Institut d'études politiques de Paris
ÉCOLE DOCTORALE DE SCIENCES PO
Programme doctoral en économie
Département d'économie
Doctorat en sciences économiques

Three Essays on Spatial Frictions

Pierre Cotterlaz

Thesis supervised by
Thierry Mayer, Professeur des universités, IEP de Paris
defended on 3rd June, 2021

Jury:
Maria Bas, Professeur des universités, CES, Université Paris 1 Panthéon-Sorbonne (Referee)
Matthieu Crozet, Professeur des universités, Université Paris-Saclay (Referee)
Anne-Célia Disdier, Professeur associé, Paris School of Economics,
and Directrice de recherche, INRAE
Cécile Gaubert, Associate Professor, University of California, Berkeley
Emeric Henry, Professeur des universités, IEP de Paris
Thierry Mayer, Professeur des universités, IEP de Paris

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, Thierry Mayer, pour ses conseils avisés et ses nombreuses suggestions qui ont grandement contribué à améliorer ces travaux. Cette thèse doit également beaucoup à Emeric Henry, dont la disponibilité et le soutien constant ont été extrêmement précieux.

J'ai aussi eu la chance de pouvoir travailler avec des coauteurs idéaux: Emanuele, toujours enthousiaste et dynamique; Arthur, bien plus patient que je ne le mérite et Etienne, jamais avare de bonnes idées. Aucun des chapitres de cette thèse n'aurait pu voir le jour sans ces collaborations.

Je remercie également Maria Bas, Matthieu Crozet, Anne-Célia Disdier, Cécile Gaubert et à nouveau Emeric Henry qui ont accepté d'être membres de mon jury. Leurs commentaires et suggestions aideront grandement à améliorer les papiers ici proposés.

Cette thèse a beaucoup bénéficié de l'environnement très favorable du département d'économie et du LIEPP, des séminaires nombreux et variés qui y ont été organisés, et des discussions aussi stimulantes qu'enrichissantes avec les professeurs. Je remercie en particulier Golvine de Rochambeau, Guy Laroque, Florian Oswald, Johannes Boehm, Julia Cagé, Thomas Chaney, Philippe Martin, Pierre-Henri Bono, Nicolas Coeurdacier, Etienne Wasmer, Jeanne Commault et Jean-Marc Robin. Je remercie également les organisateurs du Paris Trade Seminar, qui fut une excellente opportunité pour découvrir les travaux les plus récents en commerce international et donna lieu à de nombreux échanges avec les chercheurs invités. Et un grand merci aux équipes administratives du département d'économie, du LIEPP et de l'école doctorale: Pilar Calvo, Claudine Lamaze, Sandrine le Goff, Cécile Markarian, Christelle Hoteit, Samira Jebli, Juliette Seban, Hadjila Nezlioui-Serraz, et Guillaume Sarrat de Tramezaigues.

Il me faut également souligner l'apport inestimable des échanges avec mes camarades doctorants. Il est certain que ces travaux seraient bien moins aboutis sans les nombreuses rencontres qui ont émaillé les quelques années passées dans les combles. Je tiens en particulier à remercier Charles, entomologue de génie, qui des pubs de Cardiff aux terrains d'urban d'Aubervilliers a toujours su ouvrir de nouvelles perspectives sur la discipline; Ludovic "sneaky lobster" pour les innombrables discussions; Pierre D. pour ses tâcles rugueux; Florin, qui m'a tant appris sur Trump, les tabloïds britanniques et la Moldavie; Jean-Louis, sauveur du bâtiment et prévisionniste officiel du Mondial 2018, et Elisa, pour sa participation indispensable aux pauses solarium. Je me souviendrai aussi des nombreuses discussions et des bons moments passés avec Jean-Baptiste, Aseem, Joanne, Jan, Clément, Léa, Sophie, Marcos, Nicolo, Edoardo, Marcos, Paul, Florence, Lucas, Julien, Quentin, Pierre G., Odran, Assia, Eve, Jules, Stefan, Tyler, Andreea, Jean-Benoît, Guillaume C., Guillaume W., Valeria, Camille, et Riccardo.

Une nouvelle page de ma thèse s'est ouverte en 2019, et je remercie Sébastien Jean et Vincent Vicard de m'avoir fait confiance en me permettant d'intégrer le CEPII. Parce qu'il marie recherche et étude des grandes questions d'économie internationale, le CEPII est pour moi une source inépuisable de découvertes et d'apprentissages. C'est aussi un long couloir peuplé de fins connaisseurs du commerce international: merci à Alix, Cécilia, Charlotte, Cristina, Deniz, Gianluca, Houssein, Kévin et Laurence qui ont si rapidement partagé leur expertise avec moi. En poussant un peu dans le couloir, on découvre une équipe migration et une équipe macro avec lesquelles les échanges sont aussi animés que passionnants: merci donc à Anthony, Axelle, Camilo, Carl, Christophe, Erica, Isabelle, Julia et Thomas. Le CEPII ne fonctionnerait pas aussi bien sans l'aide précieuse d'Arnaud, Béatrice, Dimitri, Evguénia, Laure, Sophie, Sylvie et Véronique. Je n'oublie pas les collègues du CAE, qui, même s'ils ne sont pas au même étage, sont vite devenus des compagnons indispensables: merci à Madeleine, Florentin, Samuel, Baptiste et bien sûr Etienne à nouveau, que j'ai eu la chance de continuer à côtoyer après mon départ du département d'économie. Et merci à Deniz, Aude et Guillaume, avec qui c'est toujours un plaisir de travailler. La cantine et le Lab' se sont rapidement établis comme des rendez-vous incontournables où se tiennent des débats aussi variés qu'inattendus, dont la valeur est inappréciable.

Enfin, je remercie les nombreux amis qui n'ont pas désespéré de me voir terminer cette thèse un jour, et sans qui tout aurait été infiniment plus difficile. Merci à Cédric, infatigablement présent pour m'encourager; Lilian, juge implacable de l'enjaillement; Damien, jamais aussi désenchanté que dans la défaite; Laura, avocate redoutable et copilote indulgente; Jo, aux talents si multiples que je renonce à les recenser; Clarisse, rivale imbattable aux blind tests, et bien sûr Martin Bernard, hôte de tant de soirées mémorables. Merci également à tous les autres gagnants du Ton Air: Diana, Dédé, Thomas, Bastien, Charlotte, Marie-Sophie, Pierre, Florent, Maxime, Mégane, Barbara, et Judith, victorieuse par procuration. Je pense également à la team Rive Gauche, dont le centre de gravité s'est certes inexorablement déplacé depuis les premières années à Sciences Po, mais au sein de laquelle la bonne humeur est toujours au rendez-vous: Julie, fidèle soutien sudiste depuis mon arrivée à Paris; Caroline, pas sudiste par la géographie mais on l'oublierait presque; Sean, toujours partant pour parler des sujets les plus variés; Anne-Sophie et son l'enthousiasme communicatif; Louise, interrogatrice impitoyable mais si bienveillante; et aussi Camille, Christophe, Martin, Elsa, Aster, Cyril, Aurélie, Jean-Baptiste, Laura, Emmanuel, et le petit Jean. Et merci aux amis dignois, Caroline, Rémi, Malcolm et Vincent, car c'est dans les paysages bas-alpins que tout a commencé.

Enfin, et par-dessus tout, rien n'aurait été possible sans le soutien constant de toute ma famille. Je remercie en particulier mon frère Vincent, ma sœur Lucie et mes grand-parents. Et je ne saurais dire à quel point je suis reconnaissant envers mes parents, qui m'ont tant appris, et m'ont fait confiance de bout en bout, sans trop me demander quand tout cela serait terminé.

Contents

List of Figures

List of Tables

Introduction

Spatial frictions are key in explaining many economic phenomena. This thesis provides three pieces of evidence on the origins, prevalence and consequences of such frictions.

In the first chapter, we focus on spatial frictions in the diffusion of knowledge. We explain the puzzling persistence and stability of the spatial decay in patent citation flows by innovator networks. We establish that knowledge percolates: firms disproportionately cite new patents from prior contacts, and form links with contacts of their contacts. Embedding this percolation into a network formation model is sufficient to rationalize the negative link between aggregate knowledge flows and distance.

In the second chapter, we shed some light on the role of spatial information frictions in shaping international trade flows. We make use of the specific context of the XIXth Century, during which the creation of international news agencies facilitated the transmission of information across countries. We show that trade between a pair of countries increases when both are covered by a news agency. The reduction in information friction was therefore one of the many factors behind the First Globalization.

The last chapter investigates whether transport costs are the main component of within-country trade costs. While it is well-established that international trade costs are not limited to transport costs, evidence is much scarcer for intra-national trade flows. We use hurricane Sandy as a natural experiment shifting upwards transport costs in some areas of the US to establish that if transport costs were the sole driver of the distance elasticity of trade flows within the US, this distance elasticity would be much lower.

Chapter 1: The Percolation of Knowledge across Space

Despite considerable improvements in information and communication technologies in the past three decades, geographical distance remains a serious hindrance to knowledge diffusion. We estimate the elasticity of international patent citation flows with respect to geographical distance and show that it has remained very stable from 1980 to 2010, 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 surprising as most conventional distance-related costs, such as transport costs or tariffs, do not apply to ideas. Even more puzzling is the fact that digitization and communication technologies such as online patent search tools seem to have had no effect on knowledge diffusion as a whole.

This chapter shows that the dynamics of network formation over firms' life-cycle are key to understanding the aggregate effect of distance: young and small firms have spatially close contacts, and gradually expand their network as they grow.

Our contribution is twofold. In a first part, we identify how links form: we document a phenomenon called “triadic closure” in the economics of networks literature, in which firms disproportionately form links with firms two steps away from them (i.e. with contacts of contacts). To unveil this mechanism, we build all network links through patent citations, and implement a novel identification strategy to show the influence of the network on the probability that a link forms. We provide 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 diffusion process is reminiscent of the physics phenomenon of percolation, approaching knowledge as a fluid making its way from one inventor to another along network paths.

We causally test the influence of existing contact links on the network formation between innovators. Using previous patent citations to build contacts, we show that a firm is more likely to cite either a patent originating from a contact or cited by a contact than a similar patent from outside its close network. For identification, 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 1.5 times more likely than examiners to cite patents owned by their contacts, yet hiding some heterogeneity between small and large firms. Moreover, firms are 35% more likely to cite patents that were cited directly by their contacts. These effects are robust to a wide range of checks.

In a second part of the paper, we show the aggregate consequences of this network formation process, and in particular how it can explain the effect of distance on knowledge flows. To do so, we incorporate the above diffusion process into 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 and age increase, simply because of the time they have had to expand their network. This model delivers two predictions, relative to the firm size distribution and the relation between firm size and the distance of citations, which naturally lead to an aggregate effect of distance. 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 their own right. 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 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.

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), as well as inter-firm mobility of engineers (Almeida and Kogut, 1999; Breschi and Lissoni, 2009; Serafinelli, 2019), and input-output linkages (Carvalho and Voigtländer, 2014).

Chapter 2: Information in the First Globalization: News Agencies and Trade

Just as knowledge, information does not flow frictionless across borders. These constraints on the international diffusion of information are likely to impede trade, since knowledge of the foreign market characteristics (market size, price, trade costs, demand shifters) is of prime importance for the exporters, while for the importers, the sourcing choice is determined by the information available on price and quality from different markets.

The specific context of the XIXth century provides a unique opportunity to document the importance of information in shaping trade patterns. Indeed, this period witnessed the birth of global news agencies, which systematically collected and transmitted information across borders, so that, for the first time, news became widely available from almost all parts of the globe, with sharply reduced delays. News agencies are wholesalers of information: they gather news and sell them to governments, businesses, and newspapers. The three largest news agencies quickly syndicated into an efficient cost-sharing organization: each of them was given a monopoly over a set of countries, and in exchange committed to share information on these countries with the other news agencies. The sharing of information among the three global news agencies was truthfully enforced, since it ensured that they would stay ahead of the competition. Therefore, being covered by a global news agency meant becoming part of an international network of news sharing, including commercial news.

The development of international news agencies was deeply intertwined with the construction of an international telegraph network: news agencies relied on the telegraph to communicate and often contributed to its expansion. The telegraph was a considerable improvement upon previous technologies (physical transport of the mail on steamships, railways or horses) which had a considerably lower speed (sometimes months) and more volatile delivery times. It made communications easier, but it did not provide a centralized and reliable source of business information. In other words, in the absence of a global news agency, telegraphs reduced only the communication frictions, without affecting much the amount of information available to the public. Typically, communication was private and only benefited to the users of the telegraph themselves. On the other hand, news agencies collected, gathered and sold information that could then be accessed by anyone at a low cost. Our analysis disentangles the effects of reduced communication costs from the effects of improved information access, a distinction that previous studies were unable to make.

These two major innovations did not affect all pairs of countries simultaneously. The success of the telegraph was immediate, but the cost of the infrastructure and technical factors meant that not all countries could be quickly connected. Similarly, the global news agencies did not cover the entire world immediately. They started by sharing Europe and then extended gradually the scope of their syndication agreement through contracts struck in 1859, 1867, 1876, 1889 and 1902. This sequential entry of country pairs into the telegraph and news agencies networks is key for our identification strategy, because it allows us to estimate a panel data version of the gravity equation, meaning that on top of the usual origin and destination time varying fixed effects, we can include country-pair fixed effects, which control for any time-unvarying characteristic of the two countries.

To identify the information channel, we focus on the interaction between telegraph connections and news agency coverage: while the effect of the telegraph alone can be attributed to the sole decrease in communication costs, the interacted term specifically isolates the contribution of an improved access to news on the potential trade partner. The effect is sizable: our estimates imply that trade increases by an additional 30% when two countries are included in the global network of news diffusion, on top of being connected by a telegraph. Additionally, we corroborate previous studies that documented a positive effect on trade of the telegraph. We find that, even in the absence of coverage by a global news agency, trade flows increase by 40% when two countries become connected by a telegraph. However, news agencies, in the absence of telegraph, do not trigger any significant increase in trade, suggesting that they were unable to operate at full efficiency without an appropriate communication technology.

We then analyze the time dynamics of the effect through an event-study, and find a progressive increase in its magnitude, which slowly rises up to thirty years after the dyads are connected, a picture consistent with a slow constitution of business networks between the countries that benefited from an improved mutual access to information. Finally, we provide evidence supporting the hypothesis that the trade effect is indeed driven by an increase in the quantity of information available on foreign countries. First, we document an increase in trade volatility after the connection, in line with the findings of Steinwender, 2018. This is consistent with a better ability of traders to adapt to market conditions. Second, using data on French newspapers, we find an increase in the presence of a country in the articles once this country benefits from a telegraph connection and from news agency coverage.

While estimated from a historical event, the results are relevant to understand contemporary trade flows, since exporters still may lack the necessary information, despite considerable improvement in communication technologies. This chapter does not take a stance on the precise mechanisms through which better access to information on foreign countries affects trade. The fact that the effect keeps growing over a relatively long time horizon suggests that improved information may have affected trade through long-run channels, such as Foreign Direct Investment, human migration flows or even a convergence in cultural tastes.

Chapter 3: Trade and Transport Costs: Evidence from Hurricane Sandy

International trade flows decrease strongly when distance increases, and only part of this decrease can be attributed to transport costs. This points to the existence of other, large, "dark" trade costs, not observable but whose presence is necessary to rationalize the observed gravity patterns of trade flows (Head and Mayer, 2013). Potential sources for these frictions are diverse. They include, for example, differences in culture and tastes, a lack of mutual trust, and the spatial decay of information (as evidenced in the first two chapters). We could expect these additional dark trade costs to be lower within countries: culture and tastes are arguably more similar within a country than between countries, the spatial decay of information should be lower, and mutual trust should be higher. Additionally, tariffs and the “grey” trade costs of crossing borders (non-tariff barriers to trade) are absent. Nevertheless, in this chapter we show that only part of the distance elasticity of US intra-national trade flows can be attributed to transport costs, pointing to the existence of additional trade costs even within each country.

