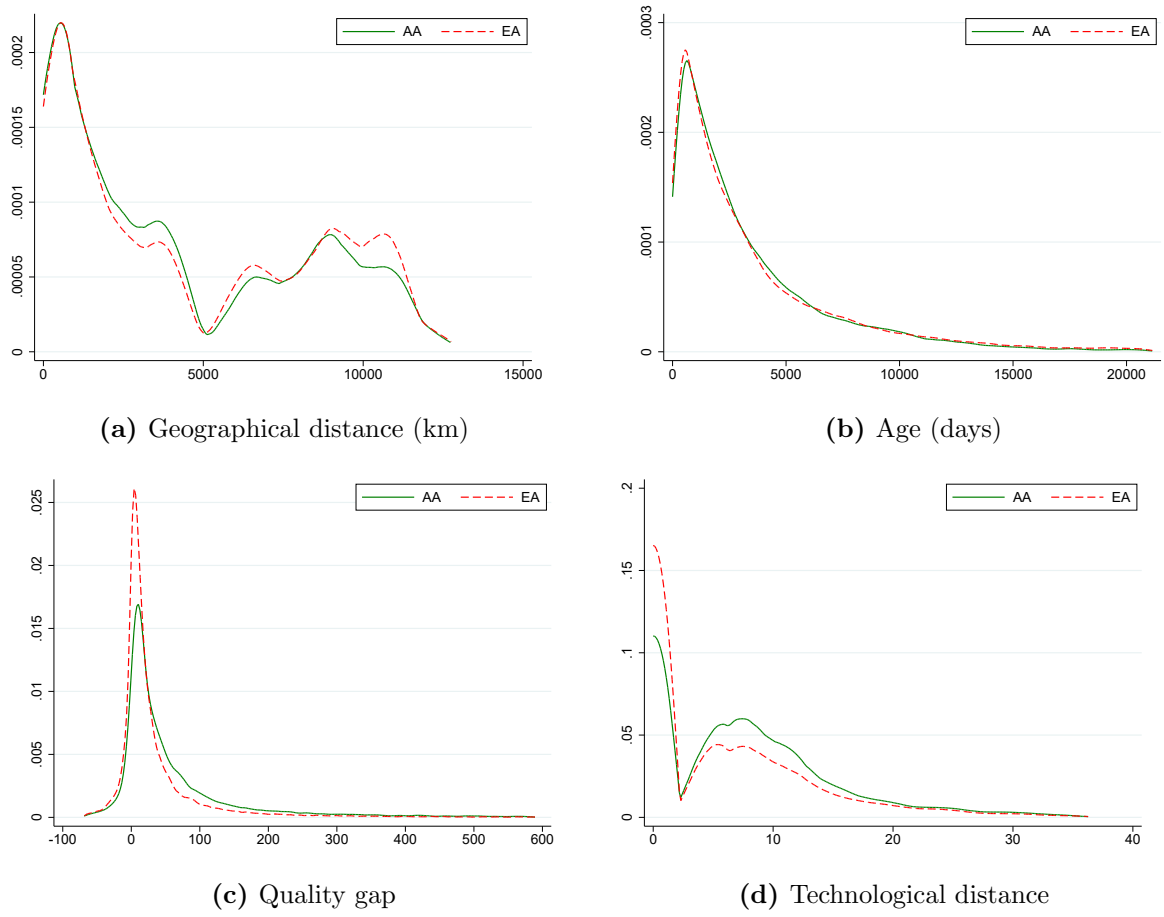
AA Citation: Citation added by the applicant; EA Citation: Citation added by an examiner. The set of studied firms is made of a randomly picked third of all firms having patented both in the initialization year and in any subsequent year.

on to build a control group, and most importantly in terms of unobservable features, in particular the relevance to the citing patent.

Our identification strategy relies on these features. Applicants will cite documents they are aware of, such as their own (self-citations are very common) and, if knowledge percolation operates, patents held by their contacts as well as patents held by contacts of their contacts. In contrast, citations added by the examiner originate from a search report posterior to the innovation process and will contain patents that are technologically relevant to the claimed invention, yet independent from the applicant's network.

In terms of observable characteristics, Figure 1.3 shows that, although not strictly equal, the empirical distributions of the four main features we observe are largely similar between examiner and applicant-added citations. Moreover, these slight differences are not a concern as long as they do not imply underlying differences in unobserved characteristics, since we control for systematic differences in quality, age, spatial or technological distance between the two groups of citations. Further detail on how these variables are constructed is provided in Appendix 1.A.

Figure 1.3: Comparison between applicant-added and examiner-added citations



(a) Geographical distance (km)

(b) Age (days)

(c) Quality gap

(d) Technological distance

(a) Upper left panel: geographical distance between the citing and the cited patent. (b) Upper right panel: age of the cited patent at the time of the citation. (c) Lower left panel: quality of the cited patent minus quality of the citing patent. (d) Lower right panel: technological distance between the citing and the cited patent. Information on the way these variables are computed is available in the Appendix. Distributions are obtained using a random sample of 0.1% of USPTO applications.

Regarding unobservable characteristics, the identifying assumptions we make are the following. We think of two key unobservable features: relevance to the citing patent, and awareness of the person making the citation. Our first assumption is that the set of patents both cited by the examiner and not cited by the applicant is similar to the set of patents cited by the applicant in

terms of unobservable characteristics. In particular, both groups are as relevant to the patented invention, and only differ on whether the applicant is aware of them or not. This assumption requires from examiners to be experts in their field, carry an extensive and independent search on relevant existing patents, and to be little influenced by past searches they may have done. To validate our approach, we match our sample with the PatEX database from USPTO's Public PAIR data, which records information about the examination process at the USPTO, notably the examiner in charge.[10] Section 1.A.2 in the Appendix shows facts supporting our assumption: on average, examiners seem to be specialized in fields, display little persistence in their behavior, and do not lose accuracy when they do cite a patent several times.[11]

Our second identifying assumption is that if a patent is not cited by the applicant, this means the applicant did not know about it. This is equivalent to assuming that applicants always have an incentive to cite any relevant patent they know, because it strengthens their application and that the examiner would find other relevant patents in any case. This is of course a simplification, and neglects the possibility for applicants to strategically withhold some citations. Lampe (2012) however shows evidence that strategic withholding is frequent, using patents already cited by applicants in the past, and cited by the examiner but not by the applicant in a subsequent application. Such phenomenon would in fact bias our estimates downwards: a citation in the past would make the cited applicant a contact, who would later receive an examiner citation but no applicant citation, going against the network effect we intend to estimate.[12]

An important point to bear in mind is that the sequentiality of the citation procedure does not threaten our identification. Indeed, we do not strictly compare applicant citations to examiner ones, but really applicant citations to their complementary set in examiner citations. That is, we compare applicant-added citations to patents cited by the examiners that have not been cited by the applicant. Therefore, our strategy is not affected by the extent of overlapping citations between the two sets, such that the influence that applicant citations may have on the decision by the examiner to cite these patents again or not is irrelevant for our purposes. We however exploit the fact that the overlap is often quite large in a robustness check, in order to test potential differences in relevance between examiner and applicant citations.

10. We match approximately 5 million USPTO applications with examiner information.

11. Moreover, as shown by Lei and Wright (2017), the fact that a thorough search has been conducted is true even for objectively weak patents, which tend to receive more attention even though they are eventually granted.

12. More generally, it is difficult to imagine a mechanism which would bias our estimates upward. It would imply for the patents originating from the applicant's network to be always relevant yet unable to limit the claims of novelty in any of the applications in the eyes of the examiner, and to be issued by firms unlikely to enter litigation. While this knife-edge alignment may occur, it seems far too restrictive to play a first-order role in our effects.

Expressing these assumptions in coherence with a discrete choice framework (the canonical McFadden et al., 1973, conditional logit model), this means that applicants face a set of N relevant patents, and are aware of k of them. Citing a patent they know costs 0 and is worth ϵ (for instance because it increases the grant probability, or because it protects it from subsequent litigation). Searching for unknown patents costs $\eta \gg \epsilon$, such that applicants always cite the k patents they know about out of N. Examiners complement the citations list with the $N - k$ remaining patents (or with any random subset of m out of the $N - k$ remaining patents both observably and unobservably equivalent to the non-cited ones).

Specification We model the citation decision of patent o towards patent d as resulting from variations of an unobserved latent variable, V_{od}, which combines both the relevance of the (potentially) cited patent d for the citing patent o, and the awareness of o for d (as in Head et al., 2018). A citation occurs as soon as the value of V_{od} exceeds a given threshold, denoted κ. In other words, defining a dummy variable C_{od} taking value 1 when patent o cites patent d:

$$P(C_{od} = 1) = P(V_{od} > \kappa)$$

The value of the latent variable depends on $\mathbf{X}_{od}$, a set of variables affecting the relevance of patent d for patent o, and on our variable of interest, L_{od}, the existence of a link between patent o and patent d's applicants:

$$V_{od} = \exp(\psi L_{od} + \boldsymbol{\beta}' \mathbf{X}_{od} + \varepsilon_{od})$$

Taking logs, the probability of o citing d writes:

$$P(C_{od} = 1) = P(-\varepsilon_{od} < \psi L_{od} + \boldsymbol{\beta}' \mathbf{X}_{od} - \ln \kappa)$$

Assuming that ε_{od} follows a logistic distribution with location parameter 0 and scale parameter 1, and denoting F the CDF of this distribution, this equation rewrites:

$$P(C_{od} = 1) = F(\psi L_{od} + \boldsymbol{\beta}' \mathbf{X}_{od} - \ln \kappa) \tag{1.2}$$

with $F(x) = (1 + e^{-x})^{-1}$, which can be estimated through maximum-likelihood. In order to neutralize any characteristics specific to the origin patent (the o specific components of $\mathbf{X}_{od}$),

we use a conditional logit estimator. Nevertheless, there are still potential confounding factors that we need to control for, as shown in Figure 1.3: the geographical distance between patent o and patent d, their technological distance, the quality of patent d, and the age of patent d at the time patent o was invented, as well as the persistence in citation behaviour.

To truly identify the effect of the network, we also control extensively for potential persistence in the citation behavior of applicants. We build a full set of dummy variables indicating in different ways whether a patent has already been cited: if patent d was cited by at least one of the assignees of o; if patent d was cited by at least one patent of one of the assignees of o before 2000 (at a time where we do not know whether the citation originates from the applicant herself or from an examiner). Similarly, we account for the fact that the assignee(s) of the cited patents may be known to the citing firm: we create a dummy turning on when at least one of the assignees of d was cited by at least one of the assignees of o, and another one indicating that at least one of the assignees of d appears on at least one patent of at least one of the assignees of o before 2000. Finally, the cited patent may already be cited by another patent of the Inpadoc family[13] of o, which is accounted for by another dummy variable.

We keep only patents applied for at the USPTO to ensure consistency of the group of potential alternatives across patents (different offices may have different behaviors in terms of examiner-added citations and have different rules for applicant-added citations).[14] Some citing patents could appear more than once in our sample, because they have several assignees belonging to the set of studied firms. We drop these duplicates and record $L_{od} = 1$ as soon as at least one of the co-assignees is linked with the destination patent. This ensures that we are left with one single observation per patent dyad (combination of citing and cited patent).

To summarize, our sample is made of the whole set of citations by our randomly selected applicants posterior to 2000 (applicant-added and examiner-added citations). Some of these citations correspond to actual knowledge transfers (the applicant-added citations), others to patents that were relevant but did not give rise to any knowledge transfer (examiner-added citations). Our dependent variable is a dummy equal to one if patent o cites patent d through an applicant-added citation, zero if d is cited only by the examiner. To test reliance on contacts' patents, we include as a regressor a dummy indicating whether an applicant of patent d is a contact of the applicant of patent o, where contacts are defined as applicants (outside of the

13. Inpadoc family is a variable provided in Patstat, which clusters patent applications referring to the same innovation, either because of renewals, resubmissions, submissions to several offices, etc.

14. Note however that, to construct the network of firms, we use patents from all patent offices, which is not an issue since we use only applicant-added citations.

1% largest) cited for the first time in year 2000. To test dependence on citations from contacts, we include a dummy indicating whether patent d had already been cited by a contact of the applicant of patent o.

1.3.2 Results

Table 1.2: Baseline results for network formation tests

	(1)	(2)	(3)	(4)	(5)
Firms		All		Small	Large
Contact	1.17^a	1.31^a	1.28^a	2.82^a	1.24^a
	(0.01)	(0.01)	(0.01)	(0.11)	(0.01)
Cited by Contact	1.73^a	2.12^a	1.67^a	1.62^a	1.68^a
	(0.02)	(0.02)	(0.02)	(0.09)	(0.02)
Age (log)			1.66^a	1.76^a	1.65^a
			(0.01)	(0.02)	(0.01)
Quality (log)			1.13^a	1.17^a	1.12^a
			(0.00)	(0.01)	(0.00)
Tech. Dist. (log)			1.02^a	1.02^c	1.02^a
			(0.00)	(0.01)	(0.00)
Geo. Dist. (log)			0.98^a	0.98^a	0.98^a
			(0.00)	(0.00)	(0.00)
Persistence Controls	✓	✓	✓	✓	✓
Orig. Pat. FE	×	✓	✓	✓	✓
Nbr of orig. applicants	7,373	7,242	6,020	3,720	2,300
Nbr of obs	6.3M	6.3M	6.1M	0.46M	5.7M

Note: Logit and conditional logit (when Orig. Pat. FE is Yes) estimations of the determinants of knowledge transfers (equation (1.2)). The sample is the set of citations of the randomly selected applicants after 2000, from and to USPTO patents. The dependent variable is a dummy equal to 1 when there is an applicant-added citation of patent d by patent o. "Contact" is a dummy equal to 1 when patent d belongs to a contact of the firm. "Cited by Contact" is a dummy equal to 1 when patent d has been cited by a contact of the firm. "Several Cit." is a dummy equal to 1 when patent d is cited several times by the origin applicant from 2000 on. Small firms are ones with less than 1000 patent applications, large firms ones with more than 1000. Coefficients are exponentiated, standard errors refer to these exponentiated coefficients (i.e. coefficients are odds ratios). Standard errors are clustered at the citing patent level in all regressions. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

Table 1.2 presents the results for our two tests of network effects conducted simultaneously, looking at the set of randomly selected firms having applied for a patent in year 2000, with coefficients expressed as odds ratios. The first column of Table 1.2 shows the result of a simple binary logit regression without controls. Column 2 displays conditional logit coefficients, which

amounts to adding fixed-effects for citing patents to the first column, while column 3 adds a set of control variables. Column 3 is our preferred specification: it controls for the fact that some citing patents have more applicant-added citations than others, as well as for any feature depending only on the citing patent: size of the applicant, etc. The coefficient associated to contacts' patents shows that applicants are approximately 30% more likely than examiners to cite patents from their contacts, which implies that applicants do rely on their network links in their citation behavior. Incidentally, this test confirms that applicant citations are a meaningful tool to proxy contacts. Focusing on links of distance 2, column 3 shows that firms are about 75% more likely than examiners to cite patents cited by their contacts. This means that we do indeed observe triadic closure in the formation of the innovators networks: applicants are disproportionately likely to form links towards contacts of contacts.[15] This property is key if one wants to link the network features to the overall effect of distance on citations.

Columns 4 and 5 split the sample respectively between small/medium sized firms (belonging to the 99 percent smallest innovators) and large firms (belonging to the 1 percent largest). It shows an interesting fact: small firms rely considerably more on existing contacts, since they are 2.75 times more likely to cite their contacts' patents than examiners. This suggests that small firms are actually much more constrained in terms of the knowledge they have access to, therefore learning about a firm which has produced a patent relevant to one of their applications makes them much more likely to rely on other inventions from that same innovator in the future. In contrast, large firms access to different sources of knowledge with less frictions, and are therefore less likely to rely on existing links. In contrast, the coefficient on patents cited by contacts has a similar magnitude across small and large firms: this means that, although the share of citations made up by contacts may decrease when firms get larger, they have the same propensity to rely on their contacts' contacts as a stepping stone to find novel sources of knowledge. Sections 1.4 and 1.5 elaborate on the differentiated roles of small and large firms, and their respective contribution to the aggregate effect of distance.

The fact that the coefficient associated to contact patents is lower than the one for patents cited by contacts in our preferred specification is driven by several factors. First, the coefficient on the former is strongly reduced by the fact the various controls we introduce for persistence, notably for repeated citations to the patent on which the link has been initialized. It is therefore

15. Interestingly, Carayol et al. (2018) find a negative effect on the probability of triadic closure in co-invention links and argue it is due to choices of avoiding redundant connections. Their paper is focused on collaborations rather than knowledge diffusion, it contradicts in no way the above result but rather complements it.

a conservative estimate of the effect of links of distance 1 on citations. Second, as columns 4 and 5 of the above table show, it hides a considerable amount of heterogeneity: large firms are much less likely to rely on contacts, but constitute the largest share of patents and citations, hence driving the coefficient down. Firm-level estimates shown in Table 1.A.4 in Appendix 1.C confirm that the coefficient on contacts is much larger than the one on contacts of contacts when the unit of observation is a firm rather than a patent. Moreover, the contact variable is defined at the applicant level, which largely dilutes the effect compared to the cited by contact variable, which is defined at the patent level: when defined at the applicant level, the second coefficient becomes smaller than the first one. Last, coefficients are odds ratio, and only give an idea of relative importance across both groups, even though citations to contacts are much more frequent than ones toward patents cited by contacts.

1.3.3 Robustness

Alternative Strategy An alternative strategy can be pursued to conduct similar tests comparing applicant-added citations to examiner-added ones, using the examiner-added citations to build false links rather than as counterfactual citations. To test reliance on contacts' patents, one may compare the probability that this group cites patents developed by applicants truly cited in 2000 (actual contacts) relative to applicants cited by examiners in 2000 (control group of contacts). Similarly, to test for indirect links, rather than assessing whether the group of interest is more likely to cite patents previously cited by contacts than its examiners, one may assess whether this group is more likely to cite patents actually cited by its contacts than patents cited by its contacts' examiners (*i.e.* examiners for its contacts' applications). This implicitly assumes that if a patent from a given applicant was relevant once to a a firm's citing patent, then other patents of the former applicant should be relevant in future citing patents.

As shown in Appendix 1.C, although coefficients are not comparable with the baseline ones (mostly because we cannot control for characteristics of the origin patent), results largely confirm the effect of the network. They show that contacts' patents are more likely to get recited than their comparison group. Similarly, patents cited by contacts get recited more than patents cited by examiners on contacts' applications.

Overlapping citations A potential threat to the identification we propose could be that patents cited by applicants and by examiners have systematically different levels of relevance.

For instance, it could be that while examiner-added citations are indeed relevant, applicant-added citations may be somewhat fictitious references. This may be particularly problematic if firms cite patents made by their contacts or cited by their contacts not because their discoveries are based on them, but only to avoid making a thorough search to find the accurate references. Yet, because it happens frequently that examiners cite a patent which was already in the list of applicant citations, it is possible to conduct the exact same test only on the examiner-added citations only, which means that our dependent variable will take the value 1 only when a cited patent belongs to the overlapping set of examiner and applicant added citations. The underlying assumption is that, contrary to our baseline strategy in which all applicant patents are considered relevant, only patents eventually cited by examiners are actually relevant.

Table 1.A.5 displayed in Appendix 1.C shows results similar to the baseline but defining our dependent variable as being both an examiner and an applicant citation, and dropping all patents which do not contain such citation. It shows that the coefficients on our variables of interest are very similar to the ones we have in the baseline regression, which alleviates the potential concern that our coefficients of interest would be upward biased if applicant citations were less relevant to the patented invention than examiner citations.

Group level results A critical point in the interpretation of our results is the extent to which assignees are correctly identified, in order to fully remove self-citations. Moreover, if firms have subsidiaries, this may mean that citations occurring between a parent company and its subsidiaries should be included in our analysis. This is a matter of concern, since links within the multinational firm have been found to be important for knowledge flows (Keller and Yeaple, 2013; Bilir and Morales, 2016), and that we want our mechanism to be valid beyond the borders of MNEs. Appendix 1.C provides more detail on how we recover information on groups. Results once group linkages are accounted for are very similar to the baseline ones.

Other robustness checks We conduct a wider range of robustness checks: we change the initialization year, the maximum size of contacts, measure our effects at the firm level. We also run Placebo regressions initializing contacts with examiner citations.

1.3.4 Spatial search of knowledge

An alternative way of approaching our test can be to mix our study of network formation through citations with the more traditional method used to emphasize local knowledge spillovers,

stemming from Jaffe et al. (1993). Indeed, firms could be practising spatial search for knowledge parallel to network search, looking for new relevant patents with a spatial bias from their existing knowledge stock. This spatial search could include various mechanisms: firms could for instance have language or cultural biases in contact formation, be more likely to go to tech fairs and shows where their contacts are located, get biased research results from search engines based on their past researches, follow some specialization operating within clusters. This means that a firm may have higher chances to form links with geographical neighbours of its contacts, without its contacts being linked to these geographical neighbours.

In such setting, a first test is similar to the initial idea of Jaffe et al. (1993), and the way of conducting it is conceptually equivalent to that of Alcácer and Gittelman (2006) and Thompson (2006). It tests whether applicants are more likely to cite patents developed geographically close to them than office examiners. A point of enquiry is whether this mechanism is different from the mechanism we test above, or whether one dominates the other when both are introduced jointly in a regression. Moreover, the approach can be followed one step further, testing if citations to geographical neighbors of contacts are more likely. Indeed, while having a purely spatial approach only allows to proxy links of distance 1 (being close to A who is close to B means one is also close to B), looking at applicants close to contacts formed through citations gives a proxy for links of distance 2. This test therefore proposes an alternative and more flexible proxy of link formation than the one used above based purely on patent citations.

For each firm of the random sample, we select the set of neighbor firms as the firms whose main location is less than 5km away.[16] We then define a dummy variable "Close to Origin" taking value one if the cited patent was made by a neighbor firm of the citing patent's assignee. Similarly, we create a dummy variable "Close to Contact" turning on when the assignee of the cited patent is close to a contact of the citing firm(s).

Table 1.3 displays coefficients associated with the variables indicating if applicants are geographically close to our set of firms, or close to their contacts. It shows that firms are indeed 12% more likely than examiners to cite patents from assignees located less than 5 km from them, which replicates the results of Alcácer and Gittelman (2006) and Thompson (2006). As column 3 shows, applicants also seem very slightly more likely to cite patents from assignees located close to their contacts. This effect however seems much weaker than the ones found on the search along the network, and disappears completely in column 5 when confronted with our

16. The main location being defined as the mode of the locations where patents have been registered for this firm.

Table 1.3: Results of the spatial test

	(1)	(2)	(3)	(4)	(5)
Close To Origin	1.054^a	1.127^a	1.119^a	1.112^a	1.118^a
	[0.023]	[0.027]	[0.027]	[0.027]	[0.027]
Close To Contact			1.033^a		0.972^a
			[0.011]		[0.010]
Contact				1.566^a	1.533^a
				[0.038]	[0.036]
Cited By Contact					1.152^a
					[0.039]
Age (log)		1.778^a	1.780^a	1.770^a	1.723^a
		[0.025]	[0.025]	[0.025]	[0.023]
Quality (log)		1.069^a	1.069^a	1.062^a	1.047^a
		[0.004]	[0.004]	[0.004]	[0.004]
Tech. Dist (log)		1.039^a	1.039^a	1.040^a	1.043^a
		[0.005]	[0.005]	[0.005]	[0.005]
Orig. Patent FE	✓	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5,767	5,512	5,512	5,512	5,512
Nbr of obs	6.6M	5.3M	5.3M	5.3M	5.3M

Note: Conditional logit estimations of the determinants of knowledge transfers (equation (1.2)). The sample is the set of citations of the randomly selected applicants after 2000, from and to USPTO patents. The dependent variable is a dummy equal to 1 when there is an applicant-added citation from patent o to patent d. "Contact" is a dummy equal to 1 when patent d belongs to a contact of the firm. "Close to Origin" indicates that patent d belongs to an applicant located less than 5 kilometers away from the origin applicant, "Cited by Contact" that patent d has been cited by a contact of the firm, and "Close To Contact" that the applicant of patent d is close to a contact of the citing applicant. Coefficients are exponentiated (i.e. odds ratios). In brackets are the standard errors, clustered at the citing patent level. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

covariates for network effects.

1.4 Theory: Network Origins of the Distance Effect

This section develops a dynamic model able to bridge our finding that knowledge percolates through a network of innovators from section 1.3, with the fact that distance hinders aggregate knowledge flows shown in section 1.2.

As established by Chaney (2018), citation flows will exhibit a negative distance elasticity as long as the following two conditions hold:

- **Condition 1:** Firm sizes follow a Pareto distribution of shape parameter λ with $\lambda > 1$.
- **Condition 2:** An increasing power function of parameter μ links the average squared distance of a firm's citations to its size.

Under these two sufficient conditions,[17] knowledge flows are negatively related to distance. These two conditions to generate an effect of distance are very intuitive. The Pareto distribution for firm sizes means that small firms are far more numerous than big firms, and the constant elasticity μ of the average squared distance of a firm's citations with respect to its size means that larger firms are able to access to more distant knowledge. If there were only large firms, no spatial effect would come out on knowledge flows intensity. However, the vast majority of firms are small and primarily cite innovators located close to them, which generates the aggregate observation that citation flows decay with distance.

Chaney (2018) also builds a dynamic network formation model which endogenizes the two above-mentioned conditions. We extend the model to adapt it to the context of knowledge. The mechanics are as follow: agents can access knowledge through their contacts, and start off with initial contacts distributed close to them. Firms then gain some new contacts as time passes, which are either the contacts of their own contacts (network search), or agents located close to their contacts (spatial search).

The idea that young and small firms initially start with localized contacts (as assumed in the model developed below) has received some empirical support: Almeida and Kogut (1997) looked at innovators in the semi-conductor industry in the US, and found that small firms were more prone to cite patents developed closer to them than big firms were. Their interpretation of the result was that start-ups were more strongly tied to local knowledge networks than big

17. Additional conditions detailed in Chaney (2018) to obtain an asymptotically constant distance elasticity are either that $\lambda < 1+\mu$ or that the PDF of citation distances of the smallest possible firm admits a finite $\left(1+2\frac{\lambda-1}{\mu}\right)$-th moment. More precisely, for distances going to infinity, these ensure that ζ tends towards $\left(1+2\frac{\lambda-1}{\mu}\right)$.

firms, which is exactly the logic at work in the model.

Model We extend the model featured in Chaney (2018) to introduce some "spatial search" additionally to the "network search", consistently with the empirical evidence provided in section 1.3. The model is the following. Time is continuous, and infinitely-lived firms are born with a growth rate γ. Space is infinite and one-dimensional ($\mathbb{R}$), so that coordinates of any location are a scalar x. When they are born, firms are endowed with a set of contacts of mass K_0, born at the same time, and distributed around them according to the distribution $k_0(x)$, which is assumed to be symmetric and to admit a finite second-moment. Each contact provides a firm with one unit of knowledge. The set of contacts of a firm of age a evolves in three ways:

- Gain via spatial search: the firm can directly find new contacts, through a random Poisson shock of parameter ρ, in each location where it already has contacts. This means that, going from age a to age $a + \mathrm{d}a$, the firm has chances to get some new contacts with the exact same spatial distribution as the contacts it already has.
- Gain via network search: a firm's existing contact may reveal one of its own contacts through a random Poisson shock of parameter β. This revealed contact in turn joins the set of the firm's contacts. A technical constraint requires that firms can only gain contacts with firms of their age.
- Loss of a contact, also through a Poisson shock of parameter δ.

Based on these three channels, the evolution of k_a, the mass of contacts at point x of an aged a firm writes:

$$\frac{\partial k_a(x)}{\partial a} = \underbrace{\rho k_a(x)}_{\text{spatial search}} \underbrace{+\beta \int_{\mathbb{R}} \frac{k_a(x-y)}{K_a} k_a(y)\mathrm{d}y}_{\text{network search}} - \underbrace{\delta k_a(x)}_{\text{contact loss}} \tag{1.3}$$

At the same time, the evolution of the overall number of contacts of a firm of age a, K_a, follows the simple ODE:

$$\frac{\partial K_a}{\partial a} = (\rho + \beta - \delta) K_a$$

with initial value K_0.

Proposition. *Shape parameters λ and μ from conditions 1 and 2 are functions of the Poisson shock parameters:*

- *The distribution of firm sizes is Pareto, with a shape parameter $\lambda = \frac{\gamma}{\rho+\beta-\delta}$;*

- *The average squared distance at which firms cite is a power function of their number of contacts, of parameter* $\mu = \frac{\beta}{\rho+\beta-\delta}$.

Proof. See Appendix 1.E. □

A wrap-up of this mechanical model is that firms will gradually be less and less affected by distance as their grow old: their average contact is further and further away. In aggregate however, because new firms are born every period with a constant growth rate, this model will imply both a Pareto size distribution and distance from contacts being an increasing function of size. In other words, if large firms cite on average further away than small firms, then citations at long distances mostly come from large firms (firms applying for many patents). This means that the greater the number of large firms compared to small ones (smaller λ), and the quicker the distance at which firms cite increases with size (larger μ), the lower the negative impact of distance on patent citations is in aggregate.

Comparative Statics Partial derivatives of the parameters of interest with respect to ρ are as follows:

$$\frac{\partial \lambda}{\partial \rho} = \frac{-\gamma}{(\rho+\beta-\delta)^2} < 0$$
$$\frac{\partial \mu}{\partial \rho} = \frac{-\beta}{(\rho+\beta-\delta)^2} < 0$$

This means that, when spatial search increases, this generates a decrease in λ, i.e. an increase in the proportion of large firms relative to small ones. It also generates a decrease in μ, implying that the difference between the distance at which big and small firms cite drops.

Similarly, partial derivatives of the parameters of interest with respect to β are:

$$\frac{\partial \lambda}{\partial \beta} = \frac{-\gamma}{(\rho+\beta-\delta)^2} < 0$$
$$\frac{\partial \mu}{\partial \beta} = \frac{\rho-\delta}{(\rho+\beta-\delta)^2} \lessgtr 0$$

Thus, the effect on the distribution of firms sizes of an increase in network search is exactly equivalent to the magnitude of the effect of an increase in spatial search: it makes the tail of the size distribution thicker, by increasing the rate at which firms get new contacts while leaving unchanged the entry rate of newborn firms. The sign of the effect of a change in β on μ is, however, undetermined, and depends on the relative importance of spatial search versus

contact destruction. If the latter does less than compensate spatial search, then the last partial derivative is positive and an increase in network search would have the consequence of increasing μ, i.e. increasing the sensitivity of citation distance to firm size. If, however, δ is large enough in front of ρ such that gains of contacts through spatial search do not compensate contact losses, then an increase in network search would negatively affect μ, similarly to an increase in spatial search. In other words, in a context in which spatial search is too weak to compensate contact loss, an increase in network search will make the distance of firms citations less dependent on the firm sizes, while the opposite effect will occur if spatial search is strong enough to compensate contact loss.

1.5 Estimation of Aggregate Predictions

The network formation model presented in the previous section (section 1.4) provides sufficient conditions for a negative elasticity of knowledge flows with respect to distance to emerge. These predictions can be directly tested in the data. We find that they hold well empirically, which gives credit to the idea that the network formation mechanism that we described underlies the spatial decay of knowledge flows.

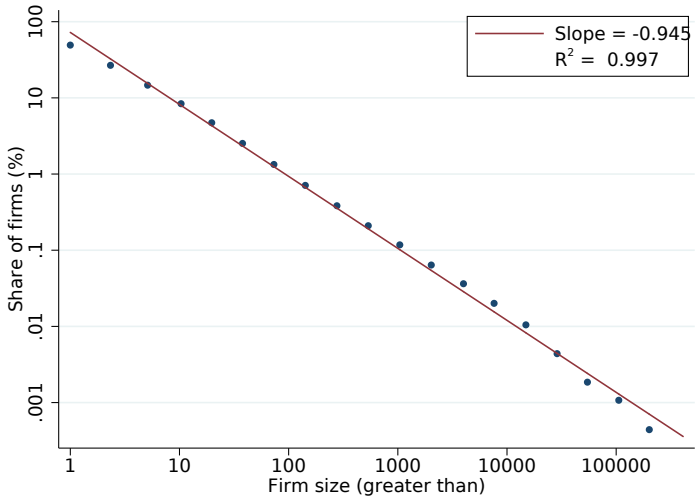
1.5.1 Pareto distribution of innovator sizes

The network formation model predicts that the distribution of firm sizes will be Pareto, i.e. that $F(K) = 1 - \left(\frac{K}{K_0}\right)^{\lambda}$, with $\lambda = \frac{\gamma}{\rho+\beta-\delta}$. We therefore check that a Pareto distribution fits our data well, and we estimate the shape parameter of this distribution, using the method introduced by Axtell (2001). We rank innovators by increasing order of size, where size is the number of patent applications in a given year[18] and distribute them in 20 bins of equal log width.[19] We compute the average size of firms in each bin, denoted K_b, and the fraction of firms of size larger than

18. Measuring size in our context is not obvious: although the closest to the model would be to assess it through the number of outward citations, the information on citations is missing for many patents, the number of citations should be very correlated with the number of patents, is very dependent on the office rules, and overall yields very noisy results. Thus, our aim is to find the best measure of firm size: we use a patent count, where all applications are included, and not only the first one, to weight "quality" in (Lanjouw et al., 1998).

19. There seems to be no established consensus regarding the number of bins that should be used: while Axtell (2001) uses 30 bins, Chaney (2018) uses 50. The specificity of patent data is that most firms in our sample have 1 or 2 patent applications in a given year; thus, if we use too many bins, some bins are actually empty because firms having 1 patent application fill more than one bin. This leads to some bins being dropped, and may lead to λ and μ being estimated with a different number of bins.

Figure 1.4: Estimation of λ.



Note: Each dot corresponds to one of the 20 bins. The x-axis gives the average size of firms in the bin (K_b) and the y-axis the share of firms that are larger than this size ($1 - F(K_b)$). Firm size is measured as the number of patent applications of the firm over the period 1980-2010.

K_b, denoted $1 - F(K_b)$. λ is estimated from:

$$\ln[1 - F(K_b)] = a - \lambda \ln(K_b) + \varepsilon_b \tag{1.4}$$

The slope of the regression line shown in Figure 1.4 corresponds to our baseline estimate of λ. The Pareto distribution fits very well our firm size data (considering the very high R-squared). Table 1.A.10 in the Appendix confirms this finding. Moreover, it shows that the obtained shape parameter is always very close to 1, whether we consider all applications or restrict the sample to first applications, left-truncate the distribution by keeping only firms larger than a certain threshold (5 patents), or measure size with the number of outward citations instead of applications. This makes the measured distribution enter the specific case of a Zipf law, a Pareto distribution with shape parameter equal to 1. In the model, this implies that the net growth of the mass of contacts should equate the growth rate of the firm population.

The economic literature has uncovered a wide class of objects following a power law, which are as diverse as city sizes, firm sizes, income distribution, the number of trades per day (Gabaix, 2016), or closer to our object of study the productivity of innovations (Ghiglino, 2012). We add the size distribution of patenting firms to this class. From the empirical standpoint, the distribution of firm sizes in general has been shown to follow a Zipf law by Axtell (2001).

Moreover, while the assumption that productivities are Pareto distributed is common in the trade literature, Nigai (2017) has shown that the left-hand side of the distribution of productivities is closer to log-normal while the right-hand side fits the Pareto distribution better. In our context, if productivity is equated to the number of innovations, it is sufficient to posit that only firms above a certain productivity threshold are able to innovate, which means left-truncating the productivity distribution, to obtain an innovator sizes distribution well described by a Zipf law. From the theoretical standpoint, it is also common to generate power laws either from random growth in size as in Gibrat (1931), or from scale-free network formation processes as in the Albert and Barabási (2002) model, which also features growth in the number of nodes, and link formation through preferential attachment (which takes the form of network search in the model we use, growth alone being insufficient to generate a scale-free network).

If we disaggregate our sample and run distinct regressions for each year, the Pareto distribution still provides an excellent fit. Figure 1.6a shows the estimated coefficient of the Pareto distribution for each year from 1980 to 2010. Note that Zipf law cannot be rejected for almost all of the period we study. Firm sizes are also found to be Pareto distributed when measured using only the patents from one single office, be it the EPO, the JPO or the USPTO. Finally, within a given technological field (IPC section), the distribution is also Pareto, but the shape parameters exhibit some mild differences across sectors.

1.5.2 Distance of citations as an increasing function of innovator size

The network model generates a second, more specific feature. It predicts that larger firms are able to access knowledge generated further away than smaller firms. More precisely, there is a positive constant elasticity of the average squared distance at which firms cite with respect to their size. To test whether this holds in our data, we rank firms in increasing order of size,[20] and construct 20 bins of equal log width. We compute the average size of firms in each bin K_b and the average squared distance,[21] denoted Δ_b, at which firms in bin b cite. μ is estimated from:

$$\ln[\Delta_b] = a + \mu \ln(K_b) + \varepsilon_b \tag{1.5}$$

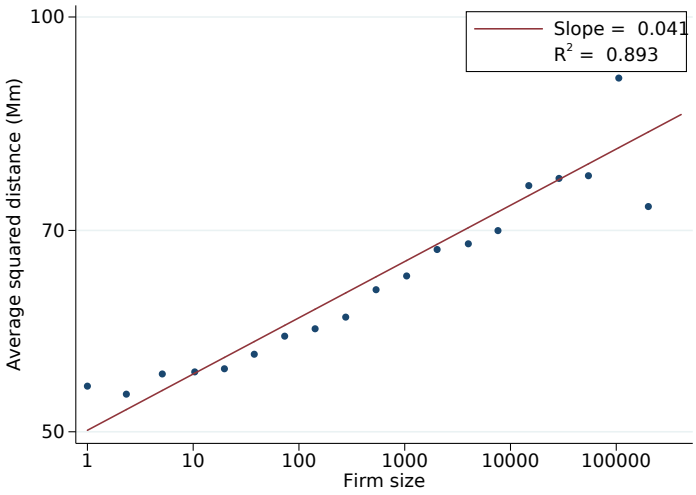
Figure 1.5 shows that the relationship between the average squared distance at which a firm

20. Size is again defined as the number of applications of the firm

21. In our baseline estimations, the distance of a citation is defined as the distance between the largest city of the country of each applicant, and intranational citations are excluded, but we show that our results hold for alternative geographic choices.

cites and its size is well described by an increasing power function (*i.e.* increasing linear in logs). To the best of our knowledge, this systematic relationship between a firm size and its ability to access more distant ideas is a novel finding in the analysis of patent citations. Table 1.A.11 shows that this positive relationship between firm size and distance of citations holds using different measures of citations, firm size, and including or not a minimal size threshold.

Figure 1.5: Estimation of μ



Note: Each dot corresponds to one of the 20 bins. The x-axis gives the average size of firms in the bin (K_b) and the y-axis the average squared geographical distance at which firms in the bin cite (Δ_b), in millions kilometers. The period we study is 1980-2010. Firm size is measured as the number of patent applications of the firm over that period. Distance is measured as the distance between the largest city of the countries of the citing and cited patents. Intranational citations and self-citations are excluded.

The positive link between firm size and their ability to access more distant knowledge does not disappear when we disaggregate the sample by patent office (see Fig. 1.A.9 in Appendix 1.D). When running separate regressions for each IPC section, μ is always found to be positive, and significant for 7 sections out of 8 (see again Fig. 1.A.9). Additionally, the estimated μ is robust to alternative ways of measuring distance (notably to switching to city to city distances instead of country distances between their capital cities). Note that the elasticity with respect to firm size of the average squared distance for citations is around half the one of exports. In other words, the ability to create links with more distant firms is less sensitive to firm size for ideas than for exports.

A complementary exercise which we conduct is to estimate μ within firms over time, meaning that we include firm fixed effects in a firm $\times$ year level regression, such that μ is estimated

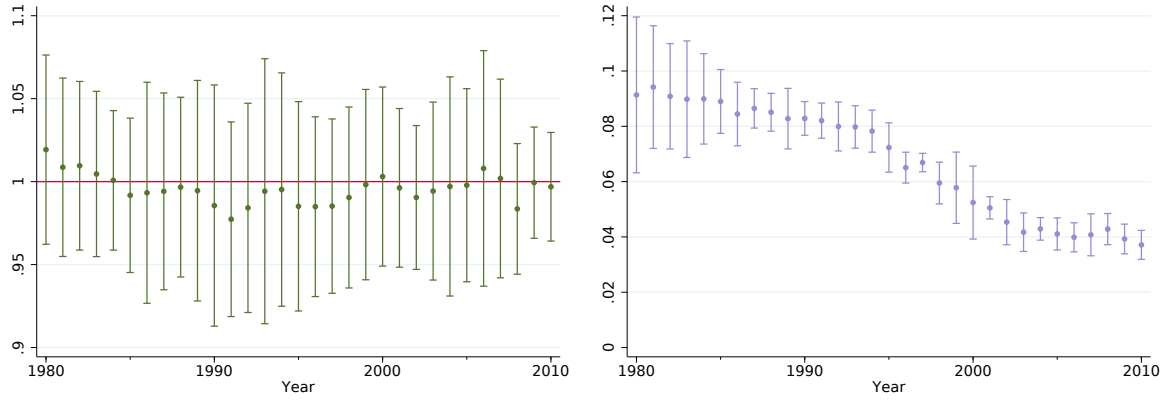
from time variations in firm size and distance of citations. Using this specification, we also find a positive significant relationship between these two variables, as shown in Table 1.A.12 in Appendix 1.D. The estimated coefficients are naturally larger, simply because they weight observations in very different ways (they give an equal weight to each firm rather than equal weights to each size bin). This holds whether we consider all citations (columns 1, 2 and 3) or restrict our attention to applicant-added citations (column 4), and is not driven by a time trend, since the inclusion of year fixed effects does not change the estimated μ. This means that firms getting bigger also tend to cite further away, while firms shrinking would tend to cite closer. Consistently with the model, we also find a positive link between the age of a firm and its size. We measure age as the number of years since the first appearance of the firm in our database, and we regress the log of the firm size on its age, with a set of firm fixed effects. We find a semi-elasticity around 0.7 (see Table 1.A.13 in Appendix 1.D). This result is robust to the introduction of year fixed effects correcting for the fact that firms tend to be older in more recent years, and have also recorded more patents, and to the restriction of the sample to firms that appear at least 10 distinct years in our database.