More precisely, we find that while the total distance elasticity of within-US trade flows is -0.84, this distance elasticity would be significantly smaller, around -0.06, if there were no other trade costs than transport costs. This result is established by making use of a natural experiment: hurricane Sandy, that hit the North-East of the US at the end of October 2012. The hurricane caused massive disruptions on the transport infrastructure, leading to a sizable increase in transport costs in the affected areas. Depending on the optimal path between each origin and destination, some dyads were more affected by these disruptions than others: dyads for which a large share of the usual optimal route goes through the affected region experienced a larger increase in transport costs than dyads for which the usual optimal path avoids the damaged area. For instance, transport costs between Los Angeles and Seattle were not affected at all, unlike transport costs between Boston and Miami, among others. We obtain a lower bound for the road distance equivalent of this change in transport costs and regress trade flows on this time varying distance. The distance effect obtained doing so is much lower than its cross-sectional counterpart, which confirms that the cross-sectional distance elasticity of trade flows captures trade costs unrelated to transport costs.

We compute the change in transport costs induced by Sandy using a least cost path algorithm. We decompose the american highway network into a grid of cells, each cell corresponding to a certain cost, and look for the path between two points that minimizes the cost. A key parameter we have to feed this algorithm with is an “overcost parameter”, which indicates by how much the cost increases in areas affected by the hurricane. This parameter is estimated using an indirect inference method, meaning that we minimize the distance between observed and predicted moments based on the structural gravity model.

The fact that the distance elasticity is not entirely attributable to transport costs still holds true when we exclude dyads for which the bilateral change in transport costs that we compute might have been less accurately determined. It also remains valid if we choose a more restrictive perimeter for the areas affected by Sandy, or if we consider different durations for the disruptions caused by the hurricane. Additionally, we provide evidence that firms did not advance or postpone their shipments because of the hurricane, which would have resulted in a downward bias of our results. However, we leave for future research the precise identification of channels through which these dark trade costs operate.

Chapter 1

The Percolation of Knowledge across Space

This chapter is co-authored with Arthur Guillouzouic (IPP)

Abstract

This paper sheds new light on the negative effect of spatial distance on knowledge flows. We show that it is rooted in the dynamics of the innovation network formation over firms' life-cycles: young and small firms have contacts near them, and progressively expand their network. Using patent citations, we show that knowledge percolates: firms disproportionately cite new patents from prior contacts, and form links with contacts of their contacts. A network formation model that builds on these facts yields two predictions which are met in the data: firm sizes follow a Pareto distribution, and larger firms cite further away. Combining these two facts naturally explains an effect of distance and implies that small firms are its main contributors.

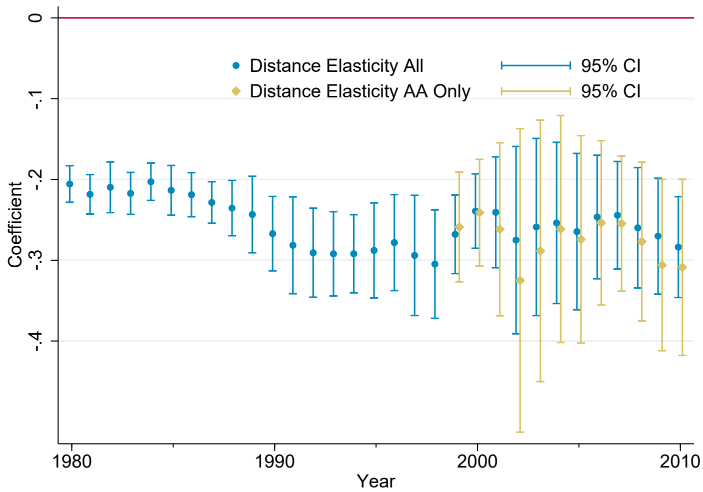
1 Introduction

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 has remained very stable around -0.3, meaning that a 10% increase in the distance between two countries is associated with a 3% decrease in citations between them over the whole period. This is surprising as most conventional distance-related costs, such as transport costs or tariffs, do not apply to ideas. Even more puzzling is the fact that digitization and communication technologies such as online patent search tools seem to have had no effect on knowledge diffusion as a whole. Several papers have argued that space hampers knowledge diffusion because knowledge travels through social links, which are spatially clustered:[1] these papers typically find that controlling for social distance decreases the effect attributed to spatial distance.

This paper shows that the dynamics of network formation over firms' life-cycle are key to under-

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

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



Note: The above figure displays Poisson pseudo ML estimates and confidence intervals obtained on the distance coefficient in structural gravity estimations - equation (1.1) -. Self-citations and intranational citations are excluded, citations from all offices are considered. Estimates plotted in blue use all citations, those plotted in yellow use only citations added by applicants. Section 2 provides more information on the specification and the data used to perform these estimations.

standing the aggregate effect of distance: young and small firms have spatially close contacts, and gradually expand their network as they grow. Our contribution is twofold. In a first part, we identify how links form: we document a phenomenon called "triadic closure" in the economics of networks literature,[2] in which firms disproportionately form links with firms two steps away from them (i.e. with contacts of contacts). To unveil this mechanism, we build all network links through patent citations, and implement a novel identification strategy to show the influence of the network on the probability that a link forms. We provide 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 diffusion process is reminiscent of the physics phenomenon of percolation, approaching knowledge as a fluid making its way from one inventor to another along network paths. In a second part of the paper, we show the aggregate consequences of this network formation process, and in particular how it can explain the effect of distance on knowledge flows. To do so, we incorporate the above diffusion process into 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 and age increase, simply because of the time they have had to expand their network. This model delivers two predictions, relative to the firm size distribution and the relation between firm size and the distance of citations, which are met in the data and naturally lead to an aggregate effect of distance.

Having a precise 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 (Aghion and Jaravel, 2015; Akcigit et al., 2018; Buera and Oberfield, 2020). Since only a small group of high income countries achieves a disproportion-

[2] Jackson and Rogers, 2007.

ate share of technological knowledge production,[3] productivity growth in other countries depends considerably on knowledge flows from those few highly innovative economies which are likely to condition technology adoption.[4] 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.

Our first contribution is to document important facts on the network formation process between innovators. We design a test for diffusion along the network links, relying on the use of examiner-added citations to build a counterfactual for what innovators would cite if they knew every relevant 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 add 60% of all citations. Therefore, the union of applicant and examiner citations allows to construct an almost ideal group of counterfactual citations in a world with frictionless knowledge diffusion. Patents cited by examiners are indeed relevant to the patented invention and are observably similar to applicant citations, but were not known by its applicant (otherwise she would have cited them). We use a snapshot of the network using patent citations made by applicants in a given year, which indicate a set of innovations they knew about, and control extensively for other citations that could have occurred in the past between two applicants. 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.

We estimate that firms are 1.5 times more likely to cite a patent belonging to one of their contacts than if it originated from outside their network. This is however strongly heterogeneous, since firms belonging to the bottom 99% of the size distribution rely twice as much on their existing network as the 1% largest innovators. Moreover, percolation really operates since this effect expands beyond direct links: we also find that the citation of a patent is 35% 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. In particular, we address the facts that citations could be strategic, that applicant and examiner-added references could have systematically different levels of relevance or that citations could be occurring within economic groups. Since the existing literature has put an emphasis on spatial proximity rather than network connection, we build a similar strategy to the above one but where search for new knowledge could be spatial. That is, firms could cite disproportionately often the innovators located around them, as well as the innovators located around their contacts. We find support for the above mechanisms, but show that that the effects are much weaker and less robust.

Our second contribution is to bridge the above findings 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, 2014a), through 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 (knowledge is

[3] 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.

[4] 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.

often assumed to be non-rival and non-excludable, in contrast with traded goods). Abstracting from trade costs, Chaney (2018a) 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 population 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. As we have shown empirically, 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. Since we also find empirical support for a purely spatial search of knowledge, we extend Chaney (2018a)'s model to allow for the possibility of "spatial search",[5] which we model as the possibility for firms to find new partners in the places where they already have a contact.

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 their own right. 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 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 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 dis-

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

tance). 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. (2019) 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 neighbors (*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 3 provides micro evidence of knowledge percolation from actual link formation between contacts, while section 4 builds a theoretical framework linking dynamic network formation with the effect of distance. Finally, section 5 empirically shows that the aggregate theoretical predictions hold on patent citation data.

2 Data and Stylized Fact

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. Surveys have given some empirical support to this assumption: 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.

Yet, patent citations are not a perfect 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 statistical noise and the potential bias induced by the presence of examiner-added citations among the citations.

Applicant and examiner citations are added according to the following procedure. 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.[6] 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 disposal.[7] 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.

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 1.10 in the Appendix). In the population from which we draw our samples (USPTO patents posterior to 2000), each patent has on average 5

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

[7] "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)

applicant-added citations, and 12 examiner-added citations (see Figure 1.12).

An interesting fact is that the sequentiality of the citation procedures (applicant then examiner) does not 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. This strongly suggests that examiner-citations are chosen completely independently of the list of patents selected by the applicant. Note that another very common phenomenon is self-citation, *i.e.* a citation pointing to a previous patent of the assignee applying for the patent. Because these citations are by nature unable to reflect knowledge transfers from outside the firm, we exclude them throughout the paper.[8]

Applicant and examiner citations' characteristics. Before exploiting differences between applicant and examiner-added citations for identification, it is natural to compare their observable characteristics. Figure 1.2 plots the distribution of four important observable characteristics for applicant and examiner added citations. These characteristics are expressed as differences between the citing and the cited patent: geographical distance (panel a), age (panel b), quality measured as citations within a technological class (panel c), and technological distance measured as the Mahalanobis distance between patents' IPC 3-digits technological classes (panel d). More detail on how these variables are constructed is provided in Appendix A.

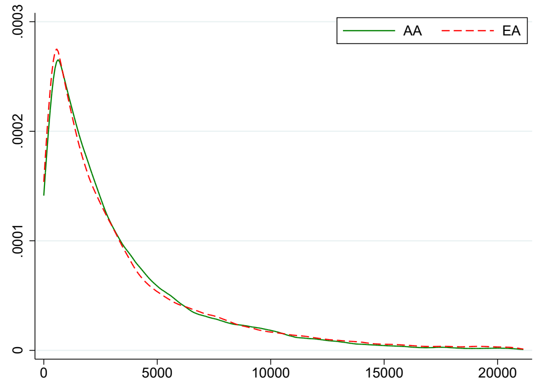
While both types of citations could potentially exhibit very different characteristics, Figure 1.2 shows that they are in fact quite similar. Patents cited by applicants are a bit more likely to originate from places geographically close to them, and are very slightly older. They are also of higher quality, yet technologically further away from their invention than patents chosen by examiners. This suggests that applicants tend to cite more salient references in the field, while examiners really look for very close references even though they may not be as good nor as well-known. For the subsequent analysis which compares both groups of citations, these slight differences are not a concern since we can control for such systematic differences. Section 3.1 discusses the assumptions we make about potential differences in unobservable characteristics between groups.

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 legal rights over the invention. Therefore, inventors are always private individuals, while the 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

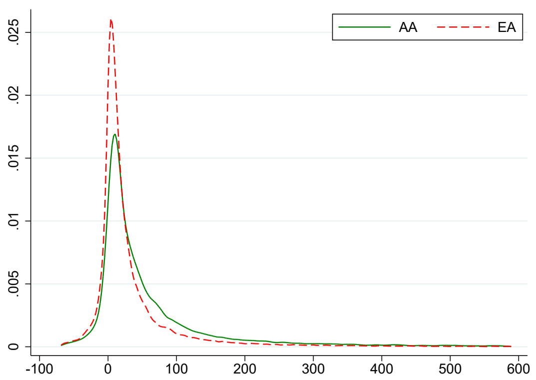
[8]We consider an outward citation to be a "self-citation" as soon as the cited and the citing patent have at least one common applicant or inventor.

Figure 1.2: Distribution of observable characteristics in applicant-added and examiner-added citations

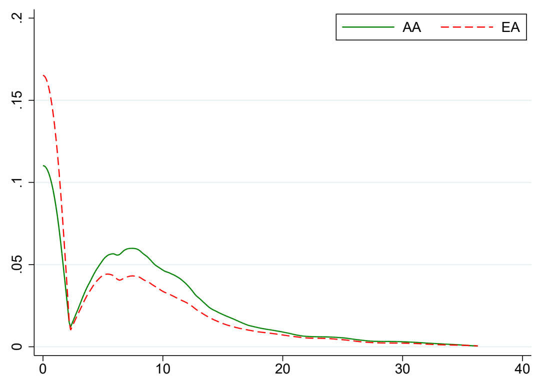
(a) Geographical distance (km)



(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. Solid green lines represent applicant-added citations, dashed red lines represent examiner-added citations.

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.[9] 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 signalled as firms in this harmonization. As a robustness check (shown in subsection 3.3), we also go a step further and conduct our analysis

[9]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.

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 an additional third of the patents, which leaves us with only few patents without country information, as illustrated in Figure 1.11.

Contacts. Our definition of contacts, which section 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 (i.e. cited by the applicant) in a given year which we use for initialization of the network. 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 underlying this restriction 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 (otherwise all the citations made by all their patent applications would have to be constructed). We however provide a sensitivity analysis to changes in this arbitrary threshold.

We then build citation links of distance 2 in the network of any given applicant A, meaning that such patents are two steps away from applicant A, having been cited by an applicant B which is a contact of applicant A. Distance 2 links therefore consist of all the applicant citations made by contacts, and define the contacts of contacts. These links are said to be directed: the fact that A cites B implies a knowledge transfer from B to A, but has no implications for transfers from A to B. Building these distance 2 links is computationally demanding. To alleviate the analysis conducted in section 3 while keeping high statistical power, 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. Our analysis includes evidence that resampling has no effect on measured coefficients.

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

As an introductory exercise, we use patent citations to study the effect of geographical distance on aggregate citation flows, in order to replicate findings by previous papers (Maurseth and Verspagen, 2002; Peri, 2005; Li, 2014). We therefore test for the existence of spatial frictions in the diffusion of knowledge by studying the sensitivity of the flows of outward patent citations (citations made by a patent, in contrast with the ones it might later receive) 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 o to country d (denoted Y_{od}) is the product of an origin specific component, Ω_o, a destination specific component, Δ_d, and a bilateral resistance term, related to the geographical distance between the countries (dist_{od}) and to unobserved factors (η_{od}).

$$Y_{od} = \Delta_d \cdot \Omega_o \cdot \text{dist}_{od}{}^{\zeta} \cdot \eta_{od} \tag{1.1}$$

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.[10] 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 1.16).

The first exercise we conduct is to estimate the elasticity with respect to distance of 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 corresponding estimates using PPML regressions.

The second exercise consists in running a series of yearly cross section estimations. This sheds 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 1.14 in the Appendix).

The negative effect of distance on the intensity of international knowledge flows is a very robust finding. In particular, as shown in Appendix B, 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, a different estimator (OLS or Mixed Pseudo-Maximum Likelihood) 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 1.15 and 1.16 as well as Tables 1.6, 1.7, 1.8 and 1.9 in the Appendix for further explanations and results).

[10]see Mayer and Zignago, 2011, `http://www.cepii.fr/cepii/fr/bdd_modele/presentation.asp?id=6`.

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

	Cit. flow			
	(1)	(2)	(3)	(4)
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: $^a : p < 0.01$; $^b : p < 0.05$; $^c : p < 0.1$

3 Micro Evidence of Networked Knowledge Search

Building on the fact that distance negatively affects knowledge flows in aggregate, we now delve into its determinants through a micro-level analysis of applicants' citation behaviour. This section aims at explaining how the network is formed between inventors. Such analysis requires information on the network of innovators. As described in section 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. 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.