Taken together, these two findings are consistent with the dynamics of the model: as firms grow older, they become larger and are able to build links with more distant firms. Interestingly, the economy described here shares similar features with ones emanating from the Schumpeterian growth theory (Aghion et al., 2015): the size distribution of firms, where size is assimilated to the number of their innovations, is highly skewed, and larger firms are older.

1.5.3 Time variations and their implications

The exercise of estimating λ and μ may also be useful in order to understand the changes undergone during our period, in an attempt to explain why these have not implied a decrease in the aggregate distance effect. The context of innovation does not meet the conditions expressed by Chaney (2018) to give a closed firm expression of the elasticity of flows with respect to distance as a function of λ and μ (notably, λ is not always above 1 as required). However, changes in one parameter keeping the other constant can be interpreted in terms of changes in the resulting ζ. For instance, decreasing μ all other things equal (including the initial distance of contacts k_0) means that the link between distance of citations and size is less stark, therefore large firms do not cite at much larger distances than small firms do, and the elasticity of flows with respect to distance should increase.

Figure 1.6: Estimates of λ and μ over time.



(a) Shape parameter of the Pareto distribution of firm sizes (λ)

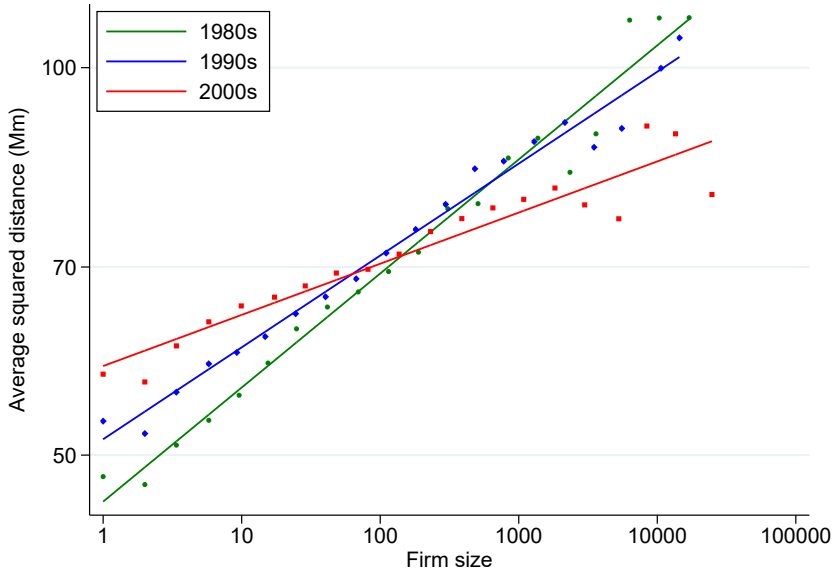
(b) Elasticity of the average squared distance of citations with respect to firm size (μ)

λ and μ are estimated from a series of cross-sectional regressions (respectively of equation (1.4) and (1.5)), one for each year. All patents are included in the sample. Firm size is measured as the number of patent applications of the firm during the year. The distance is the geographical distance between the largest city of the countries of the citing and the cited patent.

Figure 1.6a shows that the shape parameter of the firm sizes distribution has remained quite close to 1. Yet, point estimates seem to have increased slightly, going from values clearly below 1 to values close or above 1. This implies a relative increase in the number of small innovators relative to large ones. Figure 1.6b shows that the strength of the link between firm size and distance of citations has varied a lot over time. μ has strongly decreased over the years: while the elasticity was around 0.1 in the 80s, it hovers around 0.04 in the 2000s, with a strong decline occurring during the 90s. This means that the distance at which firms cite has become less and less sensitive to size.

This drop in the value of μ could be an effect of ICTs: while small firms were very constrained to get new knowledge in the 1980s, they can now find a share of the new knowledge they need through internet searches, which makes distance of citations less sensitive to size. In such a case, this would be associated with a structural change in the spatial distribution of newborn firms: the k_0 in the model would then have higher variance in later years. The alternative explanation of such a decrease in the magnitude of μ would be that big firms are now less able to access to distant knowledge, which seems hard to rationalize unless the geography of innovations has changed. Figure 1.7 argues in favor of the former: on the left-hand side of the graph, the distance at which the small firms cite has increased in the years 2000s compared to the 1980s and 1990s (while the right-hand side is estimated with some noise and difficult to interpret with certainty).

Figure 1.7: Estimates of μ by decade



Firm size is measured as the number of patent applications of the firm during the decade. The distance is the geographical distance between the largest city of the countries of the citing and the cited patent.

Explaining the stability of the aggregate distance coefficient displayed in Figure 1.1 then requires an increase in the share of small firms, i.e. an increase in the shape parameter of the Pareto distribution λ. This seems to be verified in Figure 1.6a, although standard errors are too large to reject equality of these parameters over time. The explanation for the fact that the gravity coefficients on knowledge flows have remained stable over the period would then be hiding two joint changes: the fact that small firms are less affected by distance than they used to, offset by the fact that small firms have grown relatively more numerous.

1.6 Conclusion

This paper shows that the negative effect attributed to distance on international knowledge flows can in fact be explained, at least partly, by the spatially biased network formation between innovators.

First, we empirically test the way links are formed between innovators using patent citations. Using previous citations to build contacts, we show that a firm is more likely to cite a patent originating from, or cited by at least one of its contacts than a similar patent from outside the its close network. We exploit the fact that some citations are added by applicants while others are added by the office examiners, the union of which provides us with a group of counterfactual

citations under frictionless knowledge circulation. We estimate the effect of a direct or indirect link on the likelihood of being cited by the applicant itself (versus the likelihood of being cited by the examiner). We find that firms are 1.3 times more likely than examiners to cite patents owned by their contacts, yet hiding very strong heterogeneity between small and large firms. Moreover, firms are 1.7 times as likely to cite patents that were cited directly by their contacts. This first effect is highly heterogeneous, since firms which do not belong to the top 1% of the size distribution are almost three times as likely to cite their contacts' patents. These effects are robust to a wide range of checks.

Based on this finding, we slightly extend and use the network formation model developed by Chaney (2018) to draw aggregate implications from the above phenomenon. In this model, the initial spatial clustering of an innovators' contacts tends to vanish over time since innovators gain new, more distant contacts through network search, *i.e.* through a contact's contacts, and spatial search, *i.e.* gain of new contacts that are geographically close to an existing contact. Nevertheless, the continuous arrival of new firms, which are not able to access distant knowledge, maintains an aggregate negative relationship between distance and knowledge flows.

We then show that the theoretical aggregate predictions of the model hold remarkably well in the data. The sizes of innovators - in terms of the number of patent applications - are Pareto-distributed (and even Zipf distributed), and the average squared distance at which innovators cite is an increasing power function of their size. Moreover, we find the latter relationship to hold both across firms and within firms over their lifetime. The Zipf distribution of firm sizes, as well as the systematic increasing relationship between size and distance at which firms are able to cite, are novel findings. They allow generating easily a negative effect of distance on aggregate citation flows: if small firms are far more numerous than big ones and if they cite relatively closer, the intensity of flows will naturally depend negatively from distance.

Interestingly, the network formation mechanism put forward is general enough to encompass many of the usual explanations of the localization of knowledge spillovers: it is consistent with formal R&D collaboration agreements and the natural network they generate, but also with explanations based on cultural proximity and common ethnicity (Agrawal et al., 2008; Kerr, 2008), with linkages with geographical neighbours (*e.g.* inside clusters), as well as inter-firm mobility of engineers (Almeida and Kogut, 1999; Breschi and Lissoni, 2009; Serafinelli, 2018), and input-output linkages (Carvalho and Voigtländer, 2014).

References

Aghion, P., U. Akcigit, and P. Howitt. 2015. "Lessons from Schumpeterian growth theory." *American Economic Review* 105 (5): 94–99.

Aghion, P., and X. Jaravel. 2015. "Knowledge Spillovers, Innovation and Growth." *Economic Journal,* no. 583: 533–573.

Agrawal, A., D. Kapur, and J. McHale. 2008. "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics* 64, no. 2 (September): 258–269.

Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva, and V. Sterzi. 2018. *Dancing with the stars: Innovation through interactions.* Technical report. National Bureau of Economic Research.

Akers, N. J. 2000. "The referencing of prior art documents in European patents and applications." *World Patent Information* 22 (4): 309–315.

Albert, R., and A.-L. Barabási. 2002. "Statistical mechanics of complex networks." *Reviews of Modern Physics* 74 (January): 47–97.

Alcácer, J., and M. Gittelman. 2006. "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *The Review of Economics and Statistics* 88 (4): 774–779.

Almeida, P., and B. Kogut. 1997. "The Exploration of Technological Diversity and the Geographic Localization of Innovation." *Small Business Economics* 9 (1): 21–31.

———. 1999. "Localization of knowledge and the mobility of engineers in regional networks." *Management science* 45 (7): 905–917.

Axtell, R. L. 2001. "Zipf distribution of US firm sizes." *Science* 293 (5536): 1818.

Bilir, L. K., and E. Morales. 2016. "Innovation in the Global Firm." *Journal of Political Economy:* Forthcoming.

Breschi, S., and F. Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of Economic Geography:* lbp008.

Carayol, N., L. Bergé, L. Cassi, and P. Roux. 2018. "Unintended triadic closure in social networks: The strategic formation of research collaborations between French inventors." *Journal of Economic Behavior and Organization:* forthcoming.

Carvalho, V. M., and N. Voigtländer. 2014. *Input diffusion and the evolution of production networks.* Technical report. National Bureau of Economic Research.

Catalini, C., C. Fons-Rosen, and P. Gaulé. 2018. *How Do Travel Costs Shape Collaboration?* Working Paper, Working Paper Series 24780. National Bureau of Economic Research, June.

Chaney, T. 2014. "The network structure of international trade." *The American Economic Review* 104 (11): 3600–3634.

———. 2018. "The Gravity Equation in International Trade: An Explanation." *Journal of Political Economy* 126 (1): 150–177.

Comin, D., M. Dmitriev, and E. Rossi-Hansberg. 2012. *The Spatial Diffusion of Technology.* NBER Working Papers 18534. National Bureau of Economic Research, Inc.

Comin, D., and B. Hobijn. 2010. "An Exploration of Technology Diffusion." *American Economic Review* 100 (5): 2031–59.

Duguet, E., and M. MacGarvie. 2005. "How well do patent citations measure flows of technology? Evidence from French innovation surveys." *Economics of Innovation and New Technology* 14 (5): 375–393.

EPO. 2016. *Guidelines for Examination, Part B (Guidelines for Search).* Technical report. European Patent Office.

Gabaix, X. 2016. "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30, no. 1 (February): 185–206.

Ghiglino, C. 2012. "Random walk to innovation: Why productivity follows a power law." *Journal of Economic Theory* 147 (2): 713–737.

Gibrat, R. 1931. *Les inégalités économiques: applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc: d'une loi nouvelle: la loi de l'effet proportionnel.* Librairie du Recueil Sirey.

Griffith, R., S. Lee, and J. Van Reenen. 2011. "Is distance dying at last? Falling home bias in fixed-effects models of patent citations." *Quantitative economics* 2 (2): 211–249.

Head, K., Y. A. Li, and A. Minondo. 2018. "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics." *Review of Economics and Statistics,* no. forthcoming.

Head, K., and T. Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." Chap. 3 in *the Handbook of International Economics,* 4:131–195. Elsevier.

Iaria, A., C. Schwarz, and F. Waldinger. 2018. "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science." *The Quarterly Journal of Economics* 133 (2): 927–991.

Jackson, M., and B. Rogers. 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review* 97 (3): 890–915.

Jaffe, A. B., M. Trajtenberg, and M. S. Fogarty. 2000. "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." *American Economic Review, Papers and Proceedings* 90, no. 2 (May): 215–218.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *the Quarterly journal of Economics:* 577–598.

Keller, W., and S. R. Yeaple. 2013. "The Gravity of Knowledge." *American Economic Review* 103, no. 4 (June): 1414–44.

Kerr, W. R. 2008. "Ethnic scientific communities and international technology diffusion." *The Review of Economics and Statistics* 90 (3): 518–537.

Lampe, R. 2012. "Strategic citation." *Review of Economics and Statistics* 94 (1): 320–333.

Lanjouw, J., A. Pakes, and J. Putnam. 1998. "How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data." *Journal of Industrial Economics* 46 (4): 405–32.

Lei, Z., and B. D. Wright. 2017. "Why weak patents? Testing the examiner ignorance hypothesis." *Journal of Public Economics* 148:43–56.

Li, Y. A. 2014. "Borders and distance in knowledge spillovers: Dying over time or dying with age?—Evidence from patent citations." *European Economic Review* 71:152–172.

Maurseth, P. B., and B. Verspagen. 2002. "Knowledge Spillovers in Europe: A Patent Citations Analysis." *Scandinavian Journal of Economics* 104 (4): 531–45.

Mayer, T., and S. Zignago. 2011. *Notes on CEPII's distances measures: The GeoDist database.* Working Papers. CEPII research center.

McFadden, D., et al. 1973. "Conditional logit analysis of qualitative choice behavior."

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura. 2014. "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *The Review of Economics and Statistics* 96 (5): 967–985.

Nelson, R. R., and E. S. Phelps. 1966. "Investment in humans, technological diffusion, and economic growth." *The American economic review* 56 (1/2): 69–75.

Nigai, S. 2017. "A tale of two tails: Productivity distribution and the gains from trade." *Journal of International Economics* 104:44–62.

Peri, G. 2005. "Determinants of knowledge flows and their effect on innovation." *Review of Economics and Statistics* 87 (2): 308–322.

Serafinelli, M. 2018. "Good firms, worker flows and local productivity." *Journal of Labor Economics* forthcoming.

Singh, J. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5): 756–770.

Thompson, P. 2006. "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations." *The Review of Economics and Statistics* 88 (2): 383–388.

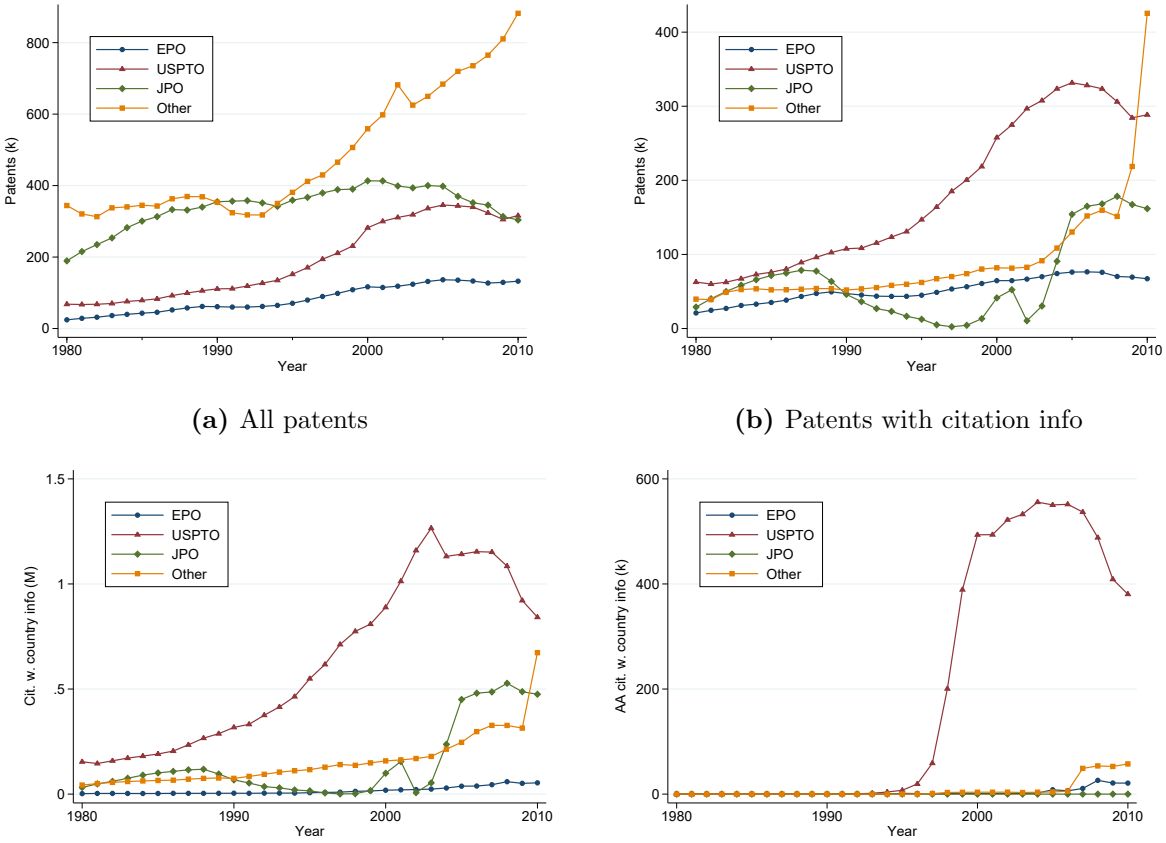
Thompson, P., and M. Fox-Kean. 2005. "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review* 95 (1): 450–460.

Appendix

1.A Technical Appendix

1.A.1 Description of the PATSTAT database

Figure 1.A.1: Number of patents/citations, decomposed by patent office



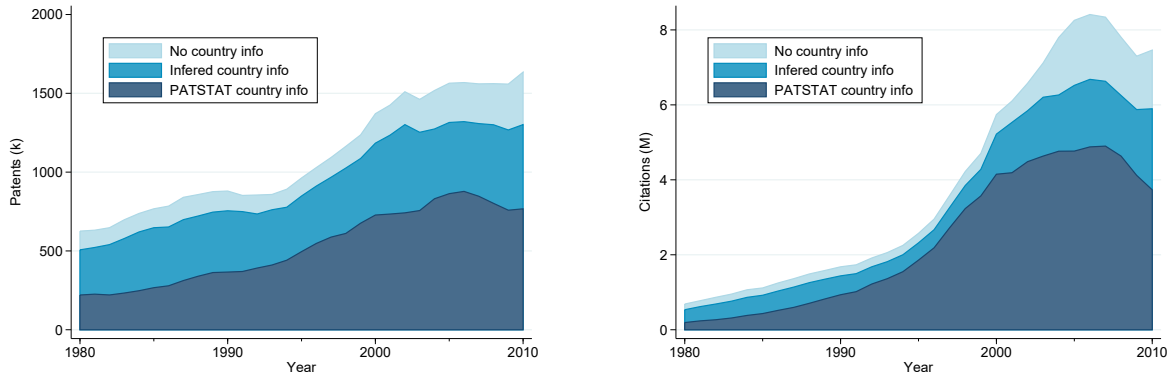
(a) All patents

(b) Patents with citation info

(c) Number of citations

(d) Applicant added citations

Figure 1.A.2: Patents/citations for which we have geographic information (country)

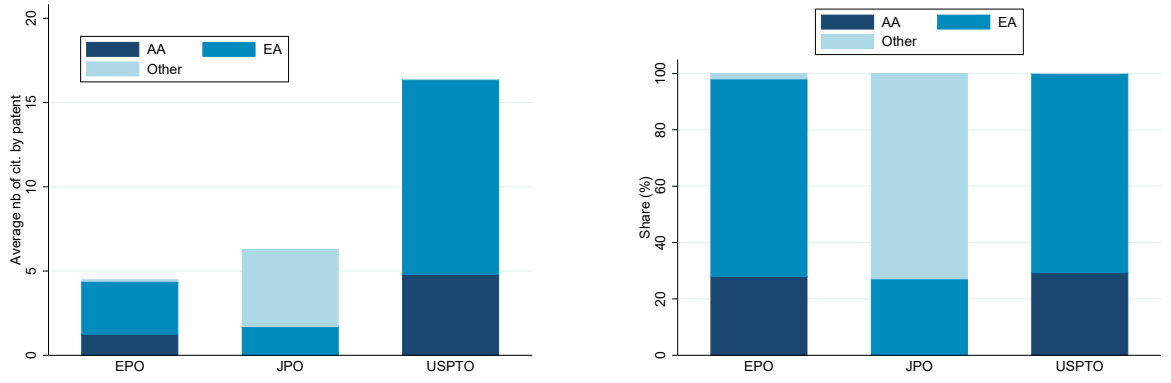


(a) Patents **(b)** Citations

Table 1.A.1: International Patent Classification, list of the sections

A	Human necessities
B	Performing operations, transporting
C	Chemistry, metallurgy
D	Textiles, paper
E	Fixed constructions
F	Mechanical engineering; lighting; heating; weapons; blasting
G	Physics
H	Electricity

Figure 1.A.3: Type of outward citations, by patent office



(a) Average number per patent **(b)** Proportion

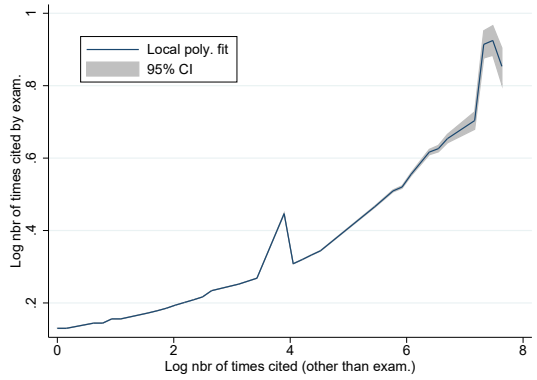
Patents with priority year posterior to 2000

1.A.2 Description of examiner citations

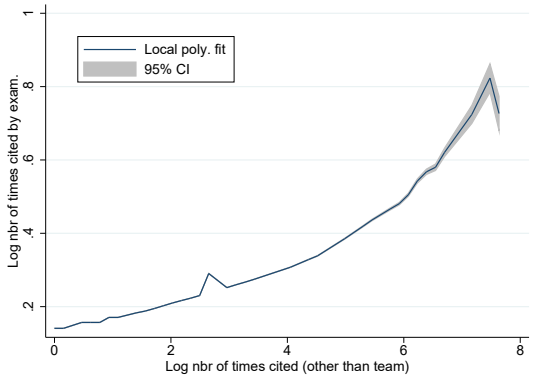
The time spent on a patent application by an examiner is substantial: after dropping very occasional examiners (less than 5 applications), the average examiner handles 40 patent applications per year, with the 95th percentile being slightly above 100, meaning that even very busy examiners deal with two applications in an average working week. This suggests that the citations added in the process of examination should have been cautiously analyzed. Similarly, examination is conducted by one person only.

Examiners appear to be very specialized in their field: keeping only the eight technological centers as they exist today (to avoid counting organizational changes as movements), 78% of examiners remained their whole career in one of the centers, while 86% of examiners handled patents for less than 4 of the 589 technological divisions called art units[22] over their career.

Figure 1.A.4: Correspondence between the number of times an examiner cites a patent and the number of times other examiners cite it.



(a) Number of times cited by an examiner vs by other examiners



(b) Number of times cited by an examiner vs by other examiners outside of her team

Their is limited habit formation in examiners' behavior. As Figure 1.A.4 shows, patents cited several times by the same examiner also tend to be cited many times by other examiners, even when we consider only the ones outside the examiner's art unit (to exclude potential peer-effects). Looking at the technological distance between the patent application assessed by the examiner and the patents she cites, as shown in Table 1.A.2, we find that the first time an examiner cites a patent, the technological distance is only 1% of a standard deviation lower, or equivalently that each additional time a patent is cited by a given examiner implies an average increase of

22. Art units are grouped generally by 10 into clusters which include fields such as "Memory access and control", "Digital and optical communications", "Immunology, Receptor/Ligands, Cytokines Recombinant Hormones, and Molecular Biology", etc.

technological distance of .4% of a standard deviation. This means that habit formation in the way examiners cite, while it may exist, implies very small losses in the relevance of citations as evidenced by our measure of technological distance.

Table 1.A.2: Technological distance in multiple citations by examiners

	Std. tech dist	
	(1)	(2)
First citation by examiner	-0.011[a]	
	(0.001)	
Rank of examiner citation		0.004[a]
		(0.001)
Examiner FE	Yes	Yes
Dest. Pat. FE	Yes	Yes
Nbr of obs	9.6M	9.6M
R-sq	0.74	0.74
VCE	Cluster Exam-Id	

Robust standard errors in parentheses

[a] p<0.01, [b] p<0.05, [c] p<0.1

Note: The sample is composed of all citations to destination patents cited more than once by the same USPTO examiner. The dependent variable is the standardized technological distance between the citing and the cited patent (Mahalanobis distance calculated on IPCs 3 digits). "First citation by examiner" is a dummy variable taking value 1 when a patent is cited for the first time by an examiner. "Rank of examiner citation" is a variable taking value n when a citation corresponds to the n^{th} time an examiner cites a patent. Standard errors are clustered at the citing patent level in all regressions. Significance levels: [a] p<0.01 [b] p<0.05 [c] p<0.1.

1.A.3 Description of the variables used in Section 1.3.

Age Age is simply the difference between the priority date of the citing patent and the priority date of the cited patent.

Quality We build a proxy for the quality of each patent by regressing the number of citations this patent received on a set of fixed effects absorbing the effects of technological classes (IPC 3 digits), priority year and office.[23] In order to use log-transformed values in the regressions, we shift all values by the absolute value of the minimum to have only positive values. This is not a problem since it is a control variable and that we do not interpret the associated coefficient.

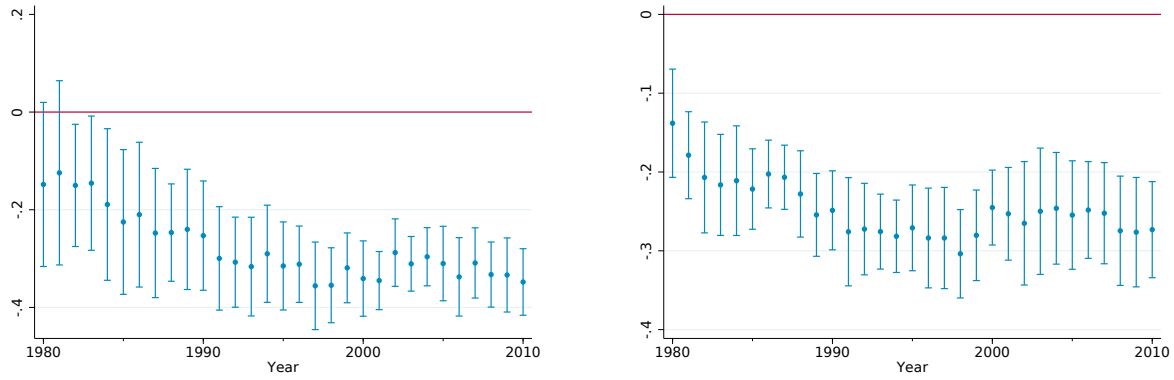
23. To include IPC 3 digits fixed effects, we need to assign a single IPC3 digit of each patent (a patent may belong to several IPC 3 digits, whereas our strategy requires that each patent is associated with one single IPC3 digit). To determine the main IPC 3 digit of a patent, we consider all the IPC 6 digits of this patent, each of which corresponding to a single IPC 3 digits, and find the mode of IPC 3 digit based on this.

Geographical Distance Spatial distance is determined based on the cities of the assignees of o and d. In the case where there are several applicants located in different cities, we take the mode of the different cities, or we randomly choose the city of one of the applicants if there is no mode.

Technological Distance Additionally to the previous variables, we build a measure of technological distance between the citing and the cited patents based on the IPC classes in which it has been filed. The origin of this approach can be traced back to the seminal paper of Jaffe et al. (1993). It has later been refined by Thompson and Fox-Kean (2005) and Murata et al. (2014). The use of Mahalanobis distance between patents as a way to calculate technological distances between them is a valid approach, as confirmed by the recent work of Bloom et al. (2013). We can also compare our treatment and control group based on this distance measure, and introduce it as a control variable in our tests.

1.B Additional Elements on Gravity Estimates

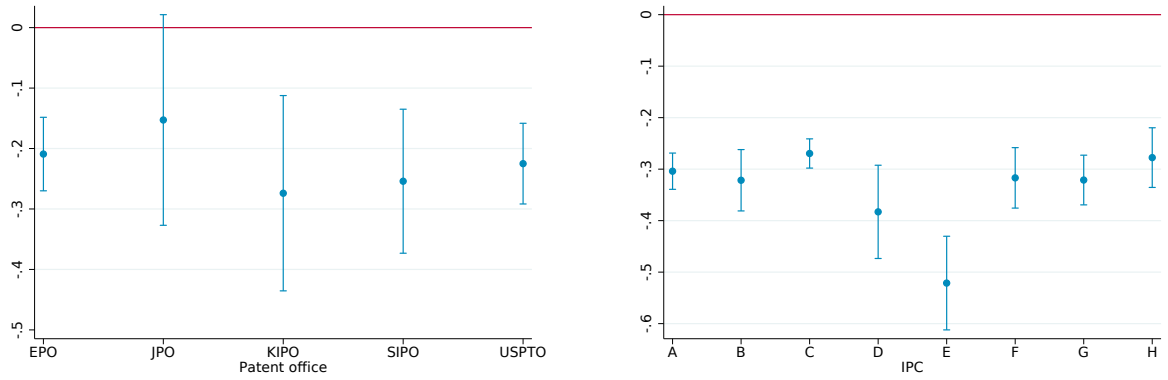
Figure 1.A.5: Evolution of the distance elasticity of citation flows over time, alternative estimators



(a) OLS **(b)** PPML on flow shares

Notes: These distance elasticities are obtained from a series of cross-sectional gravity estimations. Flow share = flow / total flow to the destination. The capped spikes depict the 95% confidence interval of each estimate.

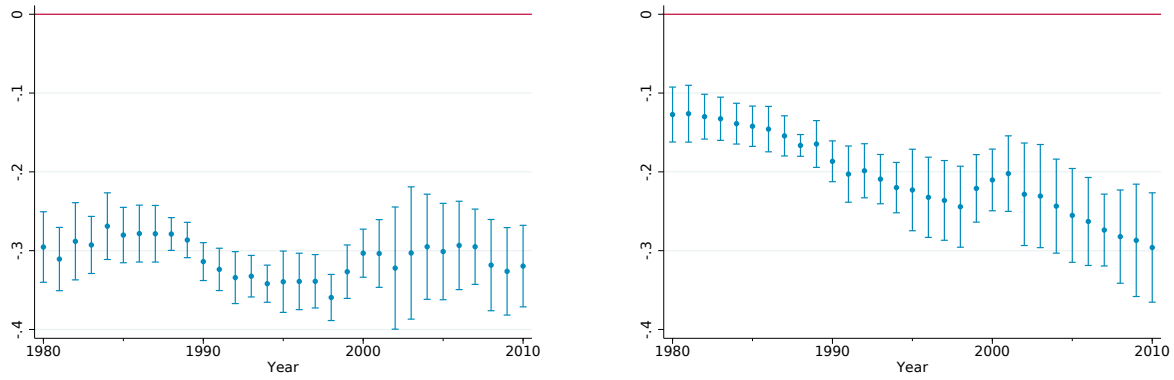
Figure 1.A.6: Distance elasticity of citation flows, sample split by patent office or by technological sector



(a) By patent office **(b)** By IPC

Notes: PPML estimates. The sample is split either according to the patent office of the citing patent (a), or to the IPC section of the citing patent (b). The capped spikes depict the 95% confidence interval of each estimate.

Figure 1.A.7: Distance elasticity of citation flows over time, alternative geographic measures



(a) "Weighted distance" **(b)** Inventor's country

Notes: (a) Instead of using the distance between the most populated city of the origin and destination country, we use a "weighted distance" where the main cities are taken into account, with a weight corresponding to their share of the country population (see Mayer and Zignago, 2011, for more info) (b) Instead of using the applicants to determine the country of a patent, we use its inventors. Inventors may be a more accurate source of information on the country of the patent if the applicant is a large firm with worldwide R&D facilities. Indeed, in this case, the country of the applicant is likely to be the country of the headquarters of the firm, while the country of the inventor will be the one of the R&D facility at which the research was conducted.

1.C Network search: Robustness

Alternative Test The sample is the set of patents that could potentially be cited, i.e. patents granted after 2000 to the true or fake contacts. The dependent variable is a dummy variable taking value one if the patent was actually cited by the random set of firms, zero otherwise. We

want to know whether this binary choice is affected by the fact that the patent belongs to a true contact as opposed to a fake one, *i.e.* whether a dummy variable indicating that the patent belongs to a true contact has a positive and significant effect. This dummy variable is defined at the citing firm × destination patent level, which is therefore the unit of observation we adopt for our analysis.

This strategy has some drawbacks compared to the baseline one. Since citations are only observed when these patents cited either by applicants or examiners get cited after 2000 by origin applicants, the sample is composed of all the cited patents rather than all the citations, which does not allow to run a conditional logit as before. Indeed, it implies conducting the tests at the citing applicant-cited patent pair level rather than at the citing patent-cited patent pair level. For this reason, we run a simple logit, with standard errors clustered at the citing applicant level. Another drawback, although minor, is that it implies conducting the two tests on very different samples (all patents from contacts, either true or control for Test 1, and all citations, either applicant or examiner added, in all contacts' applications in Test 2). To be consistent with the baseline identification strategy, in which running a conditional logit implies dropping any citation without at least one applicant added and one examiner added citation, we drop the citations which do not meet this criterion in the following tests.

Table 2.2 shows estimates of the alternative version of the tests. Columns 1 of both tables run logit regressions without controls respectively on patents from our group of studied firms. Columns 2 show the same regressions controlling for quality as well as year and technological class (1 digit) fixed effects of the cited patents. Because a citation is only observed whenever patents get cited again (i.e. when our dependent variable is equal to 1), we can only control for characteristics of the destination patent, mostly by using year and technology class fixed-effects. Results for both tests support the ones from the baseline identification strategy: applicants are more 80% more likely to cite again patents from the applicants they have truly cited than ones cited by examiners, as shown in the left panel of Table 2.2, and are also about 30% more likely to cite patents truly cited by their contacts than patents cited by examiners of their contacts, as shown in the right panel of Table 2.2.

Firm level results for network search Because large firms display lower direct network effects and make the bulk of citations, we conduct the same tests at the firm level, to provide average effects on firms rather than on patent applications. To do so, we take the mean of all

Table 1.A.3: Results, Alternative versions of tests

Alternative Test 1

	(1) Contact	(2) Pat.
Applicant Cit.	1.70^a	1.78^a
	(0.08)	(0.07)
Cited in 2000	2.31^a	2.39^a
	(0.06)	(0.07)
Dest. Quality (log)		1.03^a
		(0.01)
IPC 1d FE	×	✓
Year FE	×	✓
Applicants	Firms	Firms
Nbr of obs	3.3M	3.3M

Note: Logit estimations of the determinants of knowledge transfers (applicant-added citations). The sample is the set of patent applications from both true and false contacts of studied applicants after 2000, including examiner added citations, recorded on USPTO patents only. The dependent variable is a dummy equal to 1 when patents belonging to contacts get cited after 2000, 0 otherwise. "Applicant Cit." is a dummy equal to 1 when patent d belongs to a true contact (cited by applicant in 2000). "Cited in 2000" is a dummy equal to 1 when patent d was cited in 2000. Results are exponentiated coefficients (odds ratios). Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

Alternative Test 2

	(1) Cited by	(2) contact
Applicant Cit.	1.56^a	1.29^a
	(0.06)	(0.05)
Dest. Quality (log)		1.60^a
		(0.04)
IPC 1d FE	×	✓
Year FE	×	✓
Applicants	Firms	Firms
Nbr of obs	3.1M	3.1M

Note: Logit estimations of the determinants of knowledge transfers (applicant-added citations). The sample is the set of citations made by both true and false contacts of studied applicants after 2000, including examiner added citations, recorded on USPTO patents only. The dependent variable is a dummy equal to 1 when patents previously cited by contact get cited after 2000, 0 otherwise. "Applicant Cit." is a dummy equal to 1 when the contact cited directly patent d. Results are exponentiated coefficients (odds ratios). Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

our variables (dependent and independent) by firm over the whole period and by firm-year. We add two variables for the number of patent applications and the number of citations made. We then run a fractional logit regression, which is adapted to dependent variables resulting of a mean of realizations of a binary variable. We present results as odds ratio, which can therefore be interpreted like other robustness results.

Table 1.A.4: Network search measured at the firm level

	(1)	(2)	(3)	(4)
	Share AA			
	Firm		Firm × Year	
Contact	3.32^a	2.92^a	2.07^a	1.77^a
	(0.32)	(0.29)	(0.10)	(0.09)
Cited by Contact	1.62^a	1.41^b	2.83^a	1.96^a
	(0.28)	(0.24)	(0.24)	(0.17)
Nbr of patents (log)	1.30^a	1.24^a	1.22^a	1.49^a
	(0.03)	(0.03)	(0.01)	(0.03)
Nbr of citations (log)	0.77^a	0.80^a	0.92^a	0.77^a
	(0.02)	(0.03)	(0.01)	(0.01)
Geo. Dist. (log)		0.80^a		0.85^a
		(0.03)		(0.01)
Age (log)		1.53^a		1.57^a
		(0.06)		(0.03)
Quality (log)		1.26^a		1.24^a
		(0.04)		(0.02)
Tech. Dist. (log)		0.97		0.93^a
		(0.05)		(0.02)
Obs Unit	Firm		Firm × Year	
Persistence Controls	✓	✓	✓	✓
Year FE	×	×	✓	✓
Nbr of obs	7,365	7,236	36,318	35,822

Note: Fractional Logit estimations of the determinants of knowledge transfers (applicant-added citations), aggregated at the firm or firm and year levels. The sample is the set of randomly selected firms with patents in and after 2000, recorded on USPTO patents only. The dependent variable is the share of applicant-added citations over the period considered. Nbr of patents and Nbr of citations are sums of patent applications and citations by the firm over the period considered. All other covariates are similar to the ones found in the baseline test shown in Table 1.2, yet averaged at the relevant observation unit level. Coefficients are odds-ratios (exponentiated), standard errors refer to these exponentiated coefficients. Standard errors are clustered at the citing patent level in all regressions. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

Identification on examiner and applicant citations overlap A motivation and description of this test are given in the body of the paper.

Group level results To consolidate the definition of assignee identifiers, we match origin applicants and their contacts flagged as firms with the Orbis database. We match approximately 60% of the names we enter to firms in the database.[24] The matched firms are a priori the largest

24. We try to match 36,000 contacts for which we have information on the name, country, and which are flagged as firms in Patstat.

Table 1.A.5: Network search using citations overlapping applicant and examiner added citations

	(1)	(2)	(3)	(4)	(5)
Firms		All		Small	Large
Contact	1.12^a	1.35^a	1.30^a	3.30^a	1.26^a
	(0.01)	(0.01)	(0.01)	(0.20)	(0.01)
Cited by Contact	1.96^a	2.88^a	1.75^a	2.10^a	1.74^a
	(0.03)	(0.04)	(0.03)	(0.19)	(0.03)
Age (log)			1.79^a	2.12^a	1.77^a
			(0.01)	(0.05)	(0.01)
Quality (log)			1.34^a	1.38^a	1.34^a
			(0.00)	(0.02)	(0.00)
Tech. Dist. (log)			1.02^a	1.08^a	1.02^a
			(0.01)	(0.03)	(0.01)
Geo. Dist. (log)			0.97^a	0.98^a	0.97^a
			(0.00)	(0.01)	(0.00)
Persistence Controls	✓	✓	✓	✓	✓
Orig. Pat. FE	×	✓	✓	✓	✓
Nbr of orig. applicants	7,329	7,198	3,925	1,993	1,932
Nbr of obs	3.8M	2.4M	2.4M	158k	2.2M

Note: Logit and conditional logit estimations of the determinants of knowledge transfers (applicant-added citations), considering only citations found in the set of examiner-added citations. The sample is the set of examiner citations of the randomly selected applicants after 2000, recorded on USPTO patents only, from patents containing at least a a citation in the overlap of examiner and applicant citations. The dependent variable is a dummy equal to 1 when there is an applicant-added citation of patent d by patent o. "Contact" is a dummy equal to 1 when patent d belongs to a contact of the firm. "Cited by Contact" is a dummy equal to 1 when patent d has been cited by a contact of the firm. "Several Cit." is a dummy equal to 1 when patent d is cited several times by the origin applicant from the init. year on. Coefficients are odds-ratios (exponentiated), standard errors refer to these exponentiated coefficients. Standard errors are clustered at the citing patent level in all regressions. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

ones, thus the ones which are most likely to have subsidiaries. We spot within-group citations through firms which have the same parent company as their contacts (as of September 2018). We find that remaining within-group citations are indeed rare: we find only 0.2% of links to be within group. We also conduct the same regressions as in the baseline merging applicants and their contacts when they belong to the same group. Table 1.A.6 provided in Appendix shows that consolidating assignees and contacts at the group level does not harm our results.