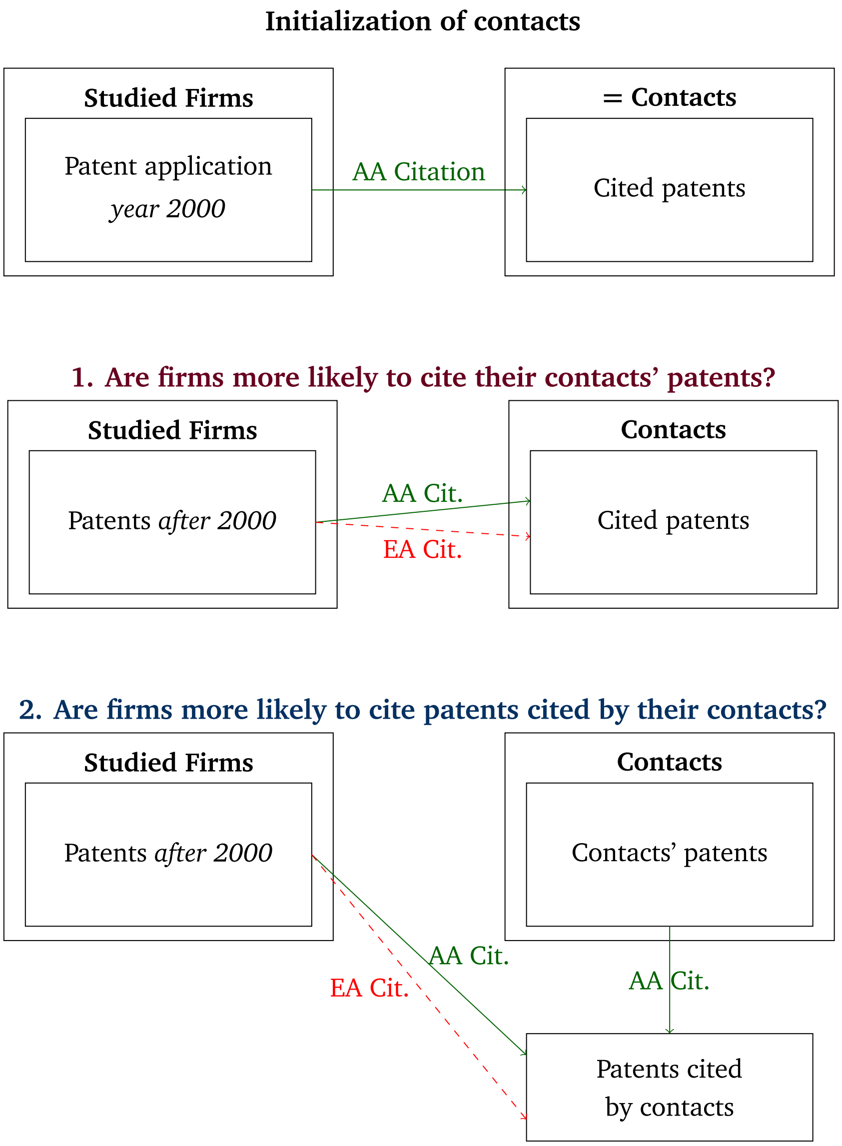
These two tests are depicted graphically in Figure 1.3, as well as through the following example. Consider patent a, which was applied for by firm A, with priority year[11] 2000. This patent cited patent b_1 applied for by applicant B. Our first test assesses whether in its subsequent patent applications, firm A is more likely to cite patent b_2, the other patent of its contact B, than similar control patents. Furthermore, 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 patent c than similar control patents.

3.1 Empirical strategy

Identification. For each patented invention, there are millions of patents that the applicant could potentially cite, which makes it computationally infeasible to consider the complete set of potential

[11]Year of the first patent applicant for an invention.

Figure 1.3: 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.

choices. In other words, the full set of patents which are relevant to an applicant's invention is 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 patents added by the patent office during the examination process constitute a credible set of potential yet non-realized citations.

As shown in section 2.1, while there is no reason to expect that applicant and examiner-added citations should be observably similar, the distribution of four important characteristics (spatial distance, age, quality difference and technological distance) of these citations are quite close and make these sets observably comparable. Moreover, the remaining differences between these two groups can easily be controlled for. 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. In a nutshell, our identification strategy relies on the assumption that the only unobserved characteristic along which patents in the two groups (applicant and examiner-added) differ is whether the applicant was aware of them or not.

Specifically, our first assumption is that all patents cited through either channel are relevant to the patented invention, and that there are no systematic differences of relevance between examiner and applicant citations. The first requirement that this assumption poses if for 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.[12] Section A.3 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.[13]

There are however several ways in which the relevance of the cited patent could actually differ between the examiner and applicant citations. Indeed, one could imagine that, if incentives for citations to be accurate are not high enough on the applicant side, they might be tempted to add irrelevant citations simply to validate their application. At the exact opposite, one could imagine that applicants systematically cite all the major references, while the set of patents cited by the examiner but not by the applicant (our set of control citations) would simply be a complement to these major references with a lower relevance. If these major references were also more likely to have been cited by the applicants in the past, we would only observe the process repeat itself, without any implications on network formation. Nevertheless, because (as mentioned in Section 2.1) citations often overlap, meaning that examiners cite patents that have already been cited by the applicant, we can conduct the very same tests as in the baseline comparing only the overlapping set to the rest of examiner-citations. In this setup, we compare the intersection of applicant and examiner citation sets to its complementary in the set of examiner citations, rather than comparing the set of applicant citations to its complementary set in the union of examiner and applicant citations. The approach taking advantage of the overlap therefore neutralizes potential differences in the relevance of citations made by examiners and applicants.[14]

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) notably showed 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. If such phenomenon is present in our data, it should however bias our estimates downwards. Indeed, a

[12]We match approximately 5 million USPTO applications with examiner information.

[13]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 when they are eventually granted.

[14]An important point to bear in mind is that the sequentiality of the citation procedure does not *per se* threaten our identification. Indeed, we do not strictly compare applicant citations to examiner ones, but really applicant-added citations to patents cited by the examiners but not 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.

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.[15] We however provide a robustness check where we reclassify as applicant citations all the citations made by examiners toward patents or firms which had been cited by the applicant in the past.

Expressing these assumptions in coherence with a discrete choice framework (the canonical McFadden, 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 has a prohibitive cost $\eta \gg \epsilon$, such that applicants always cite the k patents they know out of N. Examiners complement the citations list with the $N-k$ remaining patents (or equivalently with any random subset of m out of the $N-k$ remaining patents both observably and unobservably similar 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., 2019). 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 1.2: 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.

[15] 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.

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 of the cited patents may be known to the citing firm: we create a dummy equal to one 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[16] 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).[17] 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 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. Table 1.10 in the Appendix presents some summary statistics.

3.2 Results

Table 1.2 presents the results for our two tests of network effects, 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 logit coefficients but introduces all the control variables; column 3 is similar but is estimated with conditional logit, which amounts to adding fixed-effects for citing patents to the first column adds a set of control variables. Column 3 is our preferred specification: it accounts 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

[16]The Inpadoc family identifier 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.

[17]Note however that, to construct the network of firms, we use patents from all patent offices to ensure that the network is as comprehensive as it could be.

Table 1.2: Baseline results for network formation tests

Firms	All			Small	Large
	(1)	(2)	(3)	(4)	(5)
Contact	1.47^a	1.41^a	1.48^a	1.65^a	1.32^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
Cited by Contact	1.41^a	1.27^a	1.35^a	1.36^a	1.36^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]
Orig. Patent FE	×	×	✓	✓	✓
Dest. patent Controls	×	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5614	5576	5316	5243	53
Nbr of orig. patents	305.7k	302.1k	260.6k	130.2k	130.3k
Nbr of obs	6.62M	5.37M	5.10M	2.84M	2.26M

Note: Logit and conditional logit (when Orig. Pat. FE is checked) 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. "Orig. Patent FE" refer to conditional logit specifications. "Dest. patent Controls" include the logs of the age of the cited patent, the log of its quality, as well as of the geographical distance and the technological distance to the citing patent. "Persistence Controls" include dummy variables indicating whether a patent has already been cited by the applicant, either through an applicant citation or in general, whether the applicant has already been cited, and whether a patent of the same INPADOC patent family has already been cited, as well as a dummy accounting for whether a patent has been cited by a contact before information availability on AA and EA citations. In columns (4) and (5), the sample is halved according to the size of origin patents' largest applicant, measured as the number of applications in the sample: "Small" refers to patents applied for by applicants below the median size, "Large" refers to patents applied for by applicants above the median size. 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.

contacts' patents shows that belonging to a contact makes applicants around 1,5 times more likely to cite a patent: this 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 the fact of being cited by a contact increases the probability of a citation by around 35%. 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.[18] This property is key if one wants to link the network features to the overall effect of distance on citations.

Columns 4 and 5 estimate the test separately for two parts of the sample: we sort origin patents according to the size of their largest applicant and split the sample at the median. Patents belong-

[18] Interestingly, Carayol et al. (2019) 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.

ing to firms below the median size are tagged as belonging to "small" firms, while above average ones are tagged as "large" firms patents. These estimates show an interesting fact: small firms rely substantially more on existing contacts, since they are 65% more likely to cite their contacts' patents than examiners, compared to 32% for large firms. In other words, small firms rely twice as often on their contacts' patents as large firms. 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 size groups: 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. Therefore, the pace of network formation seems to be uniform over firms' life-cycle, but the breadth of the existing network strongly conditions firms' citations. Sections 4 and 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 contacts' patents is in the same order of magnitude as 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 various controls we introduce for persistence, notably for repeated citations to the patent on which the link has been initialized. It is therefore a very 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 substantial heterogeneity: large firms are much less likely to rely on contacts, but only a few of these firms constitute a very large share of patents and citations, hence driving the coefficient down. Firm-level estimates shown in Table 1.17 in the Appendix confirm that the absolute and the relative magnitude of coefficients varies a lot depending on what is considered to be the relevant unit of observation, and on weights it implies. While the baseline unit of observation is a citing - cited patent dyad, switching to a firm - year unit of observation weights small firms more and increases coefficients, while switching to a citing - cited firm definition seems to weight large firms more and drives coefficients down. 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.

3.3 Robustness

This subsection conducts a variety of robustness checks on our test for network search. These tests can be divided in two categories: the majority of them consists in estimating coefficients on either the same or a very similar sample as in the baseline with the same specification, which makes coefficients comparable to the baseline. However in our alternative strategy and our firm-level estimates, the tests are very close in spirit to the baseline but are they are conducted on samples with a different structure which does not allow to compare the magnitude of the coefficients we obtain. For all the tests delivering coefficients that can be compared with the baseline, Figure 1.4 plots the coefficients and confidence intervals associated to the variables **Contact** and **Cited by Contact** for our preferred specification (column 3 in Table 1.2).

Overlapping citations. As mentioned above, 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 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 1.11 displayed in the Appendix 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 might be biased if applicant citations were less relevant to the patented invention than examiner citations.

Strategic citations The idea that citations could be strategic instruments, as shown by Lampe (2012), is a valid point of concern for our identification strategy and deserves scrutiny. Lampe (2012) spots such citations through the fact that applicants have cited a patent in the past, showing that the applicant knew about it, but do not cite it in a further applicants even though the examiner cites it. In such case, this patent meets both the awareness and the relevance conditions that should perfectly predict a citation, yet it is not cited by the applicant. To handle this, we reclassify all patents meeting this criterion (having been cited by an applicant in the past and being cited by the examiner only later on) as patents cited by the applicant. This is denoted “patent definition” of strategic citations. We also go one step further, and tag as strategic any citation made by the examiner but not by the applicant toward a firm which had been cited in the past (we denote it “firm definition”).

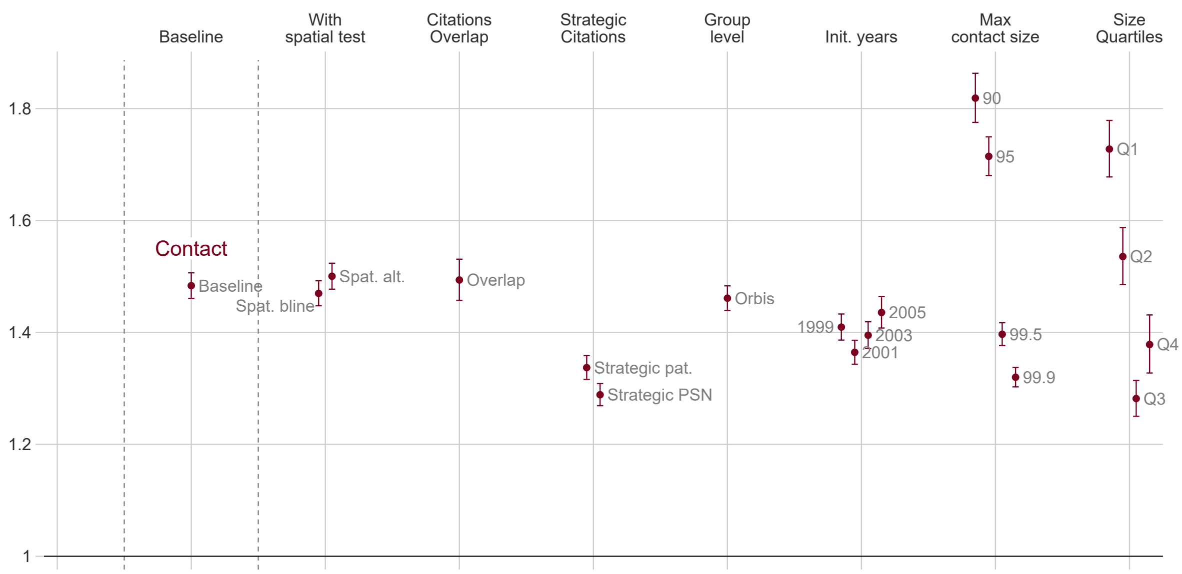
Table 1.12 in the Appendix shows coefficients calculated similarly to the baseline regressions but reclassifying citations suspected to be strategically omitted as applicant-added citations. Note that the fact of having been cited in the past, defined either at the applicant or at the firm level, is part of the set of persistence controls introduced in all the regressions we display, but has to be excluded here for collinearity reasons. This exclusion combined with the reclassification seems to slightly inflate the estimates for the second test (use of contacts of contacts) at the expense of the first one (use of contacts), but largely confirms the findings shown in the baseline regressions.

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, 2020), and that we want our mechanism to be valid beyond the borders of MNEs. Appendix

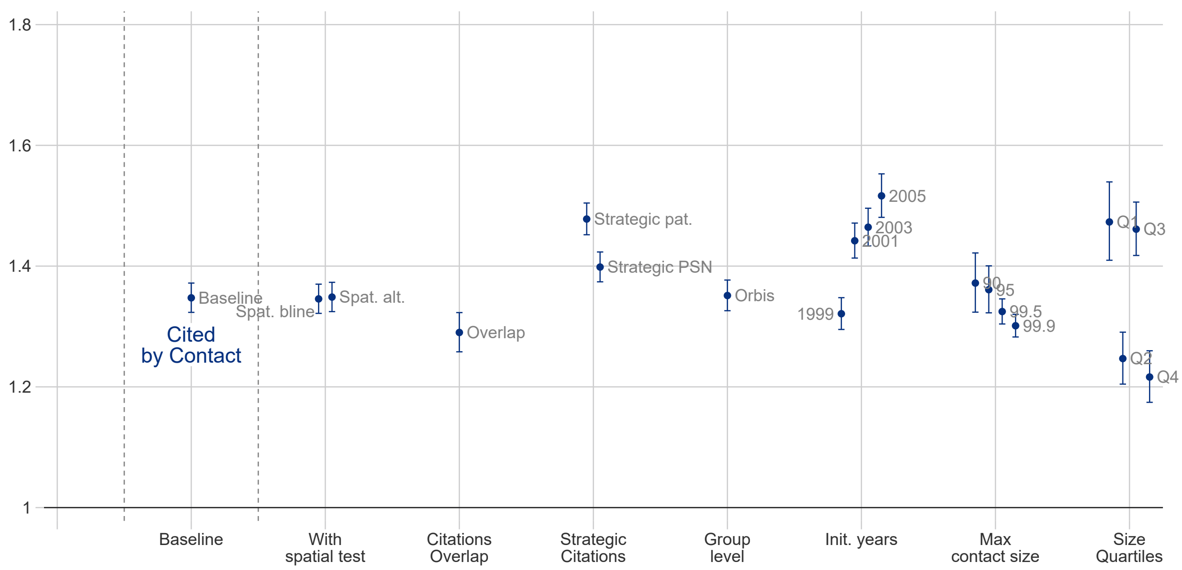
C provides more detail on how we recover information on groups. Results once group linkages are accounted for, shown in Table 1.13 in the Appendix, are very similar to the baseline ones.

Figure 1.4: Coefficients and standard errors of robustness tests

(a) Test 1: Citations toward contacts



(b) Test 2: Citations toward patents cited by contacts



Note: These figures plot exponentiated coefficients (odds-ratios) and 95% confidence intervals of our preferred specification (column 3 in Table 1.2) for various robustness tests. The corresponding tables are displayed in Appendix C.

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 the 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 firm's citing patent, then other patents of the former applicant should be relevant in future citing patents.

As shown in Table 1.14 in the Appendix, although coefficients are not comparable with the baseline ones (mostly because we cannot control for characteristics of the origin patent, which is why we do not include these coefficients in Figure 1.4), 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.

Other robustness checks. We conduct a wider range of robustness checks: we change the initialization year (Table 1.15), the maximum size of contacts (Table 1.16), measure our effects at the firm level (Table 1.17), decompose our effect by size quartile (Table 1.19). We also run Placebo regressions initializing contacts with examiner citations (Table 1.18). Additional explanations and result tables for these tests are provided in Appendix C.

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, originating 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 internet search results from search engines based on their past searches, follow some specialization operating within clusters. This means that a firm may have higher chances to form links with geographical neighbors of its contacts, without its contacts being linked to these geographical neighbors.

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 with a spatial dimension. This test therefore proposes an alternative and more flexible proxy of link formation than the one used above based purely on patent citations.

Table 1.3: Results of the spatial test

Firms	All			Small	Large
	(1)	(2)	(3)	(4)	(5)
Close Firm	1.08[a]	1.03[b]	1.05[a]	1.15[a]	1.00
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
Close to Contact		1.13[a]	1.09[a]	1.11[a]	1.07[a]
		[0.00]	[0.00]	[0.01]	[0.00]
Contact			1.49[a]	1.65[a]	1.34[a]
			[0.01]	[0.02]	[0.01]
Cited by Contact			1.50[a]	1.55[a]	1.48[a]
			[0.01]	[0.02]	[0.02]
Orig. Patent FE	✓	✓	✓	✓	✓
Dest. patent Controls	✓	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5537	5537	5537	5465	53
Nbr of orig. patents	264.8k	264.8k	264.8k	132.7k	132.2k
Nbr of obs	5.30M	5.30M	5.30M	2.96M	2.34M

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 Firm** 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 located less than 5 kilometers from a contact of the citing applicant. "Dest. patent Controls" include the logs of the age of the cited patent, the log of its quality, as well as of the technological distance to the citing patent. "Persistence Controls" include dummy variables indicating whether a patent has already been cited by the applicant, either through an applicant citation or in general, whether the applicant has already been cited, and whether a patent of the same INPADOC patent family has already been cited, as well as a dummy accounting for whether a patent has been cited by a contact before information availability on AA and EA citations. In columns (4) and (5), the sample is halved according to the size of origin patents' largest applicant, measured as the number of applications in the sample: "Small" refers to patents applied for by applicants below the median size, "Large" refers to patents applied for by applicants above the median size. 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.

For each firm in our randomly selected sample, we select the set of patents developed by neighbor firms as patents in which the location of all applicants is less than 5km away.[19] We then define a dummy variable **Close Firm** taking value one if the cited patent was made by a neighbor firm of the citing patent's assignee. We proceed in exactly the same way to define the dummy variable **Close to Contact**, meaning we select patents applied for by applicants located less than 5km away.[20] The variable is equal to one 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 (through the dummy variable **Close Firm**), or geographically close to

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

[20]Note that this step is computationally quite demanding, which explains why we select a distance of 5km, which is lower than what is usually used in this literature.

their contacts (through the dummy variable **Close to Contact**). As the color association indicates, these tests are the respective spatial equivalent of the network tests run through the variables **Contact Firm** and **Cited by Contact**. It shows that firms are indeed between 5 and 10% more likely than examiners to cite patents from assignees located less than 5 km from them depending on the specification. This replicates the results of Alcácer and Gittelman (2006) and Thompson (2006), yet with a slightly lower coefficient, which decreases once we introduce measures for network links. As columns 2 to 4 show, applicants also seem 5% more likely to cite patents from assignees located close to their contacts, and this does not seem to depend as much on origin applicants' size as the first effect. Indeed, the effect making applicants more likely to cite neighboring firms hides substantial heterogeneity in terms of size of origin applicants: while below median size patents are 15% more likely to cite patents from neighbors, this effect is null for above median size patents. Albeit consistent with our findings in the network test, *i.e.* that larger firms are less affected by distance, this may also reflect the fact that the address used for large applicants is likely to be less precise: many applications are made by headquarters or dedicated entities, hence a spatial disconnect between the patent's citations and the applicant's environment. Excluding very large firms, the coefficient of 1.15 we obtain for the increased likelihood of citations towards neighbors indicates a 15% increase in probability, which seems reasonably close to the 25% increase obtained by Thompson (2006), considering the differences in samples and methods used. Surprisingly, the introduction of spatial links does not really affect the estimated coefficients for network links.

Contrary to the baseline network results, the estimates presented above do not include the log of geographical distance between the citing and the cited patent as a control variable. This would indeed make the test considerably more demanding, as it would test for the existence of non log-linearities in distance conditional on network measures in applicants' citation behaviour. Results shown in Table 1.20 in Appendix indeed show that no such effect exists on average. This however has to be contrasted, since large firms seem to drive the effect down, while below median size patents do indeed seem to disproportionately cite close patents relative to examiners. As an additional robustness check, we conduct a strategy close to the alternative strategy for network search: we use examiner citations to construct false contacts, and compare subsequent citations towards firms close to the true contacts relative to the false ones. Table 1.22 in the Appendix shows support for the existence of link formation toward geographical neighbors of contact firms using this strategy.

4 Theory: Network Origins of the Distance Effect

To interpret the aggregate consequences of the micro-level empirical findings of section 3, a model featuring network formation along firms' life-cycle is warranted. This section develops a dynamic model able to bridge our finding that knowledge percolates through a network of innovators from section 3, with the fact that distance hinders aggregate knowledge flows shown in section 2.

To do so, we adapt a model developed in Chaney (2018a) to the context of knowledge and build it around the empirical findings shown in section 3. 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).[21]

Model. We extend the model featured in Chaney (2018a) to introduce some "spatial search" in addition to the "network search", consistently with the empirical evidence provided in section 3.4 that pure spatial forces may be at play. 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 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 joins the set of the firm's contacts. A technical constraint requires that firms can only gain contacts with firms of their cohort. This corresponds to the coefficient associated to the variable **Cited by Contact** in Table 1.2 showing that innovators do form links toward contacts of contacts.
- 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 picks some new contacts with the exact same spatial distribution as the contacts it already has. This corresponds to the coefficient associated to the variable **Close To Contact** in Table 1.3 in our estimations of the spatial test, showing that firms get new contacts with the same spatial distribution as their existing contacts.
- 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:

$$\frac{\partial K_a}{\partial a} = (\rho + \beta - \delta) K_a \tag{1.4}$$

with initial value K_0.

Proposition. *When the distribution of the stock and mass of contacts is described by equations* (1.3) *and* (1.4)*:*

- *The distribution of innovator 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* $\mu = \frac{\beta}{\rho+\beta-\delta}$.

[21] Note that 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.

Proof. See Appendix E. □

These predictions are very intuitive. In a nutshell, this model describes an environment in which firms will gradually be less and less affected by distance as they 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 and that increases in size are generated by random shocks, this model will imply a Pareto size distribution, meaning that small firms are considerably more numerous than large ones. Moreover, because new contacts are further away than old ones, the distance from contacts will be an increasing function of size.

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

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

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. In other words, adding this force to the baseline Chaney (2018a) model predicts a lower value for λ than if network search was the only way to gain new contacts, as well as a reduced relation between firm size and distance of citations.

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

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

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 innovator 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 innovator sizes, while the opposite effect will occur if spatial search is strong enough to compensate contact loss. This compares with the baseline model without spatial search in which an increase in network search has an unambiguous negative effect on μ.

Gravity equation Similarly to trade flows in Chaney (2018a), citation flows will exhibit a negative distance elasticity as long as the following two main conditions hold:

- **Condition 1:** Innovator 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.

These two conditions are exactly the predictions of the above network formation model, therefore connecting directly our micro findings of section 3 with the motivating fact that distance negatively affects aggregate knowledge flows shown in section 2. Under these two sufficient conditions,[22] knowledge flows are negatively related to distance.

Under these two conditions, as stated above, small innovative firms are considerably more numerous than large firms, and there is a systematic relation between a firm's size and the distance of the citations it makes. 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.

5 Estimation of Aggregate Predictions

The network formation model presented in the previous section (section 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.

5.1 Pareto distribution of innovator sizes

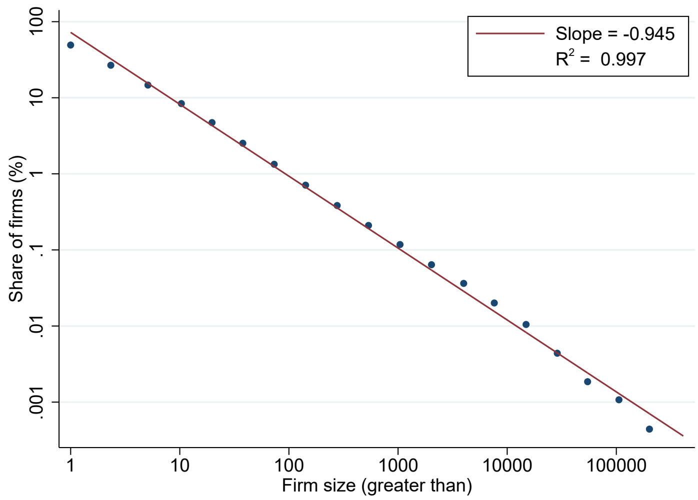
The network formation model predicts that the distribution of innovator sizes will be Pareto, i.e. that $F(K) = 1 - \left(\frac{K}{K_0}\right)^{\lambda}$, with $\lambda = \frac{\gamma}{\rho+\beta-\delta}$. We therefore check that a Pareto distribution fits our data well, and we estimate the shape parameter 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[23] and distribute them in 20 bins of equal log width.[24] We compute the

[22] Additional conditions detailed in Chaney (2018a) 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 $\left(1 + 2\frac{\lambda-1}{\mu}\right)$.

[23] 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 innovator 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).

[24] There seems to be no established consensus regarding the number of bins that should be used: while Axtell (2001) uses 30 bins, Chaney (2018a) 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 1.5: 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 ($1-F(K_b)$). Innovator size is measured as the number of patent applications of the firm over the period 1980-2010.

average size of firms in each bin, denoted K_b, and the fraction of firms of size larger than K_b, denoted $1-F(K_b)$. λ is estimated from:

$$\ln[1-F(K_b)] = a - \lambda \ln(K_b) + \varepsilon_b \tag{1.5}$$

The slope of the regression line shown in Figure 1.5 corresponds to our baseline estimate of λ. The Pareto distribution fits very well our innovator size data (considering the very high R-squared). Table 1.23 in the Appendix confirms this finding and shows results display little sensitivity to the number of bins used or the selected patent office. 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, innovator 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 more innovative firms are also more productive, 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, random growth in size typically generates log-normal distributions (Gibrat, 1931), while it is common to generate power laws from scale-free network formation processes (*e.g.* from the model of Albert and Barabási, 2002), 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 count patent applications in each year, the Pareto distribution still provides an excellent fit. Figure 1.7a 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. As shown in Figure 1.17 in the Appendix, innovator 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. Similarly, within a given technological field (IPC section), the distribution is also Pareto, but the shape parameters exhibit some mild differences across sectors.

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,[25] 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,[26] denoted Δ_b, at which firms in bin b cite. μ is estimated from:

$$\ln[\Delta_b] = a + \mu \ln(K_b) + \varepsilon_b \tag{1.6}$$

Figure 1.6 shows that the relationship between the average squared distance at which a firm 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 innovator size and its ability to access more distant ideas is a novel finding in the analysis of patent citations. Tables 1.24, 1.25 and 1.26 in the Appendix show that this positive relationship still holds with a very good fit, when the sample is restricted to USPTO patents, to applicant citations, and when the number of bins is increased.

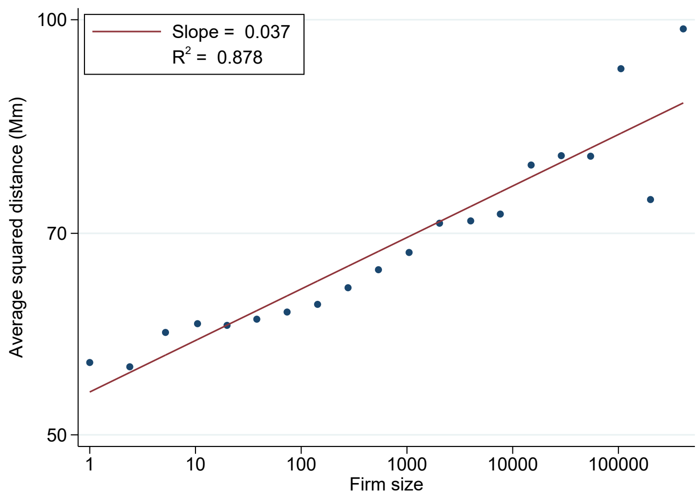
The positive link between innovator size and their ability to access more distant knowledge does not disappear when we disaggregate the sample by patent office (see Figure 1.18). When running separate regressions for each IPC section, μ is always found to be positive, and significant for 7 sections out of 8 (see again Figure 1.18). 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 innovator size of the average squared distance for citations is around half of its counterpart for exports. In other words, the ability to create links with distant firms is less sensitive to size for ideas than for exports.

A complementary exercise which we conduct is to estimate μ within firms over time, meaning that we run a simple two-way fixed effects regression at the firm × year, such that μ is estimated from time variations in innovator size and distance of citations. Using this specification, we also find a positive significant relationship between these two variables, as shown in Table 1.27 in the Appendix. This

[25] Size is again defined as the number of applications of the firm.

[26] 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.

Figure 1.6: 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. All citations over the period 1980-2010 are used. Innovator 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.

holds for samples including all patents or USPTO patents only, and whether we consider all citations (columns 1 and 3) or restrict our attention to applicant-added citations (column 2 and 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 define age as years since the first patent, and regress the log of innovator size on its age, with a set of firm fixed effects. We find a semi-elasticity around 0.03 (see Table 1.28 in the Appendix), meaning each additional year makes an innovator on average 3% larger.

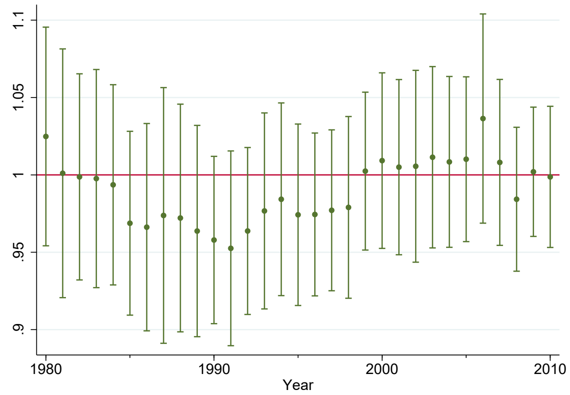
Taken together, these two findings are consistent with the dynamics of the model: as firms grow older, they become larger and are able to build links with more distant firms. Interestingly, the economy described here shares similar features with ones emanating from the Schumpeterian growth theory (Aghion 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.