Initialization year Table 1.A.7 displayed in Appendix provides results similar to the baseline, changing the initialization year from 2000 to 1999, 2001, 2003 and 2005. Results hold in all these

Table 1.A.6: Network search tests spotting groups with Orbis

	(1)	(2)	(3)	(4)	(5)
Firms		All		Small	Large
Contact	1.64^a	1.59^a	1.43^a	2.42^a	1.36^a
	(0.01)	(0.01)	(0.01)	(0.08)	(0.01)
Cited by Contact	2.04^a	2.48^a	1.74^a	1.85^a	1.74^a
	(0.02)	(0.03)	(0.02)	(0.12)	(0.02)
Age (log)			1.73^a	1.84^a	1.72^a
			(0.01)	(0.03)	(0.01)
Quality (log)			1.20^a	1.21^a	1.19^a
			(0.00)	(0.01)	(0.00)
Tech. Dist. (log)			1.02^a	1.03^b	1.02^a
			(0.00)	(0.01)	(0.00)
Geo. Dist. (log)			0.97^a	0.98^a	0.97^a
			(0.00)	(0.00)	(0.00)
Persistence Controls	✓	✓	✓	✓	✓
Orig. Pat. FE	×	✓	✓	✓	✓
Nbr of orig. applicants	7,169	7,040	5,828	3,661	2,238
Nbr of obs	6.3M	6.3M	6.1M	438k	5.7M

Note: Logit and conditional logit estimations of the determinants of knowledge transfers (applicant-added citations), merging applicants and their contacts based on the group they belong to, estimated like the baseline (equation (1.2)). The sample is the set of citations of the randomly selected applicants grouped by Orbis head of group identifier, recorded on USPTO patents only. The dependent variable is a dummy equal to 1 when there is an applicant-added citation of patent d by patent o. "Contact" is a dummy equal to 1 when patent d belongs to a contact of the firm. "Cited by Contact" is a dummy equal to 1 when patent d has been cited by a contact of the firm. "Several Cit." is a dummy equal to 1 when patent d is cited several times by the origin applicant from the init. year on. Coefficients are odds-ratios (exponentiated), standard errors refer to these exponentiated coefficients. Standard errors are clustered at the citing patent level in all regressions. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

situations, and coefficients show little sensitivity to changes in initialization year as well as to the implied re-sampling (since the random selection of a third of all applicants is conducted again for each year).

Maximum size of contacts Table 1.A.8 displayed in Appendix shows results changing the maximum size of contacts from the 99^{th} percentile of the applicant size distribution as in the baseline to the 90^{th} percentile, 95^{th}, 995^{th} millile or 999^{th} millile. Coefficients also deviate very little from the ones obtained in the baseline.

Table 1.A.7: Network search tests changing the initialization year

	(1)	(2)	(3)	(4)
Init. Year	1999	2001	2003	2005
Contact	1.16^a	1.23^a	1.23^a	1.29^a
	(0.01)	(0.01)	(0.01)	(0.01)
Cited by Contact	1.99^a	1.86^a	1.58^a	1.71^a
	(0.03)	(0.02)	(0.02)	(0.02)
Age (log)	1.68^a	1.66^a	1.60^a	1.60^a
	(0.01)	(0.01)	(0.01)	(0.01)
Quality (log)	1.12^a	1.11^a	1.14^a	1.11^a
	(0.00)	(0.00)	(0.00)	(0.00)
Tech. Dist. (log)	1.03^a	1.01^a	1.01^b	1.02^a
	(0.00)	(0.00)	(0.00)	(0.00)
Geo. Dist. (log)	0.98^a	0.98^a	0.97^a	0.97^a
	(0.00)	(0.00)	(0.00)	(0.00)
Firms	All	All	All	All
Persistence Controls	✓	✓	✓	✓
Orig. Pat. FE	✓	✓	✓	✓
Nbr of orig. applicants	5,383	5,604	5,710	5,693
Nbr of obs	6.4M	5.8M	5.5M	4.6M

Note: Logit and conditional logit estimations of the determinants of knowledge transfers (applicant-added citations), changing the initialization year compared to the baseline (equation (1.2)). The sample is the set of citations of the randomly selected applicants after the relevant init. year, recorded on USPTO patents only. The dependent variable is a dummy equal to 1 when there is an applicant-added citation of patent d by patent o. "Contact" is a dummy equal to 1 when patent d belongs to a contact of the firm. "Cited by Contact" is a dummy equal to 1 when patent d has been cited by a contact of the firm. "Several Cit." is a dummy equal to 1 when patent d is cited several times by the origin applicant from the init. year on. Coefficients are odds-ratios (exponentiated), standard errors refer to these exponentiated coefficients. Standard errors are clustered at the citing patent level in all regressions. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

Placebo tests Since the two different types of citations (AA and EA) exist at three different levels in our tests (contact initialization, citations by contacts, citations after 2000), it is easy to run Placebo tests, building false contacts or false citations by contacts. We invert applicant and examiner-added citations: this way, we build a false set of contacts as the applicants cited in 2000 by examiners, and a false set of patents cited by contacts as the patents cited by the examiners of these false contacts. We then make sure that none of these false links overlaps with the true links constructed in the above baseline strategy, which would naturally make our tests spuriously positive. We run the same regressions as the ones presented in Table 1.2. Table 1.A.9

Table 1.A.8: Network search tests changing the maximum size of contacts

	(1)	(2)	(3)	(4)	(5)
Contacts max. size (percentile)	90	95	99	99.5	99.9
Contact	1.29^a	1.29^a	1.28^a	1.28^a	1.28^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Cited by Contact	1.67^a	1.65^a	1.67^a	1.68^a	1.67^a
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Age (log)	1.67^a	1.67^a	1.66^a	1.66^a	1.66^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Quality (log)	1.14^a	1.14^a	1.13^a	1.13^a	1.12^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Tech. Dist. (log)	1.02^a	1.02^a	1.02^a	1.02^a	1.02^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Geo. Dist. (log)	0.98^a	0.98^a	0.98^a	0.98^a	0.98^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Firms	All	All	All	All	All
Persistence Controls	✓	✓	✓	✓	✓
Orig. Pat. FE	✓	✓	✓	✓	✓
Nbr of orig. applicants	6,020	6,020	6,020	6,020	6,020
Nbr of obs	6.1M	6.1M	6.1M	6.1M	6.1M

Note: Logit and conditional logit estimations of the determinants of knowledge transfers (applicant-added citations), changing the maximum size of contacts compared to the baseline (equation (1.2)). The sample is the set of citations of the randomly selected applicants after 2000, recorded on USPTO patents only. The dependent variable is a dummy equal to 1 when there is an applicant-added citation of patent d by patent o. "Contact" is a dummy equal to 1 when patent d belongs to a contact of the firm. "Cited by Contact" is a dummy equal to 1 when patent d has been cited by a contact of the firm. "Several Cit." is a dummy equal to 1 when patent d is cited several times by the origin applicant from the init. year on. Coefficients are odds-ratios (exponentiated), standard errors refer to these exponentiated coefficients. Standard errors are clustered at the citing patent level in all regressions. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

displayed in Appendix shows that the effects of our network variables disappear once controls on observable characteristics are introduced, which is the desired result.

Table 1.A.9: Results of Placebo Tests

	(1)	(2)	(3)	(4)	(5)
Firms		All		Small	Large
Contact	0.90^a	0.96^a	1.00	0.98	1.00
	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)
Cited by Contact	1.48^a	1.31^a	0.95^a	0.95	0.95^a
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
Age (log)			1.75^a	1.88^a	1.74^a
			(0.01)	(0.03)	(0.01)
Quality (log)			1.26^a	1.26^a	1.26^a
			(0.00)	(0.01)	(0.00)
Tech. Dist. (log)			1.02^a	1.03^b	1.02^a
			(0.00)	(0.01)	(0.00)
Geo. Dist. (log)			0.97^a	0.98^a	0.97^a
			(0.00)	(0.00)	(0.00)
Persistence Controls	✓	✓	✓	✓	✓
Orig. Pat. FE	✕	✓	✓	✓	✓
Nbr of orig. applicants	7,373	7,242	6,020	3,720	2,300
Nbr of obs	6.3M	6.3M	6.1M	460k	5.7M

Note: Logit and conditional logit (when Orig. Pat. FE is YES) estimations of the determinants of knowledge transfers (applicant-added citations) (equation (1.2)). Placebo versions of contacts and citations by contacts are constructed following the above-described procedure. The dependent variable is a dummy equal to 1 when there is an applicant-added citation of patent d by patent o. "Contact" is a dummy equal to 1 when patent d belongs to a contact of the firm. "Cited by Contact" is a dummy equal to 1 when patent d has been cited by a contact of the firm. "Several Cit." is a dummy equal to 1 when patent d is cited several times by the origin applicant from 2000 on. Coefficients are odds-ratios (exponentiated), standard errors refer to these exponentiated coefficients. Standard errors are clustered at the citing applicant level in all regressions. Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

1.D Additional aggregate results

1.D.1 Additional tables

Table 1.A.10: Estimates of the shape parameter of the Pareto distribution of firm size (λ)

	All patents			USPTO patents		
	Applications	Granted	Citations	Applications	Granted	Citations
	(1)	(2)	(3)	(4)	(5)	(6)
λ	-0.936^a	-0.946^a	-0.935^a	-0.973^a	-0.979^a	-0.927^a
	[0.026]	[0.034]	[0.038]	[0.045]	[0.050]	[0.043]
Nbr. of bins	9	9	9	9	9	9
R^2	0.997	0.995	0.993	0.992	0.990	0.991

Notes: Firm size is measured either as the number of patent applications of the firm over the period 1980-2010 (columns (1) and (4)), the number of granted patents (columns (2) and (5)) or the number of outward citations (columns (3) and (6)). Significance levels: $^a : p < 0.01$; $^b : p < 0.05$; $^c : p < 0.1$

Table 1.A.11: Estimates of the elasticity of the average squared distance of citations with respect to firm size (μ)

	All patents		USPTO patents		
	All Cit.	All Cit.	All Cit.	AA Cit.	EA Cit.
	(1)	(2)	(3)	(4)	(5)
μ	0.038^a	0.034^a	0.030^a	0.026^a	0.020^b
	[0.003]	[0.003]	[0.009]	[0.005]	[0.006]
Size measure	Patents	Cit.	Patents	Patents	Patents
Nbr. of bins	10	10	10	10	10
R^2	0.972	0.962	0.702	0.825	0.677

Notes: Firm size is measured as the number of patent applications of the firm over the period 1980-2010. Distance is measured as the geodesic distance between the most populated city of each country. "AA Cit." refers to the applicant added citations, while "EA Cit." refers to the examiner added citations. Significance levels: $^a : p < 0.01$; $^b : p < 0.05$; $^c : p < 0.1$

Table 1.A.12: Estimates of the elasticity of the average squared distance of citations with respect to firm size (μ), using within firm variations.

	All patents		USPTO patents		
	All Cit.	All Cit.	All Cit.	AA Cit.	EA Cit.
	(1)	(2)	(3)	(4)	(5)
μ	0.143^a	0.138^a	0.082^a	0.098^a	0.084^a
	[0.002]	[0.002]	[0.002]	[0.003]	[0.002]
Firm FE	✓	✓	✓	✓	✓
Year FE	×	✓	✓	✓	✓
Nbr. of obs	689760	689760	472211	192774	276066
R^2	0.529	0.530	0.522	0.544	0.533

Notes: Firm size is measured as the number of patent applications of the firm over the period 1980-2010. Distance is measured as the geodesic distance between the most populated city of each country. "AA Cit." refers to the applicant added citations, while "EA Cit." refers to the examiner added citations. Standard errors clustered by firm. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

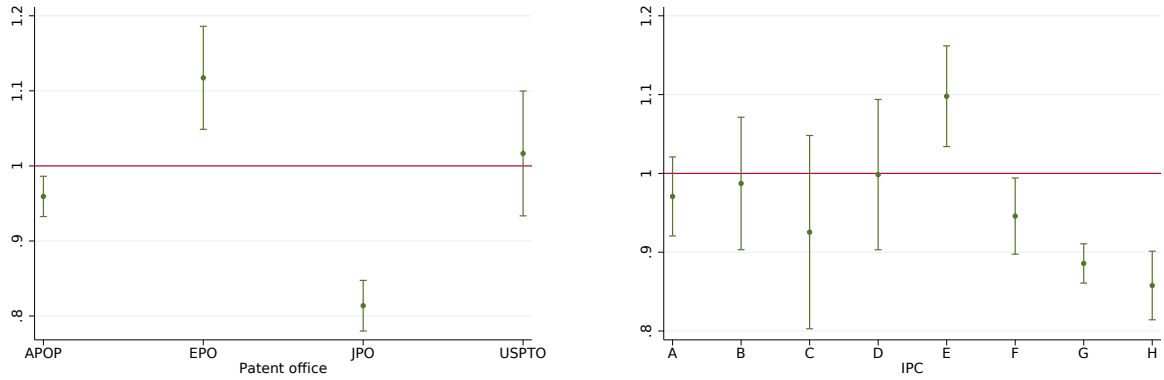
Table 1.A.13: Semi-elasticity of firm size with respect to age.

	All patents			USPTO patents		
	Applications	Granted	Citations	Applications	Granted	Citations
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.016^a	0.013^a	0.043^a	0.019^a	0.014^a	0.047^a
	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.001]
Firm FE	✓	✓	✓	✓	✓	✓
Nbr. of obs	1.981M	1.348M	689760	560276	501064	472211
R^2	0.660	0.664	0.612	0.672	0.678	0.620

Notes: Firm size is measured either as the number of patent applications of the firm over the period 1980-2010 (columns (1) and (4)), the number of granted patents (columns (2) and (5)) or the number of outward citations (columns (3) and (6)). Standard errors clustered by firm. Significance levels: ^{a} : $p < 0.01$; ^{b} : $p < 0.05$; ^{c} : $p < 0.1$

1.D.2 Additional figures

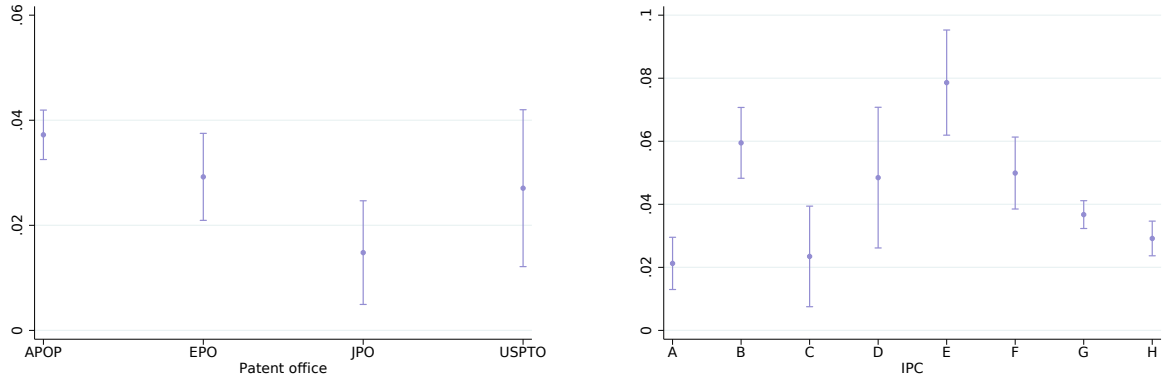
Figure 1.A.8: Shape parameter of the Pareto distribution of firm size (λ), sample split by patent office or by technological sector



(a) By patent office **(b)** By IPC Section (1 digit)

Notes: Firm size is measured as the number of patent applications of the firm over the period 1980-2010. Data is split by patent office (left-hand side) or by technological field (IPC section, right-hand side) and equation (1.4) is estimated on each sub-sample.

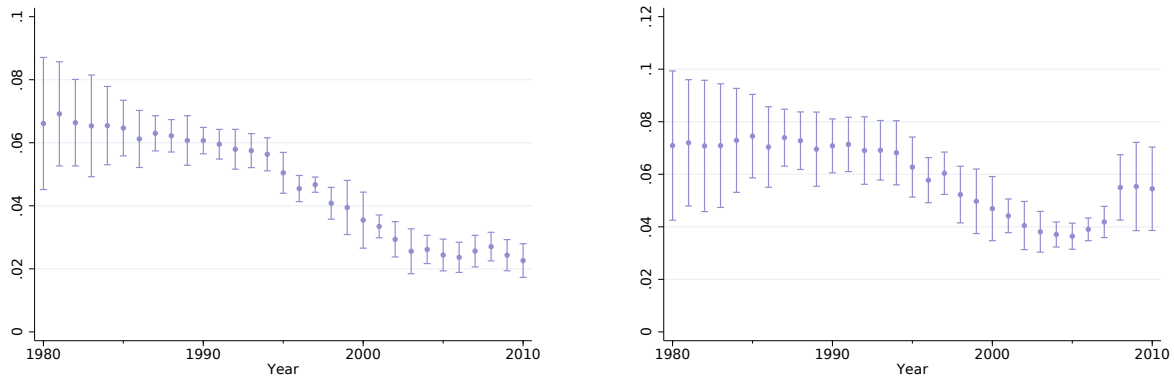
Figure 1.A.9: Elasticity of the average squared distance of citations with respect to firm size (μ), sample split by patent office or by technological sector



(a) By patent office **(b)** By IPC Section (1 digit)

Notes: Firm size is measured as the number of patent applications of the firm over the period 1980-2010. Distance is measured as the geodesic distance between the most populated city of each country. Data is split by patent office (left-hand side) or by technological field (IPC section, right-hand side) and equation (1.4) is estimated on each sub-sample.

Figure 1.A.10: Elasticity of the average squared distance of citations with respect to firm size (μ), estimated with alternative distance measures

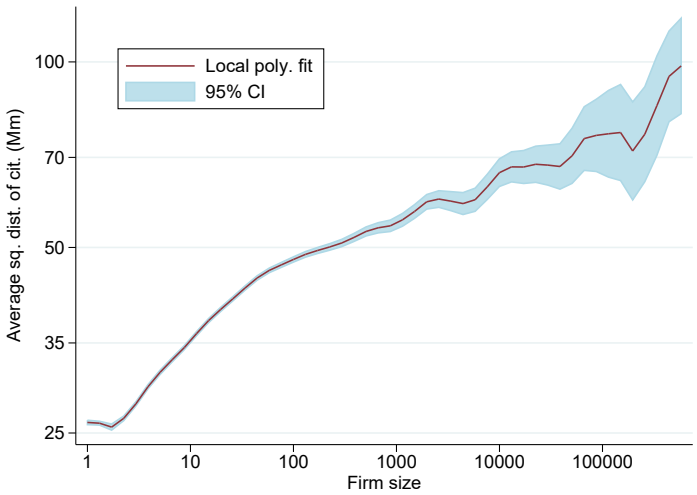


(a) Weighted distance

(b) Inventor's country

Notes: (a) Instead of using the distance between the most populated city of the origin and destination country, we use a "weighted distance" where the main cities are taken into account, with a weight corresponding to their share of the country population (see Mayer and Zignago, 2011, for more info) (b) Instead of using the applicants to determine the country of a patent, we use its inventors. Firm size is measured as the number of patent applications of the firm over the period 1980-2010.

Figure 1.A.11: Average squared distance of citations as a function of firm size



Note: Polynomial fit built on all applicants.

1.E Theory Appendix

1.E.1 Proof of Proposition 1

Proof. A solution for the ODE given in (1.E.1) is:

$$K_a = K_0 e^{(\rho+\beta-\delta)a}$$

Introduce the distribution of contacts normalized by the total number of contacts for a firm of age a: $f_a = \frac{k_a}{K_a}$. Partially differentiating this distribution with respect to a, and denoting $*$ the convolution product of two distributions yields:

$$\begin{aligned}
\frac{\partial f_a(x)}{\partial a} &= \frac{\frac{\partial k_a(x)}{\partial a} K_a - k_a(x) \frac{\partial K_a}{\partial a}}{(K_a)^2} \\
&= \frac{\left[(\rho - \delta) k_a + \beta \frac{k_a * k_a}{K_a}\right] K_a - k_a(\rho + \beta - \delta) K_a}{(K_a)^2} \\
&= \frac{\beta \left[\frac{k_a * k_a}{K_a} - k_a\right] K_a}{(K_a)^2} \\
&= \beta (f_a * f_a - f_a)
\end{aligned}$$

Using the Fourier transform of f_a yields a simple product instead of a convolution product, which yields that f_a converges towards a Laplace distribution when age grows large (Proposition 2 in Chaney, 2018).

One can then derive the endogenized conditions allowing to get a constant elasticity of flows with respect to distance. The distribution of firm sizes is simply derived from the ODE: $K_a = K_0 e^{(\rho+\beta-\delta)a}$. The relation between a firm's size and its age is $e^a = \left(\frac{K_a}{K_0}\right)^{\frac{1}{\rho+\beta-\delta}}$. With a growth rate of the firm population being equal to γ, this means that the fraction of firms having less than K contacts writes:

$$F(K) = 1 - \left(\frac{K}{K_0}\right)^{-\frac{\gamma}{\rho+\beta-\delta}}$$

Thus, the distribution of firm sizes is Pareto, with a shape parameter $\lambda = \frac{\gamma}{\rho+\beta-\delta}$.

The average squared distance at which firms cite others, Δ_a, is the second moment of the normalized density of contacts f_a. Following exactly the steps of the demonstration in Chaney (2018), $\Delta_a = \Delta_0 e^{\beta a}$. Plugging the previous expression $e^a = (\frac{K_a}{K_0})^{\frac{1}{\rho+\beta-\delta}}$, this yields:

$$\Delta(K) = \Delta_0 \left(\frac{K}{K_0}\right)^{\frac{\beta}{\rho+\beta-\delta}}$$

Thus, the average squared distance at which firms cite is a power function of their number of contacts, of parameter $\mu = \frac{\beta}{\rho+\beta-\delta}$. □

Chapter 2

Moving R&D Labs

2.1 Introduction

It is well established that agglomeration economies are large for R&D activities and cause them to be highly spatially concentrated (Audretsch and Feldman, 1996; Carlino and Kerr, 2015; Buzard et al., 2017). Yet, little is known on the mechanisms that cause this clustering. This paper proposes an explanation based on the fact that firms tend to relocate their R&D labs to take advantage of stronger local innovation networks, that generate higher knowledge spillovers.

Understanding the roots of the spatial clustering of innovative firms is essential to design efficient innovation policies. In the last two decades, most developed countries have implemented so-called cluster policies trying to imitate the success of the Silicon Valley, despite a limited understanding of the mechanisms at work and the conditions for success (Duranton, 2011). Beyond cluster policies, the question of how firms internalize knowledge spillovers in their location choices also connects with a strand of the urban economics literature trying to understand how firm sorting and agglomeration forces may explain the productivity advantage of large cities.

This paper focuses on firms relocating a R&D lab, and seeks to elicit the value of local knowledge networks through the decision of abandoning a given location jointly with choosing a new one. It therefore exploits several advantages of relocations over initial location choices to characterize the value of local knowledge spillovers and how firms maximize them. In particular, only incumbent innovative firms, which are already specialized in a technological field, can potentially engage in relocations. In contrast, initial location choices are by definition made by entrants, who could adapt the nature of their research to the spillovers surrounding them. Moreover, relocations provide a clear trade-off between the spillovers a firm currently receives in its location, and the spillovers it expects to receive in its new location.

The paper starts by providing novel stylized facts about firm relocation decisions, understood as a sharp change in the spatial distribution of their R&D workforce, and characterized using French matched employer-employee data. It links these relocations with local knowledge networks, building links using patent citation data. The paper develops two sets of facts explaining how and why innovators move. First, it shows that R&D activities are more likely to be relocated, and tend to be moved further away. Secondly, after giving evidence on the localization of knowledge spillovers in line with the literature, the paper shows that more central firms in their local innovation network are less likely to move their R&D plants. Additionally, if these firms do move, they tend to choose locations with higher expected knowledge spillovers.

Relying on these facts, the paper then builds a theoretical framework in which the amount of spillovers that a firm gets depends on its position in a local network of innovators. It focuses on the trade-off faced by innovative firms between existing connections and uncertain new ones in another location, in the presence of knowledge spillovers occurring only within a location. The use of a network models knowledge spillovers in a rich manner: it tracks precisely the flows between every pair of firms rather than assuming location-specific gains, and therefore accounts in a subtle way for heterogeneity in firms' incentives to move. Moreover, it is naturally in line with empirical analyses, since networks accurately represent usual sources of knowledge spillovers, ranging from explicit links such as R&D collaborations or patent citations to more implicit ones such as transfers of engineers, supply-chain linkages, and of course face-to-face interactions. In the model developed below, firms inherit a position on a local network which determines the spillovers they obtain from their neighbors, and decide whether to relocate or not. In such case, they break all their existing links and choose how many links to form in their new location, as well as their R&D effort.

While many papers have study firms location choices in general, the literature studying the location strategies of innovative firms is relatively sparse. Duranton and Puga (2001) build a model of plant relocations emphasizing the virtues of initially learning from diversified sources before moving to a more specialized area. The main paper with an empirical focus trying to unveil strategic location choices from innovative firms is Alcácer and Chung (2007): the authors study the entry in the US of foreign (thus large) firms, and find that these firms tend to maximize the net spillovers they receive. Alcacer and Delgado (2016) explore theoretically the empirical finding of Alcácer and Chung (2007) that firms are sensitive to the local production of basic research, but also try to minimize the knowledge leakage given away to neighboring competitors. Fitjar and Rodríguez-Pose (2017) survey firms engaged in collaboration about their location choices: they find that collaborations almost never happen by chance, while partners have often chosen their location with the potential for collaboration in mind. Catalini (2017) looks at how colocation influences collaboration and the trajectory of convergence or divergence of scientific research. Henry et al. (2017) use an endowment shock for public labs and study the effect on local private labs.

Building on Jaffe et al. (1993)'s seminal paper on the localization of knowledge spillovers, a vast empirical literature approached this fact along the lines of firm and inventor networks. Almeida and Kogut (1999) and Breschi and Lissoni (2009) focus on the mobility of inventors

across firms and the consequences these movements had on knowledge diffusion, while Singh (2005) studies the effects of co-invention links (applying jointly for a patent). Head et al. (2018) look at the effect of interpersonal links on citations of mathematics research papers. All these papers find that controlling for social proximity implies a drastic reduction on the effect of geography, suggesting that the effect of spatial proximity on knowledge diffusion is largely hiding an effect of social proximity. Kerr and Kominers (2015) also provide an analysis of innovation clusters in which firms obtain spillovers from other firms in a given radius, thus implicitly designing a complete network around them.

As mentionned above, this paper also connects with the urban economics literature studying agglomeration economies. In a seminal paper, Duranton and Puga (2001) built a model in which firms initially benefit from a diverse environment to find their ideal production process, and then move to more specialized areas, placing relocations at the core of a firm's life-cycle. This paper is also particularly related to their role on the productivity advantage of large cities, explored in Combes et al. (2012) and Behrens et al. (2014), seeking to explain why more talented individuals and more productive firms tend to locate in larger cities, making the college wage-premium increasing in city-size. Gaubert (2018) shows that half of this estimated productivity advantage of large cities is due to firm sorting into these locations, thus explicitly linking agglomeration economies and firms' strategic location choices. Davis and Dingel (2019) delve into the nature of this agglomeration force and build a model in which idea exchange is central for productivity and is more intense with more numerous and more talented partners, making large cities better places for idea exchange. The model developed below is reminiscent of these features.

The paper also contributes to the literature on games on networks, and more particularly to the strand studying the interaction between geography and a network. Jackson and Rogers (2005) analyse how a cost of forming links depending on geography contributes to generating small world networks. Helsley and Zenou (2014) study the location choice of individuals in a city based on their position in the social network and the need to commute to the center for social interactions. Using different tools than the above papers, Chaney (2014), 2018 explore the spatial features of international trade based on the assumption that trade relations form through "triadic closure". Cotterlaz and Guillouzouic (2018) show that this assumption is verified for knowledge flows, and study the consequences it has on the geography of international knowledge flows.

2.2 Motivating Facts

This section explores relocation patterns of innovative firms and links them with the local connections these firms have or may expect to have. The first subsection documents relocations, defining them as an increase of employment in at least one plant and a decrease in another over the decade I study. The second subsection starts by replicating evidence that exchanges of knowledge are very localized, and then shows how existing and expected local knowledge links influence mobility.

2.2.1 Innovative Firms Relocations

Data

To characterize firm mobility, I rely on the French matched employer-employee cross-section DADS Postes ("Déclaration annuelle de données sociales"), from 2006 to 2015. This database compiles information declared on a yearly basis by all employers (both private and public) active on the French territory. The information is provided at the job spell level, and includes detailed categories on the nature of jobs (in the 4-digits PCS classification), the hours worked and the corresponding wages, as well as information on the plant where the person works (plant identifier, municipality, main activity).

To define R&D employment, I sum full-time equivalent job spells in the 9 categories of engineers and technicians with labels mentioning explicitly R&D (see Appendix A for the included 4-digit PCS classes).

Characterizing relocations

In Duranton and Puga (2001), the authors characterize relocating firms as the ones which had both a complete plant closure and a plant creation over the same period. Nevertheless, since R&D activities may be transferred to another location without the plant being closed, I hereby follow a more flexible approach based on a spatial reallocation of labor of a certain type over a short period of time. I select R&D plants based on the criterion that they employ at least one R&D worker in an average year, and call R&D firms all firms owning a R&D plant. To assert that firms are relocating, I impose both a quantitative and a spatial criterion, which I detail below.

On the quantitative side of relocations, I define relocating innovative firms in the baseline

as the ones having both a downsized R&D plant which lost at least half of its R&D workforce, and an upsized plant, which gained at least half of its initial R&D workforce.[1] To consider sharp enough reallocations, variations are considered between two consecutive years.[2] Since the thresholds are - necessarily - arbitrary, I do robustness checks with different values.

On the spatial side of relocations, these should occur between plants which are located far enough. Indeed, a common phenomenon for any firm is to change premises within a given city because the current ones have become too small. While this phenomenon has important implications,[3] it is of little interest for our study as it arguably leaves knowledge spillovers unchanged. To select only true changes of location, I rely on INSEE's commuting zones ("Zones d'Emploi 2010", hereafter referred to as commuting zones or CZ), and remove all plants located in commuting zones appearing both in the list of downsized and of upsized plants. This way, a firm with upsized plants in Paris and Marseille, and downsized plants in Paris and Bordeaux is left with one relocation, from Bordeaux to Marseille.

Since relocations are characterized using the spatial reallocation of labor, the above criteria can be applied using different types of labor: relocations of R&D workforce (as in the above examples) can be compared to relocations of overall employment, or - to provide a closer comparison - to relocations of high-skilled workforce.[4] Note that R&D workers are a subset of high-skilled workers. As with R&D, employing a high skilled worker makes a plant a "high-skilled" (hereby HS) plant and a firm a HS firm. Thus, R&D plants and R&D firms are respective subsets of HS plants and HS firms.

Descriptive Statistics

Table 2.1 shows statistics on relocations applied to three subcategories of labor: all workers, high-skilled workers, and researchers. The shares of firms considered to be moving are very similar across the types of employment considered: innovative firms do not stand out as being particularly mobile under this descriptive criterion. Yet, high-skilled firms that move pick a plant to relocate out of an average of 9, while innovative moving firms have on average 4 plants employing researchers, implying that a R&D plant is more likely to be relocated than a HS

1. I further impose that the variation in the workforce should also be larger than a given absolute number, to avoid considering a plant going from .5 to 1 (full-time equivalent) R&D worker as an upsized plant due to its 100% increase in R&D employment.
2. For firms to be considered to be relocating, the variation should also hold true between the initial and the final year (to avoid counting the firms with intrinsically volatile employment as relocating).
3. Bergeaud and Ray, 2017.
4. i.e. the general categories 3 and 4 of the PCS-ESE classification, containing all R&D codes

Table 2.1: Descriptive Statistics of relocations

Category	All firms	HS firms	Innov. firms
Nbr of firms in category	689,691	301,571	24,064
Nbr of moving firms	15,583	8,232	634
Av. nbr of plants in cat. in 2006	6.90	9.18	3.90
Av. nbr of employees	166.4	292.1	879.9
Share of firms moving	2.26	2.73	2.63
Av. relocation distance	122.5	132.3	191.3

The above table shows descriptive statistics of partial relocations of at least 50% and one full-time employee of a plant for 3 categories of firms: all firms, firms employing high-skilled workers (HS firms), and innovative firms. A firm enters a category if at least one of its plants employs at least one of the corresponding type of workers. Firms are considered moving if they relocate a plant of their category: an innovative firm relocating a plant employing high skilled but non-R&D workers is counted as a moving HS firm but not as a moving innovative firm. Numbers of employees and plants are averages calculated over the period covered by the data, i.e. 2006-2015.

plant. Moving innovative firms are, however, much larger than the corresponding moving firms either high-skilled or general, having around five times as many employees as the latter, and 2.5 times as many as the former.

Distance of movements The most salient fact in the descriptive statistics shown in Table 2.1 is that R&D relocations tend to occur at larger distances. Table 2.2.b displays simple OLS regressions of the distance of relocations on the fact of doing R&D and numbers of plants and employees. It shows that having a R&D plant increases the distance at which a firm relocates by about 10 to 15%.

Similarly, plotting the distributions of relocation distances as displayed in Figure 2.1 shows that while most high-skilled and general employment relocations take place at short distances (with a spike around 50km), the distribution of R&D relocations is much flatter with a thicker tail.

Propensity to move To disentangle the effect of size from R&D intensity on the probability of relocating, Tables 2.2.a, and 2.3 provide results of simple logistic regressions of the probability that a firm relocates regressed on size and the fact that it does R&D. Table 2.2.a shows results on the effect of being a R&D firm (defined as having a plant employing at least one researcher) on the probability of relocating either any type of labor (column 1) or high-skilled labor (column 2). It shows that being a R&D firm does not make it more likely to move a plant in general, but makes it 70% more likely to relocate a plant employing high-skilled workers.

Figure 2.1: Distribution of relocation distances

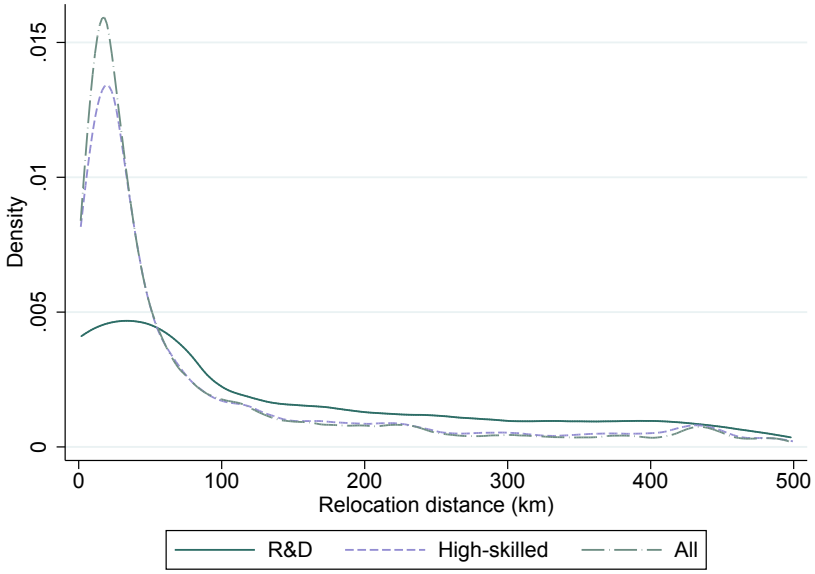


Table 2.2: Effect of being a R&D firm on the probability of defined types of relocation (panel a.) and distance at which they occur (panel b.)

a. R&D firm and probability of moving

	Firm relocates	
	(1)	(2)
R&D firm	0.97	1.68***
	(0.04)	(0.06)
Log Nbr of plants	4.03***	3.30***
	(0.06)	(0.05)
Log Nbr of employees	1.28***	1.08***
	(0.01)	(0.01)
Sample	All firms	HS firms
Type of movement	Overall	HS empl.
Nbr obs	988775	291668
Pseudo R-sq	0.229	0.200

Logit regressions of the probability of relocating a given type of labor on the fact of being a R&D firm. "R&D firm" is a dummy variable taking value 1 if the firm employs at least a R&D worker. "Log Nbr of employees" and "Log Nbr of plants" respectively are the log values of the total number of worker in an average year and number of plants a firm has over the studied period. Coefficients are **odds ratios**. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

b. R&D firm and relocation distance

	Log. Dist.	
	(1)	(2)
R&D firm	0.16***	0.09**
	(0.04)	(0.04)
Log Nbr of plants	0.21***	0.16***
	(0.02)	(0.02)
Log Nbr of employees	-0.08***	-0.07***
	(0.01)	(0.01)
Sample	All firms	HS firms
Type of movement	Overall	HS empl
Nbr obs	15000	7923
R-squared	0.01	0.01

OLS regressions of log distance at which firms relocate on the fact of being a R&D firm. "R&D firm" is a dummy variable taking value 1 if the firm employs at least a R&D worker. "Log Nbr of employees" and "Log Nbr of plants" respectively are the log values of the total number of worker in an average year and number of plants a firm has over the studied period. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

An additional element to show that R&D activities are the ones being moved can be to look at the sensitivity of the propensity to relocate to the amount of R&D employment, i.e. if having more R&D employees makes a firm more likely to relocate its employees overall, its high-skilled employees, and its R&D employees. Table 2.3 shows that the probability of relocating R&D seems very sensitive to the volume of R&D employment: tripling the number of R&D employees (more precisely multiplying it by $\exp(1) \approx 2.7$) increases only slightly the probability of moving either a HS plant or a plant in general, but strikingly almost doubles the probability that an innovative firm relocates a R&D plant.

Table 2.3: Relocations and amount of R&D employment

	Firm relocates		
	(1)	(2)	(3)
Log Nbr of R&D employees	1.08***	1.08***	1.99***
	(0.03)	(0.03)	(0.09)
Log Nbr of HS employees	0.94	1.15**	0.82*
	(0.06)	(0.08)	(0.09)
Log Nbr of employees	0.97	0.86***	1.04
	(0.05)	(0.05)	(0.10)
Log Nbr of plants	3.28***	3.01***	1.73***
	(0.11)	(0.10)	(0.07)
Sample	Inno firms	Inno firms	Inno firms
Type of movement	Overall	HS empl	R&D empl
Nbr obs	23117	23117	23117
Pseudo R-sq	0.243	0.241	0.195

Logit regressions of the effect of total R&D employment on the probability of relocating a given type of labor. "Log Nbr of R&D employees", "Log Nbr of HS employees", "Log Nbr of employees" and "Log Nbr of plants" respectively are the log values of the total number of R&D workers, HS workers, total number of worker in an average year and number of plants a firm has over the studied period. Coefficients are **odds ratios**. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

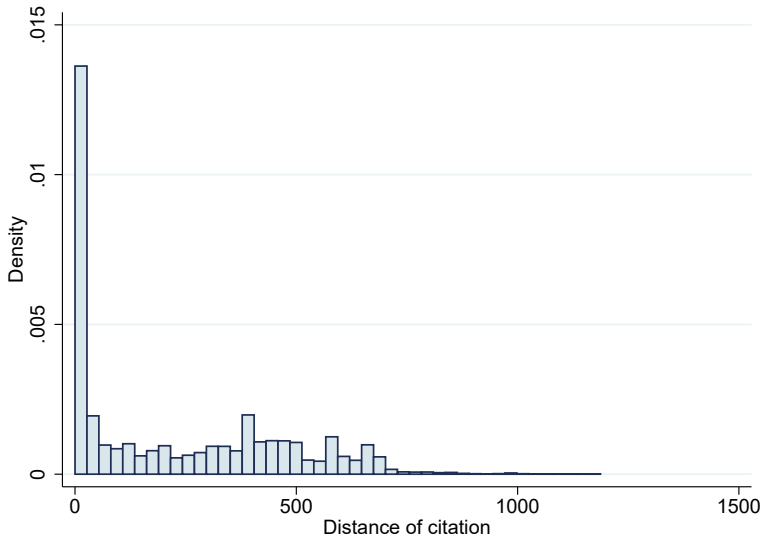
2.2.2 Localization of Innovation Networks

A natural explanation for the higher mobility of R&D firms shown in the above subsection is that they seek to take advantage of knowledge produced locally. The fact that knowledge disseminates more intensely at shorter distances has been shown in many papers following Jaffe et al. (1993)'s seminal paper. The current subsection shows evidence from France which is consistent with the literature.

Data To characterize links along which knowledge flows, I use the Patstat database provided by the EPO. In particular, I build patent application densities per geographical unit, and network links using citations added by the applicants. I match patent applicants with firm identifiers through direct matching on names as well as scrapping web-searches (more detail is provided in the Appendix).

Descriptive Statistics As Figure 2.2 shows, patent citations tend to occur disproportionately at very short distances. This is consistent with the fact that innovators tend to rely on knowledge produced around them, as shown first by Jaffe et al. (1993) and an abundant subsequent literature.

Figure 2.2: Distance of patent citations



Distribution of within-France patent citation distances

This cannot be explained solely with the fact that areas tend to specialize in technological fields. Figure 2.A.1 provided in Appendix is an example of the spatial distribution of patents filed in IPC section C, corresponding to chemistry and metallurgy, for each commuting zone of the French metropolitan territory. It shows that, while the Paris area has high levels of concentration, many regions scattered over the territory also have an intense innovative activity. If knowledge diffusion was frictionless and local innovation did not matter, one would expect citations to occur at higher distances than what we observe from Figure 2.2.

2.2.3 Plant relocations and local connections

Given that we have shown that knowledge transfers are localised, they must be flowing though local innovation networks.[5] This subsection explores how the position of a firm in its local network connects with its decision to relocate.