5.3 Discussion

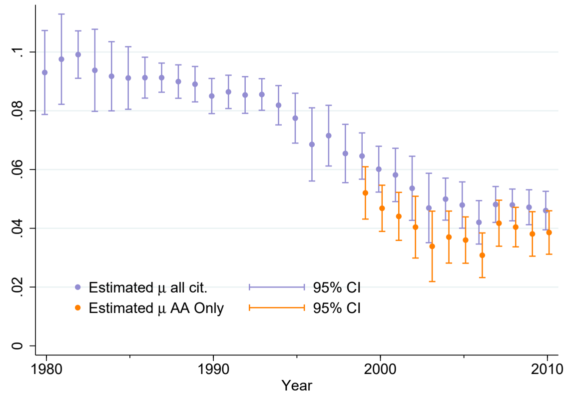
Time variations and their implications The exercise of estimating λ and μ may also be useful in order to understand the changes undergone during our period, in an attempt to explain why these have not implied a decrease in the aggregate distance effect. The context of innovation does not meet the conditions expressed by Chaney (2018a) 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

Figure 1.7: Estimates of λ and μ over time.

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



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



Note: λ and μ are estimated from a series of cross-sectional regressions (respectively of equation (1.5) and (1.6)), one for each year. All patents are included in the sample. Innovator 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. Standard errors are obtained using 100 bootstrap replications.

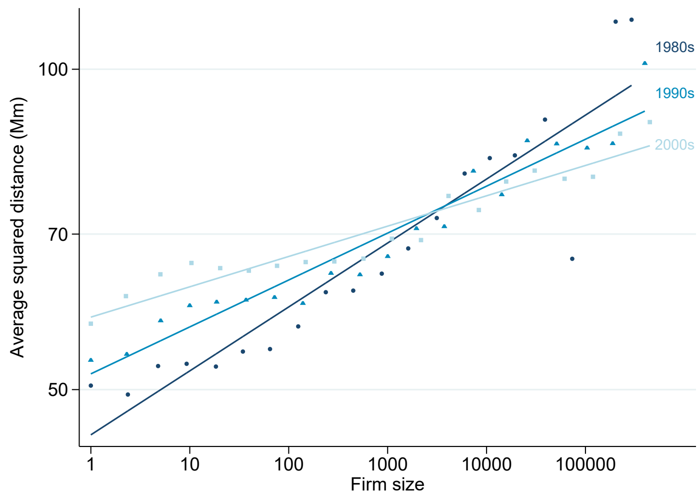
firms do not cite at much larger distances than small firms do, and the elasticity of flows with respect to distance should increase.

Figure 1.7a shows that the shape parameter of the innovator 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 1.7b shows that the strength of the link between innovator 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 1.8 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).

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 1.7a, although standard errors are too large to reject equality of these parameters over time. The explanation for the fact that the gravity coefficients on

Figure 1.8: Estimates of μ by decade



Note: Innovator size is measured as the number of patent applications made by the firm over the decade. Distance is the geographical distance between the largest city of the countries of the citing and the cited patent. For the comparison across decades to hold, all citations are used.

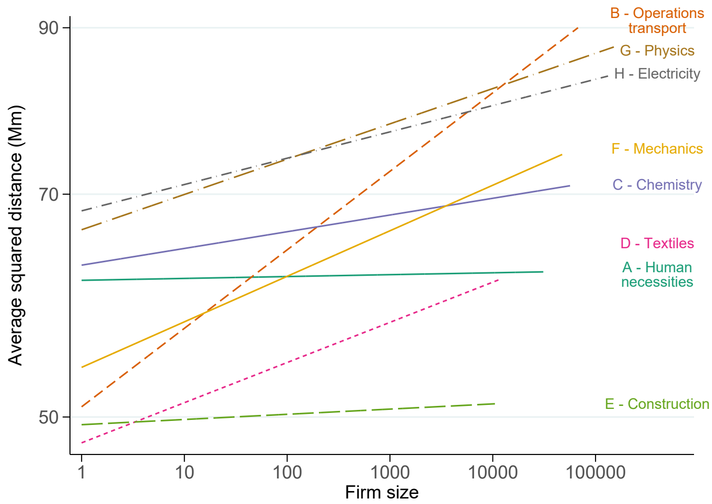
knowledge flows have remained stable over the period would then be hiding two joint changes: the fact that small firms are less affected by distance than they used to, offset by the fact that small firms have grown relatively more numerous.

Predictive power Among other reasons, the changes over time exposed above also have implications on the predictive power we can gain on the value of ζ, the elasticity of flows with respect to distance. Several reasons indeed make the model unfit to convincingly predict ζ.

The first reason originates in the convergence conditions imposed on λ and μ for the model to deliver a prediction on an asymptotically constant elasticity of flows with respect to distance, as shown in Chaney (2018a). Indeed, this constant elasticity equal to $1+2\frac{\lambda-1}{\mu}$ arises only under the conditions that $\lambda > 1$ and $\lambda < 1+\mu$. While the latter condition is often met in our data, the former is not: λ is generally not significantly different from unity, but $\lambda > 1$ can be rejected in many contexts. Additionally, since λ has to be greater than 1, the predicted value of ζ can only be larger than 1, which is both conceptually very large and at odds with what we measure.

A second reason for the poor predictive power is the fact that predicted values of zeta are conditional on a value of k_0, the dispersion of contacts for newborn firms. Yet as stated above, a decrease in μ in the data could stem either from the fact that large firms cite closer than before, or from the fact that small firms cite further away. The model fixes k_0, therefore imposing the first interpretation of a decrease in μ (resulting in an increase in the predicted effect of distance). It is however quite likely that measured variations in μ will also reflect changes in k_0, which should in contrast decrease the effect of distance. Therefore, testing the correlation between the predicted and the measured values of ζ over a dimension of variation (e.g. technological classes) would require homogeneity in k_0 across sectors. This condition is not met in the data, as Figure 1.9 shows: the intercept of μ, which should correspond to k_0 (the squared distance at which the smallest firms cite on average) is very heterogeneous across technologies. For instance, chemistry and construction appear to have very

Figure 1.9: Estimates of μ by technology



Note: Innovator size is measured as the number of patent applications made by the firm over the decade. Distance is the geographical distance between the largest city of the countries of the citing and the cited patent, where only applicant-added citations are considered.

similar values of μ, but have very different intercepts, showing that small innovators in chemistry use knowledge from more distant places than innovators in construction. Similarly, large innovators in operations and transport appear to cite as far as their counterpart in physics and electricity, but the latter technological classes exhibit a much lower μ because the small innovators in these technologies already use knowledge generated very far from them. This is noticeable since these two classes include all innovations relative to semiconductors, electric communication, digital and optical computing, meaning that small innovators in ICTs manage to escape the negative effect of distance more easily than innovators in more traditional technologies. The rise of these technologies may therefore explain the changes observed in the value of μ shown in Figure 1.8. Because of the changes shown in k_0 over time and across sectors, we are however deprived of the two most natural leeways to test the predictive power of the model on the measured values of ζ.

6 Conclusion

This paper shows that the negative effect attributed to distance on international knowledge flows can credibly be explained by the spatial pattern in the dynamics of network formation between innovators.

First, we causally test the influence of existing contact links on the network formation between innovators. Using previous patent citations to build contacts, we show that a firm is more likely to cite either a patent originating from a contact or cited by a contact than a similar patent from outside its close network. For identification, 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 1.5 times more likely than examiners to cite patents owned by their contacts, yet hiding some heterogeneity between small and large firms. Moreover, firms are 35% more likely to cite patents that were cited directly by their contacts. These effects are robust to a wide range of checks.

Based on this finding, we use and extend the network formation model developed by Chaney (2018a) 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.

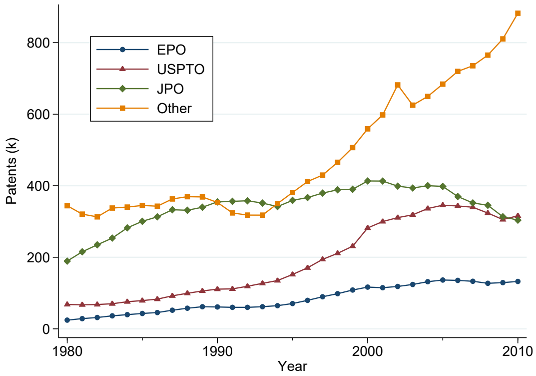
We then show that the theoretical aggregate predictions of the model hold remarkably well in the data. The sizes of innovators - measured as the number of patent applications - are Pareto-distributed (and even Zipf distributed), and the average squared distance at which innovators cite is an increasing 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 innovator 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 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 neighbors (*e.g.* inside clusters), as well as inter-firm mobility of engineers (Almeida and Kogut, 1999; Breschi and Lissoni, 2009; Serafinelli, 2019), and input-output linkages (Carvalho and Voigtländer, 2014).

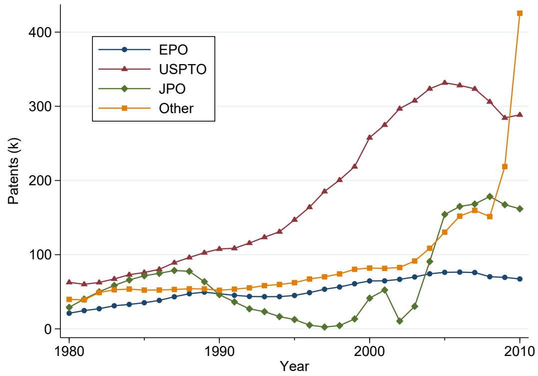
A Technical Appendix

A.1 Description of the Patstat database

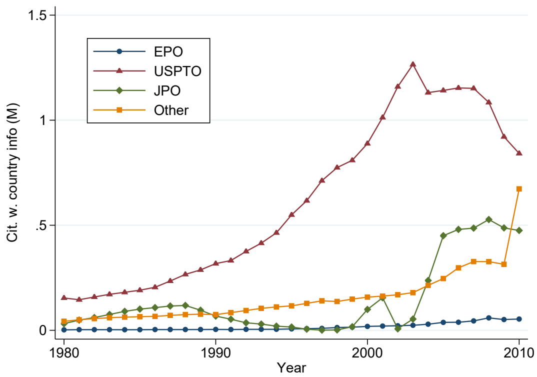
Figure 1.10: Number of patents/citations, decomposed by patent office



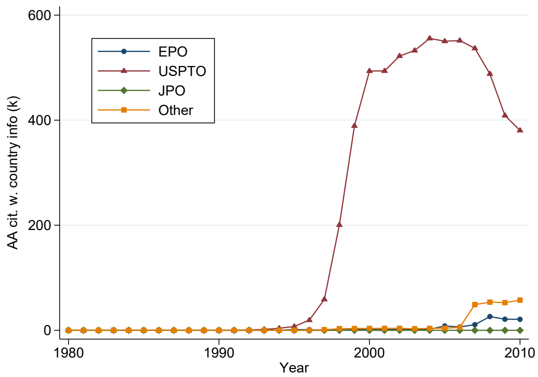
(a) All patents



(b) Patents with citation info

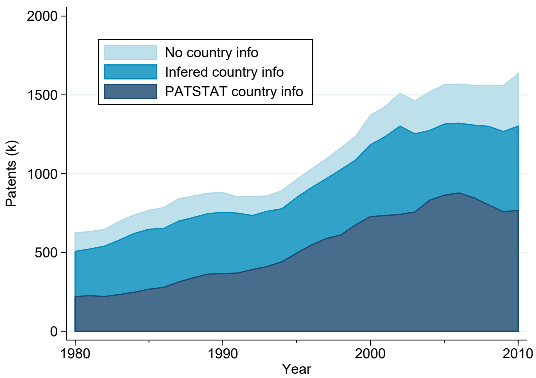


(c) Number of citations

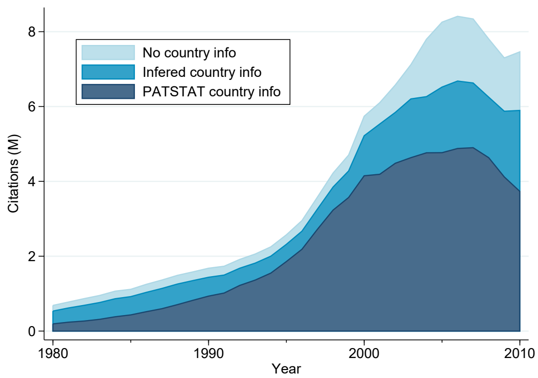


(d) Applicant added citations

Figure 1.11: Patents/citations for which we have geographic information (country)



(a) Patents

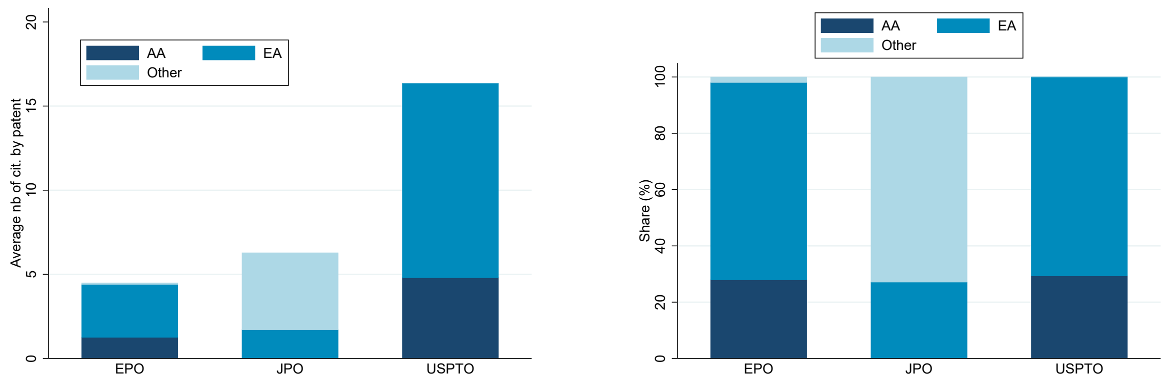


(b) Citations

Table 1.4: 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 1.12: Type of outward citations, by patent office



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

Patents with priority year posterior to 2000

A.2 Description of data handling in Section 3.

- Select randomly a third of all firms having at least a patent in 2000 and a patent after 2000;
- Form contacts of these as all assignees with less than 1000 patent applications in the database cited (AA) in their applications of the year 2000;
- Take all citations (AA and EA) made by these applicants after 2000: keep only citations within the USPTO;
- Check families: drop citations to the same patent from the same patent family;
- Define AA as "APP" in database, EA as "ISR", "SEA" or "PRS" (remove "EXA" which is added during examination so posterior, potential contact with assignee);
- do not consider as "cited by contact" if citation by origin applicant is anterior to the first time the contact cited this patent (i.e. link of the contact did not exist yet)
- drop patents that had been cited before 2000 by the origin applicants, because we don't know if these citations had been made by applicants or by examiners;
- drop if age (time between citing and cited patent application dates) is negative;

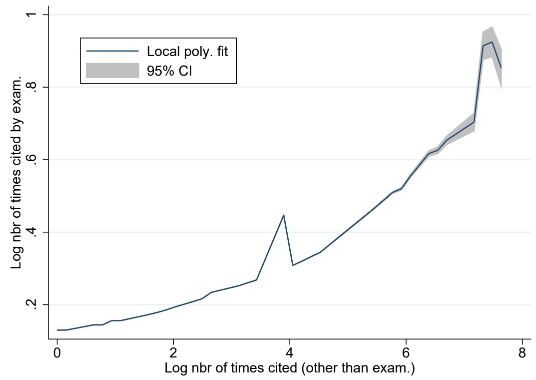
A.3 Description of examiner citations

This section examines the PatEX database from USPTO's Public PAIR data, which records information about the examination process at the USPTO, matched with our sample of USPTO patent applications from Patstat. It reveals the following findings.

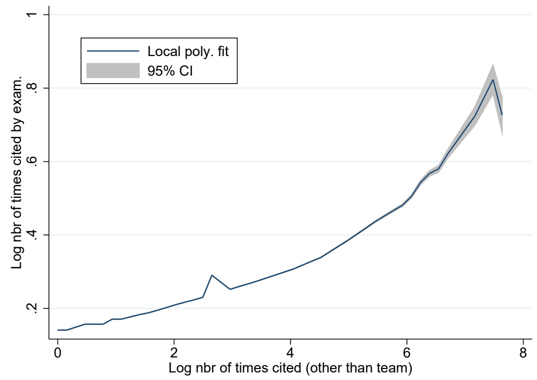
Time spent on a patent application by an examiner is substantial: after dropping very occasional examiners (the ones with less than 5 applications), the average examiner handles 40 patent applications per year, with the 95th percentile being slightly above 100, meaning that even very busy examiners deal 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[27] over their career.

Figure 1.13: 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

There appears to be limited habit formation in examiners' behavior. As Figure 1.13 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 1.5, 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 technological distance of .4% of a standard deviation. This means that, while habit formation in the way examiners cite may exist, it implies very small losses in the accuracy of citations as evidenced by our measure of

[27] 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.

Table 1.5: Technological distance in multiple citations by examiners

	Sd. 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.

A.4 Description of the variables used in Section 2.1.

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.[28] 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.

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.

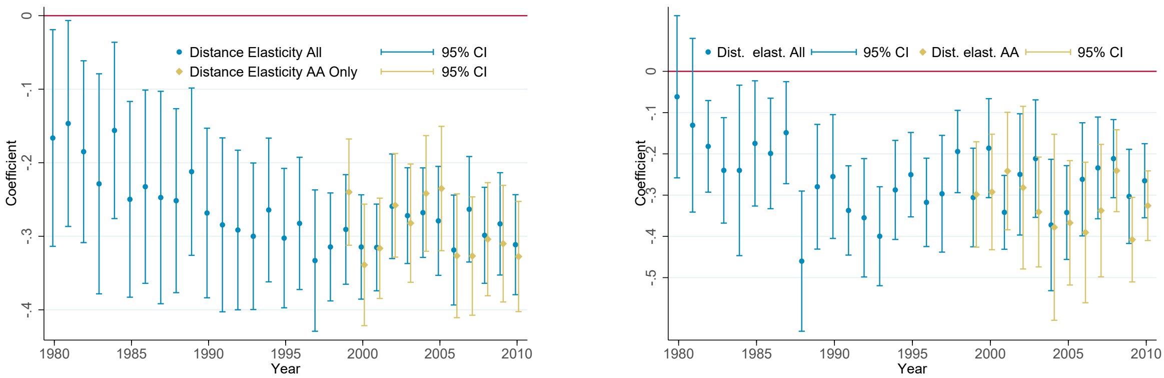
Technological Distance. Additionally to the previous variables, we build a measure of technological distance between the citing and the cited patents based on the IPC classes in which it has been filed. The origin of this approach can be traced back to the seminal paper of Jaffe et al. (1993).

[28]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.

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 work of Bloom et al. (2013).

B Additional Elements on Gravity Estimates

Figure 1.14: Evolution of the distance elasticity of citation flows over time, alternative estimators

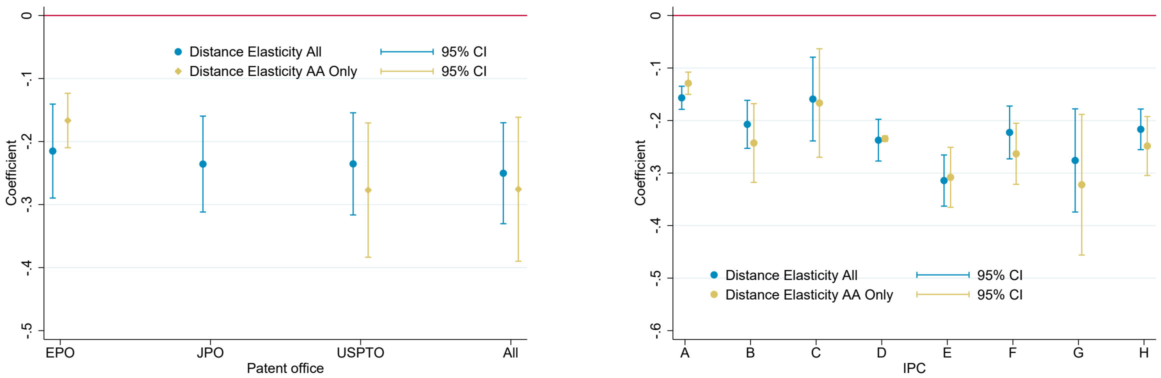


(a) OLS

(b) PPML on flow shares

Notes: OLS (a) and PPML (b) coefficients with associated 95% confidence intervals. These distance elasticities are obtained from a series of cross-sectional gravity estimations. Flow share = flow / total flow to the destination. Standard-errors are obtained through 100 bootstrap replications.

Figure 1.15: Distance elasticity of citation flows, sample split by patent office or by technological sector

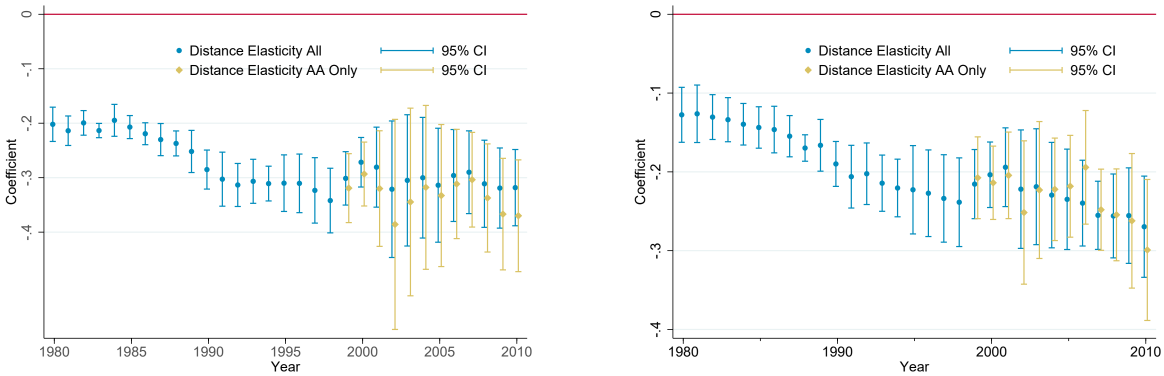


(a) By patent office

(b) By IPC

Notes: PPML coefficients and associated 95% confidence intervals. 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). Standard-errors are obtained through 100 bootstrap replications.

Figure 1.16: Distance elasticity of citation flows over time, alternative geographic measures



(a) Weighted distance

(b) Inventor country

Notes: PPML coefficients and associated 95% confidence intervals. (a) Distance is calculated relative to a barycenter where cities are weighted according to their share of the country population (see Mayer and Zignago, 2011, for more info) (b) Distance is the distance between inventors' countries. Inventors may be a more accurate source of information on the country where the patent was developed when assignees are large multinational firms with worldwide R&D facilities. Standard-errors are obtained through 100 bootstrap replications.

Table 1.6: Distance elasticity of citation flows, pooled sample.

	All patents		USPTO patents	
ζ	-0.255^a	-0.278^a	-0.235^a	-0.277^a
	[0.037]	[0.053]	[0.041]	[0.054]
Citations	All	AA	All	AA
Nb of Dyads	20592	19038	20306	18614

Notes: PPML coefficients and associated 95% confidence intervals. Standard errors are calculated with 100 bootstrap replications. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

Table 1.7: Distance elasticity of citation flows (ζ) in the pooled sample, using different distance measures.

ζ	-0.255^{a}	-0.278^{a}	-0.306^{a}	-0.335^{a}	-0.407^{a}	-0.441^{a}
	[0.037]	[0.053]	[0.039]	[0.057]	[0.018]	[0.021]
Citations	All	AA	All	AA	All	AA
Dist.	Main city		Weighted		Main city	
Intra nat. cit.	No	No	No	No	Yes	Yes
Nb of Dyads	20592	19038	20306	18763	20735	19170

Notes: PPML coefficients and associated 95% confidence intervals. Standard errors are calculated with 100 bootstrap replications. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

Table 1.8: Distance elasticity of citation flows (ζ) in the pooled sample, using different estimators.

	All patents			USPTO patents		
ζ	-0.322^{a}	-0.255^{a}	-0.288^{a}	-0.315^{a}	-0.235^{a}	-0.289^{a}
	[0.034]	[0.037]	[0.030]	[0.034]	[0.041]	[0.030]
Citations	All	All	All	All	All	All
Estimator	OLS	PPML	MPML	OLS	Poisson	MPML
Nb of Dyads	5849	20592	20592	5635	20306	20306

Notes: Coefficients and associated 95% confidence intervals. Standard errors are calculated with 100 bootstrap replications. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

Table 1.9: Distance elasticity of citation flows (ζ) in the pooled sample, using different offices.

	(1) All offices	(2) USPTO	(3) EPO	(4) JPO
ζ	-0.255^{a}	-0.235^{a}	-0.215^{a}	-0.236^{a}
	[0.037]	[0.041]	[0.038]	[0.039]
Citations	All	All	All	All
Nb of Dyads	20592	20306	16186	7396

Notes: Coefficients and associated 95% confidence intervals. Standard errors are calculated with 100 bootstrap replications. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

C Network and spatial searches: Robustness

Table 1.10: Summary statistics for network and spatial search tests

	Applicant citations				Examiner citations			
	Mean	S.d.	1st dec.	9th dec.	Mean	S.d.	1st dec.	9th dec.
Main covariates								
Contact	0.097	0.297	0.000	0.000	0.038	0.192	0.000	0.000
Cited by Contact	0.151	0.358	0.000	1.000	0.053	0.224	0.000	0.000
Cited by Contact before 2000	0.161	0.368	0.000	1.000	0.061	0.240	0.000	0.000
Close Firm	0.015	0.122	0.000	0.000	0.017	0.130	0.000	0.000
Close to Contact	0.298	0.458	0.000	1.000	0.252	0.434	0.000	1.000
Persistence controls								
Firm already cited by applicant	0.189	0.392	0.000	1.000	0.140	0.347	0.000	1.000
Firm already cited	0.546	0.498	0.000	1.000	0.486	0.500	0.000	1.000
Patent already cited by applicant	0.253	0.435	0.000	1.000	0.081	0.273	0.000	0.000
Patent already cited before 2000	0.160	0.366	0.000	1.000	0.072	0.259	0.000	0.000
Patent family already cited	0.382	0.486	0.000	1.000	0.153	0.360	0.000	1.000
Dest. patent controls								
Ln(Age)	8.056	0.947	6.899	9.186	7.667	1.105	6.332	8.976
Ln(Quality)	3.870	1.464	2.030	5.611	3.336	1.500	1.450	5.114
Ln(TechnoDist)	1.158	1.276	0.000	2.825	1.117	1.269	0.000	2.803
Ln(GeoDist)	7.388	2.162	4.940	9.246	7.475	2.283	4.755	9.273
Observations	4,778,259				5,117,996			

C.1 Network search robustness

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

Table 1.11: Network search using citations overlapping applicant and examiner added citations

Firms	**All**			**Small**	**Large**
	(1)	(2)	(3)	(4)	(5)
Contact	1.28^a	1.21^a	1.49^a	1.76^a	1.25^a
	[0.01]	[0.01]	[0.02]	[0.03]	[0.02]
Cited by Contact	1.35^a	1.18^a	1.29^a	1.37^a	1.25^a
	[0.02]	[0.01]	[0.02]	[0.03]	[0.02]
Orig. Patent FE	×	×	✓	✓	✓
Dest. patent Controls	×	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5605	5513	3609	3542	52
Nbr of orig. patents	305.2k	290.2k	111.9k	56.6k	55.3k
Nbr of obs	3.96M	3.11M	2.03M	1.14M	0.89M

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 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.

Strategic citations. A motivation and description of this test are given in the body of the paper.

Table 1.12: Network search tests reclassifying potentially strategic citations.

(a) Patent definition

Firms	**All**			**Small**	**Large**
	(1)	(2)	(3)	(4)	(5)
Contact	1.34^a	1.24^a	1.34^a	1.45^a	1.22^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
Cited by Contact	2.07^a	1.64^a	1.48^a	1.49^a	1.48^a
	[0.02]	[0.01]	[0.01]	[0.02]	[0.02]
Orig. Patent FE	×	×	✓	✓	✓
Dest. patent Controls	×	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5813	5774	5499	5427	53
Nbr of orig. patents	306.4k	302.8k	255.7k	127.8k	127.9k
Nbr of observations	6.64M	5.39M	4.95M	2.76M	2.19M

(b) Firm definition

Firms	**All**			**Small**	**Large**
	(1)	(2)	(3)	(4)	(5)
Contact	1.20^a	1.14^a	1.29^a	1.37^a	1.20^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
Cited by Contact	1.75^a	1.45^a	1.40^a	1.39^a	1.43^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]
Orig. Patent FE	×	×	✓	✓	✓
Dest. patent Controls	×	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5813	5774	5495	5424	53
Nbr of orig. patents	306.4k	302.8k	254.4k	126.6k	127.8k
Nbr of observations	6.64M	5.39M	4.99M	2.77M	2.22M

Note: Logit and conditional logit estimations of the determinants of knowledge transfers (applicant-added citations), reclassifying potentially strategic citations using the patent-level definition, 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.

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.[29] The matched firms are a priori the largest 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 1.13 shows that consolidating assignees and contacts at the group level does not harm our results.

Table 1.13: Network search tests spotting groups with Orbis

Firms	**All**			**Small**	**Large**
	(1)	(2)	(3)	(4)	(5)
Contact	1.41^a	1.37^a	1.46^a	1.64^a	1.30^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
Cited by Contact	1.39^a	1.22^a	1.35^a	1.31^a	1.38^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]
Orig. Patent FE	×	×	✓	✓	✓
Dest. patent Controls	×	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5596	5560	5296	5224	53
Nbr of orig. patents	306.4k	302.8k	261.1k	130.8k	130.3k
Nbr of obs	6.63M	5.38M	5.11M	2.85M	2.26M

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.

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

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

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 both 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 1.14: Results, Alternative versions of tests

Alternative Test 1

	(1)	(2)
	Contact Pat.	
Cited by applicant	1.59^{a}	1.33^{a}
	[0.06]	[0.05]
Cited in 2000	0.63^{a}	0.44^{a}
	[0.01]	[0.01]
Dest. Quality (log)		1.54^{a}
		[0.01]
IPC 1d FE	×	✓
Year FE	×	✓
Nbr of obs	1.90M	1.90M

Note: Logit estimations of the determinants of knowledge transfers (applicant-added citations). The sample is the set of patent applications from both true contacts (firms cited by applicants in 2000) and false contacts (firms cited by the contacts' examiners) of studied applicants after 2000, recorded on USPTO patent applications only. The dependent variable is a dummy equal to 1 when patents belong to a true contact, 0 when they belong to a false contact . **Cited by applicant** is a dummy equal to 1 when patent *d* is cited after 2000. "Cited in 2000" is a dummy equal to 1 when patent *d* was cited in 2000. "IPC 1d FE" are dummy variables for 1 digit IPC patent classes, "Year FE" are year dummy variables. Results are exponentiated coefficients (odds ratios). Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

Alternative Test 2

	(1)	(2)
	Cited by contact	
Cited by applicant	1.89^{a}	1.65^{a}
	[0.05]	[0.05]
Contact Pat.	1.70^{a}	1.39^{a}
	[0.03]	[0.02]
Dest. Quality (log)		1.46^{a}
		[0.01]
IPC 1d FE	×	✓
Year FE	×	✓
Nbr of obs	8.33M	8.33M

Note: Logit estimations of the determinants of knowledge transfers (applicant-added citations). The sample is the set of patent applications truly cited (*i.e.* cited by applicants) and falsely cited (cited by examiners) of studied applicants' contacts after 2000, recorded on USPTO patent applications only. The dependent variable is a dummy equal to 1 when patents are true contact citations, 0 otherwise. **Cited by applicant** is a dummy equal to 1 when patent *d* is cited after 2000. "Contact Pat." controls for the fact that cited patents might also belong to actual contacts. "IPC 1d FE" are dummy variables for 1 digit IPC patent classes, "Year FE" are year dummy variables. Results are exponentiated coefficients (odds ratios). Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

Table 1.14 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 1.14, 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 1.14.