Centrality and propensity to move

A first point of enquiry is whether firms enjoying better connections in their current location are less likely to move their R&D plants. To answer this question, I build links either using the number of patents developed locally, or the number of citations to patents developed locally. I construct four variables accounting for centrality in the local network. The first variable is a measure of potential links in that location based on patents produced locally. To build this measure for each innovator, I use the share of its patents filed in each IPC class as weights, and sum the number of patents developed in these categories in a location in the past 5 years. This accounts for how intense innovative activity is in one's fields of interest in each location. For the two other measures, I calculate centralities using the network of citations (applicant-added) in France[6] within commuting zones. The first measure is simply the outdegree centrality, *i.e* the number of firms cited in the CZ. The second measure is the Bonacich centrality, a measure which takes indirect links into account and will be extensively used in the theoretical section. I add a last measure, which is simply the Bonacich centrality weighting links by a proxy of the quality of cited patents[7] as a weight for the value of the links, implicitly assuming that citations to higher quality patents yield higher spillovers.[8]

Table 2.4 presents logit regressions using the decision of an innovative firm to relocate a R&D plant or not as the dependent variable. It shows that all three measures correlate negatively with the probability of moving: for instance, column 1 implies that roughly tripling the number of patents in one's technological and geographical areas decreases by (1 - .86 =) 14% the probability that it decides to relocate a R&D plant. Similarly in column 2, tripling (more precisely multiplying by $\exp(1) \approx 2.7$) the number of cited firms in the firm's commuting zone makes it

5. Cotterlaz and Guillouzouic (2018) provide a complete analysis of network formation between innovators and its spatial consequences

6. I keep patents which have inventors located in France, or, in the absence of such information, a French assignee.

7. This measure is the residual of a regression predicting the number of citations received by a patent, using technological class and year fixed effects, as in Cotterlaz and Guillouzouic (2018).

8. The parameter used to calculate the Bonacich centralities in the regressions shown hereafter is .1 (corresponding to ϕ in the theory section), which means in the model that a firm benefits from 10% of the R&D effort of its neighbors.

Table 2.4: Probability of moving and local centrality

	Firm relocates			
	(1)	(2)	(3)	(4)
Log. Potential Local Links	0.86***			
	(0.05)			
Log. Local Degree Centrality (non-weighted)		0.51**		
		(0.16)		
Log. Local Bonacich Centrality (non-weighted)			0.08**	
			(0.09)	
Log. Local Bonacich Centrality (weighted)				0.09**
				(0.10)
Log. Nbr employees	1.04	1.03	1.03	1.03
	(0.04)	(0.04)	(0.04)	(0.04)
Log. Nbr R&D employees	1.13**	1.13**	1.12**	1.12**
	(0.06)	(0.06)	(0.06)	(0.06)
Log. Nbr plants	3.45***	3.41***	3.48***	3.47***
	(0.36)	(0.34)	(0.34)	(0.34)
Several patenting locations	3.28***	1.82***	1.65**	1.62**
	(1.34)	(0.42)	(0.35)	(0.34)
Nbr obs	24064	24064	24064	24064
Pseudo r-sq	0.0916	0.0895	0.0895	0.0893

Logit regressions at the firm level, coefficients are **odds ratios**, standard errors of these odds ratios are in parentheses. The dependent variable is a dummy taking value 1 if a firm is considered to be moving a R&D lab based on the baseline definition. The first four regressors are different centrality measures. Log. Nbr employees is the number of employees the firm has in an average year, Log Nbr R&D employees is the same for R&D employees. Log Nbr plants is the number of plants the firm has in total over the period. Several patenting locations is a dummy variable taking value 1 if the number of commuting zones in which the firm has filed patents is greater than 1. *** p<0.01, ** p<0.05, * p<0.1

roughly half as likely that it decides to move a R&D plant. Note that the coefficients are not comparable across columns because the ranges of the regressors are very different (Bonacich centralities range from 1 to less than 5, meaning that tripling the variable implies a shift from the very bottom to the very top of the distribution, hence the very small odds ratio). Taking very simple network positions provides a rough interpretation of the coefficients shown in columns (3) and (4): they imply that changing the centrality of a firm from 1 (completely isolated firm) to exp(1) $\approx$ 2.7 (a firm connected to 17 firms with no other neighbors) divides the probability that it moves by roughly 20. Table 2.A.6 shown in Appendix emphasizes the fact that local centrality is really the determinant of relocations: taking the same measures of centrality computed over the whole country yields unsignificant effects on the probability of relocating.

Expected centrality and mobility

If firms value knowledge spillovers, they should also consider the links they can expect to form in the location they are moving to. To evaluate this, I construct the same measure for potential local links as in the previous subsection, for all commuting zones. This measures the number of patents issued in technological fields close to those of the moving firm, and captures how interesting the location is to the moving firm.

Then, I use the sample of moving innovative firms which could be matched with patents (i.e. the ones for which I am able to infer the technological area), crossed with all the locations where they could potentially move (for a total of around 150 firms), and create a dummy equal to 1 when they did move to that location, and 0 otherwise. I regress this dummy on firm-level covariates using a conditional logit estimator. The conditional logit structure is standard to estimate location choices across several units where one or several of them are chosen, and is conceptually equivalent to introducing firm fixed-effects in a logit type regression.

Table 2.5 confirms that increasing the potential links a firm could acquire in a location makes it more likely to be chosen for a relocation. Indeed, roughly tripling the number of relevant patents for a mover in a commuting zone doubles the odds that it will be chosen as the destination.

Table 2.5: Potential linkages and location chosen for relocation

	ZE chosen			
	(1)	(2)	(3)	(4)
EZ potential links	1.97***	1.88***		
	(0.06)	(0.13)		
EZ potential links (quality weighted)			1.96***	1.86***
			(0.06)	(0.13)
Empl. Zone FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
Nbr obs	30692	30692	29690	29690
Pseudo r-sq	0.219	0.327	0.218	0.327

Conditional logit regressions, coefficients are *odds ratios*. The dependent variable is the fact for a mover of having an upsized R&D plant in the EZ over the period. Regressors are potential links measures, weighted or not by quality, as decribed in the body of the text. *** p<0.01, ** p<0.05, * p<0.1

Importantly, this result is robust to the introduction of commuting zone fixed-effects, which capture any time-invariant characteristics of these areas such as universities, tax cuts or subsidies

for innovative firms, etc. It is also robust to the introduction of a quality weight on surrounding patents, with nearly equal coefficients as in the non-weighted specification.

2.2.4 Discussion

The above section has uncovered two main facts which help describe how relocation choices by innovative firms relate to the value of local knowledge spillovers.

The first fact developed in subsection 2.2.1 is that firms doing R&D are more mobile than other firms employing high-skilled labor. This higher mobility is found at the intensive margin: the distance at which R&D plants are moved is much larger than other types of plants. It also appears at the extensive margin, both through the fact that in the group of firms employing high-skilled labor, firms doing R&D are more likely to move; and that within firms doing R&D, the ones employing more R&D workers are also more likely to move.

The second fact, explored in subsection 2.2.3, is that firms are sensitive to existing and expected local links in their relocation choices. This is shown along two dimensions. First, the propensity of a firm to relocate is negatively correlated with its existing local connections. Second, a focus on movers shows that they choose to relocate to areas in which they can expect to form a denser network.

In what follows, I build a theoretical framework relying on these novel findings. It formalizes the choice firms have to make about the location of their R&D activities in the presence of strategic complementarities between their own R&D and that of their neighbors. It embeds the ideas that firms benefit only from the effort of neighbors they are locally connected to, and that firms have uncertainty about the quality of research produced by other firms outside of their own location.

2.3 Model

2.3.1 Setup

In the game described below, players start off with a location and a set of links to other players. They get spillovers from these neighbors, through a fraction of their effort.

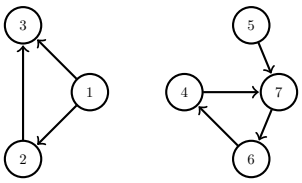
The network $\mathcal{G}$ is an ordered pair $< V, E >$ with a finite set of nodes $V = \{1, \ldots, n\}$ representing players and a set of links E connecting these nodes. It is a directed graph, represented by a $n \times n$ non-symmetric adjacency matrix $\mathbf{G}$. Its elements g_{ij} indicate whether there exists a

link from node i to node j, corresponding to a knowledge transfer from j to i.

As shown in Figure 2.2, citations are mainly found at very short distances. This is translated in a stylized way in the model through the fact that geography strictly constrains the network structure: links may only exist within a given location. There are two locations, meaning that the network is composed of two disconnected subgraphs (an example of which is shown in Figure 2.3) and the adjacency matrix $\boldsymbol{G}$ representing the overall network is block-diagonal:

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{G_1} & \boldsymbol{(0)} \\ \boldsymbol{(0)} & \boldsymbol{G_2} \end{pmatrix}$$

Figure 2.3: Example of two disconnected subgraphs



Payoffs Players have linear-quadratic gross payoffs, which depend on their effort and the efforts of connected players such that:

$$\pi_i = x_i - \frac{1}{2}x_i^2 + \phi_k \sum_{j \in G_k} x_i x_j g_{ij}$$

where firm i is located in subgraph G_k with spillovers intensity ϕ_k, π is its payoff, x the efforts, and g_{ij} are its network links to other firms j. This payoff form has often been used in network games applied to R&D such as König et al. (2018).

A direct consequence of this payoff structure is that the standard result exposed in Ballester et al. (2006) applies: the equilibrium effort of players is equal to their Bonacich centrality in the network.

Definition 1 (Bonacich centrality)**.** *Let $\boldsymbol{G}^p$ be the power p of the adjacency matrix. Its elements $g_{ij}^{[p]}$ record the number of walks of length p between two nodes i, j.*[9] *Define*

$$\boldsymbol{M}(\boldsymbol{G}, \phi) = \sum_{p=0}^{\infty} \phi^p \boldsymbol{G}^p = (\boldsymbol{I} - \phi \boldsymbol{G})^{-1}$$

the sum of walks of length p discounted by the intensity of network effects (< 1) to the power p.

9. i.e. any sequence of nodes starting with i, ending with j and containing $p + 1$ elements.

The (unweighted) vector of Bonacich centralities is then written as:

$$\boldsymbol{b_u}(\boldsymbol{G}, \phi) = \boldsymbol{M}(\boldsymbol{G}, \phi).\boldsymbol{u}$$

with $\boldsymbol{u}$ *the vector of ones.*

The Bonacich centrality has become a classic centrality measure in the games on networks literature,[10] accounting not only for the number of neighbors a node has, but also for its indirect links. Thus, using this centrality measure and following Ballester et al. (2006) yields the ensuing proposition.

Proposition 1 (from Ballester et al. (2006))**.** *For any subgraph* k *with* $\phi_k < \frac{1}{\lambda_{\max}(\boldsymbol{G_k})}, \forall k \in \{1, 2\}$*, there exists a unique interior Nash equilibrium for activity levels equal to the unweighted Bonacich centrality of each player:*

$$\forall k \in \{1, 2\},\ \forall i \in \boldsymbol{G_k},\ x^*_{i|\boldsymbol{G_k}} = b_{u,i}(\phi_k, \boldsymbol{G_k}) \tag{2.1}$$

And the gross payoffs of agents are given by:

$$\forall k \in \{1, 2\},\ \forall i \in \boldsymbol{G_k},\ \pi^*_{i|\boldsymbol{G_k}} = \frac{b_{u,i}(\phi_k, \boldsymbol{G_k})^2}{2} \tag{2.2}$$

where $b_{u,i}(\phi_k, \boldsymbol{G_k})$ *denotes the Bonacich centrality of player* i *in subgraph* k*.*

Expected payoffs Let $\mathcal{I}(-k)$ be the information any player i in subgraph k has about the other subgraph, such that all players in a subgraph have the same information set. Denote $\boldsymbol{G}^{+i}_{\boldsymbol{-k}}$ the receiving subgraph, containing incumbents plus player i.

In this setup, players only differ in their position in the network. Because all nodes from a given subgraph have the same information set, they have the same expected payoffs in the other subgraph. Consequently, firms within a subgraph differ only in their decision to move or not.

Thus, denote $\mathbb{E}[\pi_i(\boldsymbol{G}^{+i}_{\boldsymbol{-k}}) \mid \mathcal{I}(-k)] - c(k)$ the expected payoff a player i belonging to subgraph k has about moving to the subgraph $-k$, conditional on its information set about subgraph $-k$, net of a moving cost which is only a function of the origin subgraph. For given parameter values and initial graph structure, $\mathbb{E}[\pi_i(\boldsymbol{G}^{+i}_{\boldsymbol{-k}}) \mid \mathcal{I}(-k)] - c(k)$ does not depend on i, and takes only two values: the expected payoff of players in subgraph k if they were to move to subgraph $-k$, and

10. Notably Helsley and Zenou, 2014; König et al., 2018; Calvó-Armengol et al., 2009.

the converse.

2.3.2 Stable networks

Stability concept To help the subsequent analysis, I introduce the novel concept of *intergraph stability*. A graph composed of two disconnected subgraphs is said to be intergraph stable if no firm has higher expected payoffs of moving to the other subgraph than realized payoffs in the current one. A more formal definition follows.

Definition 2 (Intergraph Stability)**.** *A graph* $\boldsymbol{G} = (\boldsymbol{G_1}, \boldsymbol{G_2})$ *is intergraph stable if*

$$\forall i \in G_1, \pi_i(\boldsymbol{G_1}) > \mathbb{E}[\pi_i(\boldsymbol{G_2^{+i}}) \mid \mathcal{I}(2)]$$
$$\forall j \in G_2, \pi_j(\boldsymbol{G_2}) > \mathbb{E}[\pi_j(\boldsymbol{G_1^{+j}}) \mid \mathcal{I}(1)]$$

This stability concept substantially differs from the concept often used in network formation problems, namely pairwise stability, as introduced by Jackson and Wolinsky (1996),[11] or Nash stability in the case of directed networks. Stability is here defined at the location-level rather than at the link-level, and therefore does not feature strategic choices of specific partners and the associated stability of links that players form.

This subsection characterizes intergraph stable settings, in which no player would be better off in the other subgraph. It shows that the intergraph stability of any graph of this type can be characterized knowing only the lowest Bonacich centralities in each subgraph and the two values taken by expected payoffs.

Rank agents by increasing centrality, such that agent 1_1 is the agent with lowest centrality out of the n_1 agents in graph $\boldsymbol{G_1}$, and 1_2 is the agent with lowest centrality out of the n_2 agents in graph $\boldsymbol{G_2}$.

Proposition 2. $(\boldsymbol{G_1}, \boldsymbol{G_2})$ *is intergraph stable iff both* 1_1 *is better off staying in* $\boldsymbol{G_1}$ *and* 1_2 *is better off staying in* $\boldsymbol{G_2}$*.*

This proposition provides a simple characterization of intergraph stability: the least central agent in each subgraph will always have the highest propensity to move, because its opportunity

11. Pairwise stability characterizes situations of undirected network formation in which players need to agree for a link to be formed. An intergraph stable network is then a network in which no couple of players would be better off forming a link between them, and no player has an incentive to unilaterally sever a link.

cost of moving is the lowest.[12] As a consequence, knowing the minimum centrality and the value of expected payoffs for both subgraphs is sufficient to characterize the stability of the network.

2.3.3 Game with simultaneous relocations

Relying on the above conditions for subgraphs to be stable, this subsection explores a game where firms are allowed to move simultaneously and characterizes equilibria.

Timing Start from an arbitrary network structure with two disconnected subgraphs in which firms are initially distributed, have linear quadratic payoffs and limited information on the other subgraph, following in every aspect the setting described above. Firms play the following two-stage game:

- **Stage 1:** Firms simultaneously choose whether to move and, in such a case, how many (costly) links l to form;
- **Stage 2:** Firms simultaneously choose their optimal effort x (R&D activity) based on the new network structure.

Moving cost A direct consequence of the imposed network structure is that changing location implies severing all existing links within the original location. Beyond this opportunity cost, moving firms incur two explicit costs: a fixed cost κ (e.g. changing facilities, abandoning prior investments), and a link formation cost $c \cdot l_i^2$ which is quadratic in l_i, the number of links formed (degree) in the new network.[13]

Information Before the game starts, firms in $\boldsymbol{G_k}$ get aggregate information about the structure of the other subgraph: they know $\bar{b}^{G_{-k}}$, the average Bonacich centrality in $\boldsymbol{G_{-k}}$. This assumption restricts the information players have over connections in the other subgraph. Because of the linear-quadratic payoff structure, Proposition 1 applies, and the average Bonacich centrality is simply equal to the average R&D activity in the other location.

12. A direct consequence of Proposition 2 is that if $\exists\, i \geq 1$ such that i_k would profitably move to $\boldsymbol{G_{-k}}$, then $\forall\, j < i$, j_k would also move to $\boldsymbol{G_{-k}}$.

13. Convexity in the cost of links only avoids trivial equilibria, in which a firm deciding to move would always form a link with every other firm in its new location: since firms have to pay a fixed cost κ of moving, plus a unit cost per link they decide to form, linear linking costs imply that the marginal cost of a link will always be smaller than the average cost of a link.

Thus, a firm i moving from subgraph k to subgraph $-k$ receives the following net payoffs:

$$x_{i|\boldsymbol{G_{-k}}} - \frac{1}{2}x_{i|\boldsymbol{G_{-k}}}^2 + \phi_{-k}\sum_{j\in N_i} x_i x_j g_{ij} - \kappa - cl_i^2$$

Equilibria The unique equilibrium of the second stage of the game is straightforward from Proposition 1 displayed in the above subsection: firms' equilibrium effort level is equal to the Bonacich centrality in their subgraph.

Turning to the first stage, a crucial result in order to characterize equilibrium relocation decisions is the effect that movers have on incumbents in their new subgraph. From the structure of the game, incumbent firms are not allowed to revise their linkages and form links towards entrants. Consequently, all the walks involving the entrants are the ones starting from them, which can be formalized through the following lemma.

Lemma 1. *In a given subgraph, the activity levels of incumbent firms are unaffected by entrants.*

This property ensures that potentially moving firms do not internalize any effect of their own entry on the average activity of the subgraph they are about to join.

Based on Lemma 1, a firm moving from $\boldsymbol{G_1}$ to $\boldsymbol{G_2}$ would get the expected net payoff

$$x_{i|\boldsymbol{G_2}} - \frac{1}{2}x_{i|\boldsymbol{G_2}}^2 + \phi_2 x_{i|\boldsymbol{G_2}} l_i \bar{b}^{G_2} - c.l_i^2 - \kappa \tag{2.3}$$

where $l_i\bar{b}^{G_2}$ is the expected effort of new neighbors.

Therefore, the first order condition is:

$$\frac{\partial \pi_{i|\boldsymbol{G_2}}}{\partial x_{i|\boldsymbol{G_2}}} = 1 - x_{i|\boldsymbol{G_2}} + \phi_2 l_i \bar{x}^{G_2} = 0$$

Which yields a firm i's ex ante best response activity level in case it moves from $\boldsymbol{G_1}$ to $\boldsymbol{G_2}$:

$$x_{i|\boldsymbol{G_2}}^{BR} = 1 + \phi_2 l_i \bar{b}^{G_2} \tag{2.4}$$

Plugging (2.4) into (2.3), this yields a simple second-degree polynomial function in l_i, the movers' choice variable:

$$\frac{1}{2}(\phi_2^2 - c)l_i^2 + \phi_2 \bar{x}^{G_2} l_i - \kappa - \frac{1}{2}[b_{u,i}(\phi_1, G_1)^2 - 1] \tag{2.5}$$

Clearly, this expression only depends on i through $b_{u,i}(\phi_1, G_1)$, which shifts i's incentive to

move up or down. Therefore, there exists a number of links $l^*_{G_2}$ which maximizes the above expression for any firm in a given subgraph.[14] In other words, firms are heterogeneous in their opportunity cost of moving, but make the exact same decision once they decide to move.

A coordination problem arises once it appears that the network is not intergraph stable, because the network structure is common knowledge within a subgraph. Firms can anticipate movements of other firms in their own subgraph, and the resulting changes in their own centrality which depend on how tightly connected they are to movers.

Thus, several equilibria may exist in this situation in which some firms simultaneously decide to move and do not have *ex post* incentives to deviate. Typically, under reasonably low moving costs, the situation in which all firms move is always an equilibrium: a deviating firm would remain alone and isolated in its original subgraph, which can never be profitable if a connected firm has found it profitable to move in the first place. Using the definition of stability explained in the previous subsection, all equilibria have to be *ex post* stable: no firm should want to change subgraph at the end of the game. Similarly, any intergraph stable network with the corresponding vector of effort levels is an equilibrium. This directly yields the following proposition.

Proposition 3. *In any PBE, the resulting network* $\boldsymbol{G}$ *is intergraph stable and the vector of R&D efforts is the vector of Bonacich centralities* $\boldsymbol{b_u(\phi, G)}$.

Notably, there is one equilibrium which is of particular interest, namely the equilibrium with the minimal number of movements, in which only firms that *ex ante* had higher expected payoffs in the other subgraph than in theirs end up moving (*i.e.* that move irrespective of their neighbors' decisions). In what follows, I show that such an equilibrium always exists, and has special properties: it is both welfare-maximizing and Pareto-optimal.

Welfare-maximizing equilibrium An important property of Bonacich centrality is that it weakly decreases for a given node when any link in the subgraph is severed, even if that node is not directly affected by the severance of this link. Therefore, a firm's centrality in a given subgraph decreases in the number of firms moving away, which means that its effort and payoff also weakly decrease as other firms decide to move. Thus, if moving is a best-response for a firm conditional to a given set of other firms moving from its own subgraph, it will remain so if more firms are added to this set. Thus, when movements are simultaneous, there is no benefit

14. $\forall i$, a maximum is reached in $\frac{\phi_2 \bar{x}^{G_2}}{2c-\phi_2^2(\bar{x}^{G_2})^2}$, so $l^*_{G_2} = \lfloor \frac{\phi_2 \bar{x}^{G_2}}{2c-\phi_2^2(\bar{x}^{G_2})^2} \rfloor$ or $l^*_{G_2} = \lfloor \frac{\phi_2 \bar{x}^{G_2}}{2c-\phi_2^2(\bar{x}^{G_2})^2} \rfloor + 1$.

to a firm of its neighbors moving: movers produce a negative externality on the subgraph they leave while they produce no externality on the subgraph they join. Therefore, the equilibrium with the minimum possible number of movements is welfare-maximizing: forcing an additional movement would harm the firm and its origin subgraph without benefiting incumbents in the receiving subgraph.

Describing this equilibrium requires a measure of dependence of a node's centrality to each node in its subgraph. Lemma 2 extends the intercentrality concept in Ballester et al. (2006) and establishes new individual Bonacich centralities if a firm leaves.

Lemma 2. *The activity level of a player j in subgraph k after player i has left the subgraph is:*

$$b_{u,j}(G_k^{-i}, \phi_k) = b_{u,j}(G_k, \phi_k) - \frac{m_{ji}(G_k, \phi_k)}{m_{ii}(G_k, \phi_k)} b_{u,i}(G_k, \phi_k) \tag{2.6}$$

where the m_{ij} are, again, the self or non self loops, ie the terms of the matrix $\boldsymbol{M} = (\boldsymbol{I} - \phi \boldsymbol{G})^{-1}$.

To show existence of an equilibrium in which firms do not all move, start from an inter-graph stable network, and let ϕ_2 be hit by a positive shock before the game starts, making $\boldsymbol{G_2}$ more attractive. Again, order firms by increasing centrality in each subgraph. For the sake of simplicity, since firms in this context may only move from $\boldsymbol{G_1}$ to $\boldsymbol{G_2}$, the subgraph index is dropped.

Although firm 2 has the second lowest centrality in $\boldsymbol{G_1}$ before the game starts, hence the second highest *ex ante* expected gains from moving, it may not remain so once 1 moves. Indeed, the centrality of other nodes may be more dependent on paths going through 1, implying larger centrality drops caused by 1's exit for these firms. It is however possible to construct iteratively a new ranking of firms in $\boldsymbol{G_1}$, $\{o_e\}_{e \in \{1,\dots,n_1\}}$, which is a simple permutation of $\{1, \dots, n_1\}$, such that firm e would have the lowest centrality in $\boldsymbol{G_1}$ conditional on firms up to o_{e-1} moving.

Denoting $f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2)$ the part of the expected payoffs for $\boldsymbol{G_1}$ firms from moving to $\boldsymbol{G_2}$ that is independent of $\boldsymbol{G_1}$ network structure yields the following proposition.

Proposition 4 (Welfare-maximizing equilibrium)**.** *$\forall e \in \{1, ..., n_1 - 2\}$, there exists a welfare-maximizing Bayesian perfect equilibrium, in which firms $\{o_1, \dots, o_e\}$ move to $\boldsymbol{G_2}$, form $l^*_{G_2}$ links each with incumbent $\boldsymbol{G_2}$ firms, after what all firms choose an effort level equal to their Bonacich*

centrality iff

$$\begin{cases} f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2) - \kappa - \frac{1}{2}\left\{[b_{u,o_e}(\boldsymbol{G_1}^{-\{o_1,\ldots,o_{e-1}\}}, \phi_1)]^2 - 1\right\} > 0 \\ f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2) - \kappa - \frac{1}{2}\left\{[b_{u,o_{e+1}}(\boldsymbol{G_1}^{-\{o_1,\ldots,o_e\}}, \phi_1)]^2 - 1\right\} < 0 \end{cases} \tag{2.7}$$

i.e. iff firm o_e finds it profitable to move but firm o_{e+1} does not.[15]

In Appendix, I show that this welfare-maximizing equilibrium is also Pareto-optimal.

2.4 Sequential relocations and cascading effects

While the above section studies a game in which movements are simultaneous, a more realistic setting is one where movements are sequential, such that potential movers observe past movements and update their information about the other location. In such setting, the expected quality of potential neighbors changes in a subgraph when firms start joining in sequentially. Notably, it could decrease if joining firms form few links and end up having a lower Bonacich centrality/activity than the average. A real world analogy would be that compared to the original Silicon Valley setting from which all the leading tech companies emerged, many firms have joined, among which a number of them probably have a low potential and are not high-value contacts. This makes the signal on whether other firms should try to join the Silicon Valley ambiguous, because the neighbors they will end up having may not be particularly valuable even though the area as a whole is successful. This would be a balancing force, because if firms with low connection levels decide to leave and form only a limited set of links, the location from which firms initially exit will see its average quality increase, and may in turn become attractive to firms in the receiving location. The ensuing section explores this question.

2.4.1 Environment

Timing Start again from an intergraph stable network, in which firms in each subgraph are myopic and learn the average centrality in the other subgraph before the game starts, and let ϕ_2 be hit by a positive shock. Slightly modify the previous timing to the following:

1. Every round, a firm from $\boldsymbol{G_1}$ is randomly picked and may decide to move, in which case it chooses how many links l to form;

15. For $e = n_1 - 1$, only the first part of the condition needs to be filled for all firms to move. For $e = 1$, if the first part of the condition is not met, the graph is intergraph stable.

2. All firms update their optimal effort level x.

3. Firms collect profits and update their belief about $\bar{x}^{\boldsymbol{G}_{-k}}$ (average effort in the other subgraph).

Discussion A fundamental change between the sequential and the simultaneous moves environments is that in the sequential case, moving firms at a later round may end up being linked to movers from earlier rounds. This may generate *cascade* phenomena (as in the financial networks studied in Elliott et al., 2014). Indeed, the exit of a small set of firms leads to a centrality loss while it may increase the average centrality in the competing network. This may lead to the exit of more firms from their initial network, whereas these firms had no incentive to move in the simultaneous moves context. The opposite phenomenon may also occur: because peripheral firms are the most likely to leave, their exit may translate into a positive effect on the average activity in their original network, which may be a stronger effect than the one they have on the receiving subgraph.

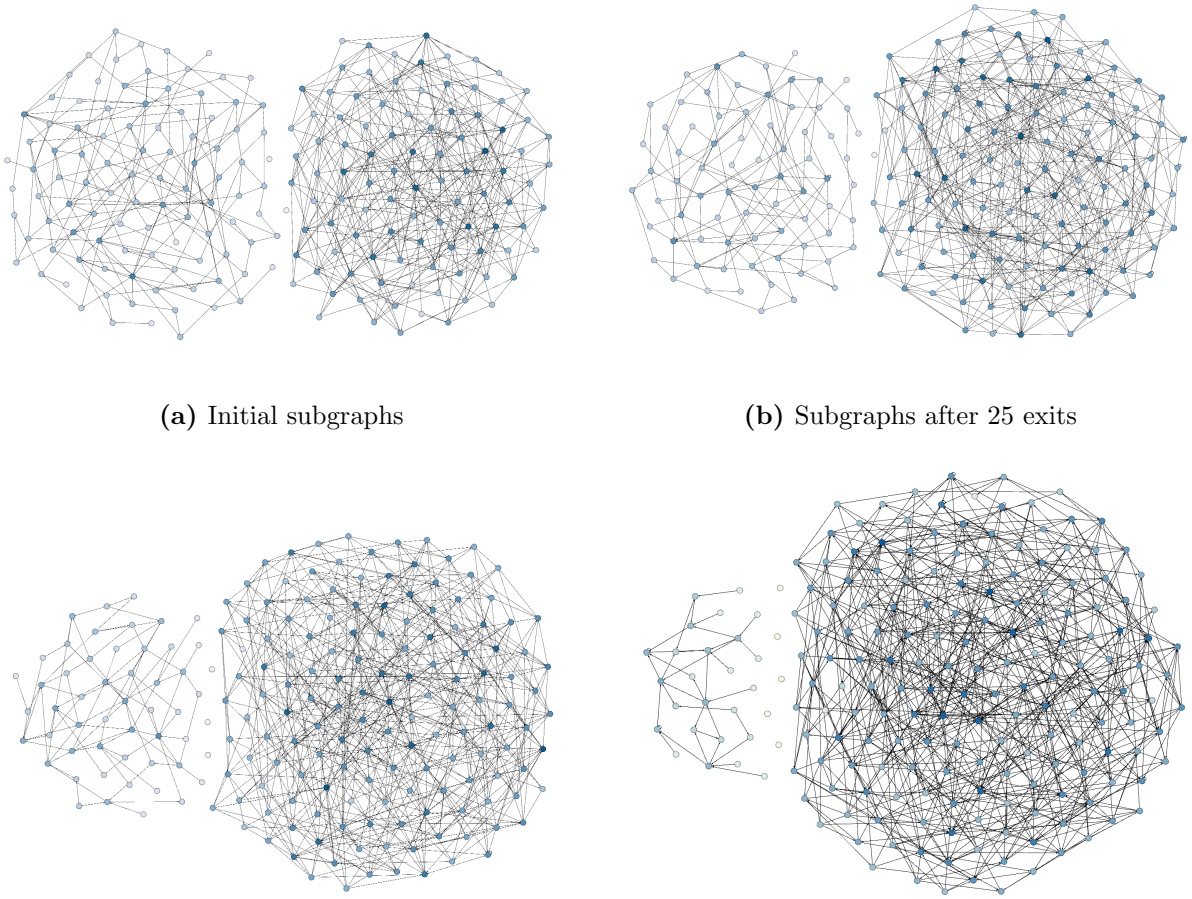
In contrast, a caveat of this framework to analyze dynamic phenomena is that only movers are allowed to revise their links, while incumbent firms cannot form connections with entrants. Nevertheless, provided the initial graph is Nash stable (no firm is better off severing or creating a link), a setting allowing firms to change their linkages after a movement has taken place is likely to enhance the cascade. Indeed, firms in the receiving networks forming links with the entrant should increase the average centrality of the receiving subgraph, while firms in the exited subgraph would have no incentive creating new links, resulting in a decrease in the average centrality of the subgraph. Consequently, this framework can be conceived as a lower bound of the extent of a cascade phenomenon taking place between two graphs.

Note that total activity in $\boldsymbol{G_1}$ decreases, and total activity in $\boldsymbol{G_2}$ always increases when a firm leaves $\boldsymbol{G_1}$ for $\boldsymbol{G_2}$. Overall, we may fall into four different cases:

1. ($\bar{x}^{\boldsymbol{G_1}} \searrow, \bar{x}^{\boldsymbol{G_2}} \nearrow$): The exiting firm decreases average activity in $\boldsymbol{G_1}$ and creates enough links as to increase it in $\boldsymbol{G_2}$: in this case the process is strongly amplified for the next step (making it more likely that another firm will move);

2. ($\bar{x}^{\boldsymbol{G_1}} \nearrow, \bar{x}^{\boldsymbol{G_2}} \searrow$): The exiting firm increases average activity in $\boldsymbol{G_1}$ and decreases it in $\boldsymbol{G_2}$: this is an interesting case based on the fact that firms with low centrality are more likely to move. Their exit could foster an increase in the average activity in $\boldsymbol{G_1}$, and make

Figure 2.4: Example of a cascade: following a spillovers shock, the left-hand side location empties as firms move to the right-hand side one.



(a) Initial subgraphs

(b) Subgraphs after 25 exits

(c) Subgraphs after 50 exits

(d) Subgraphs after 75 exits

it more attractive. Coincidentally, it may be that the cost of links is high enough for the moving firm to form few links or that it is unlucky enough to end up being linked only with very peripheral firms, such that this move decreases average activity in $\boldsymbol{G_2}$. This case is the converse of case 1, and could even trigger a reverse cascade if firms from $\boldsymbol{G_2}$ were allowed to move: then firms of $\boldsymbol{G_2}$ would be better off moving to $\boldsymbol{G_1}$ even though their subgraph was the one benefiting from a positive spillovers shock;

3. $(\bar{x}^{\boldsymbol{G_1}} \searrow, \bar{x}^{\boldsymbol{G_2}} \searrow)$: The exiting firm decreases average activity in $\boldsymbol{G_1}$ but creates few links which also implies a decrease in the average activity in $\boldsymbol{G_2}$: then the sign of the effect of this step on the next step of the process is undetermined, as the relative magnitude of both effects will lead the process either to be amplified or weakened;

4. $(\bar{x}^{\boldsymbol{G_1}} \nearrow, \bar{x}^{\boldsymbol{G_2}} \nearrow)$: The exiting firm creates enough links to increase average activity in

$\boldsymbol{G_2}$ but was peripheral enough as to also increase average activity in $\boldsymbol{G_1}$: this is again undetermined.

The following subsection characterizes the above-listed cases.

2.4.2 Cascades

Effect of a move on the exited subgraph The overall effect of the exit of any firm in $\boldsymbol{G_1}$ can be assessed using the intercentrality concept defined in Ballester et al. (2006).

Definition 3 (from Ballester et al. (2006))**.** *For directed graphs (*i.e. *non symmetric adjacency matrices), the intercentrality is defined as:*

$$c_i(\boldsymbol{G}, a) = b_{u,i}(\boldsymbol{G}, a)\frac{\sum_{j=1}^{n} m_{ji}(\boldsymbol{G}, a)}{m_{ii}(\boldsymbol{G}, a)} \tag{2.8}$$

The intercentrality is the total number of walks that hit i, so it is the sum of i's Bonacich centrality, and i's contribution to other players' Bonacich centralities. Therefore, we know that removing firm 1 from $\boldsymbol{G_1}$ means:

$$\sum_{j=2}^{n_1} b_{u,j}(\boldsymbol{G_1}^{-1}, \phi_1) = \sum_{j=1}^{n_1} b_{u,j}(\boldsymbol{G_1}, \phi_1) - c_1(\boldsymbol{G_1}, \phi_1) \tag{2.9}$$

Proposition 5. *For i's exit to have a negative effect on average activity in $\boldsymbol{G_1}$, the following condition needs to be met:*

$$\bar{x}^{\boldsymbol{G_1}} < c_i(\boldsymbol{G_1}, \phi_1)$$

i.e. the intercentrality of player i (contribution to the network's aggregate activity) must be higher than the average activity in this network.

In the Ballester et al. (2006) context, the argmax of the intercentrality yields the node that will maximally decrease total activity when removed (it is the argmin of total activity net of this measure). In our context, the intercentrality of the exiting node tells us the exact importance of this node for the subgraph it belonged to before the exit. Therefore, for the effect of a move on the average centrality to be negative, it should be that the contribution of the mover to all centralities outweighs the effect of its exit on the average centrality of the subgraph (*i.e.* the fact that the mean centrality now is calculated over $n_1 - 1$ players rather than n_1).

Effect of a move on the receiving subgraph Based on Lemma 1, the activity vector of firms in the receiving subgraph, indexing by n_2+1 the firm moving from $\boldsymbol{G_1}$ to $\boldsymbol{G_2}$, writes

$$(\boldsymbol{x}^{G_2'})^* = \begin{pmatrix} b_{u,1}(\boldsymbol{G_2}, \phi_2) \\ \vdots \\ b_{u,n_2}(\boldsymbol{G_2}, \phi_2) \\ 1 + \phi_2 \sum_{j \in N_{n_2+1}} b_{u,j}(\boldsymbol{G_2}, \phi_2) \end{pmatrix} \tag{2.10}$$

As shown in Lemma 1, the activity of incumbent firms in the receiving subgraph remains unchanged, and the activity of the entrant can be expressed simply through a sum of its new neighbors' centralities discounted with a factor ϕ_2.

Amplifying effect of an exit Assembling the above results indicates which of the cases exposed earlier we will fall into after a movement, based on the graph structure, the parameter values, and the firm's new neighbors.

Proposition 6. *The cascade is strengthened after i's exit if either:*

- *Average activity has increased in $\boldsymbol{G_2}$ and has decreased in $\boldsymbol{G_1}$ (denoted case 1 in subsection 2.4.1), which happens if both $n_2 + \sum_{i=1}^{n_2}(\phi_2 n_2 g_{n_2+1,i} - 1) b_{u,i}(\boldsymbol{G_2}, \phi_2) > 0$ and $\bar{x}^{\boldsymbol{G_1}} < c_i(\boldsymbol{G_1}, \phi_1)$;*

- *Average activity has decreased in both subgraphs but decreased more in subgraph 1 (case 3), which happens if $\frac{\bar{x}^{\boldsymbol{G_1}} - c_i(\boldsymbol{G_1}, \phi_1)}{n_1 - 1} < \frac{n_2 + \sum_{i=1}^{n_2}(\phi_2 n_2 g_{n_2+1,i} - 1) b_{u,i}(\boldsymbol{G_2}, \phi_2)}{n_2(n_2+1)} < 0$;*

- *Average activity has increased in both subgraphs but increased more in subgraph 1 (case 4), which happens if $\frac{n_2 + \sum_{i=1}^{n_2}(\phi_2 n_2 g_{n_2+1,i} - 1) b_{u,i}(\boldsymbol{G_2}, \phi_2)}{n_2(n_2+1)} > \frac{\bar{x}^{\boldsymbol{G_1}} - c_i(\boldsymbol{G_1}, \phi_1)}{n_1 - 1} > 0$;*

The conditions for the cascade to be weakened are exactly opposite to the above ones.

Proof. Assembling results from previous subsections. □

The above proposition shows that transitions from one step to the other of the cascade can be described using only simple conditions on the structure of the graph.

2.4.3 Long-run Equilibrium

This subsection intends to outline a situation in which the cascade is likely to stop, and a long-run equilibrium emerges in which all possible movements have occurred. In every round, when

a set $\{r_1, \ldots, r_e\}$ of firms has moved, there exists a unique optimal number of links, common to every $\boldsymbol{G_1}$ firm, here denoted $l^*_{G_2+\{r_1,\ldots,r_e\}}$. The expected centrality of the first entrant writes

$$E(x_{r_1}) = 1 + \phi_2 l^*_{G_2} \bar{x}^{G_2}$$

Since firms learn about the average activity in the other subgraph before the game starts, the expected centrality of the e^{th} moving firm in the sequence for incumbent firms in the exited subgraph can be written

$$E(x_{r_e}) = 1 + \phi_2 l^*_{G_2+\{r_1,\ldots,r_{e-1}\}} \bar{x}^{G_2+\{r_1,\ldots,r_{e-1}\}}$$

Thus, at any point of the sequence, firms in $\boldsymbol{G_1}$ take their moving decisions based on the following expected average activity in $\boldsymbol{G_2}$

$$\begin{aligned}
&E(\bar{x}^{G_2+r_1}) = \frac{n_2 \bar{x}^{G_2} + 1 + \phi_2 l^*_{G_2} \bar{x}^{G_2}}{n_2 + 1} \\
&E(\bar{x}^{G_2+\{r_1,r_2\}}) = \frac{(n_2+1)E(\bar{x}^{G_2+r_1}) + 1 + \phi_2 l^*_{G_2+r_1} \bar{x}^{G_2}}{n_2+2} = \frac{n_2 \bar{x}^{G_2} + 2 + \phi_2 (l^*_{G_2} + l^*_{G_2+r_1}) \bar{x}^{G_2}}{n_2+2} \\
&\vdots \\
&E(\bar{x}^{G_2+\{r_1,\ldots,r_e\}}) = \frac{n_2 \bar{x}^{G_2} + e + \phi_2 (\sum_{k=1}^{e-1} l^*_{G_2+\{r_1,\ldots,r_k\}}) \bar{x}^{G_2}}{n_2 + e}
\end{aligned}$$

This means that with every additional movement, average activity sums up to a linear function of the initial average activity in the receiving subgraph and the number of links movers decide to create in each moving round, over the number of firms in that subgraph after the last movement.

Proposition 7 (Sequential Moves Equilibrium)**.** *When the number of rounds played is large enough,* $\forall e \in \{2, ..., n_1 - 2\}$*, there exists a unique Bayesian perfect equilibrium in which firms* $\{r_1, \ldots, r_e\}$ *have moved to* $\boldsymbol{G_2}$*, and all others will stay in* $\boldsymbol{G_1}$ *whatever the additional number of rounds iff*

$$\begin{cases}
f\left[E(\bar{x}^{G_2+\{r_1,\ldots,r_{e-1}\}}), l^*_{G_2+\{r_1,\ldots,r_{e-1}\}}, \phi_2\right] - \kappa - \frac{1}{2}\left[b_{u,r_e}(\boldsymbol{G_1}^{-\{r_1,\ldots,r_{e-1}\}}, \phi_1)^2 - 1\right] > 0 \\
f\left[E(\bar{x}^{G_2+\{r_1,\ldots,r_e\}}), l^*_{G_2+\{r_1,\ldots,r_e\}}, \phi_2\right] - \kappa - \frac{1}{2}\left[b_{u,r_{e+1}}(\boldsymbol{G_1}^{-\{r_1,\ldots,r_e\}}, \phi_1)^2 - 1\right] < 0
\end{cases}$$

Additionally, for $e = n_1 - 1$, only the first part of the condition needs to be filled for all firms to move in the long-run. For $e = 1$, if the first part of the condition is not met, the graph is

intergraph stable.

While Proposition 7 provides a result that is very close to the one presented for the simultaneous moves case in Proposition 4, it adds a fundamental feature through the role of changes in expected average activity in the other subgraph, which will greatly condition the number of moves. Indeed, as shown in the overall results, if average activity decreases in $\boldsymbol{G_2}$ as firms move because the cost of forming new links is high, this may result in a lower number of movers than in the simultaneous moves case. If however the average activity increases as firms move, this could imply a self-reinforcing cascade with a much larger number of movers than in the simultaneous moves case, if not a unique equilibrium in which all firms moves.

2.5 Conclusion

This paper has sought to shed light on the importance of local knowledge networks for innovative firms through the study of their relocation choices. Relocations are an interesting phenomenon to answer the wider question of how firms internalize knowledge spillovers in their location choices, because relocating firms are facing a clear trade-off between their current and their future location, and which already have a defined technological field.

I study relocations in a flexible way, looking at plant-level employment changes in opposite directions in different cities. Based on this definition, I show that being an innovative firm (*i.e.* employing at least a researcher in an average year) makes a firm 70% more likely to relocate its high-skilled workforce. Similarly, tripling the number of R&D employees a firm has makes it twice as likely that it relocates its R&D employment over the period. I also show that the distance of R&D relocations is far larger than that of other types of employment. These facts suggest that R&D activities are more mobile than others, which can be linked to the importance of localized knowledge spillovers for such activities. The paper gives evidence on the localization of knowledge networks on France in line with the literature, and then shows that firms that are less central in their local innovation network are more likely to relocate R&D activities. What is more, these moving firms are more likely to pick areas in which their potential for collaborations and local linkages is higher.

Relying on these empirical facts, the paper then develops a theoretical framework which intends to link relocations to local knowledge networks. It builds a model in which firms inherit a position in a network with two disconnected subgraphs, and may sever their links to join the

other part of the network. They then choose the level of R&D effort that maximizes their payoff. To characterize equilibria, I first introduce a novel stability concept called intergraph stability, where no firm wants to move to the other subgraph, and show that it can be characterized simply knowing the minimum Bonacich centrality in each subgraph and the expected payoff in the other one. Furthermore, in a game where firms first choose simultaneously whether to relocate or not, and then choose their optimal R&D effort, all PBEs can be characterized using this stability concept. I then prove existence and describe extensively the welfare-maximizing and Pareto-optimal equilibrium, which is the one with the minimal number of movements.

The paper then turns to the analysis of a similar game in which the timing is changed to allow for sequential movements: a firm is chosen randomly every round and decides whether to move or not, after what all firms adjust their R&D efforts. In such setting, it describes the conditions for a movement to either strengthen or weaken a *cascade* phenomenon by either increasing or decreasing the incentives to move of the firms located in both the exited and the receiving subgraph. It then shows the existence of a long-term equilibrium in a setting where only firms from one subgraph are picked to move. Depending on the links that movers end up forming with incumbents, equilibria with both a larger and a smaller number of movements than in the simultaneous moves game may be achieved.

This paper leaves several points of inquiry for further research. Indeed, it would be interesting to causally estimate how an exogenous shock to the network structure in a given location affects relocations. Nevertheless, such an empirical setting is hard to find: it requires that the shock affects the network structure strongly enough to foster relocations, yet without affecting the characteristics of the location where it hits to allow identification.

References

Alcácer, J., and W. Chung. 2007. "Location Strategies and Knowledge Spillovers." *Management Science* 53 (5): 760–776.

Alcacer, J., and M. Delgado. 2016. "Spatial organization of firms and location choices through the value chain." *Management Science* 62 (11): 3213–3234.

Almeida, P., and B. Kogut. 1999. "Localization of knowledge and the mobility of engineers in regional networks." *Management science* 45 (7): 905–917.

Audretsch, D. B., and M. P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (3): 630–640.

Ballester, C., A. Calvó-Armengol, and Y. Zenou. 2006. "Who's Who in Networks. Wanted: The Key Player." *Econometrica* 74 (5): 1403–1417.

Behrens, K., G. Duranton, and F. Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy* 122 (3): 507–553.

Bergeaud, A., and S. Ray. 2017. *Adjustment Costs and Factor Demand: New Evidence From Firms' Real Estate.* Working papers 641. Banque de France.

Breschi, S., and F. Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of Economic Geography:* lbp008.

Buzard, K., G. A. Carlino, R. M. Hunt, J. K. Carr, and T. E. Smith. 2017. "The agglomeration of American R&D labs." *Journal of Urban Economics* 101:14–26.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou. 2009. "Peer Effects and Social Networks in Education." *Review of Economic Studies* 76 (4): 1239–1267.

Carlino, G., and W. R. Kerr. 2015. "Chapter 6 - Agglomeration and Innovation." In *Handbook of Regional and Urban Economics,* edited by J. V. H. Gilles Duranton and W. C. Strange, 5:349–404. Handbook of Regional and Urban Economics. Elsevier.

Catalini, C. 2017. "Microgeography and the direction of inventive activity." *Management Science* 64 (9): 4348–4364.

Chaney, T. 2014. "The network structure of international trade." *The American Economic Review* 104 (11): 3600–3634.

———. 2018. "The Gravity Equation in International Trade: An Explanation." *Journal of Political Economy* 126 (1): 150–177.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection." *Econometrica* 80 (6): 2543–2594.

Cotterlaz, P., and A. Guillouzouic. 2018. *The Percolation of Knowledge across Space.* Technical report. Working paper.

Davis, D. R., and J. I. Dingel. 2019. "A Spatial Knowledge Economy." *American Economic Review* 109 (1): 153–70.

Duranton, G. 2011. "California Dreamin': The Feeble Case for Cluster Policies." *Review of Economic Analysis* 3 (1): 3–45.

Duranton, G., and D. Puga. 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products." *American Economic Review* 91 (5): 1454–1477.

Elliott, M., B. Golub, and M. Jackson. 2014. "Financial Networks and Contagion." *American Economic Review* 104 (10): 3115–53.

Fitjar, R. D., and A. Rodríguez-Pose. 2017. "Nothing is in the Air." *Growth and Change* 48 (1): 22–39.

Gaubert, C. 2018. "Firm Sorting and Agglomeration." *American Economic Review* 108 (11): 3117–53.

Head, K., Y. A. Li, and A. Minondo. 2018. "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics." *Review of Economics and Statistics,* no. forthcoming.

Helsley, R., and Y. Zenou. 2014. "Social networks and interactions in cities." *Journal of Economic Theory* 150 (C): 426–466.

Henry, E., J. Monras, and G. Spanos. 2017. *The public sector multiplier in R&D.* Technical report.

Jackson, M. O., and B. W. Rogers. 2005. "The economics of small worlds." *Journal of the European Economic Association* 3 (2-3): 617–627.

Jackson, M. O., and A. Wolinsky. 1996. "A strategic model of social and economic networks." *Journal of Economic Theory* 71 (1): 44–74.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *the Quarterly journal of Economics:* 577–598.

Kerr, W. R., and S. D. Kominers. 2015. "Agglomerative forces and cluster shapes." *Review of Economics and Statistics* 97 (4): 877–899.

König, M. D., X. Liu, and Y. Zenou. 2018. "R&D Networks: Theory, Empirics and Policy Implications." *Review of Economics and Statistics,* no. forthcoming.

Singh, J. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5): 756–770.

Appendix

2.A Data

2.A.1 R&D employment

Table 2.A.1: Selected job categories (PCS) corresponding to R&D activities

Engineers	"383A"	R&D engineers in electricity and electronics
	"384A"	R&D engineers in mechanics and metal-work
	"385A"	R&D engineers in transformation industries (agro-industry, chemistry, metal-work, heavy material)
	"386A"	R&D engineers in other industries
	"388A"	R&D engineers in information technology
Technicians	"473B"	R&D and fabrication methods technicians for electricity, electromechanics and electronic industries
	"474B"	R&D and fabrication method technician in mechanical construction and metal-work
	"475A"	R&D and production methods technician for transformation industries
	"478A"	Study and development technicians in IT.

2.A.2 Patent applicants - Firms matching

To match patents applicants with SIREN firm identifiers, I follow two complementary methods:

- I match the names of applicants with firm names taken from Sirene 2017, simplifying the strings with the algorithms provided by the NBER patent project.
- Additionally, I use a web-scrapping method, in the spirit of Autor et al. (2016): I send requests to a search-engine with the applicant name in the patent database, and the constraint that the search be conducted on websites listing all firms declared in France (societe.com, verif.com, infogreffe.fr, bilansgratuits.fr and score3.fr). These websites have

the advantage of containing SIREN identifiers directly in the url associated to each firm. I then use validation across these websites to avoid having many false positives.

2.B Additional Tables and Figures

2.B.1 Desc. Stat on mobility

Table 2.A.2: Descriptive Statistics: Full Relocations (as in Duranton and Puga, 2001)

Category	All firms	HS firms	Innov. firms
Nbr of firms in category	689,691	301,571	24,064
Nbr of moving firms	14,616	7,146	421
Av. nbr of plants in cat. in 2006	7.14	9.73	4.01
Av. nbr of employees	172.1	308.4	954.1
Share of firms moving	2.12	2.37	1.75
Av. relocation distance	121	130.5	168.7

The above table shows descriptive statistics of full relocations of plants with at least one full-time employee of the relevant category for 3 categories of firms: all firms, firms employing high-skilled workers (HS firms), and innovative firms. A firm enters a category if at least one of its plants employs at least one of the corresponding type of workers. Firms are considered moving if they relocate a plant of their category: an innovative firm relocating a plant employing high skilled but non-R&D workers is counted as a moving HS firm but not as a moving innovative firm. Numbers of employees and plants are averages calculated over the period covered by the data, i.e. 2006-2015.

Table 2.A.3: Descriptive Statistics: Relocations of at least 50% and at least two full-time employees of each category

Category	All firms	HS firms	Innov. firms
Nbr of firms in category	688,201	299,855	23,942
Nbr of moving firms	13,185	6,512	494
Av. nbr of plants in cat. in 2006	7.71	10.51	4.30
Av. nbr of employees	192.5	353.1	1047
Share of firms moving	1.92	2.17	2.06
Av. relocation distance	122.1	135.1	183.4

The above table shows descriptive statistics of partial relocations of at least 50% and two full-time employees of a plant for 3 categories of firms: all firms, firms employing high-skilled workers (HS firms), and innovative firms. A firm enters a category if at least one of its plants employs at least one of the corresponding type of workers. Firms are considered moving if they relocate a plant of their category: an innovative firm relocating a plant employing high skilled but non-R&D workers is counted as a moving HS firm but not as a moving innovative firm. Numbers of employees and plants are averages calculated over the period covered by the data, i.e. 2006-2015.

Table 2.A.4: Descriptive Statistics: Relocations of at least 25% and at least one full-time employees of each category

Category	All firms	HS firms	Innov. firms
Nbr of firms in category	689,691	301,571	24,064
Nbr of moving firms	16,958	9,375	773
Av. nbr of plants in cat. in 2006	6.72	8.29	3.80
Av. nbr of employees	163.2	276.1	844.3
Share of firms moving	2.46	3.11	3.21
Av. relocation distance	123.6	136	208.9

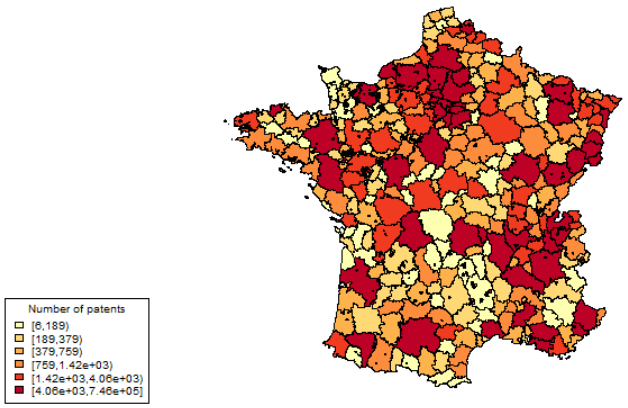
The above table shows descriptive statistics of partial relocations of at least 25% and one full-time employee of a plant for 3 categories of firms: all firms, firms employing high-skilled workers (HS firms), and innovative firms. A firm enters a category if at least one of its plants employs at least one of the corresponding type of workers. Firms are considered moving if they relocate a plant of their category: an innovative firm relocating a plant employing high skilled but non-R&D workers is counted as a moving HS firm but not as a moving innovative firm. Numbers of employees and plants are averages calculated over the period covered by the data, i.e. 2006-2015.

Table 2.A.5: Descriptive Statistics: Relocations of at least 75% and at least one full-time employees of each category

Category	All firms	HS firms	Innov. firms
Nbr of firms in category	689,691	301,571	24,064
Nbr of moving firms	15,061	7,644	540
Av. nbr of plants in cat. in 2006	7.03	9.49	3.93
Av. nbr of employees	169.6	301.2	875.2
Share of firms moving	2.18	2.53	2.24
Av. relocation distance	121.3	129.9	179.4

The above table shows descriptive statistics of partial relocations of at least 75% and one full-time employee of a plant for 3 categories of firms: all firms, firms employing high-skilled workers (HS firms), and innovative firms. A firm enters a category if at least one of its plants employs at least one of the corresponding type of workers. Firms are considered moving if they relocate a plant of their category: an innovative firm relocating a plant employing high skilled but non-R&D workers is counted as a moving HS firm but not as a moving innovative firm. Numbers of employees and plants are averages calculated over the period covered by the data, i.e. 2006-2015.

Figure 2.A.1: Distribution of patenting in chemestry and metalurgy over the French metropolitan territory



Number of patents filed in IPC section C "Chemestry and metalurgy" since 2000 per commuting zone. The scale goes from light yellow (as few as 6 patent applications) to dark red (as many as 75k patent applications).

Table 2.A.6: Effect of global versus local centrality on the probability of moving

	(1)	(2)	(3)	(4)
	Moving			
Log. Total Bonacich Centrality (non-weighted)	0.75			
	(0.19)			
Log. Total Bonacich Centrality (weighted)		0.74		
		(0.19)		
Log. Local Bonacich Centrality (non-weighted)			0.08**	
			(0.09)	
Log. Local Bonacich Centrality (weighted)				0.09**
				(0.10)
Log. Total nbr employees	1.03	1.03	1.03	1.03
	(0.04)	(0.04)	(0.04)	(0.04)
Log. nbr R&D employees	1.12**	1.12**	1.12**	1.12**
	(0.06)	(0.06)	(0.06)	(0.06)
Log. nbr plants	3.41***	3.41***	3.48***	3.47***
	(0.34)	(0.34)	(0.34)	(0.34)
Several patenting locations	1.60**	1.60**	1.65**	1.62**
	(0.38)	(0.38)	(0.35)	(0.34)
Nbr obs	24064	24064	24064	24064
Pseudo r-sq	0.0878	0.0879	0.0895	0.0893

Logit regressions at the firm level, coefficients are **odds ratios**, standard errors of these odds ratios are in parentheses. The dependent variable is a dummy taking value 1 if a firm is considered to be moving a R&D lab based on the baseline definition. The first four regressors are different centrality measures. Log. Total nbr employees is the number of employees the firm has in an average year, Log nbr R&D employees is the same for R&D employees. Log nbr plants is the number of plants the firm has in total over the period. Log nbr zones with R&D is the number of commuting zones in which the firm has filed patents. *** p<0.01, ** p<0.05, * p<0.1

2.C Proofs

Proof of Lemma 1: For incumbent $\boldsymbol{G_2}$ firms, the Bonacich centrality equating their R&D effort can be dichotomized between self-loops and non self-loops, *i.e.* loops going from them to themselves and from them to other players respectively. Since no link going from another node to the entrant (denoted n_2+1) can exist because of the absence of links across locations, all the loops going from n_2+1 to any other i are exactly the walk from n_2+1 to its neighbors concatenated with all the walks going from the neighbors to node i. These can be expressed as the centralities of the neighbors. The activity levels in the subgraph receiving an entrant firm can be written as a function of the centralities in the graph before one entry as follows:

$$(\boldsymbol{x}^{G_2'})^* = \begin{pmatrix} b_{u,1}(\boldsymbol{G_2}, \phi_2) \\ \vdots \\ b_{u,n_2}(\boldsymbol{G_2}, \phi_2) \\ 1 + \phi_2 \sum_{j \in N_{n_2+1}} b_{u,j}(\boldsymbol{G_2}, \phi_2) \end{pmatrix} \tag{2.11}$$

Proof of Lemma 2 Following Ballester et al. (2006), the effect on any node j of the exit of player i, denoting $b_{u,ji}(\boldsymbol{G}, a)$ the part of j's centrality conditional to i's presence, implies the following:

$$\begin{aligned} b_{u,ji}(\boldsymbol{G}, a) &\equiv b_{u,j}(\boldsymbol{G}, a) - b_j(\boldsymbol{G}^{-i}, a) \\ &= \sum_{k=1}^{n} [m_{jk}(\boldsymbol{G}, a) - m_{jk}(\boldsymbol{G}^{-i}, a)] \\ &= \sum_{k=1}^{n} \frac{m_{ji}(\boldsymbol{G}, a) m_{ik}(\boldsymbol{G}, a)}{m_{ii}(\boldsymbol{G}, a)} \\ &= \frac{m_{ji}(\boldsymbol{G}, a)}{m_{ii}(\boldsymbol{G}, a)} b_{u,i}(\boldsymbol{G}, a) \end{aligned}$$

Therefore, the activity level of a player j after player i has left is:

$$b_{u,j}(\boldsymbol{G_1^{-i}}, \phi_1) = b_{u,j}(\boldsymbol{G_1}, \phi_1) - b_{u,ji}(\boldsymbol{G_1}, \phi_1) = b_{u,j}(\boldsymbol{G_1}, \phi_1) - \frac{m_{ji}(\boldsymbol{G_1}, \phi_1)}{m_{ii}(\boldsymbol{G_1}, \phi_1)} b_{u,i}(\boldsymbol{G_1}, \phi_1) \tag{2.12}$$

Proof of Proposition 4 Following Proposition 2, provided that only one firm from $\boldsymbol{G_1}$ is willing to move, it has to be firm 1, with the lowest centrality. In such a case, based on Lemma

2, the vector of centralities in $\boldsymbol{G_1}$ changes as follows:

$$\begin{pmatrix} b_{u,1}(\boldsymbol{G_1},\phi_1) \\ \vdots \\ b_{u,n_1}(\boldsymbol{G_1},\phi_1) \end{pmatrix}_{(n_1\times 1)} \quad -1 \text{ exits} \rightarrow \quad \begin{pmatrix} b_{u,2}(\boldsymbol{G_1^{-1}},\phi_1) \\ \vdots \\ b_{u,n_1}(\boldsymbol{G_1^{-1}},\phi_1) \end{pmatrix}_{(n_1-1\times 1)} = \begin{pmatrix} b_{u,2}(\boldsymbol{G_1},\phi_1) - \frac{m_{21}(\boldsymbol{G_1},\phi_1)}{m_{11}(\boldsymbol{G_1},\phi_1)} b_{u,1}(\boldsymbol{G_1},\phi_1) \\ \vdots \\ b_{u,n_1}(\boldsymbol{G_1},\phi_1) - \frac{m_{n_1 1}(\boldsymbol{G_1},\phi_1)}{m_{11}(\boldsymbol{G_1},\phi_1)} b_{u,1}(\boldsymbol{G_1},\phi_1) \end{pmatrix} \tag{2.13}$$

Construct iteratively a new ranking of firms' centralities in $\boldsymbol{G_1}$ after firm 1 has left, $\{o_e\}_{e\in\{1,\dots,n_1\}}$, which is a simple permutation of $\{1,\dots,n_1\}$, such that firm e would have the lowest centrality in $\boldsymbol{G_1}$ conditional to firms up to o_{e-1} moving.

More formally $\forall e \in \{2,\dots,n_1\}$,

$$o_e = \underset{i\in\{o_e,\dots,o_{n_1}\}}{\operatorname{argmin}} b_i(\boldsymbol{G_1},\phi_1) - \sum_{k=o_1}^{o_e-1} \frac{m_{ik}(\boldsymbol{G_1},\phi_1)}{m_{kk}(\boldsymbol{G_1},\phi_1)} b_{u,k}(\boldsymbol{G_1},\phi_1)$$

Denote $f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2) = \frac{1}{2}\phi_2^2 (l^*_{G_2})^2 + l^*_{G_2}\phi_2\bar{x}^{G_2} - c(l^*_{G_2})^2$ the part of the expected payoffs for $\boldsymbol{G_1}$ firms from moving to $\boldsymbol{G_2}$ that is independent of $\boldsymbol{G_1}$ network structure. Provided it is the only firm to move, firm $i \in G_1$ moves iff

$$f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2) - \kappa - \frac{1}{2}[b_{u,i}(\phi_1, G_1)^2 - 1] > 0$$

The first stage equilibrium when firms move simultaneously follows from introducing the above ranking in this condition.

Proof Pareto optimality of equilibrium with simultaneous moves Any firm moving from $\boldsymbol{G_1}$ to $\boldsymbol{G_2}$ gets expected centrality equating $1+\phi_2 l^*_{G_2}\bar{b}^{G_2}$, which means expected net profits equal to $\frac{(1+\phi_2 l^*_{G_2}\bar{b}^{G_2})^2}{2} - \kappa - c(l^*_{G_2})^2$. Since the highlighted equilibrium is the one featuring the minimal number of movements, any non-moving firm in this equilibrium gets profits equal to $\frac{[b_{u,i}(G_1^{\{o_1,\dots,o_e\}},\phi_1)]^2}{2}$. The firm with the lowest centrality among these non movers is firm o_{e+1}. Therefore, the highlighted equilibrium only needs to provide the highest possible payoffs to o_{e+1} to be Pareto-optimal (indeed, all movers have maximal payoffs since this is a criterion for them to move, and all non movers have maximal payoffs not moving if o_{e+1} has maximal payoffs not moving). Moreover, the payoffs of non-movers in G_1 may only decrease if more firms move, and their expected payoff from moving does not depend of how many firms move. Additionally, as ensured by Lemma 1, G_2 firms' centralities, hence payoffs, are independent of moving decision from G_1 firms. Thus, firm o_{e+1} cannot get higher payoffs either from moving,

or from having other G_1 firms move. This holds for any firm in $o_{e+1}, \ldots, o_{n_1}$. Therefore, the highlighted equilibrium, which features the lowest possible number of movers, is Pareto-optimal.

Proof of Proposition 5 The average activity after removing firm n_1 in the new subgraph is:

$$\bar{x}^{\boldsymbol{G_1'}} = \frac{n_1}{n_1 - 1}\bar{x}^{\boldsymbol{G_1}} - \frac{c_i(\boldsymbol{G_1}, \phi_1)}{n_1 - 1} \tag{2.14}$$

Or equivalently, the effect of i's exit on average activity in network $\boldsymbol{G_1}$ is

$$\bar{x}^{\boldsymbol{G_1'}} - \bar{x}^{\boldsymbol{G_1}} = \frac{\bar{x}^{\boldsymbol{G_1}} - c_i(\boldsymbol{G_1}, \phi_1)}{n_1 - 1}$$

Detailed version equilibrium with simultaneous moves

- $\forall e \in \{1, ..., n_1 - 2\}$, iff

$$\begin{cases} f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2) - \kappa - \frac{1}{2}\left\{\left[b_{u,o_e}(\boldsymbol{G_1}, \phi_1) - \sum_{k=0}^{e-1} \frac{m_{o_e o_k}(\boldsymbol{G_1}, \phi_1)}{m_{o_k o_k}(\boldsymbol{G_1}, \phi_1)} b_{u,o_k}(\boldsymbol{G_1}, \phi_1)\right]^2 - 1\right\} > 0 \\ f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2) - \kappa - \frac{1}{2}\left\{\left[b_{u,o_{e+1}}(\boldsymbol{G_1}, \phi_1) - \sum_{k=0}^{e} \frac{m_{o_{e+1} o_k}(\boldsymbol{G_1}, \phi_1)}{m_{o_k o_k}(\boldsymbol{G_1}, \phi_1)} b_{u,o_k}(\boldsymbol{G_1}, \phi_1)\right]^2 - 1\right\} < 0 \end{cases}$$

  there exists a Bayesian perfect equilibrium in which firms $\{o_1, \ldots, o_e\}$ move to $\boldsymbol{G_2}$, form $l^*_{G_2}$ links each with incumbent $\boldsymbol{G_2}$ firms, then all firms choose an effort level equal to their Bonacich centrality.

- Iff

$$f(\bar{x}^{G_2}, l^*_{G_2}, \phi_2) - \kappa - \frac{1}{2}\left\{\left[b_{u,o_{n_1-1}}(\boldsymbol{G_1}, \phi_1) - \sum_{k=0}^{n_1-2} \frac{m_{o_{n_1} o_k}(\boldsymbol{G_1}, \phi_1)}{m_{o_k o_k}(\boldsymbol{G_1}, \phi_1)} b_{u,o_k}(\boldsymbol{G_1}, \phi_1)\right]^2 - 1\right\} > 0$$

  there exists a unique Bayesian perfect equilibrium in which all $\boldsymbol{G_1}$ firms move to $\boldsymbol{G_2}$, form $l^*_{G_2}$ links each with incumbent $\boldsymbol{G_2}$ firms, then all firms choose an effort level equal to their Bonacich centrality.

- Otherwise, the network is intergraph stable and the BPE is trivial.

Remainder detailed version sequential equilibrium

- Iff

$$f[\frac{n_2\bar{x}^{G_2}+n_1-2+\phi_2\bar{x}^{G_2}(\sum_{k=1}^{n_1-2}l^*_{G_2+\{1,\dots,n_1-2\}})}{n_2+n_1-2}, l^*_{G_2+\{1,\dots,n_1-2\}}, \phi_2] - \kappa$$
$$-\frac{1}{2}\left\{\left[b_{u,o_{n_1-1}}(\boldsymbol{G_1},\phi_1) - \sum_{k=0}^{n_1-2}\frac{m_{o_{n_1}o_k}(\boldsymbol{G_1},\phi_1)}{m_{o_ko_k}(\boldsymbol{G_1},\phi_1)}b_{u,o_k}(\boldsymbol{G_1},\phi_1)\right]^2 - 1\right\} > 0 \tag{2.15}$$

  there exists a unique Bayesian perfect equilibrium in which all $\boldsymbol{G_1}$ firms eventually move to $\boldsymbol{G_2}$.

- Otherwise, the network is intergraph stable and the BPE is trivial.

Chapter 3

The Public Sector and the Distribution of Economic Activity

This chapter is a joint work with Emeric Henry & Joan Monras.

3.1 Introduction

The public sector accounts for a large share of total employment in most countries, more than 20% on average in OECD countries. Two particular features make this sector stand out. First, part of its activities are aimed at producing local public goods, and it thus has the potential to create particularly high spillovers on other sectors. Second, in many countries (for instance most countries in Europe), wages are set centrally and are uniform across locations that differ considerably in their cost of living.[1] In this paper, we show theoretically and empirically, using comprehensive data on public and private employment in France, that these two features are key to explain the data and are important determinants of the distribution of overall economic activity across locations.

In a standard spatial equilibrium model where the public sector is like any other, where workers are free to move across sectors and across locations, and where wages are set competitively (Rosen, 1974; Roback, 1982), we show in the first part of the paper that wages will be equalized across sectors within a city and indirect utilities will be equalized across locations. If the productivity varies across locations in the private sector, the most productive cities will have larger populations, pay higher wages, and have a larger ratio of workers in the private sector compared to the public sector. Despite this being the canonical model to study the distribution of economic activity across space – usually cities or regions –, many of these predictions do not hold in the data.

In the second part of our paper, we show that most of the predictions of the canonical spatial equilibrium model do not hold in the French data when the public sector is taken into account. The first prediction that is violated is that wages are *not* equalized across sectors within each location. Using a new and comprehensive data set on public and private employment for France, we show that that while wages in the private sector increase steeply with productivity, with an elasticity in the order of 0.1 which is in line with prior literature (Combes and Gobillon, 2014), they barely increase in the public sector. This highlights that the formal wage setting rules in the public sector are indeed respected and hence there is no access to extra source of bonuses that is used to compensate for the higher housing costs that usually prevail in the most productive cities. This means that *real* wages in the public sector are *lower* in productive cities, which potentially distorts downwards the size of the public sector in the most productive cities.

1. Germany imposes uniformity at the smaller administrative unity of landers, while in the US, federal blue-collar workers' wages are ruled by the Federal Wage System, and adjusted according to each local wage area.

The second feature that empirically departs from the baseline model is the fact that the public sector generates positive spillovers toward the private sector. A direct way to document that there *are* such spillovers is to study whether an exogenous increase in public employment leads to a significant increase or decrease in private employment. For this, we use the large decrease in public employment in France over the period 2008-2012 implemented by President Sarkozy, where the rule was not to replace one out of two public employee retirements. While this target was set nationally, it affected different locations differentially, based on the importance of the public sector at baseline. Using this strategy, we document that places suffering higher losses in public employment undergo a temporary decrease in private sector employment, with an elasticity above .2. Contrary to previous literature on the subject, we find that public employment losses affect negatively both tradable and non-tradable employment, although at different time horizons: non-tradables react immediately, while we observe the effect on tradables with a 5 years lag. We interpret the negative effect on the tradable sector as evidence that the public sector produces spillovers, and affects local private employment beyond the mere local multiplier (Moretti, 2010) going through the demand for locally-produced non-tradable goods of civil-servants.

In the third part of the paper, we extend our baseline model to incorporate, first, fixed nominal wages in the public sector following Monras (2019), and, second, spillovers from the public to the private sectors. We show that fixed nominal wages in the public sector tend to increase the size, the cost of living, and the wages in the private sector in the most relative to the least productive cities. The intuition for this results is simple. The fixed wages in the public sector limit the extent to which workers from the public sector are attracted to the most productive locations. As a result, the share of private sector workers is larger than it would be in a model where wages in the public sector are set competitively. In fact, the lower presence of public workers when public wages are fixed makes the most productive cities more attractive to private workers. Part of this higher appeal is reflected in more private workers, part in higher private sector wages, and part in higher housing costs.

The second theoretical result that we highlight is that spillovers from the public to the private sectors have the opposite effect on the distribution of economic activity than that of uniform nominal wages. We model spillovers as a factor that increases productivity in the private sector as a function of the number of public sector workers per inhabitant in each location. More productive cities have a disproportionate number of private sector workers, and the production

of public goods is therefore spread out over more individuals. This implies that spillovers are smaller in the more productive cities, thus being a factor that dampens differences based on productivities.

This paper relates to several strands of the literature. Several papers have taken interest in how centralized wage setting influences performance and the labor market. Propper and van Reenen (2010) study if pay regulation in the health sector in the UK has consequences on the performances of that sector in the places where the outside wage is higher. They measure performance through the estimated number of additional deaths of heart attacks, and find strong negative effects on hospital quality from having a nationally set wage lower than the outside wage. Britton and Propper (2016) use a similar setup focusing on teachers and find negative effects on school quality. Boeri et al. (2019) study the labor market and productivity effects of wage equalization across regions induced by collective bargaining agreements, contrasting the cases of Italy and Germany. They estimate high employment and earnings costs of the Italian system where nominal wages tend to be equalized relative to the German one. This paper also fits in a larger literature interested in spatial misallocation, see for example the papers by Albouy (2009), Fajgelbaum et al. (2016), Eeckhout et al. (2014), or Hsieh and Moretti (2017). These papers tend to emphasize larger or smaller than optimal city sizes. Sector size within cities is not their object of study.

Another strand of the literature has studied the effects of the location of public sector workers on private employment. In particular, Faggio and Overman (2014) estimate the local labor market effects of employment variation in the public sector, using a shift-share instrument (i.e. the initial share of the public sector times the national growth trend of the public). They find evidence of positive effects of public sector workers on the non-tradable private sector, however offset almost completely by negative effects on employment in tradables. Using a similar strategy on the Spanish case, Jofre-Monseny et al. (2019) find that public employment crowds-in private employment, with effects concentrated in the non-tradable sector. Faggio (2019) finds similar effects using the relocation of public sector workers in the UK. Becker et al. (2013) look at the private labor market effects of the location of global public good producers, through the displacement of the German capital city from Berlin to Bonn. They build a spatial equilibrium model with potential productivity and amenity spillovers of the public sector on the private one. They estimate the model and find little overall effects of the relocation of government workers on local labor markets, because they crowd-out private workers from the housing market. These

papers, however, do not consider the distinctive feature that in many countries wages in the public sector are fixed.

A different yet convergent branch of the literature has focused on the existence of local fiscal multipliers, and emphasized sizable effects of additional endowment of federal resources on economic activity (Nakamura and Steinsson, 2014; Suárez Serrato and Wingender, 2016), with a multiplier ranging between 1.5 and 2. Using a similar methodology, Suárez Serrato and Wingender (2014) build a spatial equilibrium model and show that part of this effect goes through spillovers that public jobs exert on local amenities.

The paper is organized as follows. The next section presents a theoretical benchmark, where the public sector is just another sector with fixed wages. We describe our data in Section 3.3. Section 3.4 presents evidence that the public sector has two important differences relative to other sectors, which are related to its wage setting mechanism and the spillovers it produces. Section 3.5 studies a model with an influential yet constrained public sector. Section 3.6 concludes.

3.2 A baseline spatial equilibrium model with two sectors

We start by considering a benchmark model where the public sector behaves as an additional sector, with no specificities compared to the private sector. In particular wages can freely adjust, workers can freely move across sectors and locations and there are no spillovers from the public to the private sector. This is the model implicitly considered in a large part of the literature that studies the distribution of economic activity across locations, see the seminal contributions of Rosen (1974) and Roback (1982).

Specifically we study the following model. There are two cities indexed by $i \in \{1, 2\}$ and two sectors indexed by $j \in \{Pub, Pri\}$, the public and the private sectors. We denote the number of workers in sector j in city i by N_i^j and wages by w_i^j. Productivities in the private sector A_i vary by city, and we assume that $A_1 > A_2$, i.e. city 1 is the most productive city. We assume productivity in the public sector B is the same across locations.

The production technology is a CES function that combines land and labor. Land is a fixed factor. The public sector has the same production function as the private sector, with the same local labor demand elasticity – which we assume for simplicity – and maximizes profits.[2] This implies that the inverse labor demand is

2. We abstract potential preferences for and hence demand and prices for the public sector.

$$w_i^{pri} = A_i(N_i^{pri})^{-1/\eta}$$
$$w_i^{pub} = B_i(N_i^{pub})^{-1/\eta}$$

Consumers in the model spend a fraction α of their income in housing and a fraction $1-\alpha$ in the private sector good – which is freely traded at no cost. We assume this to be the numeraire. We denote by p_i is the price of housing in city i. Workers spend a fixed share α of income on housing and housing supply is assumed inelastic in each city. This implicitly defines housing prices:

$$\frac{\alpha}{p_i}(w_i^{pri}N_i^{pri} + w_i^{pub}N_i^{pub}) = H_i.$$

The key feature of this benchmark model is that there is free mobility of workers across sectors and across locations. Indifference between sectors implies that the wage in each location is equalized across sectors:

$$A_i(N_i^{pri})^{-1/\eta} = B_i(N_i^{pub})^{-1/\eta}$$

Indifference between locations implies that the indirect utility of workers is the same across locations:

$$\frac{A_1(N_1^{pri})^{-1/\eta}}{p_1^{\alpha}} = \frac{A_2(N_2^{pri})^{-1/\eta}}{p_2^{\alpha}}$$

Combining these conditions with the labor market and housing market clearing condition yields the following results.

Proposition 8. (a) *Wages are identical in the public and private sectors in both cities:*

$$\ln w_i^{pri} = \ln w_i^{pub}.$$

(b) *The ratio of private to public workers is increasing in the gap in productivities*

$$\ln \frac{N_1^{pri}}{N_1^{pub}} - \ln \frac{N_2^{pri}}{N_2^{pub}} = \eta(\ln A_1 - \ln A_2).$$

(c) *More productive cities have larger populations*

$$\ln N_1 - \ln N_2 = \frac{1-\alpha}{1-\alpha+\alpha\eta} \ln\left(\frac{A_1^\eta + B^\eta}{A_2^\eta + B^\eta}\right),$$

(d) *Wages and house prices are higher in more productive cities*

$$\ln w_1 - \ln w_2 = \frac{\alpha}{1-\alpha+\alpha\eta} \ln\left(\frac{B^\eta + A_1^\eta}{B^\eta + A_2^\eta}\right),$$

$$\ln p_1 - \ln p_2 = \frac{1}{\alpha}(\ln w_1 - \ln w_2).$$

Proposition 8.(a) is obtained directly from the free mobility of workers across sectors. Proposition 8.(b) reflects the productivity advantage of city 1 with respect to city 2 in the the private sector. This will make city 1 specialize in the private sector. The intuition for Proposition 8.(c) and (d) is that on average city 1 is more productive than city 2. It arises from the fact that productivity in the public sector is the same across cities, while the productivity in the private sector is higher in city 1. In line with standard spatial equilibrium models, the higher productivity enables to sustain higher population levels, or higher wages or higher housing prices. How much of this higher productivity is observed in population, wages, or housing prices depends crucially on the share of labor in production and the share of housing in consumption.

Hence, a standard spatial equilibrium model has (at least) two clear implications. First, if workers are free to choose the sector that they want to work on (as people are legally able to do in most countries), wages should be equalized across sectors within locations and hence the wage gradient across locations should be the same for each sector. Second, exogenous increases in public sector employment (in the model, exogenous increases in B) should not affect private employment, as highlighted in Proposition 8.(b). In the next section, we examine whether these two predictions are in line with the empirical evidence.

3.3 Data

3.3.1 Employment Data

To measure public and private employment, we rely on information from the Déclarations Annuelles des Données Sociales (DADS). The DADS is built from mandatory social security contribution records reported by firms operating in France, and it is provided by the French National

Institute for Statistics and Economic Studies (INSEE). For every year, we observe every job spell within France, which is defined at the worker-plant level. For every spell, we have some basic information on the employee's characteristics (such as age and gender), occupation, salary, and the number of hours employed. For every plant, we have information on the industry and municipality (*commune*) it operates in, as well as the industry and sector (private or public) of the parent firm. Finally, for every job spell we also have some limited information on the contract associated with the job, such as whether it is a part-time or full-time contract, and whether it has a fixed or unlimited duration.

We focus on the ten years period between 2006 and 2015. The data for the private sector is available for the whole period. For the public sector (except the army), it is available only for the years 2009-2015. Some functions are however missing from 2009: we therefore have to exclude some functions from our analysis to establish comparisons over time. This limits our analysis to 85.2% of the fonction publique de l'Etat.[3] For earlier years, we obtained specific data on the public sector contained in the Fichier Général de l'État (FGE) to complete the fonction publique de l'État. Unfortunately, this data is of lesser quality than the information on public sector workers contained in the DADS, meaning we cannot draw robust comparisons with subsequent data.

Because our interest is mainly on the local public goods provided by the public sector, we sometimes focus on a set of core functions. They exclude very centralized functions such as central administration, as well as from functions attached to the fonction publique de l'État but grouped in one or few places (for instance national shipyards, various public agencies, national parks administration). Our baseline definition of core functions includes teachers in pre-primary, primary, secondary and higher education, hospitals, medical centers and elder care, social action, police and firefighters.[4]

3.3.2 Wages

We take information for wages (net of social security contributions) at the job spell level from the DADS Postes. To measure the local wage premium associated with each area in each sector (private or public), we regress individual log hourly wages on age, age squared, a gender dummy,

3. In 2015, this amounts to 1.658 million out of the 1.944 million full-time equivalent workers in the FPE. The main functions missing from the data in 2009 are superior education (around 200 thousand workers) and museums (10 thousand workers).

4. Formally, we include all workers of the public sector in establishments with NAF (sector) codes starting with 851 (pre-primary education), 852 (primary education), 853 (secondary education), 854 (higher education), 8424 (police), 8425 (firefighters), 86 (hospitals), 87 (social and medical housing), 88 (social action).

as well as a full set of position, industry and contract dummies, and a commuting zone dummy.[5] We then recover these commuting zone fixed-effects, and normalize the minimum to zero. Thus, the sectoral local wage premia we use can be interpreted as the local percent deviation from the sector minimum in the hourly wage.

3.3.3 Local productivity estimates

Financial information on private firms is contained in the FICUS and FARE balance-sheet datasets produced by the French tax administration (DGFiP). They report accounting data at the firm level, such as gross value-added, sales, gross operating system, profits, employment and paid wages. Value-added is the excess value of the firm's production from the value of intermediate consumptions, not considering the taxes and subsidies the firm may be submitted to.

To measure productivity in each commuting zone, we aggregate gross value-added from firms in the area. To handle firms with plants in multiple commuting zones, we take the spatial distribution of their workforce from the DADS (i.e. the share of hours worked in each commuting zone), and simply split their total value-added according to this share. We then divide by the number of full-time equivalent workers in the area to obtain a measure of value-added per worker. We take this measure as a proxy of local productivity.

3.3.4 Rents and housing prices

The main source we use on rents throughout the paper originates from a large web-scrapped database, which records all rental annoucements posted in 2016 on the two main websites from rental adds in France. From this database, we use municipality level fixed-effects from a standard regression of the proposed rent on the characteristics of the house.[6] We then take the mean of these fixed-effects weighted by population to obtain a measure of rents in the commuting zone. We also obtained the same type of municipality level fixed-effects for house prices, originally calculated from the "Bases Notaires".[7]

5. The contract variable takes 14 distinct values, the position variable follows the 4 digits PCS-ESE classification and takes 429 different values (including 324 in the public sector), the industry variables follows the 5 digits NAF rev.2 classification and takes 723 values (including 124 in the public sector).

6. These fixed-effects are extracted from a regression of the log of rents per square meter on powers of the log of surface, the floor, the date of construction of the building when available, whether it is a single unit, whether it is furnished.