Initialization year. Table 1.15 provides results similar to the baseline, changing the initialization year from 2000 to 1999, 2001, 2003 and 2005. Results hold in all these 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).

Table 1.15: Network search tests changing the initialization year

Initialization year	**1999**	**2000**	**2001**	**2003**	**2005**
	(1)	(2)	(3)	(4)	(5)
Contact	1.41^a	1.48^a	1.36^a	1.39^a	1.44^a
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Cited by Contact	1.32^a	1.35^a	1.44^a	1.46^a	1.52^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]
Orig. Patent FE	✓	✓	✓	✓	✓
Dest. patent Controls	✓	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	6143	5316	4749	3865	3149
Nbr of orig. patents	286.7k	260.6k	236.6k	187.3k	138.9k
Nbr of obs	5.66M	5.10M	4.61M	3.53M	2.45M

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.

Maximum size of contacts. Table 1.16 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 1.16: Network search tests changing the maximum size of contacts

Max contacts' size percentile	**90**	**95**	**99**	**99.5**	**99.9**
	(1)	(2)	(3)	(4)	(5)
Contact	1.82^a	1.71^a	1.48^a	1.40^a	1.32^a
	[0.02]	[0.02]	[0.01]	[0.01]	[0.01]
Cited by Contact	1.37^a	1.36^a	1.35^a	1.32^a	1.30^a
	[0.02]	[0.02]	[0.01]	[0.01]	[0.01]
Orig. Patent FE	✓	✓	✓	✓	✓
Dest. patent Controls	✓	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5316	5316	5316	5316	5316
Nbr of orig. patents	260.6k	260.6k	260.6k	260.6k	260.6k
Nbr of obs	5.10M	5.10M	5.10M	5.10M	5.10M

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.

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 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, in Table 1.17.

Table 1.17: Network search measured at the firm level

Level	Firm-Year		Firm to firm	
	(1)	(2)	(3)	(4)
Share of contact patents	2.34^a	2.26^a	1.10^a	1.07^a
	[0.19]	[0.18]	[0.00]	[0.00]
Share of patents cited by contact	1.32^a	1.16	1.08^a	1.04^a
	[0.12]	[0.11]	[0.00]	[0.00]
Orig. firm Controls	×	×	✓	✓
Year FE	✓	✓	×	×
Dest. patent Controls	×	✓	×	✓
Persistence Controls	✓	✓	✓	✓
Nbr of observations	24.61k	24.53k	971.85k	933.18k

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.

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

Table 1.18: Results of Placebo Tests

Firms	All			Small	Large
	(1)	(2)	(3)	(4)	(5)
Fake Contact	1.00	0.99	1.00	1.00	1.00
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Falsely Cited by Contact	0.82^a	0.76^a	0.72^a	0.78^a	0.69^a
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Orig. Patent FE	×	×	✓	✓	✓
Dest. patent Controls	×	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5614	5576	5316	5243	53
Nbr of orig. patents	305.7k	302.1k	260.6k	130.2k	130.3k
Nbr of obs	6.62M	5.37M	5.10M	2.84M	2.26M

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.

Split sample by quartiles. This table presents estimation results of our preferred specification for the network test on the sample split by quartiles, similarly to split sample estimations around the median presented in columns (4) and (5) in table 1.2.

Table 1.19: Baseline network test decomposed by size quartiles

Firm size	**≤ Q1**	**> Q1 ≤ median**	**> median ≤ Q3**	**> Q3**
	(1)	(2)	(3)	(4)
Contact	1.73^a	1.53^a	1.28^a	1.38^a
	[0.03]	[0.03]	[0.02]	[0.03]
Cited by Contact	1.47^a	1.25^a	1.46^a	1.22^a
	[0.03]	[0.02]	[0.02]	[0.02]
Orig. Patent FE	✓	✓	✓	✓
Dest. patent Controls	×	✓	✓	✓
Persistence Controls	✓	✓	✓	✓
Nbr of orig. firms	5018	202	43	10
Nbr of orig. patents	67.1k	63.1k	68.0k	62.3k
Nbr of observations	1.59M	1.25M	1.30M	0.96M

Note: 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 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.

C.2 Spatial search robustness

Introducing the log of distance as a control variable. This version of the test adds the log of distance to the destination patent controls, as is the case for the baseline network search test.

Table 1.20: Results of the spatial test controlling for the log of geographical distance

Firms	All			Small	Large
	(1)	(2)	(3)	(4)	(5)
Close Firm	0.98^c	0.94^a	0.97^b	1.05^b	0.93^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
Close to Contact		1.12^a	1.08^a	1.11^a	1.06^a
		[0.00]	[0.00]	[0.01]	[0.00]
Contact			1.48^a	1.63^a	1.33^a
			[0.01]	[0.02]	[0.01]
Cited by Contact			1.49^a	1.54^a	1.47^a
			[0.01]	[0.02]	[0.02]
Orig. Patent FE	✓	✓	✓	✓	✓
Dest. patent Controls	✓	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5502	5502	5502	5430	53
Nbr of orig. patents	261.2k	261.2k	261.2k	130.8k	130.3k
Nbr of obs	5.12M	5.12M	5.12M	2.86M	2.26M

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 Firm** 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 located less than 5 kilometers from a contact of the citing applicant. "Dest. patent Controls" include the logs of the age of the cited patent, the log of its quality, as well as of the **geographical distance** and of the technological distance to the citing patent. "Persistence Controls" include dummy variables indicating whether a patent has already been cited by the applicant, either through an applicant citation or in general, whether the applicant has already been cited, and whether a patent of the same INPADOC patent family has already been cited, as well as a dummy accounting for whether a patent has been cited by a contact before information availability on AA and EA citations. In columns (4) and (5), the sample is halved according to the size of origin patents' largest applicant, measured as the number of applications in the sample: "Small" refers to patents applied for by applicants below the median size, "Large" refers to patents applied for by applicants above the median size. 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.

Changing the definition of spatial proximity for a patent. This version of the test switches from the requirement for a patent to be considered geographically close to an applicant that all of the patent applicants are close, to the requirement that at least one of these applicants is close.

Table 1.21: Results of the spatial test with alternative definition

Firms	All			Small	Large
	(1)	(2)	(3)	(4)	(5)
Close Firm	1.12^a	1.11^a	1.11^a	1.13^a	1.10^a
	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]
Close to Contact		1.03^a	0.98^a	0.96^a	0.98^a
		[0.00]	[0.00]	[0.00]	[0.01]
Contact			1.52^a	1.72^a	1.35^a
			[0.01]	[0.02]	[0.01]
Cited by Contact			1.51^a	1.55^a	1.48^a
			[0.01]	[0.02]	[0.02]
Orig. Patent FE	✓	✓	✓	✓	✓
Dest. patent Controls	✓	✓	✓	✓	✓
Persistence Controls	✓	✓	✓	✓	✓
Nbr of orig. firms	5537	5537	5537	5465	53
Nbr of orig. patents	264.8k	264.8k	264.8k	132.7k	132.2k
Nbr of obs	5.30M	5.30M	5.30M	2.96M	2.34M

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 Firm** 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 located less than 5 kilometers from a contact of the citing applicant. "Dest. patent Controls" include the logs of the age of the cited patent, the log of its quality, as well as of the technological distance to the citing patent. "Persistence Controls" include dummy variables indicating whether a patent has already been cited by the applicant, either through an applicant citation or in general, whether the applicant has already been cited, and whether a patent of the same INPADOC patent family has already been cited, as well as a dummy accounting for whether a patent has been cited by a contact before information availability on AA and EA citations. In columns (4) and (5), the sample is halved according to the size of origin patents' largest applicant, measured as the number of applications in the sample: "Small" refers to patents applied for by applicants below the median size, "Large" refers to patents applied for by applicants above the median size. 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.

Alternative strategy for the spatial search test This version of the spatial search test mimics the alternative strategy pursued for the network search test: it compares the frequency of citations after 2000 toward patents developed by applicants located close to true contacts, to those toward patents developed by applicants located close to false contacts, defined as applicants cited by examiners in 2000. The sample is therefore a set of origin applicant - cited patent dyads, with the dependent variable being a dummy variable taking value 1 when the patent is located close to a true contact, and the covariate of interest being a dummy variable taking value 1 when the patents gets cited by the origin applicant after 2000.

Table 1.22: Results of the alternative strategy for the spatial test

	(1)	(2)
	Close to contact	
Cited by applicant	1.36^{a}	1.37^{a}
	[0.08]	[0.08]
CBC	2.06^{a}	2.14^{a}
	[0.14]	[0.15]
Contact Pat.	1.40^{a}	1.41^{a}
	[0.07]	[0.06]
Dest. Quality (log)		0.92^{a}
		[0.02]
IPC 1d FE	×	✓
Year FE	×	✓
Nbr of observations	1.42M	1.42M

Note: Logit estimations of the determinants of knowledge transfers (applicant-added citations). The sample is the set of patent applications truly cited (*i.e.* cited by applicants) and falsely cited (cited by examiners) of studied applicants' contacts after 2000, recorded on USPTO patent applications only. The dependent variable is a dummy equal to 1 when patents are true contact citations, 0 otherwise. **Cited by applicant** is a dummy equal to 1 when patent d is cited after 2000. "Contact Pat." controls for the fact that cited patents might also belong to actual contacts. "IPC 1d FE" are dummy variables for 1 digit IPC patent classes, "Year FE" are year dummy variables. Results are exponentiated coefficients (odds ratios). Significance levels: ^{a} p<0.01 ^{b} p<0.05 ^{c} p<0.1.

D Additional aggregate results

D.1 Additional tables

Table 1.23: Estimates of the shape parameter of the Pareto distribution of innovator size (λ), changing the number of bins.

	All patents			USPTO patents		
λ	-0.945^{a}	-0.959^{a}	-0.969^{a}	-0.991^{a}	-1.012^{a}	-1.002^{a}
	[0.015]	[0.016]	[0.014]	[0.026]	[0.019]	[0.019]
Nbr. of bins	20	50	100	20	50	100
Size measure	Pat. app					

Note: Innovator size is measured as the number of patent applications of the firm over the period 1980-2010. Standard errors are calculated with 100 bootstrap replications. Significance levels: $^{a}: p < 0.01$; $^{b}: p < 0.05$; $^{c}: p < 0.1$

Table 1.24: Estimates of the elasticity of the average squared distance of citations with respect to innovator size (μ).

	All patents		USPTO patents	
μ	0.037^{a}	0.029^{a}	0.024^{a}	0.025^{a}
	[0.002]	[0.003]	[0.002]	[0.003]
Nbr. of bins	20	20	20	20
Citations	All	AA	All	AA
R^2	0.878	0.785	0.626	0.675

Notes: Innovator 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. Standard errors are calculated with 100 bootstrap replications. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

Table 1.25: Estimates of the elasticity of the average squared distance of citations with respect to innovator size (μ) using all citations, changing the number of bins.

	All patents			USPTO patents		
μ	0.037^{a}	0.039^{a}	0.039^{a}	0.024^{a}	0.028^{a}	0.029^{a}
	[0.003]	[0.003]	[0.003]	[0.002]	[0.003]	[0.003]
Citations	All	All	All	All	All	All
Nbr. of bins	20	50	100	20	50	100
R^2	0.878	0.867	0.704	0.626	0.603	0.571

Notes: Innovator 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. Standard errors are calculated with 100 bootstrap replications. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

Table 1.26: Estimates of the elasticity of the average squared distance of citations with respect to innovator size (μ) using applicant-added citations, changing the number of bins.

	All patents			USPTO patents		
μ	0.029^a	0.031^a	0.032^a	0.025^a	0.028^a	0.027^a
	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Nbr. of bins	20	50	100	20	50	100
Citations	All	All	All	All	All	All
R^2	0.785	0.780	0.603	0.675	0.603	0.550

Notes: Innovator 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. Standard errors are calculated with 100 bootstrap replications. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

Table 1.27: Estimates of the elasticity of the average squared distance of citations with respect to innovator size (μ), using within firm variations.

	All patents		USPTO patents	
μ	0.049^a	0.031^a	0.035^a	0.035^a
	[0.001]	[0.002]	[0.002]	[0.002]
Firm FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Citations	All	AA	All	AA
Nbr obs.	689781	210931	192911	192911
Nbr firms	152805	50175	46251	46251

Notes: Innovator 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 are clustered by at the firm level. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

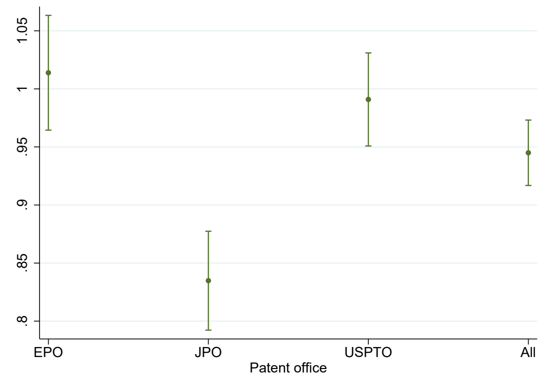
Table 1.28: Semi-elasticity of innovator size with respect to age.

	All patents		USPTO patents	
Age	0.088^a	0.028^a	0.066^a	0.029^a
	[0.001]	[0.000]	[0.001]	[0.001]
Firm FE	×	✓	×	✓
Nbr firms	471554	177034	316388	118078
Nbr obs.	1.132e+06	837709	767152	568842

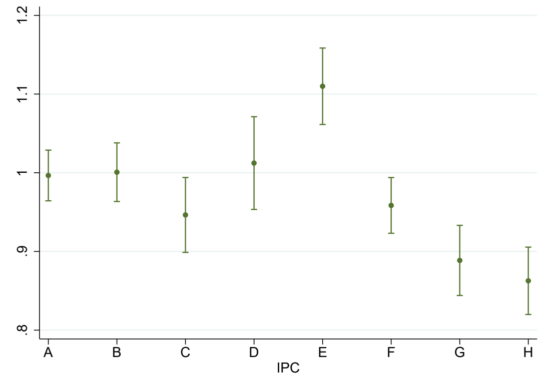
Notes: Innovator size is measured as the number of patent applications of the firm over the period 1980-2010. Distance is measured as the geodesic distance between the most populated city of each country. Age is measured as time since the first patent application. Standard errors are clustered by at the firm level. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

D.2 Additional figures

Figure 1.17: Shape parameter of the Pareto distribution of innovator size (λ), sample split by patent office or by technological sector



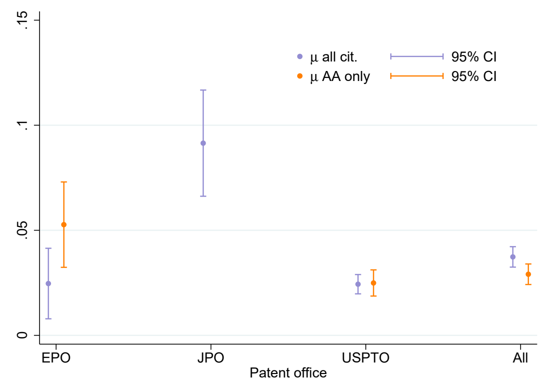
(a) By patent office



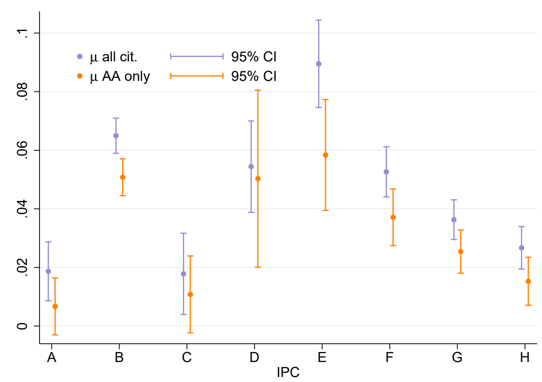
(b) By IPC Section (1 digit)

Notes: Innovator 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 (1.5) is estimated on each sub-sample. Standard-errors are obtained through 100 bootstrap replications.