7. The rents database is decribed in more detail in Chapelle and Eyméoud (2017). We warmly thank Guillaume Chapelle for giving us access to the above sources.

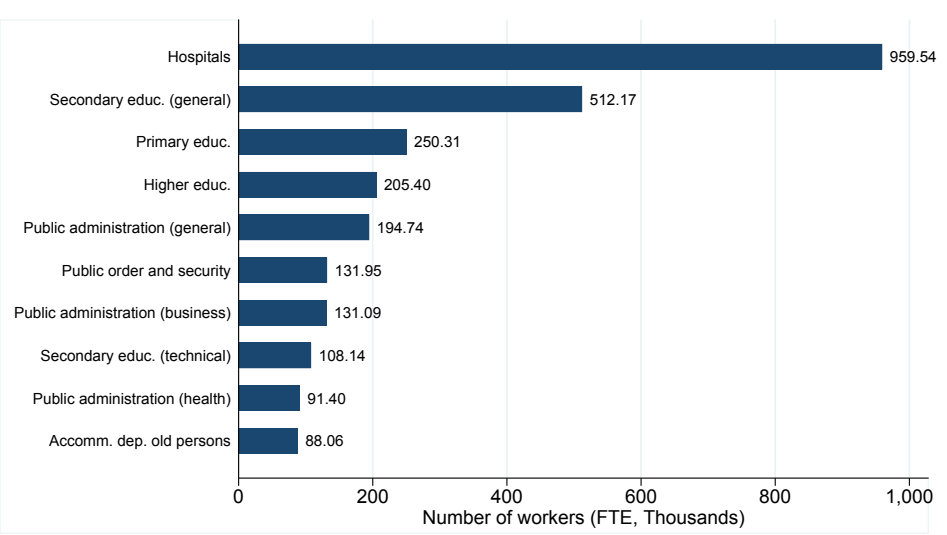
3.4 Empirical evidence

3.4.1 Institutional setting and summary statistics

The public sector in France employs more than 5 million people (i.e roughly 23% of total employment), divided in 3 categories: the fonction publique de l'État (national public servants, sometimes abbreviated FPE), the fonction publique hospitalière (hospitals) and the fonction publique territoriale (local civil servants). Wages in the fonction publique de l'État and hospitalière are set centrally and are theoretically uniform, except for a bonus called indemnité de résidence that can range from 0 to 3% based on a list of cities which was set up after the Second World War and not updated in depth since 1985.

The fonction publique de l'État covers many of the basic services. First, it includes professors, in elementary, primary and secondary schools as well as higher education. This is in fact the largest group (putting hospitals aside) as represented in Figure 3.1. The other large occupations are police and administration. Figure 3.A.1 shown in Appendix presents the same population divided by type of job rather than establishment.

Figure 3.1: Composition of the national public sector

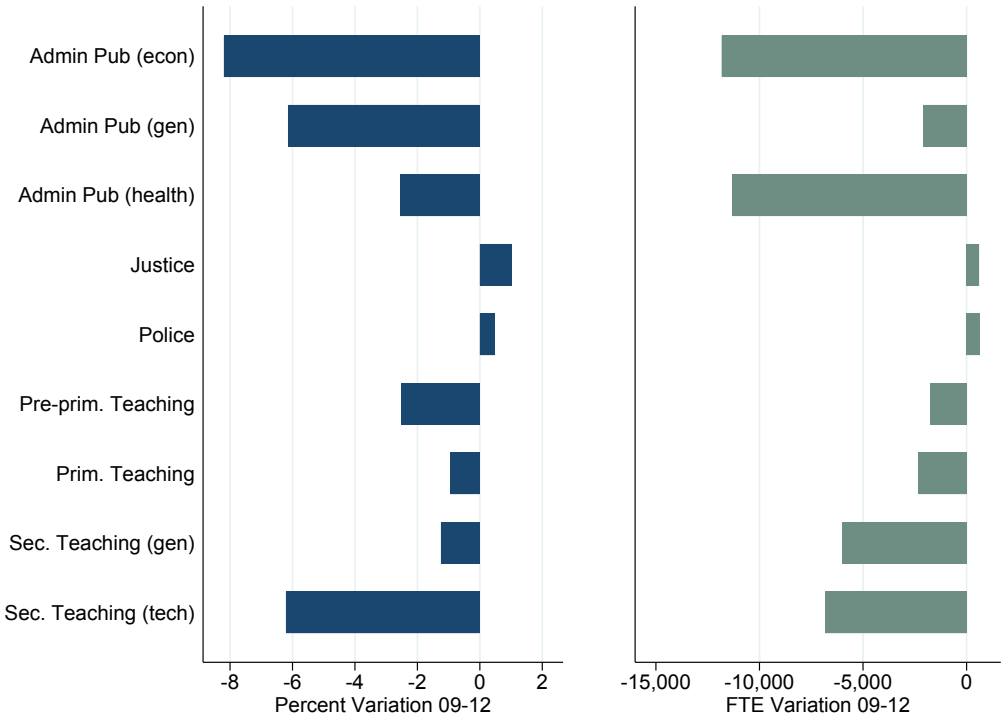


Notes: This figure presents the number of workers within the national public sector in France, expressed in full-time equivalents, divided by the function of the establishment they work in. This includes the fonction publique hospitalière, here comprised only of hospitals, and the function publique de l'État.

The Révision Générale des Politiques Publiques (RGPP) was launched in July 2007 under the presidency of Nicolas Sarkozy, and aimed at reforming in depth the "Fonction Publique de l'État". The general publicized goal was to modernize the state, but one of its key purposes was in fact a reduction of public employment. From 2008 to 2012, one out of two retirements was

not replaced, amounting to an alledged reduction of employment of 144,000 over these 4 years. The variation we observe in our data is actually lower, amounting to a loss of around 60,000 workers. Several factors could explain this difference: we observe only three years of reform, the army was affected by large job cuts but is absent from our data, and official counts usually record employment of ministries only, while our definition of public employment is wider and would not count transfers out of the scope of ministries as losses.[8] The variation by type of occupation due to the RGPP is presented in Figure 3.2. It shows that the largest drops occurred in the administration and in secondary teaching.

Figure 3.2: Variation in FPE employment during the RGPP



The figure presents variations between 2009 and 2011 by type of public sector establishment (NAF). The left panel presents relative variations, while the right panel presents variations in level (number of full-time equivalent jobs).

3.4.2 Cross-sectional evidence

As shown in Section 3.2 one of the predictions of the baseline model is that a) wages in the private and public sectors should increase in the same way with local productivity measures, b) the ratio of private sector workers to public sector workers should increase strongly with

8. We consider a person to be employed by the FPE whenever the juridical category of either its employer or its workplace depends on the FPE.

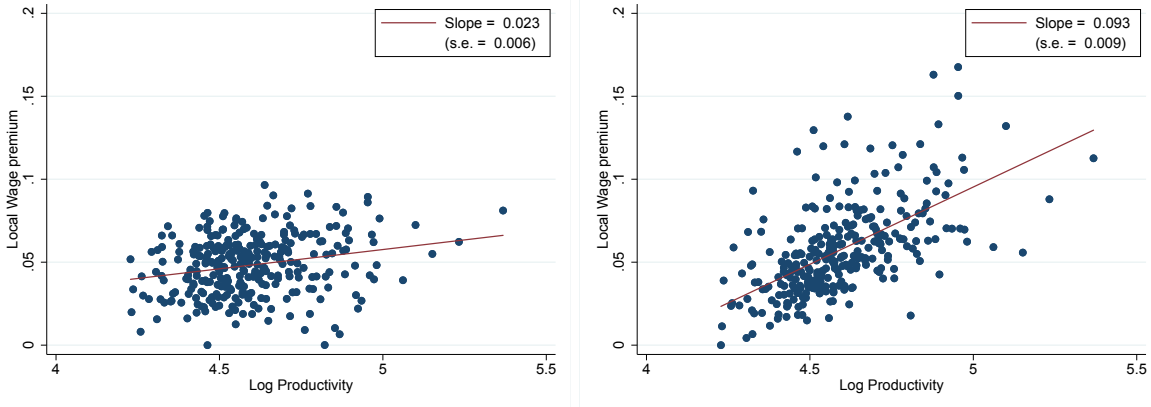
local productivity, and c) Population and housing prices should also strongly increase with productivity. In this section we show that some of these predictions do not hold in the data. We argue that in this subsection that one of the main reasons why the data is not consistent with the model in the cross-section is that wages in the public sector are not set competitively.

Private and public sector wages

The first striking fact is that equalization of wages across sectors in each city, expressed in Proposition 8.(a), is clearly not satisfied. This is what we should expect from the institutional framework, but we need to check that it indeed holds in the data.

As shown in Figure 3.A.4, wages are steeply increasing in the private sector as a function of productivity, with an elasticity in the order of 9.3%, while they are barely increasing in the public sector. The elasticity in the public sector is in the order of 2.3% , i.e. less than the maximum bonus of 3% that is awarded based on local housing costs. This fits with the idea that the list of cities where public servants can benefit from this bonus was barely revised since the second world war and is thus not well aligned with current rents (nor productivities). This result also indicates that wages cannot be adjusted using other bonuses to adjust for local living costs.

Figure 3.3: Wages in the public and private sector



(a) Public sector (nat.) **(b)** Private sector

Notes: This figure presents local wage premia measured as described in Section 3.3.2 for the public sector (left panel) and the private sector (right panel) as functions of local log productivity. Public sector here means all the national public sector (Fonction Publique de l'Etat and Fonction Publique Hospitalière), but excludes the local public sector and State-owned firms. Private sector here means all private-owned businesses. Log productivity is the natural logarithm of value-added per worker in the commuting zone.

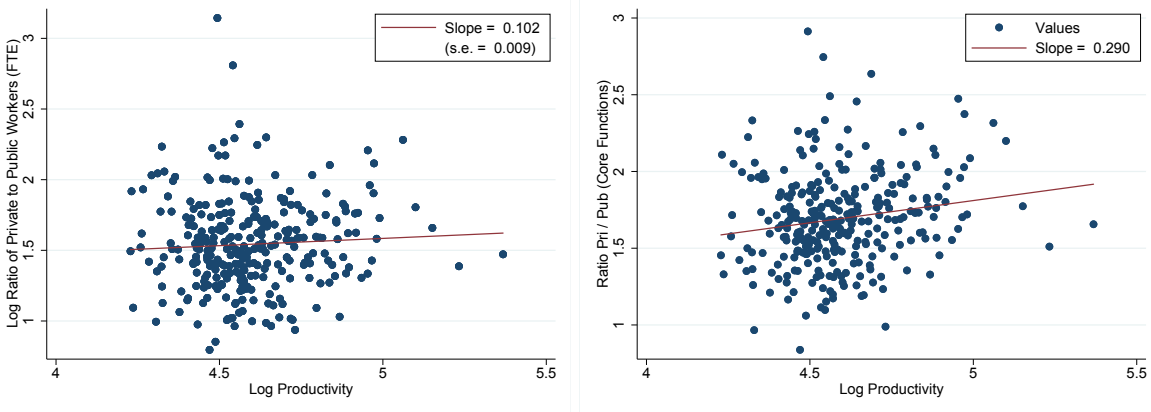
Given that French workers are free to choose what sector to work on, Figure 3.A.4 suggests that either there are large differences in real wages of public sector employees across locations relative to private sector employees, or that on the job amenities – not included in our baseline model – are significantly higher in the most productive cities. We think that the latter is unlikely given the uniformity of tasks that the public sector performs across locations.

Relative size of private and public sectors

In the baseline model, we obtain a sharp prediction for the relative size of the private and public sectors across locations. Given that wage-setting differs from what is assumed in the baseline model, this prediction may not hold in the data either.

The size of the private sector relative to the public sector is only weakly increasing in local productivity. As visible in Figure 3.4, the elasticity is at most 0.3 when we consider the core functions of the public sector and as low as 0.1 when we consider all public sector employees.

Figure 3.4: Relative size of the private and public sector as a function of productivity



(a) Public / Private employees ratio

(b) Public in core functions / Private employees ratio

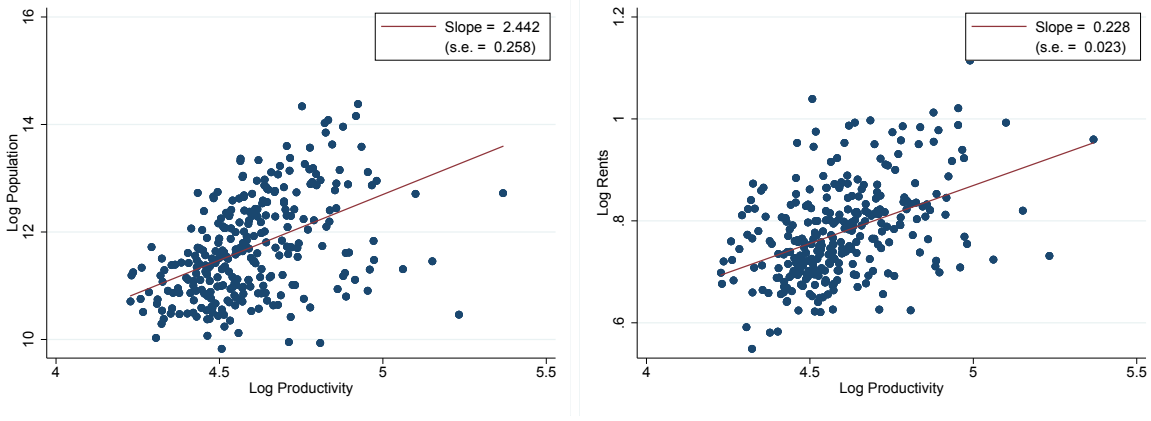
Notes: These graphs present the ratio between the number of public sector workers and the number of private sector workers (all in full-time equivalent), plotted against local log productivity. The left panel considers all public sector workers, while the right panel restricts the public sector to our definition of core functions. Log productivity is the natural logarithm of value-added per worker in the commuting zone.

At first sight we may be tempted to relate the weak relationship between the private to public sector ratio of workers to the fact that wages are not set competitively in the public sector as argued in subsection 3.4.2. However, as we show more formally in Section 3.5, if we

introduce fixed wages in the public sector into our baseline model we obtain that the relative size of the private sector should be larger than when wages are set freely. Hence, in that model we would obtain an even stronger relationship between the ratio of private to public employees and local productivity. This suggests that there are other forces that the baseline model misses that may help to explain the data.

Population size and housing prices

Figure 3.5: Total population and housing prices as a function of productivity



(a) Population **(b)** Rents

Notes: These graphs present total population in a commuting zone as a function of productivity (panel a) and rents (panel b). Log productivity is the natural logarithm of value-added per worker in the commuting zone. Log rents are measured as the natural logarithm of an population weighted mean in the commuting zone of rents cleaned from the houses characteristics.

A final aspect that the baseline model highlights is that population and housing prices should be increasing in local productivity. We show these relationships in Figure 3.5. Panel (a) shows that the elasticity between productivity and population is around 2.4. Reversing the graphs to express productivity as a function of population, this estimate is comparable to estimates in the literature that try to explain local agglomeration forces of population on productivity and obtain estimates that range from 0.04 to 0.1 (Combes and Gobillon, 2014) – we put productivity in the x-axis since this is the exogenous measure in our baseline model. Panel (b) shows the relationship between housing prices and local productivity. It displays a similar pattern as with population: rents strongly increase in productivity, with a measured elasticity of .23.

3.4.3 Spillovers: the RGPP experiment

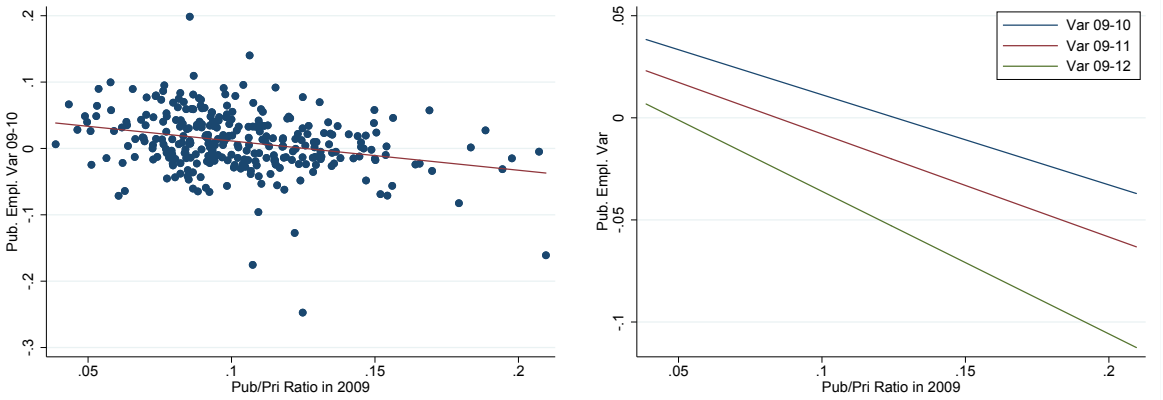
We provide evidence that the public sector generates local spillovers to the private sector, using the non-replacement of one out of two retiring civil servants in the "Fonction Publique de l'État" between 2008 and 2012 as part of the RGPP reform, and study the effects it had on the activity of the private sector. Because our database starts only in 2009 while the beginning of non-replacements was in 2008, we do not observe the full reform period and focus on job cuts from 2009 onward.

Instrument

To evaluate the local effect caused by public employment losses, we need places to be differentially affected by the reform, along a dimension that was fixed ex ante and determined the exposure of an area to the reform without determining future private employment growth. Indeed, we want to abstract from the part of the variation in public employment which is correlated with variations in private employment or population. Typically, two phenomena may coexist jointly with the reform and bias our estimates of the effect of public sector variations in the simple OLS. On the one hand, public employment may increase relatively in areas where the needs for public services have increased, therefore following recent growth in private employment, which is likely to be positively correlated with the future trend in employment growth. On the other hand, the government may be reluctant to cut public employment in declining places.

The year 2009 was quite peculiar for the public sector: while it was supposed to be the key year of the reform because of a historical high in the predicted number of retirees in the public sector (around 70 thousands), the actual number of retirees was actually much lower (around 55 thousands), meaning that recruitment was far too large to meet the one out of two rule for retirees' replacement. Therefore, we do not observe losses at the national level nor on average in a commuting zone between 2009 and 2010, which therefore resembles more to a reallocation of public employment over the territory than an actual reduction. A dimension that nonetheless stands out in public employment variations during the reform period is the fact that places starting with a high endowment of public sector relative to the private sector had to contribute more to the national effort of reducing public employment. The existence of spatial inequalities in such ratio is consistent with the facts shown in Figure 3.4: fixed wages in the public sector contribute to generating an unbalanced distribution of civil servants over the territory. Yet, the

Figure 3.6: A visual representation of the estimation first-stage

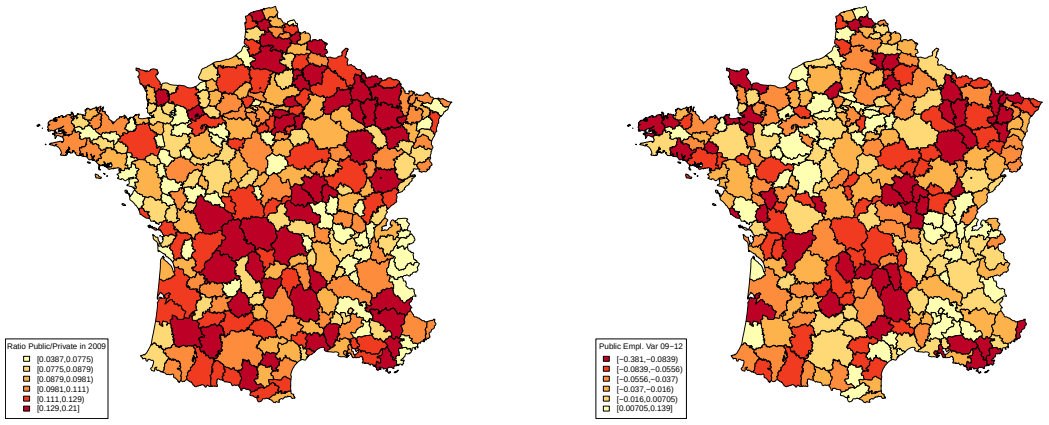


(a) Scatter of 2009-10 public employment variation against the instrument.

(b) Linear fits of public employment variation between 2009 and the next three years against the instrument.

Notes: Panel (a) shows a scatter plot of the variation in public employment (Δ log) between 2009 and 2010 as a function of the public to private ratio in 2009. The slope of the fitted line is -.44 (standard error is .09). Panel (b) shows the fitted line of public employment variation between 2009 and three subsequent years. The blue line is the fitted line relative to the variation 2009-2010 shown in panel (a), the red line is that of the variation 2009-2011 (slope is -.50, s.e. is .11), the green line shows the variation between 2009 and 2012 (slope is -.70, s.e. is .12).

Figure 3.7: Maps of public employment variation and public/private employment ratio



(a) Public (FPE)/Private employment ratio in 2009

(b) Public employment variations (Δ log) between 2009 and 2012

These figures represent maps of commuting zones in metropolitan France (excluding Corsica). In panel (a), commuting zones are colored according to their public to private employment ratio in 2009, where dark red means a high value (at least .13), and light yellow a low value (less than .08). In panel (b), areas are colored according to their variation in public employment, where dark red means larger losses (at least - 8%) and lighter colors mean lower losses (or even increases by at least .7 % for light yellow areas).

variance in the private to public employment ratio once differences in productivity are accounted for is still considerable, which could simply come from a strong path-dependence in the size of the public sector in each city. We calculate this ratio as the number of people in the area employed in the FPE (part of the public sector exposed to the reform) over the number of people employed in the private sector. Conditional on variables which should cause differences in this ratio based on the model exposed below (typically, the wage premium difference between the private and the public sector), this provides us with an instrumental variable which predicts the treatment intensity of the reform in an area.

Figure 3.6 shows the variations of public employment during the reform plotted against the instrument. Figure 3.6.b shows that most of the spatial differences were introduced between 2009 and 2010: later years of the reform keep cutting public sector jobs, but do not considerably increase the slope between public employment variation and the *ex ante* public/private employment ratio. Because the reform was staggered over several years and the later years still introduced some spatial differences (as shown in Figure 3.6.b, the slope of the 2009 to 2012 variation is around one and half times that of 2009 to 2010), we present results in two ways. In the baseline, we look at variations from 2009 to years after 2010 (i.e. after areas were treated differentially), but we present in the Appendix the analogous results for the effect of public employment over the full reform period on private employment in years after 2012 (after the reform ended). Tables 3.A.2 and 3.A.3 presented in Appendix show the first stage of our 2SLS IV regressions, on the public employment variation respectively between 2009 and 2010, and 2009 and 2012. The Kleinbergen-Paap F-stat, used to judge the strength of the instrument, ranges between around 25 and 40 depending on the specifications, meaning we can reject that the instrument might be weak.

A usual test of the exclusion restriction of the instrumental variable is whether it correlates with the outcome variable considered in years preceding the shock. Tables 3.A.4 and 3.A.5 presented in Appendix show such standard Placebo shock using pre-trends: conditional on our control variables, the instrument does not predict our outcome variable considered in the years prior to the shock, either when used in a 2SLS regression or directly in a reduced-form estimation. This confirms that places with relatively higher levels of public to private employment did not have different private employment trends conditional on our control variables than places with lower such levels before the shock hit.

There are several caveats regarding our empirical design. First, following the argument

we develop in the next section, a higher public to private ratio influences private employment through higher spillovers. While the theoretical argument we make is static, it should have dynamic implications. Therefore, a higher public to private employment ratio should imply a higher private employment growth. This force would lead us to underestimate the true effect of public employment variation, since a higher public to private ratio leads to larger public employment losses. A second major caveat is of course the fact that we miss the first year of the reform. While the 2009 to 2010 yearly variation seems very distinctive, it could be that the reform implied a similar variation between 2008 and 2009, meaning that we overestimate our effect (it could arguably also have been applied more strictly in places with lower public to private employment ratios, although this seems unlikely). The existing data on the period makes it very difficult to be conclusive on that side. While 2008 to 2009 public sector variations in that data source also strongly correlate negatively with the public to private ratio (calculated from either database), the range of variations seems far too wide, with places losing up to 43% of their public employment over a single year. Considering the above-mentioned correlation, it seems likely that the reform exposure followed the same spatial pattern in the first year of the reform, meaning that places with more civil servants compared to private workers underwent larger negative public sector variations. This means we probably overestimate the magnitude of our effects: we therefore interpret our results only qualitatively, and do not take a stance on the underlying value of the public sector multiplier on private sector activity. Regardless of that value, the fact that the public sector does generate spillovers on the private sector emerges unambiguously in our results.

Because our period of study was marked by a deep economic crisis, implying large negative variations in private employment concentrated on some sectors (for instance construction), all regressions conducted throughout this section include as a covariate Bartik (1991)-type shocks, calculated over the same period as the dependent variable, which account for the an area's private employment exposure to country-wide sectoral changes over the period.[9]

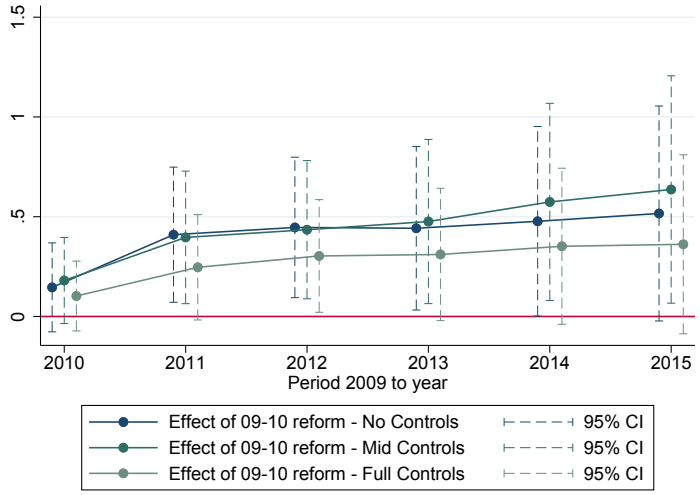
In what follows, we present results for three different specifications: one with minimal controls (called "No controls" throughout), one with a set of controls for the characteristics of an area at the time of the shock (called "Mid controls" throughout), and one controlling also for potential differences in pre-period variations in population and employment (called "Full controls"). In

9. Bartik shocks are calculated using NAF 5 digits industry codes, calculating shares of private employment by commuting zone with the DADS. As has become common practice and is formally advised by Goldsmith-Pinkham et al. (2018), we use leave-one-out growth rates, meaning that the national trend in each industry for each commuting zone excludes its own contribution to that share.

the "No controls" strategy, the only variable other than our dependent variable and our variable of interest is the Bartik shock. The "Mid controls" strategy is our preferred specification: it includes variables which are likely to influence our instrumental variable for the reasons explained above (logs of population, public private wage ratio, productivity, rents, and the share of the population living in rural municipalities). The "Full controls" strategy adds control variables accounting for the trajectory of the areas before the shock, namely the population trend between 1999 and 2009, and private employment trend between 2006 and 2009. It seems however too restrictive since our variable of interest then correlates negatively with the outcome variable in the pre-shock period. Table 3.A.1 shown in Appendix provides summary statistics of the variables used in the regressions.

Results

Figure 3.8: IV regression results of public variation between 2009 and 2010 on private employment subsequent years.



This figure plots coefficients and confidence intervals of the second-stage of 2SLS regressions of the Δ log of private employment between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2010, instrumented by the public to private ratio in 2009. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

Table 3.1 presents the main results for private employment taken in the 2009 to 2013 period, using the public employment variation induced by the reform only between 2009 and 2010. Column (1) presents the first stage of the regression, and displays a F-Stat of 24 implying that the instrument is strong. Column (2) shows the results of a simple reduced-form regression of the

Table 3.1: Results of the main regressions for the 2009 to 2013 period

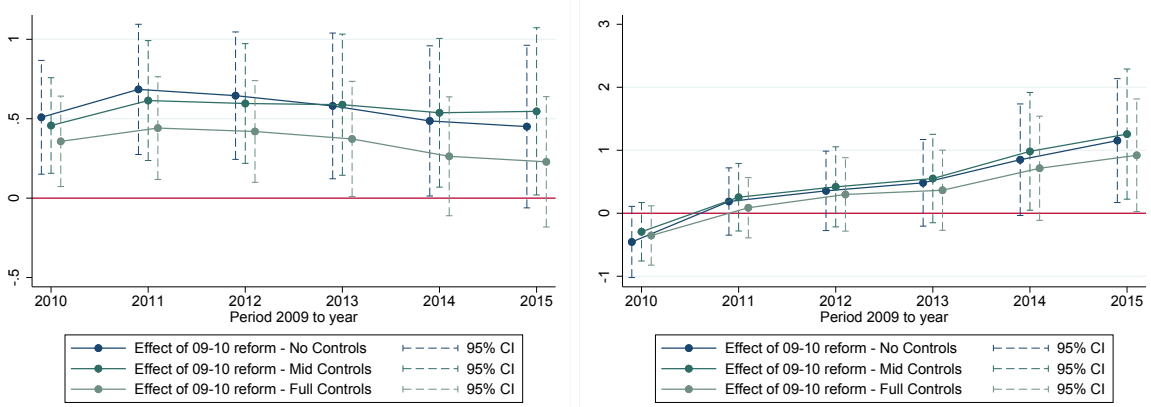
	(1)	(2)	(3)	(4)
Dep. Var.:	Δ log Pub Empl.	Δ log Pri Empl.		
Period:	09-10	09-13		
Estimate	1st stage	RF IV	OLS	2SLS
Pub (FPE)/Pri Ratio 2009	-0.48***	-0.27***		
	(0.10)	(0.10)		
Δ log Public Empl. 09-10			0.07	0.48**
			(0.05)	(0.21)
Bartik Shock	-0.09	0.71***	0.67***	0.61***
	(0.13)	(0.16)	(0.16)	(0.18)
Public/Private Wage ratio	-0.06	-0.07	-0.04	-0.02
	(0.09)	(0.09)	(0.09)	(0.09)
Log Rents	-0.02	0.06	0.07	0.04
	(0.04)	(0.05)	(0.05)	(0.04)
Log Productivity	0.02	-0.01	-0.01	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)
Log Population	0.01	0.01*	0.01	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Share of Pop. in rural area	0.01	0.06**	0.05**	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
VCE	Robust	Robust	Robust	Robust
N obs	296	296	296	296
R-squared	0.101	0.117	0.0978	.
1st stage KP F-stat	23.97	.	.	23.97

This table presents the results for the main regressions of interest, using the 2009 to 2010 public employment variation, and the 2009 to 2013 private employment variation, using the "mid-controls" set. Column (1) presents the result of the first stage of the 2SLS IV regression. Column (2) presents the reduced-form IV regression, meaning that the dependent variable (private employment variation) is regressed directly on the IV and controls. Column (3) presents the naive OLS. Column (4) presents the second-stage of our 2SLS IV regression.

dependent variable on our instrument, which displays a significant coefficient. Columns (3) and (4) display the results respectively of the naive OLS and of our 2SLS IV strategy. As expected, the variation in the OLS is very imprecisely estimated, but the IV regression is associated with a large and significant coefficient. Figure 3.8 shows the baseline IV estimates obtained with our three different specifications. For each specification, the x-axis shows the final year of the period through which we consider private employment variations. The plotted coefficients correspond to the coefficient associated to our variable of interest (public employment variation between 2009 and 2010) instrumented by the public to private employment ratio in 2009, and the dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added. The figures shows a positive effect of public employment variations on private employment, meaning that places undergoing larger

public job cuts faced a larger decrease in private employment in the subsequent years. This effect is most vivid in the years directly following the shock. In later years, point estimates remain stable but the estimation becomes less precise, suggesting that some places may recover from the treatment while others do not. Although they are more noisy and smaller in magnitude, the simple OLS estimates presented in Figures 3.A.5 and 3.A.6 in the Appendix provide a similar picture: they show short-run effects, which remain positive but become insignificant a couple of years after the shock.

Figure 3.9: IV regression results of public employment variation (Δ log) between 2009 and 2010 on non-tradable and tradable sectors.



(a) Non-tradables. **(b)** Tradables.

These figures plot coefficients and confidence intervals of the second-stage of 2SLS regressions of the Δ log of private employment (in non-tradables in panel a, in tradables in panel b) between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2010, instrumented by the public to private ratio in 2009. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

The effect on overall private employment however hides a strong dimension of heterogeneity across sectors. In particular, the mechanisms through which tradable and non-tradable sectors are affected strongly differ. The non-tradable sector will suffer from decreases in local demand, as with any local employment drops in another sector, through a local multiplier mechanism (Moretti, 2010; van Dijk, 2016). The tradable sector should however not suffer from such local consumption effect, but is more likely to be affected by spillovers from the public sector (it does not rely on local demand, arguably has larger underlying differences in productivity, and should benefit from a lower pressure on the housing market). Our estimates for non-tradables, presented in Figure 3.9.a, are in line with prior literature on local multipliers as well as with the literature on the effects of public sector employment (Faggio and Overman, 2014; Faggio, 2019;

Jofre-Monseny et al., 2019): we observe large short-run effects of public employment drops in non-tradables, which decrease and become less significant at the end of the period, suggesting at least a partial recovery in part of the places most affected. Our results on the tradable sector presented in Figure 3.9 however differ strongly from the existing literature. While some papers find public employment crowds out the tradable sector through increased prices in the housing sector, we find on the contrary that public employment job cuts strongly negatively affected employment in tradables. This effect seems to occur in the longer-run, and the time-window we observe does not allow to state whether it is temporary or not. As shown in Figures 3.A.8, 3.A.9 and 3.A.10 in the Appendix, these results carry-through when we restrict tradables to manufacturing, and when consider public job cuts in the 2009-2012 period on subsequent years. The fact that the measured elasticity is much greater on the tradables than on the non-tradables is not worrying, since the tradable sector is on average much smaller than the non-tradables.

3.5 A spatial equilibrium model of the public sector with nominal rigidities and spillovers

Two key features of the public sector are emphasized in the preceding analysis. First, wages in the public sector are essentially independent of the local characteristics and fixed across the country. Second, the public sector has a spillover effect on the private sector. We introduce in turn these two features in the model. They have opposing effects on the distribution of economic activity. Spillovers tend to reduce the dispersion of activity while geographical rigidity of wages tends to increase it. These two features are introduced in the following way:

Spillovers

The size of the public sector in city i influences the productivity of the private sector. Specifically productivity A_i is multiplied by a factor $\left(\frac{N_i^{pub}}{N_i}\right)^{\beta}$, where β measures the strength of spillovers. The implicit assumption is that, the larger the population, the less each inhabitant can benefit from the addition of an extra public worker.

Geographical rigidity of wages

The wage is assumed fixed across locations in the public sector. Wage rigidity in just one sector means that the other sectors absorb the excess supply, see for instance the minimum wage

literature on covered and uncovered sectors (Zenou, 2009). To deviate from this literature and make the model consistent with the wage setting mechanism in France, we assume instead that a worker entering the public sector does not know in which city she will be appointed. The probability of being sent to city 1 depends on the relative size of the public sectors in city 1 and city 2, i.e on $\frac{N_1^{pub}}{N_2^{pub}+N_1^{pub}}$. Indifference between sectors will thus pin down wages in the private sector, as expressed in the following relation that states that expected indirect utilities need to be equalized across cities:

$$\frac{w_i}{p_i^\alpha} = \frac{N_1^{pub}}{N_1^{pub}+N_2^{pub}}\frac{\bar{w}}{p_1^\alpha} + \frac{N_2^{pub}}{N_1^{pub}+N_2^{pub}}\frac{\bar{w}}{p_2^\alpha}$$

This equation says that the real wage in city i has to be equal to the expected indirect utility when working in the public. Expected indirect utility in the public sector is given by the probability of working in each city times the wage (which is independent of the city) and the price of housing in the city.

An immediate consequence of this equation is that *realized* real wages in the public sector are not necessarily equalized across space. Some public workers are lucky and get assigned to good cities, while others are assigned to worse ones. In practice, some administrations make this decision less random by giving some decision power to workers with more seniority. We abstract from this for simplicity in this model.

3.5.1 Model with spillovers

Given the existence of spillovers, the inverse labor demand curves in the two sectors become:

$$w_i^{pri} = A_i \left(\frac{N_i^{pub}}{N_i}\right)^\beta (N_i^{pri})^{-1/\eta}$$
$$w_i^{pub} = B_i \left(\frac{N_i^{pub}}{N_i}\right)^\beta (N_i^{pub})^{-1/\eta}$$

The fact that spillovers affect the private and public sector in the same way implies that the indifference between sectors within a city is unaffected compared to the benchmark model. The indifference between locations is however altered, since cities will benefit from different levels of

spillovers in equilibrium. This indifference can be written:

$$\frac{A_1\left(\frac{N_1^{pub}}{N_1}\right)(N_1^{pri})^{-1/\eta}}{p_1^{\alpha}} = \frac{A_2\left(\frac{N_2^{pub}}{N_2}\right)(N_2^{pri})^{-1/\eta}}{p_2^{\alpha}}$$

which implies:

$$\frac{A_1}{A_2}\frac{\left(\frac{N_1^{pub}}{N_1}\right)^{\beta}}{\left(\frac{N_2^{pub}}{N_2}\right)^{\beta}} = \left(\frac{N_1^{pri}}{N_2^{pri}}\right)^{1/\eta}\left(\frac{p_1}{p_2}\right)^{\alpha}$$

Cities with relatively less public sector workers will benefit from smaller spillovers and this will tend to decrease the gap between the most and least productive cities as expressed in the following result.

Proposition 9. *In a model with spillovers:*

(a) *The ratio of private to public workers does not depend on spillovers*

$$\ln\frac{N_1^{pri}}{N_1^{pub}} - \ln\frac{N_2^{pri}}{N_2^{pub}} = \eta(\ln A_1 - \ln A_2).$$

(b) *The gap in population between the most and least productive city decreases with the strength of spillovers β*

$$\ln N_1 - \ln N_2 = \frac{1-\alpha-\eta\beta(1-\alpha)}{1-\alpha+\alpha\eta}\ln(\frac{A_1^{\eta}+B^{\eta}}{A_2^{\eta}+B^{\eta}}).$$

(c) *The gap in wages and house prices between the most and least productive city decreases with the strength of spillovers β*

$$\ln\frac{w_1}{w_2} = \frac{\alpha(1-\eta\beta)}{1-\alpha+\alpha\eta}\ln(\frac{B^{\eta}+A_1^{\eta}}{B^{\eta}+A_2^{\eta}}),$$

$$\ln\frac{p_1}{p_2} = \frac{1}{\alpha}\ln\frac{w_1}{w_2}.$$

Workers can move freely across sectors and given that spillovers affect both sectors in the same way, Proposition 9.(a) shows that the gap in the ratio of private to public sector workers

is increasing in the gap in productivities, as in the baseline model, but does not depend on the spillover parameter β.

Spillovers however affect the other relations. As indicated in Proposition 9.(b), more productive cities have larger population since they can offer higher wages. This gap in population is decreasing in spillovers. City 1, with a larger population, receives less spillovers since, according to result (a), the ratio of public to private is the same in the two cities and given that population is larger in the more productive city, spillovers are smaller. This in turn decreases the gap between cities, both in terms of population and in terms of wages.

3.5.2 Model with geographical rigidity of wages

We now introduce the second key characteristic of the public sector, the fact that wages do not vary across cities. We consider the case where there are no spillovers, the combination with spillovers is studied in the following section.

The first key consequence is that, given that the wage in the public sector is equal across cities, the size will also be equalized: $N_1^{pub} = N_2^{pub}$.

The indifference condition between the public and the private sector is then given by:

$$\begin{aligned}\frac{w_i}{p_i^\alpha} &= \frac{N_1^{pub}}{N_1^{pub}+N_2^{pub}}\frac{\bar{w}}{p_1^\alpha} + \frac{N_2^{pub}}{N_1^{pub}+N_2^{pub}}\frac{\bar{w}}{p_2^\alpha} \\ &= \frac{1}{2}\left(\frac{1}{p_1^\alpha}+\frac{1}{p_2^\alpha}\right)\end{aligned}$$

So that:

$$w_1 = \frac{\bar{w}}{2}\frac{p_1^\alpha + p_2^\alpha}{p_2^\alpha}$$

In the previous sections we considered the traded good to be the numeraire. Here instead it's very convenient to have $p_1^\alpha + p_2^\alpha$ as the numeraire, so that the indifference conditions becomes:

$$w_1 = \frac{1}{2}\frac{\bar{w}}{p_2^\alpha}$$

Overall we obtain the following result:

Proposition 10. *In a model with geographical rigidity of wages in public sector*

(a) *The size of public sector is the same in both cities.*

$$N_i^{pub} = N_j^{pub}$$

(b) *Ratio of private to public larger in more productive cities, and the gap is wider than when wages in the public sector are flexible*

$$ln\frac{N_1^{pri}}{N_1^{pub}} - ln\frac{N_2^{pri}}{N_2^{pub}} = \eta\left(lnA_1 - lnA_2\right) + \eta\alpha\left(lnp_1 - lnp_2\right).$$

(c) *The gap in population between most and least productive cities is wider than when wages in the public sector are flexible*

$$\begin{aligned} ln(N_1) - ln(N_2) = \frac{-\alpha\eta}{1+\alpha\eta-\alpha}\left(ln(A_1^\eta + (\beta_{12})^{\eta-1}B^\eta) - ln(A_2^\eta + (\beta_{21})^{\eta-1}B^\eta)\right) \\ + \left(ln(A_1^\eta + (\beta_{12})^{\eta}B^\eta) - ln(A_2^\eta + (\beta_{21})^{\eta}B^\eta)\right). \end{aligned}$$

(d) *The gap in private wages and housing prices between most and least productive cities is wider than when wages in the public sector are flexible*

$$\ln\frac{w_1}{w_2} = \frac{\alpha}{1-\alpha+\alpha\eta}\ln(\frac{B^\eta + (\beta_{12})^{\eta-1}A_i^\eta}{B^\eta + (\beta_{21})^{\eta-1}A_j^\eta})$$

$$\ln\frac{p_1}{p_2} = \frac{1}{\alpha}\ln\frac{w_1}{w_2}$$

Proposition 10.(a) states that the size of the public sector is the same in both cities. This is a direct consequence of the fact wages and productivity do not vary across cities, so that the decision on employment will be identical.

Proposition 10.(a) states that the gap between the ratio of private to public workers widens when wages are rigid in the public sector. With fixed wages, public sector workers in city i can spend less on housing than if wages were flexible. The housing price is thus lower, and more private workers need to move to city i to clear the market. The relative size of private to

public in city i compared to j thus grows. In turn this implies that the gap in population and housing prices between the most and the least productive cities increases. Overall, the rigidity in wages has the opposite effect on the distribution of economic activity compared to the effect of spillovers. In the next section we study the balance between these different effects.

3.6 Conclusion

This article shows, both theoretically and empirically, that the location of public sector workers producing local public goods strongly influences the location of economic activity. Using comprehensive data on both public and private employment in France, we show that two distinctive features make the public sector stand out compared to the standard spatial equilibrium model.

First, we show that the fact that wages are centrally set with limited leeway to adjust them to local prices imply that wages are not equalized with those of the private sector within a city. This means a higher productivity in the private sector puts upward pressure on wages, in turn decreasing its attractiveness to public sector workers, who are better off in places with lower rents thus lower productivity (keeping amenities constant). We show that this might explain part of the imbalances between the respective numbers of workers in the public and in the private sector across the French territory.

Second, we show that the location of civil servants influences local economic activity through the spillovers they produce on the private sector. We use large public job cuts which occured during the period 2008 to 2012 and affected areas differently to highlight the implied losses in private employment in subsequent years. Consistently with the literature, we find that places suffering larger losses also lost employment in the non-tradable sector, which is easily explained by the drop in local demand for non-tradable goods. In contrast with the literature however, we find that places affected more strongly by the policy also suffer large losses in the tradable sector, which can only be rationalized through the existence of positive spillovers from the public sector to the tradable sector.

We then integrate the above findings in a spatial equilibrium model with two sectors, where public sector workers are sent in a city at random, have a fixed wage, and generate positive spillovers on the other sector. We show that the fixed-wage structure, because the wage-constrained sector can then consume less housing, tends to increase the gap in labor market size between low and high productivity cities. In contrast, because a higher productivity in the

private sector crowds out some public workers, the existence of spillovers affecting positively the private sector lowers this gap.

References

Albouy, D. 2009. "The Unequal Geographic Burden of Federal Taxation." *Journal of Political Economy.*

Bartik, T. J. 1991. "Who Benefits from State and Local Economic Development Policies?" *Books from Upjohn Press. W.E. Upjohn Institute for Employment Research.*

Becker, S., S. Heblich, and D. Sturm. 2013. *The impact of public employment: evidence from Bonn.* Technical report.

Boeri, T., A. Ichino, E. Moretti, and J. Posch. 2019. *Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces.* Working Paper, Working Paper Series 25612. National Bureau of Economic Research, February.

Britton, J., and C. Propper. 2016. "Teacher pay and school productivity: Exploiting wage regulation." *Journal of Public Economics* 133 (C): 75–89.

Chapelle, G., and J.-B. Eyméoud. 2017. "Can big data increase our knowledge of the rental market." *Sciences Po mimeo.*

Combes, P.-P., and L. Gobillon. 2014. "The Empirics of Agglomeration Economics." *Handbook of Regional and Urban Economics.*

Van Dijk, J. J. 2016. "Local employment multipliers in U.S. cities." *Journal of Economic Geography* 17, no. 2 (April): 465–487.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny. 2014. "Spatial Sorting." *Journal of Political Economy* 122(3):554–620.

Faggio, G. 2019. "Relocation of public sector workers: evaluating a place-based policy." *Journal of Urban Economics,* no. forthcoming.

Faggio, G., and H. Overman. 2014. "The effect of public sector employment on local labour markets." Spatial Dimensions of Labor Markets, *Journal of Urban Economics* 79:91–107.

Fajgelbaum, P., J. Morales E. Suarez-Serrato, and O. Zidar. 2016. "State Taxes and Spatial Misallocation." *mimeo.*

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift. 2018. *Bartik instruments: What, when, why, and how.* Technical report. National Bureau of Economic Research.

Hsieh, C.-T., and E. Moretti. 2017. "Housing Constraints and Spatial Misallocation." *mimeo.*

Jofre-Monseny, J., J. I. Silva, and J. Vázquez-Grenno. 2019. "Local labor market effects of public employment." *Regional Science and Urban Economics,* no. forthcoming.

Monras, J. 2019. "Minimum Wages and Spatial Equilibrium: Theory and Evidence." *Journal of Labor Economics,* no. Forthcoming.

Moretti, E. 2010. "Local Multipliers." *American Economic Review P&P* 100, no. 2 (May): 373–77.

Nakamura, E., and J. Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104 (3): 753–92.

Propper, C., and J. van Reenen. 2010. "Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance." *Journal of Political Economy* 118 (2): 222–273.

Roback, J. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6):1257–1278.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82:34–55.

Suárez Serrato, J. C., and P. Wingender. 2014. "Estimating the Incidence of Government Spending." Mimeo (December).

———. 2016. *Estimating Local Fiscal Multipliers.* Working Paper, Working Paper Series 22425. National Bureau of Economic Research, July.

Zenou, Y. 2009. *Urban Labor Economics.* Cambridge University Press.

Appendix

3.A Proofs

Proof of Proposition 8

Indifference between sectors gives:

$$A_i(N_i^{pri})^{-1/\eta} = B_i(N_i^{pub})^{-1/\eta}$$

And hence:

$$\frac{A_i}{B_i} = \frac{(N_i^{pri})^{1/\eta}}{(N_i^{pub})^{1/\eta}}$$

Indifference between locations implies:

$$\frac{A_1(N_1^{pri})^{-1/\eta}}{p_1^{\alpha}} = \frac{A_2(N_2^{pri})^{-1/\eta}}{p_2^{\alpha}}$$

which implies:

$$\frac{A_1}{A_2} = \frac{(N_1^{pri})^{1/\eta}}{(N_2^{pri})^{1/\eta}}\frac{p_1^{\alpha}}{p_2^{\alpha}}$$

Labor market clearing is given by:

$$N_i^{pri} + N_i^{pub} = N_i$$

$$N_1 + N_2 = 1$$

Combining labor market clearing and the indifference between sectors we obtain that:

$$\frac{A_i}{B} = \frac{(N_i - N_i^{pub})^{1/\eta}}{(N_i^{pub})^{1/\eta}}$$

And so:

$$(\frac{A_i}{B})^{\eta} N_i^{pub} = (N_i - N_i^{pub})$$

And

$$N_i^{pub} = \frac{1}{1 + (\frac{A_i}{B_i})^{\eta}} N_i$$

$$N_i^{pub} = \frac{B_i^{\eta}}{B_i^{\eta} + A_i^{\eta}} N_i$$

And that:

$$N_i^{pri} = \frac{A_i^{\eta}}{B_i^{\eta} + A_i^{\eta}} N_i$$

These conditions mean that the share of local employment in each local sector depends on the relative productivity between the two sectors.

From the indifference condition between regions we obtain that:[10]

$$(\frac{B^{\eta} + A_1^{\eta}}{B^{\eta} + A_2^{\eta}})^{\frac{1-\alpha}{1+\alpha(\eta-1)}} = (\frac{N_1}{N_2})$$

This implies result (c)

$$\ln N_1 - \ln N_2 = \frac{1-\alpha}{1+\alpha(\eta-1)} (\ln(B^{\eta} + A_1^{\eta}) - \ln(B^{\eta} + A_2^{\eta}))$$

Note now that $\frac{1-\alpha}{1+\alpha(\eta-1)} > 0$ if and only if:

$$\frac{1}{\alpha} > (1 - \eta)$$

And this always holds since $\alpha \in (0, 1)$ and $\eta > 0$.

We can now obtain the wages prevailing in equilibrium:

10. See appendix.

$$w_i^{pri} = w_i^{pub}$$

And:

$$\frac{w_1}{w_2} = \frac{A_1}{A_2}\left(\frac{\frac{A_1^\eta}{B^\eta + A_1^\eta}N_1}{\frac{A_2^\eta}{B_2^\eta + A_2^\eta}N_2}\right)^{-1/\eta}$$

Hence:[11]

$$\frac{w_1}{w_2} = \left(\frac{B^\eta + A_1^\eta}{B^\eta + A_2^\eta}\right)^{\frac{1}{\eta}\left(1-\frac{1-\alpha}{1+\alpha(\eta-1)}\right)}$$

This expression can be reexpressed as result (d).

Proof of Proposition 9

The inverse labor demand curve in the private sector is given by:[12]

$$w_i^{pri} = A_i \left(\frac{N_i^{pub}}{N_i}\right)^{\beta} (N_i^{pri})^{-1/\eta}$$

The inverse labor demand curve in the public sector is given by:

$$w_i^{pub} = B_i \left(\frac{N_i^{pub}}{N_i}\right)^{\beta} (N_i^{pub})^{-1/\eta}$$

Indifference between sectors gives:

$$A_i \left(\frac{N_i^{pub}}{N_i}\right)^{\beta} (N_i^{pri})^{-1/\eta} = B \left(\frac{N_i^{pub}}{N_i}\right)^{\beta} (N_i^{pub})^{-1/\eta}$$

And hence:

$$\frac{A_i}{B} = \frac{(N_i^{pri})^{1/\eta}}{(N_i^{pub})^{1/\eta}}$$

11.

$$\frac{w_1}{w_2} = \left(\frac{B^\eta + A_1^\eta}{B^\eta + A_2^\eta}\right)^{1/\eta}\left(\frac{N_1}{N_2}\right)^{-1/\eta}$$

Hence:

$$\frac{w_1}{w_2} = \left(\frac{B^\eta + A_1^\eta}{B^\eta + A_2^\eta}\right)^{1/\eta}\left(\left(\frac{B^\eta + A_1^\eta}{B^\eta + A_2^\eta}\right)^{\frac{1-\alpha}{1+\alpha(\eta-1)}}\right)^{-1/\eta}$$

12. This can be easily obtained from a CES production function that combines land and labor. Land is a fixed factor.

Indifference between locations implies:

$$\frac{A_1\left(\frac{N_1^{pub}}{N_1}\right)(N_1^{pri})^{-1/\eta}}{p_1^{\alpha}} = \frac{A_2\left(\frac{N_2^{pub}}{N_2}\right)(N_2^{pri})^{-1/\eta}}{p_2^{\alpha}}$$

which implies:

$$\frac{A_1}{A_2}\frac{\left(\frac{N_1^{pub}}{N_1}\right)^{\beta}}{\left(\frac{N_2^{pub}}{N_2}\right)^{\beta}} = \frac{(N_1^{pri})^{1/\eta}}{(N_2^{pri})^{1/\eta}}\frac{p_1^{\alpha}}{p_2^{\alpha}}$$

Labor market clearing is given by:

$$N_i^{pri} + N_i^{pub} = N_i$$

$$N_1 + N_2 = 1$$

Combining labor market clearing and the indifference between sectors we obtain that:

$$N_i^{pub} = \frac{B^{\eta}}{B^{\eta} + A_i^{\eta}} N_i$$

And that:

$$N_i^{pri} = \frac{A_i^{\eta}}{B^{\eta} + A_i^{\eta}} N_i$$

From the indifference condition between cities we obtain that:

$$\frac{A_1}{A_2}\frac{\left(\frac{N_1^{pub}}{N_1}\right)^{\beta}}{\left(\frac{N_2^{pub}}{N_2}\right)^{\beta}} = \frac{(N_1^{pri})^{1/\eta}}{(N_2^{pri})^{1/\eta}}\frac{p_1^{\alpha}}{p_2^{\alpha}}$$

hence:

$$\frac{A_1}{A_2}\frac{(\frac{B^{\eta}}{B^{\eta}+A_1^{\eta}})^{\beta}}{(\frac{B^{\eta}}{B^{\eta}+A_2^{\eta}})^{\beta}} = \frac{(\frac{A_1^{\eta}}{B^{\eta}+A_1^{\eta}}N_1)^{1/\eta}}{(\frac{A_2^{\eta}}{B^{\eta}+A_2^{\eta}}N_2)^{1/\eta}}\frac{p_1^{\alpha}}{p_2^{\alpha}}$$

Now from

$$\alpha\left(\frac{N_i^{pub}}{N_i}\right)^{\beta}(A_i(N_i^{pri})^{1-1/\eta} + B(N_i^{pub})^{1-1/\eta}) = p_i$$

$$\alpha(\frac{B^\eta}{B^\eta+A_i^\eta})^\beta(A_i(\frac{A_i^\eta}{B^\eta+A_i^\eta}N_i)^{1-1/\eta}+B(\frac{B_i^\eta}{B^\eta+A_i^\eta}N_i)^{1-1/\eta})=p_i$$

$$\alpha N_i^{1-\frac{1}{\eta}}(\frac{B^\eta}{B^\eta+A_i^\eta})^\beta(A_i(\frac{A_i^\eta}{B^\eta+A_i^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_i^\eta})^{1-1/\eta})=p_i$$

hence

$$\alpha^\alpha N_i^{\alpha(1-\frac{1}{\eta})}(\frac{B^\eta}{B^\eta+A_i^\eta})^{\alpha\beta}(A_i(\frac{A_i^\eta}{B^\eta+A_i^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_i^\eta})^{1-1/\eta})^\alpha=p_i^\alpha$$

So:

$$\frac{p_1^\alpha}{p_2^\alpha}=\frac{N_1^{\alpha(1-\frac{1}{\eta})}(\frac{B^\eta}{B^\eta+A_1^\eta})^{\alpha\beta}(A_1(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_1^\eta})^{1-1/\eta})^\alpha}{N_2^{\alpha(1-\frac{1}{\eta})}(\frac{B^\eta}{B^\eta+A_2^\eta})^{\alpha\beta}(A_2(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_2^\eta})^{1-1/\eta})^\alpha}$$

And thus:

$$\frac{A_1}{A_2}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^\beta}{(\frac{B^\eta}{B^\eta+A_2^\eta})^\beta}=\frac{(\frac{A_1^\eta}{B^\eta+A_1^\eta}N_1)^{1/\eta}}{(\frac{A_2^\eta}{B^\eta+A_2^\eta}N_2)^{1/\eta}}\frac{N_1^{\alpha(1-\frac{1}{\eta})}(\frac{B^\eta}{B^\eta+A_1^\eta})^{\alpha\beta}(A_1(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_1^\eta})^{1-1/\eta})^\alpha}{N_2^{\alpha(1-\frac{1}{\eta})}(\frac{B^\eta}{B^\eta+A_2^\eta})^{\alpha\beta}(A_2(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_2^\eta})^{1-1/\eta})^\alpha}$$

$$\frac{A_1}{A_2}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^\beta}{(\frac{B^\eta}{B^\eta+A_2^\eta})^\beta}=(\frac{N_1}{N_2})^{1/\eta}\frac{(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1/\eta}}{(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1/\eta}}\frac{N_1^{\alpha(1-\frac{1}{\eta})}(\frac{B^\eta}{B^\eta+A_1^\eta})^{\alpha\beta}(A_1(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_1^\eta})^{1-1/\eta})^\alpha}{N_2^{\alpha(1-\frac{1}{\eta})}(\frac{B^\eta}{B^\eta+A_j^\eta})^{\alpha\beta}(A_2(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_2^\eta})^{1-1/\eta})^\alpha}$$

$$\frac{A_1}{A_2}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^\beta}{(\frac{B^\eta}{B^\eta+A_2^\eta})^\beta}=(\frac{N_1}{N_2})^{\frac{1}{\eta}+\alpha(1-\frac{1}{\eta})}\frac{(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1/\eta}}{(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1/\eta}}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^{\alpha\beta}(A_1(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_1^\eta})^{1-1/\eta})^\alpha}{(\frac{B^\eta}{B^\eta+A_2^\eta})^{\alpha\beta}(A_2(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_2^\eta})^{1-1/\eta})^\alpha}$$

$$\frac{A_1}{A_2}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^\beta}{(\frac{B^\eta}{B^\eta+A_2^\eta})^\beta}=(\frac{N_1}{N_2})^{\frac{1}{\eta}+\alpha(1-\frac{1}{\eta})}\frac{(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1/\eta}}{(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1/\eta}}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^{\alpha\beta}(\frac{A_1^\eta+B^\eta}{(B^\eta+A_1^\eta)^{(1-1/\eta)}})^\alpha}{(\frac{B^\eta}{B^\eta+A_2^\eta})^{\alpha\beta}(A_2(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1-1/\eta}+B(\frac{B^\eta}{B^\eta+A_2^\eta})^{1-1/\eta})^\alpha}$$

$$\frac{A_1}{A_2}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^\beta}{(\frac{B^\eta}{B^\eta+A_2^\eta})^\beta}=(\frac{N_1}{N_2})^{\frac{1}{\eta}+\alpha(1-\frac{1}{\eta})}\frac{(\frac{A_1^\eta}{B^\eta+A_1^\eta})^{1/\eta}}{(\frac{A_2^\eta}{B^\eta+A_2^\eta})^{1/\eta}}\frac{(\frac{B^\eta}{B^\eta+A_1^\eta})^{\alpha\beta}(A_1^\eta+B^\eta)^{\alpha/\eta}}{(\frac{B^\eta}{B^\eta+A_2^\eta})^{\alpha\beta}(A_2^\eta+B^\eta)^{\alpha/\eta}}$$

$$\frac{(B^\eta)^\beta}{(B^\eta)^\beta} = (\frac{N_1}{N_2})^{\frac{1}{\eta}+\alpha(1-\frac{1}{\eta})} \frac{(B)^{\eta\alpha\beta}(A_1^\eta + B^\eta)^{\frac{\alpha}{\eta}+\beta-\alpha\beta-\frac{1}{\eta}}}{(B)^{\eta\alpha\beta}(A_2^\eta + B^\eta)^{\frac{\alpha}{\eta}+\beta-\alpha\beta-\frac{1}{\eta}}}$$

$$1 = (\frac{N_1}{N_2})^{\frac{1}{\eta}+\alpha(1-\frac{1}{\eta})} (\frac{A_1^\eta + B^\eta}{A_2^\eta + B^\eta})^{\frac{\alpha}{\eta}+\beta-\alpha\beta-\frac{1}{\eta}}$$

And so:

$$\ln N_1 - lnN_2 = \frac{\alpha\beta + \frac{1}{\eta} - \frac{\alpha}{\eta} - \beta}{\frac{1}{\eta} + \alpha(1-\frac{1}{\eta})} \ln(\frac{A_1^\eta + B^\eta}{A_2^\eta + B^\eta})$$

$$\ln N_1 - \ln N_2 = \frac{1 - \alpha - \eta\beta(1-\alpha)}{1 + \alpha(\eta - 1)} \ln(\frac{A_1^\eta + B^\eta}{A_2^\eta + B^\eta})$$

Now to wages:

$$\frac{w_1}{w_2} = \frac{A_1 \left(\frac{N_1^{pub}}{N_1}\right)^\beta (N_1^{pri})^{-1/\eta}}{A_2 \left(\frac{N_2^{pub}}{N_2}\right)^\beta (N_2^{pri})^{-1/\eta}}$$

$$\frac{w_1}{w_2} = \frac{A_1(\frac{B^\eta}{B^\eta + A_1^\eta})^\beta (\frac{A_1^\eta}{B^\eta + A_i^\eta} N_1)^{-1/\eta}}{A_2(\frac{B^\eta}{B^\eta + A_2^\eta})^\beta (\frac{A_2^\eta}{B^\eta + A_2^\eta} N_2)^{-1/\eta}}$$

$$\frac{w_1}{w_2} = \frac{A_1(\frac{B^\eta}{B^\eta + A_1^\eta})^\beta (\frac{A_1^\eta}{B^\eta + A_1^\eta})^{-1/\eta}}{A_2(\frac{B_j^\eta}{B^\eta + A_2^\eta})^\beta (\frac{A_2^\eta}{B^\eta + A_2^\eta})^{-1/\eta}} (\frac{N_1}{N_2})^{-\frac{1}{\eta}}$$

$$\frac{w_1}{w_2} = \frac{(\frac{B^\eta}{B^\eta + A_1^\eta})^\beta (\frac{1}{B^\eta + A_1^\eta})^{-1/\eta}}{(\frac{B^\eta}{B^\eta + A_2^\eta})^\beta (\frac{1}{B^\eta + A_2^\eta})^{-1/\eta}} (\frac{N_1}{N_2})^{-\frac{1}{\eta}}$$

$$\frac{w_1}{w_2} = \frac{B^{\eta\beta}}{B^{\eta\beta}} (\frac{B^\eta + A_1^\eta}{B^\eta + A_2^\eta})^{\frac{1}{\eta} - \beta - \frac{1}{\eta} \frac{1-\alpha-\eta\beta(1-\alpha)}{1+\alpha(\eta-1)}})$$

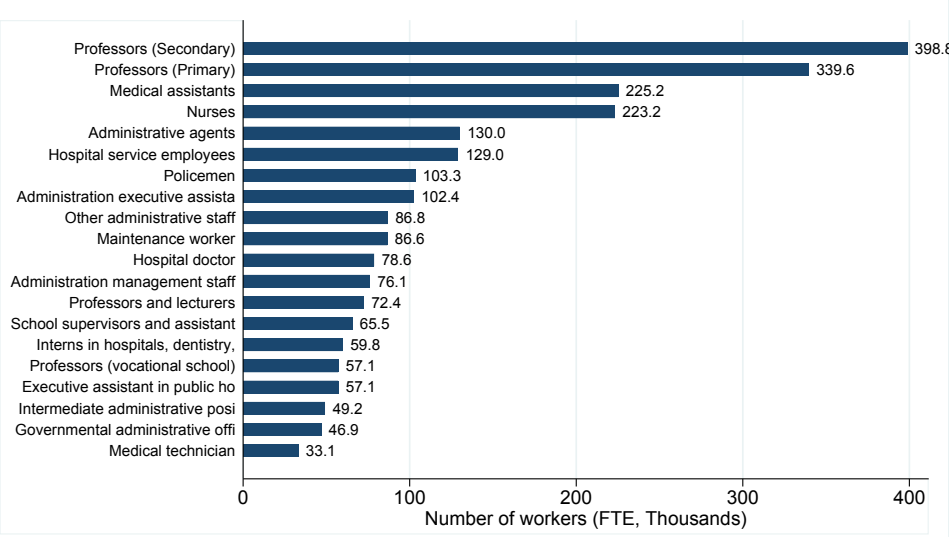
So:

$$\ln \frac{w_1}{w_2} = \frac{\alpha(1-\eta\beta)}{1 + \alpha(\eta - 1)} \ln(\frac{B^\eta + A_1^\eta}{B^\eta + A_2^\eta})$$

3.B Additional figures and tables

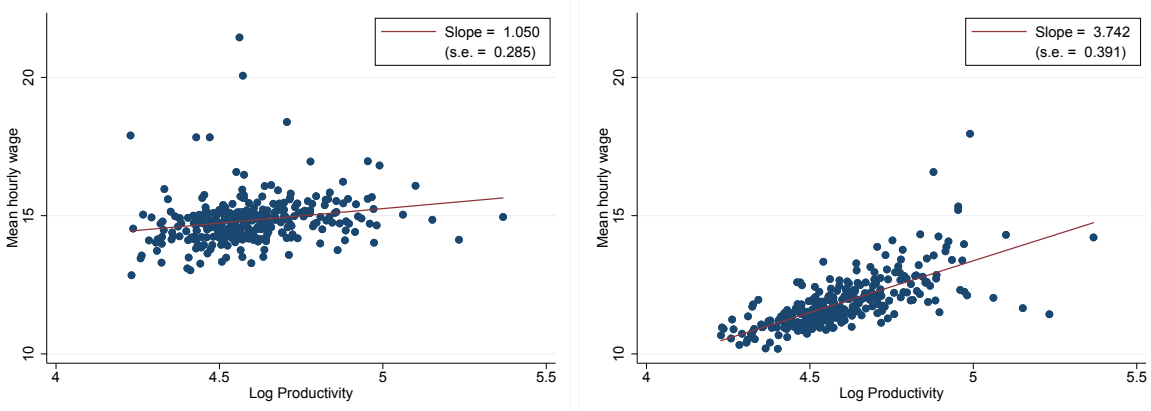
3.B.1 On cross-sectional evidence

Figure 3.A.1: Composition of the national public sector (by type of job)



Notes: This figure presents the number of workers within the national public sector in France, expressed in full-time equivalents, divided by the function of the worker. It includes both the fonction publique hospitalière and the function publique de l'État.

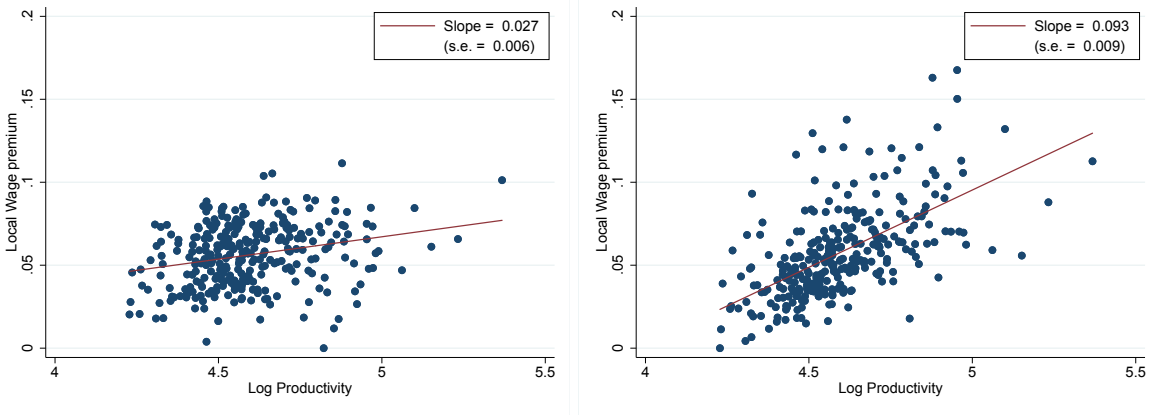
Figure 3.A.2: Mean wages in the public and private sector



(a) Public sector (nat.) **(b)** Private sector

Notes: This figure presents mean hourly wage in each commuting zone measured as described in Section 3.3.2 for the public sector (left panel) and the private sector (right panel) as functions of local log productivity. Public sector here means all the national public sector (Fonction Publique de l'Etat and Fonction Publique Hospitalière), but excludes the local public sector and State-owned firms. Private sector here means all private-owned businesses. Log productivity is the natural logarithm of value-added per worker in the commuting zone.

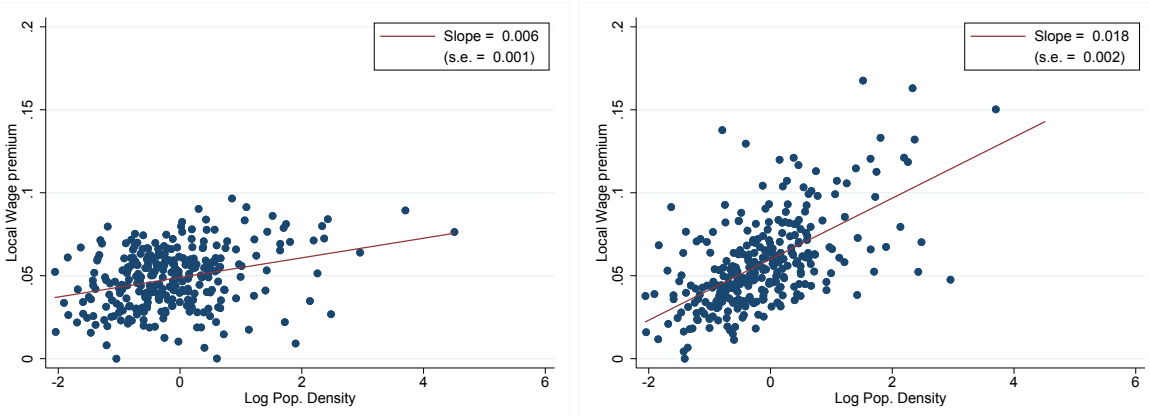
Figure 3.A.3: Wages in the public (core functions) and private sector as a function of the private sector productivity



(a) Public sector (nat.) **(b)** Private sector

Notes: This figure presents local wage premia measured as described in Section 3.3.2 for the public sector restricted to its core functions (left panel) and the private sector (right panel) as functions of local log productivity. Public sector here means all the national public sector (Fonction Publique de l'Etat and Fonction Publique Hospitalière), but excludes the local public sector and State-owned firms. Private sector here means all private-owned businesses. Log productivity is the natural logarithm of value-added per worker in the commuting zone.

Figure 3.A.4: Wages in the public and private sector as a function of population



(a) Public sector (nat.) **(b)** Private sector

Notes: This figure presents local wage premia measured as described in Section 3.3.2 for the public sector (left panel) and the private sector (right panel) as functions of local log productivity. Public sector here means all the national public sector (Fonction Publique de l'Etat and Fonction Publique Hospitalière), but excludes the local public sector and State-owned firms. Private sector here means all private-owned businesses. Log productivity is the natural logarithm of value-added per worker in the commuting zone.

3.B.2 On the RGPP reform

Table 3.A.1: Summary statistics of the variables used in the regressions

	Mean	Median	St.dev.	min	max
Number of private sector workers (FTE)	49434.92	19453.88	143760.85	2770.75	2279658.16
Number of public sector workers (FTE)	5626.42	2349.75	14926.49	342.06	231836.44
Δ log Private Empl. 09-10	-0.01	-0.01	0.02	-0.11	0.12
Δ log Private Empl. 09-12	-0.00	-0.00	0.03	-0.17	0.11
Δ log Private Empl. 09-15	0.00	0.00	0.06	-0.22	0.15
Δ log Public Empl. 09-10	0.01	0.01	0.05	-0.25	0.20
Δ log Public Empl. 09-11	-0.01	-0.00	0.06	-0.29	0.18
Δ log Public Empl. 09-12	-0.04	-0.03	0.07	-0.34	0.13
Number of workers in tradables	15945.96	6700.55	52852.19	531.42	857170.65
Δ log tradables Empl. (09-10)	-0.02	-0.02	0.06	-0.43	0.34
Δ log tradables Empl. (09-12)	-0.03	-0.03	0.07	-0.29	0.32
Δ log tradables Empl. (09-15)	-0.05	-0.05	0.11	-0.71	0.30
Number of workers in non-tradables	33375.97	13364.94	91181.34	1856.29	1421624.95
Δ log non-tradables Empl. (09-10)	-0.01	-0.01	0.03	-0.11	0.26
Δ log non-tradables empl. (09-12)	0.01	0.01	0.03	-0.14	0.10
Δ log non-tradables empl. (09-15)	0.02	0.02	0.05	-0.17	0.17
Number of workers in manufacturing	9170.35	5406.87	14063.46	252.76	166992.83
Number of workers in non-manufacturing	40151.58	14313.05	131296.33	1930.38	2111802.77
Pub (FPE) /Pri Ratio in 2009	0.10	0.10	0.03	0.04	0.21
Public/Private Wage ratio	0.04	0.05	0.03	-0.07	0.11
Log Rents	0.78	0.76	0.09	0.55	1.11
Log Productivity	4.46	4.43	0.18	4.04	5.21
Log Population	11.71	11.59	0.95	9.83	15.60
Share of Pop. in rural area	0.38	0.39	0.20	0.00	0.94
Population pre-trend (1999-2009)	0.06	0.05	0.05	-0.07	0.24
Private empl. (etp) Var. (pct) 06-09	0.07	0.07	0.05	-0.13	0.25
Observations	297				

Table 3.A.2: First stage of the 2SLS IV regression using the 2009-10 variation

	(1)	(2)	(3)
Dep. Var.: Δ log Public Empl. 09-10			
Pub (FPE) / Pri Ratio in 2009	-0.441***	-0.487***	-0.499***
	(0.093)	(0.097)	(0.100)
Bartik Shock		-0.065	-0.026
		(0.152)	(0.152)
Public/Private Wage ratio		-0.061	-0.058
		(0.089)	(0.092)
Log Rents		-0.021	-0.005
		(0.044)	(0.047)
Log Productivity		0.021	0.019
		(0.018)	(0.018)
Log Population		0.005	0.006
		(0.003)	(0.004)
Share of Pop. in rural area		0.010	0.014
		(0.022)	(0.024)
Population pre-trend (1999-2009)			-0.048
			(0.064)
Private Empl. (etp) Var. (pct) 06-09			-0.009
			(0.079)
R-squared	0.0794	0.0997	0.102
VCE	Robust	Robust	Robust
N obs	297	296	296
KP F-stat	22.65	25.07	25.11

Notes: This table presents the results of the first-stage of our 2SLS regression, i.e. the regression of our covariate of interest (Δ log of Public Employment 09-10) on the exluded instrument and other covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.A.3: First stage of the 2SLS IV regression using the 2009-12 variation

	(1)	(2)	(3)
Dep. Var.: Δ log Public Empl. 09-12			
Pub (FPE) / Pri Ratio in 2009	-0.698***	-0.798***	-0.736***
	(0.117)	(0.122)	(0.129)
Bartik Shock		0.318	0.286
		(0.200)	(0.212)
Public/Private Wage ratio		-0.182	-0.172
		(0.164)	(0.161)
Log Rents		0.063	0.017
		(0.057)	(0.061)
Log Productivity		0.036	0.040
		(0.025)	(0.025)
Log Population		0.010*	0.010*
		(0.005)	(0.005)
Share of Pop. in rural area		0.033	0.031
		(0.033)	(0.034)
Population pre-trend (1999-2009)			0.170*
			(0.090)
Private Empl. Var. 06-09			-0.066
			(0.097)
R-squared	0.0886	0.159	0.170
VCE	Robust	Robust	Robust
Controls	None	Static	Stat. and dyn.
N obs	297	296	296
KP F-stat	35.59	42.53	32.77

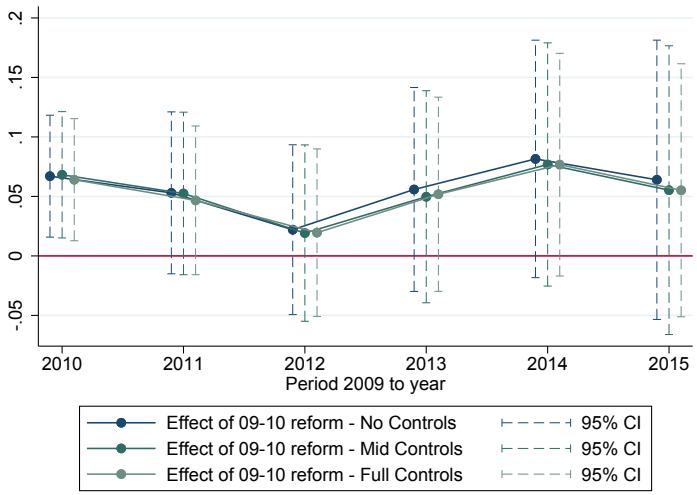
Notes: This table presents the results of the first-stage of our 2SLS regression, i.e. the regression of our covariate of interest (Δ log of Public Employment 09-12) on the exluded instrument and other covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.A.4: Second stage 2SLS IV regression table of pre-treatment outcome variations

	(1)	(2)	(3)	(4)
Dep. Var.: Δ log Private Empl.				
Period	06-07	07-08	08-09	06-09
Δ log Public Empl. 09-10	0.053	-0.074	-0.214	0.053
	(0.207)	(0.161)	(0.134)	(0.207)
Bartik Shock	0.951***	0.756***	0.630***	0.951***
	(0.126)	(0.113)	(0.116)	(0.126)
Public/Private Wage ratio	0.207***	0.138**	0.108*	0.207***
	(0.076)	(0.067)	(0.062)	(0.076)
Log Rents	0.032	0.013	0.017	0.032
	(0.044)	(0.037)	(0.033)	(0.044)
Log Productivity	0.019	0.012	0.005	0.019
	(0.017)	(0.016)	(0.014)	(0.017)
Log Population	0.010***	0.008***	0.006***	0.010***
	(0.003)	(0.003)	(0.002)	(0.003)
Share of Pop. in rural area	0.075***	0.077***	0.081***	0.075***
	(0.021)	(0.018)	(0.016)	(0.021)
VCE	Robust	Robust	Robust	Robust
N obs	296	296	296	296
1st stage KP F-stat	18.57	19.57	19.46	18.57

Notes: This table presents the results of the second-stage of 2SLS IV regressions of our dependent variable in the pre-treatment period (Δ log of Private Employment over the indicated period) on the Δ log of Public Employment 09-10 instrumented by the public private ratio in 2009 and other covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 3.A.5: OLS regression results of public variation between 2009 and 2010 on subsequent years.



This figure plots coefficients and confidence intervals of the second-stage of 2SLS regressions of the Δ log of private employment between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2010. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

Table 3.A.5: Reduced-form IV OLS regression table of pre-treatment outcome variations

	(1)	(2)	(3)	(4)
Dep. Var.: Δ log Private Empl.				
Period	06-07	07-08	08-09	06-09
Pub (FPE) /Pri Ratio in 2009	-0.02	-0.04	0.10	-0.02
	(0.03)	(0.03)	(0.06)	(0.09)
Bartik Shock	0.81***	0.71**	0.67***	0.94***
	(0.18)	(0.29)	(0.11)	(0.11)
Public/Private Wage ratio	0.05	0.03	0.12**	0.20***
	(0.04)	(0.03)	(0.06)	(0.08)
Log Rents	0.03*	-0.00	0.02	0.03
	(0.02)	(0.02)	(0.03)	(0.04)
Log Productivity	0.00	0.01	0.00	0.02
	(0.01)	(0.01)	(0.01)	(0.02)
Log Population	0.00**	0.00***	0.01**	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Share of Pop. in rural area	0.01	-0.00	0.08***	0.08***
	(0.01)	(0.01)	(0.02)	(0.02)
R-squared	0.191	0.167	0.302	0.334
VCE	Robust	Robust	Robust	Robust
N obs	296	296	296	296

Notes: This table presents the results of OLS regressions of our dependent variable in the pre-treatment period (Δ log of Private Employment over the indicated period) on the exluded instrument and other covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.A.6: Second stage 2SLS regression table of post-treatment outcome variations following the 09-10 shock

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.: Δ log Private Empl.						
Period	09-10	09-11	09-12	09-13	09-14	09-15
Δ log Public Empl. 09-10	0.18	0.40**	0.43**	0.48**	0.57**	0.64**
	(0.11)	(0.17)	(0.18)	(0.21)	(0.25)	(0.29)
Public/Private Wage ratio	-0.05	-0.01	0.01	-0.02	-0.03	-0.01
	(0.05)	(0.06)	(0.07)	(0.09)	(0.10)	(0.11)
Log Rents	-0.02	-0.01	0.01	0.04	0.08	0.08
	(0.02)	(0.03)	(0.04)	(0.04)	(0.05)	(0.06)
Log Productivity	-0.01	-0.01	-0.01	-0.02	-0.03	-0.03
	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)
Log Population	0.00	0.00	0.00	0.00	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Share of Pop. in rural area	0.01	-0.01	-0.00	0.01	0.05**	0.05*
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Bartik Shock	0.68***	0.53***	0.54***	0.61***	0.67***	0.80***
	(0.13)	(0.16)	(0.20)	(0.18)	(0.15)	(0.14)
VCE	Robust	Robust	Robust	Robust	Robust	Robust
N obs	296	296	296	296	296	296
1st stage KP F-stat	24.39	23.54	25.07	23.97	23.88	24.12

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of the second-stage of our baseline 2SLS IV regressions of our dependent variable (Δ log of Private Employment over the indicated period) on the Δ log of Public Employment 09-10 (instrumented by the public private ratio in 2009) and other covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.A.7: OLS reduced-form IV regression table of post-treatment outcome variations following the 09-10 shock

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.: Δ log Private Empl.						
Period	09-10	09-11	09-12	09-13	09-14	09-15
Pub (FPE) /Pri Ratio in 2009	-0.09*	-0.19***	-0.21***	-0.23***	-0.28***	-0.31***
	(0.05)	(0.07)	(0.07)	(0.09)	(0.10)	(0.12)
Public/Private Wage ratio	-0.06	-0.03	-0.02	-0.04	-0.06	-0.05
	(0.05)	(0.06)	(0.07)	(0.08)	(0.09)	(0.11)
Log Rents	-0.02	-0.02	0.01	0.03	0.06	0.07
	(0.02)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)
Log Productivity	-0.01	-0.00	-0.00	-0.01	-0.01	-0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Log Population	0.00*	0.00	0.00	0.01	0.01*	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Share of Pop. in rural area	0.01	-0.00	0.00	0.01	0.06**	0.06**
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Bartik Shock	0.66***	0.46***	0.52***	0.57***	0.63***	0.76***
	(0.13)	(0.13)	(0.18)	(0.16)	(0.13)	(0.12)
VCE	Robust	Robust	Robust	Robust	Robust	Robust
N obs	296	296	296	296	296	296
R-sq	0.114	0.0806	0.103	0.110	0.125	0.174

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

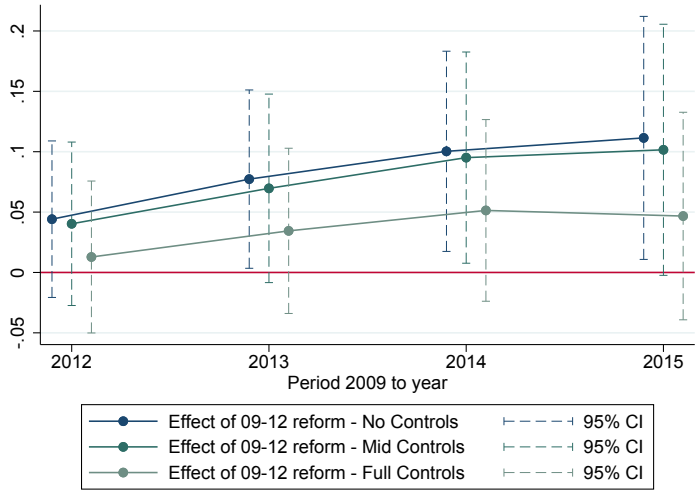
Notes: This table presents the results of OLS regressions of our dependent variable (Δ log of Private Employment over the indicated period) on the exluded instrument (the public private ratio in 2009) and other covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.A.8: Second stage 2SLS regression table of post-treatment outcome variations following the 09-12 shock

	(1)	(2)	(3)	(4)
Dep. Var.: Δ log Private Empl.				
Period	09-12	09-13	09-14	09-15
Δ log Public Empl. 09-12	0.27***	0.29**	0.34**	0.38**
	(0.10)	(0.12)	(0.13)	(0.16)
Public/Private Wage ratio	0.03	0.01	-0.00	0.02
	(0.07)	(0.09)	(0.10)	(0.11)
Log Rents	-0.01	0.02	0.04	0.04
	(0.04)	(0.04)	(0.05)	(0.06)
Log Productivity	-0.01	-0.02	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.03)
Log Population	0.00	0.00	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.01)
Share of Pop. in rural area	-0.01	0.01	0.05*	0.05*
	(0.02)	(0.02)	(0.02)	(0.03)
Bartik Shock	0.43**	0.52***	0.58***	0.71***
	(0.18)	(0.16)	(0.13)	(0.12)
VCE	Robust	Robust	Robust	Robust
N obs	296	296	296	296
1st stage KP F-stat	42.53	41.62	41.50	41.71

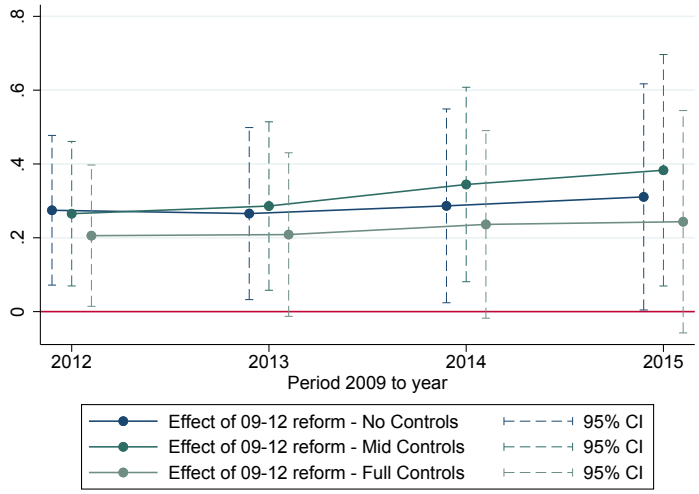
Notes: This table presents the results of the second-stage of our baseline 2SLS IV regressions of our dependent variable (Δ log of Private Employment over the indicated period) on the Δ log of Public Employment 09-12 (instrumented by the public private ratio in 2009) and other covariates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 3.A.6: OLS regression results of public variation between 2009 and 2012 on subsequent years.



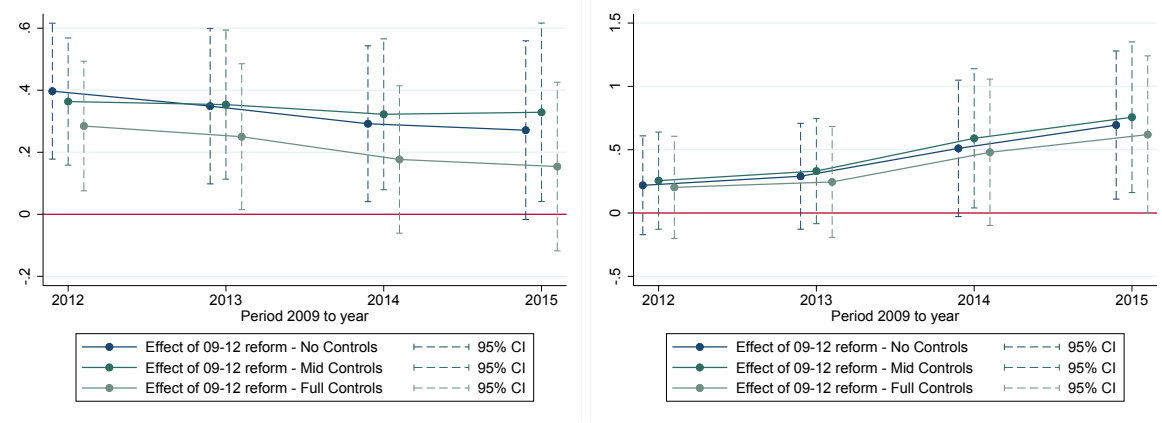
This figure plots coefficients and confidence intervals of OLS regressions of the Δ log of private employment between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2012. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

Figure 3.A.7: IV regression results of public variation between 2009 and 2012 on private employment in subsequent years.



This figure plots coefficients and confidence intervals of the second-stage of 2SLS regressions of the Δ log of private employment between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2012, instrumented by the public to private ratio in 2009. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

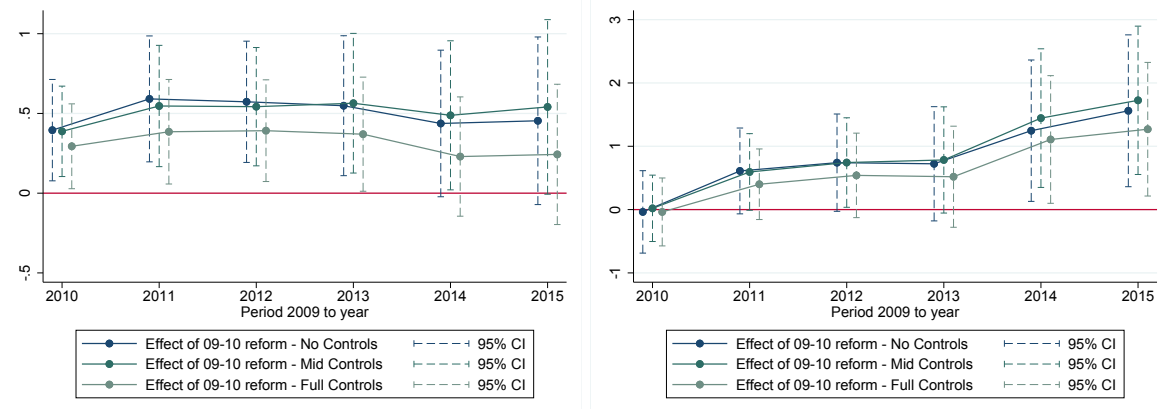
Figure 3.A.8: IV regression results of public variation between 2009 and 2012 on non-tradable and tradable sectors.



(a) Employment in non-tradables. **(b)** Employment in tradables.

These figures plot coefficients and confidence intervals of the second-stage of 2SLS regressions of the Δ log of private employment (in non-tradables in panel a, in tradables in panel b) between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2012, instrumented by the public to private ratio in 2009. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

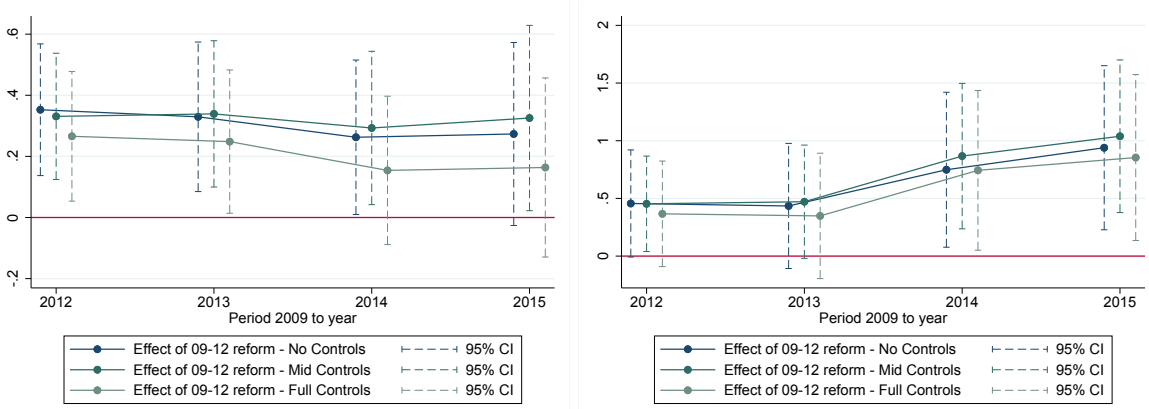
Figure 3.A.9: IV regression results of public variation between 2009 and 2010 on non-manufacturing and manufacturing sectors.



(a) Employment in non-manufacturing. **(b)** Employment in manufacturing.

These figures plot coefficients and confidence intervals of the second-stage of 2SLS regressions of the Δ log of private employment (in non-manufacturing in panel a, in manufacturing in panel b) between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2010, instrumented by the public to private ratio in 2009. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

Figure 3.A.10: IV regression results of public variation between 2009 and 2012 on non-manufacturing and manufacturing sectors.



(a) Employment in non-manufacturing.

(b) Employment in manufacturing.

These figures plot coefficients and confidence intervals of the second-stage of 2SLS regressions of the Δ log of private employment (in non-manufacturing in panel a, in manufacturing in panel b) between 2009 and the year indicated on the x-axis on the Δ log of public employment between 2009 and 2012, instrumented by the public to private ratio in 2009. The y-axis shows the value of the point-estimates and confidence interval bounds. Dots correspond to point-estimates, dashed brackets to their 95 % confidence interval. Each color corresponds to a different specification, where lighter blue means more control variables are added and the darkest blue means no controls other than Bartik shocks are added.

Résumé

Biens Publics Locaux

et Géographie de l'Activité Économique

La géographie de l'activité économique frappe par sa distribution extrêmement inégale. Certaines zones sont florissantes et attirent toujours plus d'entreprises et de travailleurs, atteignant des niveaux de concentration humaine extrêmement élevés. À l'inverse, d'autres zones dotées en apparence des mêmes avantages propres à leur localisation sont peu densément peuplées ou déclinantes. Ces inégalités spatiales suggèrent que le fait d'être situé dans un environnement dense est nécessaire pour les firmes, bien au delà des avantages naturels des localisations, et en dépit des coûts de congestion induits par la concentration. L'analyse de telles économies d'agglomération remonte à Marshall (1890), et la majorité des travaux réalisés par l'économie urbaine s'est attachée soit à mesurer leur force, soit à en comprendre et détailler les mécanismes sous-jacents. Duranton and Puga (2004) proposent ainsi une typologie de ces forces. Ils distinguent trois mécanismes principaux générant l'agglomération : les externalités de partage, relatives à la possibilité de bénéficier d'infrastructures et de fournisseurs en commun avec d'autres firmes, les externalités d'appariement, relatives à a meilleure qualité des appariements employeur-employé dans les bassins d'emploi plus denses, et les externalités de connaissance, relatives au partage de connaissance s'opérant entre acteurs localisés dans la même zone. Une façon différente d'analyser la question de l'agglomération peut être de considérer que, en fixant les avantages intrinsèques d'une localisation, les travailleurs et les firmes se localisent là où ils espèrent bénéficier de biens publics locaux de meilleure qualité. Si certains biens publics peuvent être considérés comme globaux, c'est-à-dire qu'ils affectent tous les agents d'une économie avec la même intensité (comme la défense ou les institutions démocratiques), de nombreux biens publics sont en revanche locaux, c'est-à-dire que seuls les agents localisés suffisamment près de leur lieu de production sont susceptibles de les consommer. Cette thèse s'intéresse aux mécanismes générant de l'hétérogénéité spatiale dans la production de biens publics locaux, en se focalisant sur deux mécanismes de ce type.

Dans le premier cas, le bien public étudié est la connaissance technologique. Elle est considérée comme un bien public du fait des coûts de transaction élevés qui pèsent sur sa diffusion. En effet, la connaissance devrait théoriquement être un bien public global : elle est à la fois non-rivale (la consommation de connaissance par un agent ne diminue pas l'usage que les autres peuvent en faire) et non-exclusive (il n'est pas possible d'empêcher un agent d'avoir recours à de la connaissance). Cependant, les études empiriques s'intéressant à la diffusion de connaissance sont marquées par deux faits saillants. En premier lieu, bien qu'on puisse s'attendre à une dissémination parfaite dans l'espace, la connaissance se diffuse mieux à faible distance. En

second lieu, les liens sociaux et économiques noués entre les agents sont un facteur favorisant pour la diffusion de contenus technologiques. L'introduction d'un réseau dans l'analyse permet de réconcilier la contradiction apparente entre la prédiction théorique que la diffusion de la connaissance devrait s'opérer sans frictions géographiques, et l'observation empirique que ces frictions existent. Le réseau permet ainsi une représentation naturelle dans laquelle la connaissance se diffuse librement le long des liens existants, mais la formation de liens est couteuse, et ce coût s'accroit avec la distance entre les agents, à mesure que la probabilité d'interactions sociales diminue. Le fait que la connaissance soit un bien public local a des conséquences particulièrement marquées : l'innovation est le principal moteur de la croissance économique (Aghion and Howitt, 1992; Aghion and Jaravel, 2015), et les écarts d'adoption technologique expliquent une part importante des différences de niveaux de vie entre pays (Comin and Hobijn, 2010). Si la représentation de la diffusion de la connaissance au sein d'un réseau est intuitive, sa mesure s'avère particulièrement ardue. En effet, la corrélation entre le réseau et les caractéristiques géographiques des agents est souvent tellement élevée que leur mesure séparée et l'identification de ce que chaque dimension cause est souvent impossible (Topa and Zenou, 2015). Ceci complique largement les tentatives d'isoler des liens causaux entre le réseau ou la géographie et n'importe quelle variable d'intérêt, et ce d'autant plus que le réseau est généralement partiellement ou complètement inobservable pour l'économètre. En dépit des difficultés de mesure, il est important de comprendre si ce qui influence fondamentalement l'innovation des agents est le fait d'être géographiquement proches les uns des autres, ou socialement proches. Cela a notamment des conséquences sur les politiques optimales de soutien à l'innovation : doivent-elles s'attacher à faciliter la diffusion des innovations en incitant les firmes à se regrouper dans une même zone, ou investir dans la création de programmes de recherche communs entre ces agents ?

Le second mécanisme analysé dans cette thèse génère de l'hétérogénéité spatiale dans la production d'un bien public par le secteur public. Dans la plupart des pays développés, le secteur public représente une part importante de l'emploi total[13] et est chargé d'assumer des fonctions bien plus larges que les simples fonctions régaliennes de l'État. Les fonctions d'enseignement et formation, santé, services sociaux, transport ou énergie dépendent souvent directement ou indirectement du secteur public. Toutes ces fonctions sont intrinsèquement locales: elles ne peuvent être consommées que près du lieu où elles sont produites. Elles sont néanmoins cruciales pour le bon fonctionnement de l'activité économique. Contrairement à la connaissance, elles sont

13. Entre 15 et 25% de l'emploi total dans la plupart des pays européens. Source OECD.stats, 'Government at a glance', 2017.

sujettes à congestion lorsque la population d'une zone augmente. Par conséquent, une production hétérogène de ces biens publics locaux devrait profondément influencer la géographie de l'activité économique.

Cette thèse s'intéresse à la manière dont les deux mécanismes développés précédemment, via l'hétérogénéité qu'ils créent dans la disponibilité des biens publics, sont déterminants pour la distribution spatiale de l'activité économique. Elle explore les causes d'une telle hétérogénéité, les conséquences que cela a sur les choix de localisation des agents économiques et plus généralement sur la géographie économique. Les deux premiers chapitres de la thèse se concentrent sur l'interaction entre les réseaux d'innovateurs, la géographie des flux de connaissance, et les choix de localisation des firmes innovantes. Plus précisément, le premier chapitre s'attache à comprendre comment les liens entre innovateurs affectent la géographie des flux de connaissance représentés par les citations de brevets. Le second chapitre étudie à quel point cette géographie contrainte des flux de connaissance influence les choix de relocaliser des centres de recherche-développement pour les entreprises. Le troisième chapitre se concentre sur un problème différent, dans le cadre duquel le bien public local étudié est produit par le service public. À cause de contraintes institutionelles (une égalisation nominale des salaires sur le territoire), toutes les zones ne bénéficient pas du même niveau de bien public. Ainsi, si le secteur public produit des externalités positives sur le secteur privé, le mécanisme étudié devrait avoir des conséquences sur la distribution de la taille et de la productivité des villes.

Chapitre 1 Le premier chapitre de cette thèse s'intitule "La Percolation de la connaissance dans l'espace", et se base sur un travail commun avec Pierre Cotterlaz. Il vise à expliquer comment les innovateurs sont informés de l'existence des idées développées par les autres en dépit de frictions importantes dans la diffusion de la connaissance. Il explique comment ces dynamiques d'apprentissage peuvent fournir une explication à l'évolution surprenante de la géographie des flux de connaissance.

Ce travail trouve sa source dans une vaste littérature empirique montrant que les flux de connaissance sont marqués par un fort biais spatial. Cette branche de la littérature remonte à l'article fondateur de Jaffe et al. (1993), qui utilise les citations de brevets et montre qu'un inventeur a plus de chances de citer des brevets développés autour de lui que d'autres brevets similaires mais développés dans des zones plus lointaines. Même si ce fait stylisé a suscité a un vif débat (Thompson and Fox-Kean, 2005; Henderson et al., 2005), de nombreux articles ont depuis

confirmé ce résultat, à la fois à un niveau très fin (Thompson, 2006; Murata et al., 2014) et sur des flux de citations internationales ou interrégionales (Bottazzi and Peri, 2003; Peri, 2005; Li, 2014). Le mécanisme à l'origine de ce biais spatial n'a néanmoins pas été formellement identifié. S'il semble intuitivement causé par la plus faible probabilité d'interactions sociales à des distances plus élevées, la difficulté à séparer les interactions sociales de la dimension géographique dans les données n'a été surmontée que récemment. Plusieurs articles se concentrent sur une mesure partielle de la proximité sociale et montrent que l'effet de la distance sur les flux de connaissance est réduit lorsque cette mesure est prise en compte (Singh, 2005; Breschi and Lissoni, 2009; Agrawal et al., 2008; Kerr, 2008). Head et al. (2018) montrent plus rigoureusement l'effet causal des liens sociaux sur les flux de connaissance : en se concentrant sur le cas des chercheurs en mathématiques, ils montrent que les liens hérités de la carrière des chercheurs (direction de thèse, collègues, etc.) détermine quels articles les chercheurs connaissent et donc citent, ce qui explique une grande part de l'effet de la distance sur les citations.

En se basant sur les travaux exposés précédemment, ce chapitre étudie les dynamiques de formation des liens entre innovateurs, et la manière dont ils sont liés à l'effet aggrégé de la distance que l'on observe. Il caractérise l'évolution des liens qu'une firme possède avec d'autres innovateurs, qui représentent son stock de connaissance, au cours de son cycle de vie. Pour ce faire, il étudie l'effet de ce qu'un agent connait à un point donné dans le temps sur les probabilités de citations futures. En pratique, nous utilisons les citations de brevets comme trace du stock de connaissance des firmes, et nous comparons les citations faites pour la suite pour mesurer son évolution. Nous initialisons le stock de connaissance sur une année donnée, et définissons comme contacts les innovateurs cités dans cette année, et dont la firme étudiée a donc connaissance. Nous comparons ensuite, pour chaque demande de brevet, les citations réalisées à un ensemble de citations attendues si le déposant du brevet avait eu accès à l'ensemble de la connaissance disponible. Afin d'isoler un tel ensemble de brevets pertinents et donc de citations attendues, nous utilisons les citations ajoutées par les experts des offices de brevets en charge de l'évaluation des demandes, qui ont pour mission de compléter la liste de citations faites par le demandeur. Nous montrons que les firmes acquièrent des nouvelles connaissances graduellement, via les contacts de leurs propres contacts. Ce fait stylisé est utilisé pour construire un modèle de diffusion basé sur Chaney (2018), dans lequel les dynamiques de formation des liens ont pour conséquence d'exprimer l'effet de la distance subi par les firmes comme une fonction de leur âge. Les petites firmes naissent avec un stock de connaissance générée par des firmes proches

d'elles, et étendent progressivement ce stock à des destinations plus lointaines au cours de leur vie. Cette caractéristique de la distribution spatiale des contacts permet d'expliquer pourquoi les flux de connaissance aggrégés sont restés affectés par la distance de manière aussi constante. Le modèle génère en effet deux prédictions, qui sont exceptionnellement bien vérifiées dans les données. D'une part, la distribution des tailles des innovateurs suit une loi de Pareto (et même plus précisément une loi de Zipf). D'autre part, une relation log-linéaire lie la taille des firmes et la distance moyenne (au carré) à laquelle elles citent.

Ce résultat peut produire une ébauche d'explication à un fait incompris : pourquoi l'effet de la distance mesuré en agrégé est-il resté constant depuis les années 1980 ? Cette période a en effet été marquée par l'avènement des technologies de l'information et de la communication, supposées faciliter grandement la diffusion de la connaissance et limiter l'effet de la distance via des outils comme Google Patents, qui rendent chaque demande de brevet accessible immédiatement. Ce chapitre propose un élément d'explication: si l'effet de la distance n'a pas décru, le lien entre la taille d'une firme et la distance de ses citations s'est lui largement amoindri. Des solutions possibles pour expliquer le mystère évoqué précédemment est ainsi que l'effet de la distance subi par les firmes prises individuellement a diminué, mais a été compensé par des changements dans la distribution ou dans la composition de ces firmes.

Chapitre 2 Le deuxième chapitre, intitulé "Mobilité des établissements de recherche-développement", considère comme donnés les réseaux locaux d'innovation analysés dans le chapitre précédent. Il vise à montrer, à la fois empiriquement et théoriquement, à quel point les firmes internalisent l'existence de tels réseaux locaux et les externalités technologiques qu'elles peuvent espérer en recevoir dans leurs choix de localisation. L'analyse trouve sa source dans la nécessité de lier deux faits solidement établis de la géographie des activités d'innovation : d'une part, le fait que la connaissance se diffuse mieux à des distances faibles ; d'autre part, le fait que les activités d'innovation sont plus concentrées que l'activité économique en général (Audretsch and Feldman, 1996; Buzard et al., 2017). Si les externalités technologiques locales fournissent une raison supplémentaire de se localiser au même endroit que d'autres entreprises du même type, peu d'articles se sont concentrés sur les choix de localisation des innovateurs en particulier. Ce chapitre étudie donc les choix de relocalisation d'établissements de recherche-développement, les externalités technologiques que les firmes abandonnent en quittant une zone et celles qu'elles espèrent recevoir dans la zone qu'elles rejoignent.

La compréhension de ce phénomène est nécessaire à la conception de politiques publiques d'innovation efficaces. En effet, tentant de reproduire le succès de la Silicon Valley, la plupart des pays développés ont tenté via des subventions ou des crédits d'impôts massifs de générer des regroupements géographiques[14]. Ces politiques ont été critiquées (parfois durement) par les économistes, tant à cause des effets souvent perçus comme faibles au regard de l'investissement public consenti que de leurs fondations théoriques incomplètes (Glaeser and Gottlieb, 2008; Duranton, 2011). Ceci justifie le développement d'une compréhension plus fine du lien entre les choix de localisation ou de relocalisation des firmes et les externalités technologiques qu'elles reçoivent de leur environnement direct.

La première partie de ce chapitre décrit le phénomène de mobilité tel qu'il apparait dans les données. J'utilise pour établir cette description les données DADS Postes, qui contiennent l'information exhaustive au niveau de l'individu sur l'emploi salarié en France, et caractérise les firmes innovantes comme celles employant des salariés dont le poste est explicitement libellé comme employé de recherche-développement. J'étudie ensuite leur propension à réallouer leurs activités de recherche sur le territoire, et la distance à laquelle ces réallocations s'opèrent. Deux faits saillants émergent de l'analyse : les firmes innovantes ont plus de chances de relocaliser un établissement que la moyenne des firmes employant des travailleurs hautement qualifiés, et lorsqu'elles bougent, elles le font à de plus grandes distances. J'apparie ces données de firmes à des données de brevets, et je construis une mesure de centralité dans le réseau d'innovation local pour chaque firmes, à partir des citations de brevets réalisées vers des innovations développées par des firmes géographiquement voisines. Cette mesure indique l'attractivité des zones que les firmes sont susceptibles de rejoindre et de celles qu'elles quittent, au regard de leur spécialisation technologique. L'analyse révèle qu'être plus densément connectée à ses voisines rend une firme moins susceptible de relocaliser ses activités de recherche. Par ailleurs, les firmes mobiles rejoignent des zones où elles peuvent espérer bénéficier de plus grandes externalités technologiques.

La seconde partie du chapitre se base sur les faits précédents, et formalise le problème auquel font face les entreprises innovantes cherchant à optimiser les externalités technologiques dont elles bénéficient. Pour lier leurs connections à leurs voisins avec leur choix de localisation, je construis un modèle théorique basé sur des outils développés dans le champ des "jeux sur réseaux" (Jackson and Zenou, 2015). Le modèle s'abstrait en revanche de concepts d'organisation industrielle classiques tels qu'une fonction de production réaliste ou une structure de compétition : cela

14. Par exemple en France, la politique des "pôles de compétitivité" a impliqué environ 2,5 milliards d'euros de dépense publique entre 2009 et 2014 (European Commission, 2016).

est nécessaire pour garder le modèle suffisamment simple pour pouvoir résoudre le problème auquel les firmes sont confrontées. Dans ce modèle, les firmes naissent avec un ensemble de liens dirigés vers certains de leurs voisins géographiques, par lesquels ils exraient de la connaissance. Au sein d'une localisation, les firmes diffèrent par leur centralité, une mesure de la densité de leur réseau local. Ces firmes ont un niveau d'information limité sur le niveau de connaissance dont elles seraient capables de bénéficier dans l'autre localisation. Je commence par étudier la stabilité, entendue dans ce contexte comme le fait qu'aucune firme n'ait une utilité espérée plus élevée dans l'autre localisation. Je montre que cette propriété du réseau dépend simplement de la différence entre la centralité la plus faible dans une localisation et la moyenne des centralités dans l'autre. J'étudie ensuite un jeu à deux étapes : les firmes peuvent dans un premier temps choisir de se relocaliser si cela leur garantit des profits espérés plus élevés, puis choisissent leur niveau d'effort de recherche-développement optimal. Je décris les équilibres du jeu si les choix de relocalisation se produisent de manière simultanée, qui sont équivalents aux situations précédemment classées comme stables. Je décris plus précisément un équilibre particulier, qui est à la fois Pareto-optimal et maximisateur de surplus. Je développe ensuite une analyse similaire avec un processus de mouvement séquentiel : il montre dans quelles conditions des mouvements en cascade peuvent se produire.

Chapitre 3 Le troisième chapitre de cette thèse, intitulé "Service public et distribution spatiale de l'activité économique", fait partie d'un projet commun avec Émeric Henry et Joan Monras. Il cherche à expliquer à quel point la localisation des fonctionnaires, qui produisent un bien public dont bénéficient firmes et travailleurs, est déterminant pour la distribution spatiale de l'activité du secteur privé. En utilisant des données exhaustives sur l'emploi public et privé en France, nous montrons deux caractéristiques originales qui font dévier l'analyse du modèle d'équilibre spatial standard.

La récente crise des "gilets jaunes" en France a remis au centre du débat public la question des inégalités territoriales d'accès au service public. Ceci fait écho aux propositions politiques polarisantes de suppressions massives d'emplois publics. Cependant, les travaux existants sur les effets générés par la localisation de tels emplois au niveau local sont récents, et n'en donnent pas une vision univoque. Notamment, il n'existe pas de réponse claire à la question de savoir s'ils excluent ou attirent des travailleurs du secteur privé dans les endroits où ils sont localisés. Une part de cette incertitude provient du fait que les expériences naturelles ayant pu être exploitées

consistent dans des relocalisations de fonctionnaires chargés de la production d'un bien public global, comme des employés d'administrations ministérielles ou des militaires (Becker et al., 2013; Jofre-Monseny et al., 2019; aus dem Moore and Spitz-Oener, 2012), plutôt que des biens publics locaux générés par des écoles, des hopitaux ou des commissariats. Le travail présenté dans ce chapitre tente de formaliser et d'illustrer les effets que l'on peut attendre de biens publics locaux générés par le service public, dans le cas particulier où les salaires des fonctionnaires sont fixés à l'échelle nationale et nominalement égaux sur l'ensemble du territoire. Il montre que les conclusions d'une telle analyse dévient du cadre standard d'équilibre spatial à deux secteurs (Rosen, 1974; Roback, 1982).

Le chapitre commence par souligner l'existence d'une fixation centralisée des salaires dans le secteur public, avec des possibilités d'ajustements locaux limitées, qui implique des salaires différents d'un secteur à l'autre au sein d'une ville. Il se place dans le cadre stylisé habituel d'équilibre spatial où les travailleurs du secteur privé sont parfaitement mobiles et doivent être marginalement indifférents entre les villes. Dans ce cadre, les utilités indirectes des travailleurs du secteur public ne sont pas égalisées entre les villes. Dans le cas français, ceci est illustré par certains faits connus, comme par exemple la tendance des fonctionnaires en début de carrière à être localisés dans les endroits les moins demandés avant de pouvoir choisir plus librement leur affectation. Sachant qu'une productivité du secteur privé plus élevée augmente la tension sur le marché du logement pour tous les travailleurs dans la ville, les employés du public ont intérêt à être localisés dans des endroits où les loyers, et donc la productivité, sont moindres. Cette analyse propose donc une explication aux importants déséquilibres spatiaux observés entre les effectifs du secteur public et du secteur privé entre zones du territoire français.

Le second fait important par lequel la localisation des employés du secteur public influence l'activité économique locale est la présence d'externalités générée en faveur du secteur privé. Nous exploitons des coupes d'emplois publics massives ayant eu lieu entre 2008 et 2012 (le non-remplacement d'un départ à la retraite sur deux pendant la RGPP) et touché les endroits de manière diverse pour identifier les variations d'emploi privé induites par ces pertes dans les années suivantes. En accord avec la littérature existante sur le sujet, nous trouvons que les endroits ayant subi le plus durement la réforme ont perdu relativement plus d'emploi privé dans les services locaux, ce qui s'explique aisément par la baisse de demande induite pour ces services locaux, et est désormais bien établie dans la littérature (Moretti, 2010; Faggio and Overman, 2014; Faggio, 2019). Nous trouvons en revanche, contrairement aux articles précédemment cités,

que les endroits ayant subi le plus durement la réforme ont également subi des pertes importantes dans les secteurs de biens échangeables (notamment l'industrie manufacturière). Ceci indique que le secteur public générait des externalités positives vers le secteur des biens échangeables.

Les résultats évoqués précédemment sont ensuite intégrés dans un modèle d'équilibre spatial avec deux secteurs, dans lequel les travailleurs publics sont envoyés dans une des localisations aléatoirement, ont un salaire fixe, et produisent une externalité positive vers le secteur privé. L'analyse montre que la structure de salaires fixes tend à accroitre l'écart de taille entre les villes très productives et les villes peu productives. Ceci s'explique par le fait que le secteur contraint par la régulation salariale peut dépenser moins sur le marché du logement que si les salaires étaient fixés de manière flexible, ce qui augmente moins la tension sur le marché du logement dans les villes les moins productives, augmentant encore l'écart entre villes de productivité différente. Les externalités positives produites par le service public ont l'effet inverse. Une productivité plus élevée dans le secteur privé attire les salariés du secteur privé aux dépens de ceux du public, et génère de la congestion sur le bien public local produit par le public. Ainsi, l'existence d'externalités affectant positivement le secteur privé diminue l'écart entre villes les plus productives et villes les moins productives.

References

Aghion, P., and P. Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60 (2): 323–51.

Aghion, P., and X. Jaravel. 2015. "Knowledge Spillovers, Innovation and Growth." *Economic Journal,* no. 583: 533–573.

Agrawal, A., D. Kapur, and J. McHale. 2008. "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics* 64, no. 2 (September): 258–269.

Audretsch, D. B., and M. P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (3): 630–640.

Becker, S., S. Heblich, and D. Sturm. 2013. *The impact of public employment: evidence from Bonn.* Technical report.

Bottazzi, L., and G. Peri. 2003. "Innovation and spillovers in regions: Evidence from European patent data." *European economic review* 47 (4): 687–710.

Breschi, S., and F. Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of Economic Geography:* lbp008.

Buzard, K., G. A. Carlino, R. M. Hunt, J. K. Carr, and T. E. Smith. 2017. "The agglomeration of American R&D labs." *Journal of Urban Economics* 101:14–26.

Chaney, T. 2018. "The Gravity Equation in International Trade: An Explanation." *Journal of Political Economy* 126 (1): 150–177.

Comin, D., and B. Hobijn. 2010. "An Exploration of Technology Diffusion." *American Economic Review* 100 (5): 2031–59.

Duranton, G. 2011. "California Dreamin': The Feeble Case for Cluster Policies." *Review of Economic Analysis* 3 (1): 3–45.

Duranton, G., and D. Puga. 2004. "Micro-foundations of urban agglomeration economies." Chap. 48 in *Handbook of Regional and Urban Economics,* 1st ed., edited by J. V. Henderson and J. F. Thisse, 4:2063–2117. Elsevier.

European Commission. 2016. *Smart Guide to Cluster Policy.* Technical report.

Faggio, G. 2019. "Relocation of public sector workers: evaluating a place-based policy." *Journal of Urban Economics,* no. forthcoming.

Faggio, G., and H. Overman. 2014. "The effect of public sector employment on local labour markets." Spatial Dimensions of Labor Markets, *Journal of Urban Economics* 79:91–107.

Glaeser, E. L., and J. D. Gottlieb. 2008. "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity* 2008 (1): 155–253.

Head, K., Y. A. Li, and A. Minondo. 2018. "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics." *Review of Economics and Statistics,* no. forthcoming.

Henderson, R., A. Jaffe, and M. Trajtenberg. 2005. "Patent citations and the geography of knowledge spillovers: A reassessment: Comment." *American Economic Review* 95 (1): 461–464.

Jackson, M. O., and Y. Zenou. 2015. "Games on networks." In *Handbook of game theory with economic applications,* 4:95–163. Elsevier.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *the Quarterly journal of Economics:* 577–598.

Jofre-Monseny, J., J. I. Silva, and J. Vázquez-Grenno. 2019. "Local labor market effects of public employment." *Regional Science and Urban Economics,* no. forthcoming.

Kerr, W. R. 2008. "Ethnic scientific communities and international technology diffusion." *The Review of Economics and Statistics* 90 (3): 518–537.

Li, Y. A. 2014. "Borders and distance in knowledge spillovers: Dying over time or dying with age?—Evidence from patent citations." *European Economic Review* 71:152–172.

Marshall, A. 1890. *Principles of economics.* Digireads. com Publishing.

Aus dem Moore, J. P., and A. Spitz-Oener. 2012. *Bye bye, GI-The impact of the US military drawdown on local german labor markets.* Technical report. Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Moretti, E. 2010. "Local Multipliers." *American Economic Review P&P* 100, no. 2 (May): 373–77.

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura. 2014. "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *The Review of Economics and Statistics* 96 (5): 967–985.

Peri, G. 2005. "Determinants of knowledge flows and their effect on innovation." *Review of Economics and Statistics* 87 (2): 308–322.

Roback, J. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(6):1257–1278.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82:34–55.

Singh, J. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5): 756–770.

Thompson, P. 2006. "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations." *The Review of Economics and Statistics* 88 (2): 383–388.

Thompson, P., and M. Fox-Kean. 2005. "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review* 95 (1): 450–460.

Topa, G., and Y. Zenou. 2015. "Neighborhood and network effects." In *Handbook of regional and urban economics,* 5:561–624. Elsevier.

List of Figures

1.1 Elasticity of international patent citation flows with respect to distance, over time. . . . 12
1.2 Design of the tests . . . 24
1.3 Comparison between applicant-added and examiner-added citations . . . 25
1.4 Estimation of λ. . . . 39
1.5 Estimation of μ . . . 41
1.6 Estimates of λ and μ over time. . . . 43
1.7 Estimates of μ by decade . . . 44
1.A.1 Number of patents/citations, decomposed by patent office . . . 51
1.A.2 Patents/citations for which we have geographic information (country) . . . 52
1.A.3 Type of outward citations, by patent office . . . 52
1.A.4 Correspondence between the number of times an examiner cites a patent and the number of times other examiners cite it. . . . 53
1.A.5 Evolution of the distance elasticity of citation flows over time, alternative estimators . . . 55
1.A.6 Distance elasticity of citation flows, sample split by patent office or by technological sector . . . 56
1.A.7 Distance elasticity of citation flows over time, alternative geographic measures . 56
1.A.8 Shape parameter of the Pareto distribution of firm size (λ), sample split by patent office or by technological sector . . . 67
1.A.9 Elasticity of the average squared distance of citations with respect to firm size (μ), sample split by patent office or by technological sector . . . 67
1.A.10 Elasticity of the average squared distance of citations with respect to firm size (μ), estimated with alternative distance measures . . . 68

1.A.11 Average squared distance of citations as a function of firm size 68

2.1 Distribution of relocation distances 78
2.2 Distance of patent citations 80
2.3 Example of two disconnected subgraphs 85
2.4 Example of a cascade: following a spillovers shock, the left-hand side location empties as firms move to the right-hand side one. 94
2.A.1 Distribution of patenting in chemestry and metalurgy over the French metropolitan territory 106

3.1 Composition of the national public sector 122
3.2 Variation in FPE employment during the RGPP 123
3.3 Wages in the public and private sector 124
3.4 Relative size of the private and public sector as a function of productivity . . . 125
3.5 Total population and housing prices as a function of productivity 126
3.6 A visual representation of the estimation first-stage 128
3.7 Maps of public employment variation and public/private employment ratio . . 128
3.8 IV regression results of public variation between 2009 and 2010 on private employment subsequent years. 131
3.9 IV regression results of public employment variation (Δ log) between 2009 and 2010 on non-tradable and tradable sectors. 133
3.A.1 Composition of the national public sector (by type of job) 149
3.A.2 Mean wages in the public and private sector 149
3.A.3 Wages in the public (core functions) and private sector as a function of the private sector productivity 150
3.A.4 Wages in the public and private sector as a function of population 150
3.A.5 OLS regression results of public variation between 2009 and 2010 on subsequent years. 154
3.A.6 OLS regression results of public variation between 2009 and 2012 on subsequent years. 159
3.A.7 IV regression results of public variation between 2009 and 2012 on private employment in subsequent years. 159

3.A.8 IV regression results of public variation between 2009 and 2012 on non-tradable and tradable sectors. 160
3.A.9 IV regression results of public variation between 2009 and 2010 on non-manufacturing and manufacturing sectors. 160
3.A.10 IV regression results of public variation between 2009 and 2012 on non-manufacturing and manufacturing sectors. 161

List of Tables

1.1 Estimates of the distance elasticity of citation flows (ζ). 22
1.2 Baseline results for network formation tests 29
1.3 Results of the spatial test 34
1.A.1 International Patent Classification, list of the sections 52
1.A.2 Technological distance in multiple citations by examiners 54
1.A.3 Results, Alternative versions of tests 58
1.A.4 Network search measured at the firm level 59
1.A.5 Network search using citations overlapping applicant and examiner added citations 60
1.A.6 Network search tests spotting groups with Orbis 61
1.A.7 Network search tests changing the initialization year 62
1.A.8 Network search tests changing the maximum size of contacts 63
1.A.9 Results of Placebo Tests 64
1.A.10 Estimates of the shape parameter of the Pareto distribution of firm size (λ) . . 65
1.A.11 Estimates of the elasticity of the average squared distance of citations with respect to firm size (μ) 65
1.A.12 Estimates of the elasticity of the average squared distance of citations with respect to firm size (μ), using within firm variations. 66
1.A.13 Semi-elasticity of firm size with respect to age. 66

2.1 Descriptive Statistics of relocations 77
2.2 Effect of being a R&D firm on the probability of defined types of relocation (panel a.) and distance at which they occur (panel b.) 78
2.3 Relocations and amount of R&D employment 79
2.4 Probability of moving and local centrality 82
2.5 Potential linkages and location chosen for relocation 83

2.A.1 Selected job categories (PCS) corresponding to R&D activities 103
2.A.2 Descriptive Statistics: Full Relocations (as in Duranton and Puga, 2001) . . . 104
2.A.3 Descriptive Statistics: Relocations of at least 50% and at least two full-time employees of each category 104
2.A.4 Descriptive Statistics: Relocations of at least 25% and at least one full-time employees of each category 105
2.A.5 Descriptive Statistics: Relocations of at least 75% and at least one full-time employees of each category 105
2.A.6 Effect of global versus local centrality on the probability of moving 107

3.1 Results of the main regressions for the 2009 to 2013 period 132
3.A.1 Summary statistics of the variables used in the regressions 151
3.A.2 First stage of the 2SLS IV regression using the 2009-10 variation 152
3.A.3 First stage of the 2SLS IV regression using the 2009-12 variation 153
3.A.4 Second stage 2SLS IV regression table of pre-treatment outcome variations . . 154
3.A.5 Reduced-form IV OLS regression table of pre-treatment outcome variations . . 155
3.A.6 Second stage 2SLS regression table of post-treatment outcome variations following the 09-10 shock 156
3.A.7 OLS reduced-form IV regression table of post-treatment outcome variations following the 09-10 shock 157
3.A.8 Second stage 2SLS regression table of post-treatment outcome variations following the 09-12 shock 158