Figure 1.18: Elasticity of the average squared distance of citations with respect to innovator size (μ), sample split by patent office or by technological sector



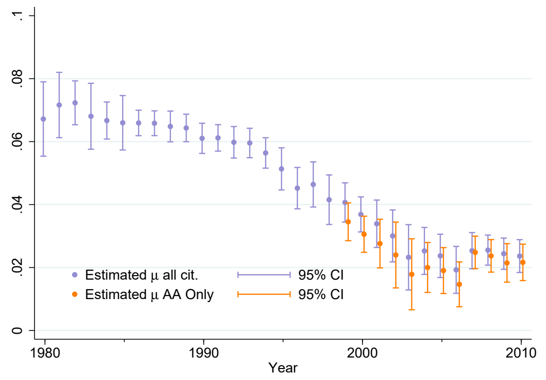
(a) By patent office



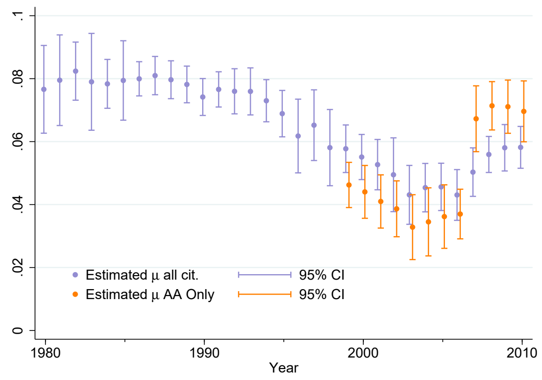
(b) By IPC Section (1 digit)

Notes: Innovator 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 (1.5) is estimated on each sub-sample. Standard-errors are obtained through 100 bootstrap replications.

Figure 1.19: Elasticity of the average squared distance of citations with respect to innovator size (μ), estimated with alternative distance measures



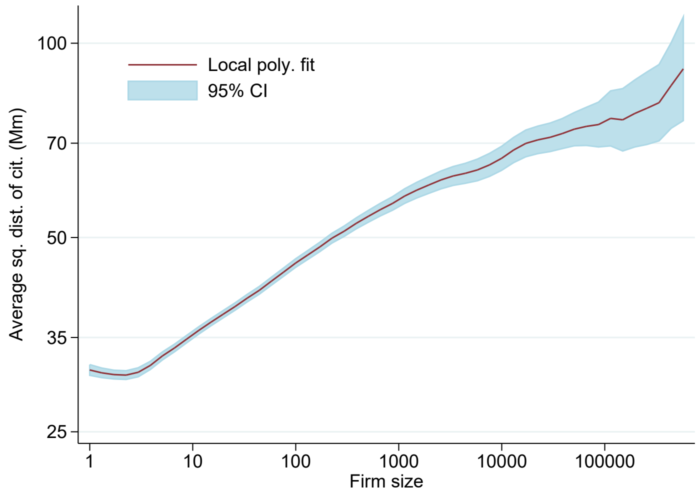
(a) Weighted distance



(b) Inventor's country

Notes: (a) Distance is calculated relative to a barycenter where cities are weighted according to their share of the country population (see Mayer and Zignago, 2011, for more info) (b) Distance is the distance between inventors' countries. Innovator size is measured as the number of patent applications of the firm over the period 1980-2010. Standard-errors are obtained through 100 bootstrap replications.

Figure 1.20: Average squared distance of citations as a function of innovator size



Note: Polynomial fit between firm size and average squared distance of citations, built on all applicants.

E Theory Appendix

Proof of Proposition 1

A solution for the ODE given in (E) 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:

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

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, 2018a).

One can then derive the endogenized conditions allowing to get a constant elasticity of flows with respect to distance. The distribution of innovator 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 innovator 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 (2018a), $\Delta_a = \Delta_0 e^{\beta a}$. 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

Information in the First Globalization: News Agencies and Trade

This chapter is co-authored with Etienne Fize (CAE)

Abstract

This paper documents the effect of information frictions on trade using a historical large-scale improvement in the transmission of news: the emergence of global news agencies. The information available to potential traders became more abundant, was delivered faster and at a cheaper price between countries covered by a news agency. Exploiting differences in the timing of telegraph openings and news agency coverage across pairs of countries, we are able to disentangle the pure effect of information from the effect of a reduction in communication costs. Panel gravity estimates reveal that bilateral trade increased by 30% more for pairs of countries covered by a news agency and connected by a telegraph than for pairs of countries simply connected by a telegraph.

1 Introduction

Over the course of the XIXth century, international trade flows experienced an unprecedented rise. Recent contributions (Steinwender, 2018; Juhász and Steinwender, 2018) highlighted the key role of the telegraph in explaining this "First Globalization" (on top of the "usual suspects", namely changes in trade policy and lower transport costs[1]). However, an improvement in the communication technology, such as the one studied in these papers, could affect international trade through two channels. Firstly, it reduces search frictions: it's less costly to find buyers or sellers for a specific product, and to coordinate with them. Secondly, it increases the quantity and quality of information available to potential traders. This information channel seems to have been overlooked by the literature, despite its expected relevance as a determinant of export and import decisions: for the exporters, knowledge of the foreign market characteristics (market size, price, trade costs, demand shifters) is of prime importance, while for the importers, the sourcing choice is determined by the information available on price and quality from different markets.

[1] See for instance Estevadeordal et al. (2003) and Pascali (2017)

The specific context of the XIXth century provides a unique opportunity to disentangle the two above-described mechanisms. Indeed, this period witnessed the birth of global news agencies, which systematically collected and transmitted information across borders, so that, for the first time, news became widely available from almost all parts of the globe, with sharply reduced delays. News agencies are wholesalers of information: they gather news and sell them to governments, businesses, and newspapers.[2] On top of more conventional subjects such as diplomacy and politics, providing their customers with commercial news was an important part of the business model of these news agencies. Moreover, the three largest news agencies quickly syndicated into an efficient cost-sharing organization: each of them was given a monopoly over a set of countries, and in exchange committed to share information on these countries with the other news agencies. The sharing of information among the three global news agencies was truthfully enforced, since it ensured that they would stay ahead of the competition. Therefore, being covered by a global news agency meant becoming part of an international network of news sharing, including commercial news.

The development of international news agencies was deeply intertwined with the construction of an international telegraph network: news agencies relied on the telegraph to communicate and often contributed to its expansion. The telegraph represented a considerable improvement upon previous communication technologies,[3] allowing for shorter and less volatile transmission delays. Even though it made communications easier, it did not provide a centralized and reliable source of business information. Indeed, telegraphic messages were private, and it was therefore possible to restrict access to some chosen users. On the other hand, news agencies collected, gathered and sold information that could then be accessed by anyone at a low cost: the end of XIXth century was the period of penny papers and mass media consumption. In other words, in the absence of a global news agency, telegraphs reduced only the communication frictions, without much effect on the amount of information available to the public. We use this distinction between easily enforceable restrictions of access on the telegraph communications and the quasi-public nature of newspaper information to disentangle the effects of reduced communication costs from the effects of improved information access, a separation that previous studies were unable to make.

The telegraph and the news agencies did not cover all pairs of countries simultaneously. The success of the telegraph was immediate, but the cost of the infrastructure and technical factors implied that not all the countries could be quickly connected. Similarly, the global news agencies did not expand their operations to the entire world immediately. They started by sharing Europe and then gradually increased the scope of their syndication agreement through contracts struck in 1859, 1867, 1876, 1889 and 1902. This sequential entry of country pairs into the telegraph and news agencies networks is key for our identification strategy, because it allows us to estimate a panel data version of the gravity equation, meaning that on top of the usual origin and destination time varying fixed effects, we can include country-pair fixed effects, which control for any time-unvarying characteristic of the two countries.

News agencies collected and sold information in a country based on the expected profits from

[2]Newspapers were the main channel of information at the time, they became widely available, and relatively cheap, so that any information published in a newspaper can be considered as public information. In the very competitive environment that prevailed on the press market, it would have been too expensive for these numerous newspapers to collect news from abroad on their own, so they had to rely on news agencies to obtain these news.

[3]Before the invention of the telegraph, mail had to be transported carrying on steamships, railways or horses, which in some situations implied months of delay

serving this country, which in turn may be related to future trade flows. To alleviate the concerns arising from this situation, we use the inclusion of the country into the global news agencies syndication agreement as our measure of news agency coverage. We argue that it accurately reflects the entry in the global news sharing network, since it creates a clear incentive for a global news agency to start covering the country: before being granted exclusivity on a market, a global news agency could decide to sell and collect news, but the exploitation was less likely to be profitable in a competitive situation than in an organized monopoly. Moreover, the extensions of the agreement never concerned one single country, they always included groups of countries, usually of the same geographic region, which suggests that the date of entry into the agreement had more to do with negotiations between the news agencies than with expected trade flows, and that many countries started being covered as a by-product of the agreement rather than for their anticipated economic importance.

Our approach to capture the effect of the information channel is therefore to focus on the interaction between telegraph connections and news agency coverage: while the effect of the telegraph alone can be attributed to the sole decrease in communication costs, the interacted term specifically isolates the contribution of an improved access to news between the two covered countries. The effect is sizable: we estimate the value of trade flows to increase by an additional 30% when two countries are included in the global network of news diffusion, on top of being connected by a telegraph. Our results also corroborate estimates from previous studies that documented a positive effect on trade of the telegraph: we find that, even in the absence of coverage by a global news agency, trade flows increase by 40% when two countries become connected by a telegraph. However, news agencies, in the absence of telegraph, do not trigger any significant increase in trade, suggesting that they were unable to operate at full efficiency when an appropriate communication technology was not available.

These results hold when we control for colonial ties, and data on French tariffs suggest that they are not driven by correlated changes in trade policy. Moreover, we use indirect connections as sources of plausibly more exogenous variations in our explanatory variables. For country pairs that became indirectly linked after the opening of a telegraph line, as for instance Chile and Egypt in 1874, the timing of opening is unlikely to be related to bilateral expected trade flows. Similarly, for the news agencies, if potential traders lobby to include some countries in the cartel agreement based on a perceived high trade potential, it should mostly happen in country pairs in which one of the countries is the headquarter of the news agency, since it is easier to influence a domestic firm than a foreign one. Coverage of other country pairs, as for instance Argentina and New-Zealand in 1876, can be seen as an unintended by-product of the expansion. Leaving aside the direct telegraph connections and the country pairs where one country hosts the headquarter of the news agency, we still find a significant positive trade effect of the interaction between news agency and telegraph.

We then analyze the time dynamics of the effect through an event-study, and find a progressive increase in its magnitude, which slowly rises up to thirty years after the dyads are connected, a picture consistent with a slow constitution of business networks between the countries that benefited from an improved access to information on each others. Finally, we provide evidence supporting the hypothesis that the trade effect is indeed driven by an increase in the quantity of information available on foreign countries. First, we document an increase in trade volatility after the connection, in line with the findings of Steinwender, 2018. This is consistent with a better ability of traders to adapt to market conditions. Second, using data on French newspapers, we find an increase in the presence of a country in the articles once this country benefits from a telegraph connection and from a news agency coverage.

This paper builds on a large literature documenting the effect of information frictions on various economic outcomes, starting with a reduction in price dispersion (Jensen, 2007; Ejrnæs and Persson, 2010). Linking this literature with international trade, Allen (2014) shows that, on top of the usual effect on price dispersion, a decrease in information frictions positively affects the volume of trade. More recent approaches to incorporate information frictions in trade models include Dasgupta and Mondria (2018) and Lenoir et al. (2020). Empirically, a negative link between trade flows and communication costs is established by Malgouyres et al. (2020), who make use of the sequential arrival of high speed internet in French cities. Unlike our approach, these papers however think of information frictions either as search costs or as a mixture of search costs and constrained information set, and do not attempt to distinguish these two dimensions.

Our work highlights the importance in explaining export decisions of the information set available to firms, a dimension that the literature recently started exploring. Indeed, trade decisions crucially rely on the expectation the firm forms about the profits it will earn by serving the foreign market. While perfect foresight is commonly assumed (firms perfectly predict the profits from exporting), Dickstein and Morales (2018) acknowledge that firms actually use only a restricted set of observable variables to assess their expected profits, and study the consequences of enlarging this set of variables. Consistently with our results, they find that more information results in more exports.

Other contributions have explored the economic consequences of the expansion of the telegraph network in the XIXth century. Our results complement the work of Juhász and Steinwender (2018), who document an increase in trade flows following the opening of telegraph lines, especially for goods whose characteristics are easily codifiable, suggesting that these openings decreased communication costs. They are also in line with recent work by **W2020**, who finds a positive effect of telegraph lines connecting the UK with other countries on financial capital flows from the UK to these countries. He shows that this effect is mainly driven by the “newspaper channel”, without explicitly considering the role of news agencies. We also build on the seminal paper of Steinwender (2018), who demonstrates that the transatlantic telegraph led to price convergence in the cotton market, and to a better adaptation to demand shocks. She mentions the importance of the “Reuter’s telegram” provided to traders in the exchanges, but is unable to separate the contribution of these telegrams from the one of other communications allowed by the telegraph. Additionally, her work focuses on UK-US flows while we adopt a broader scope, which allows to further control for potential endogeneity by separating direct and indirect connections.

The importance of the press as a vector of public information is the subject of a large literature. However, to our knowledge, there is no previous work on the importance of news agencies as providers of valuable economic information. The literature on news agencies has focused on industrial organization aspects of the news agencies syndication agreement. Wolff (1991) describes the historical evolution of the news agency industry. Bakker (2014) explains how their business model was designed to answer the specificity of the news market: it provided a solution to the Arrow information paradox (buyers want to know the information in order to determine how much they are willing to pay, but once the information is revealed, they don't need to pay anymore).

In the next section, we present the historical context of our analysis, and our data. Section 3 details our main results, followed by some robustness checks in section 4. In section 5, we provide evidence that the effect we find can indeed be attributed to an increase in the circulation of information.

2 Context and data

2.1 Context

The creation of an international telegraph network The first telegraph system implemented at a significant scale is the optical telegraph of Claude Chappe, which used blades or paddles to transmit information between towers (semaphore telegraph). It was mostly adopted in France where, in the first half of the XIXth century, all major cities were linked together, but had little commercial use and international connections. This is why we focus on the electrical telegraph, that appeared in the 1830s.[4], and quickly became the dominant communication technology, as witnessed by the spectacular growth of the network, both domestically[5] and internationally. The total length of telegraphic wires reached 7 millions km in 1913, and the number of messages exchanged followed a similar pattern, rising sharply from 28 millions telegrams sent in 1865 to 528 millions in 1913.

For long-distance lines, the preferred technology was the submarine cable, because it avoided having to negotiate with countries crossed by the line. The first successful international submarine cable linked France and the UK in 1851. Governments quickly recognized the importance of the telegraph for international communications and soon started cooperating within multinational instances: the German-Austrian Telegraph Union in 1851, the Western European Telegraph Union in 1855, and finally the International Telecommunication Union (ITU) in 1865, that still operates today as a UN agency. International conventions were adopted to ensure the smoothness of transmissions. For instance, in 1865, the Morse code was chosen for all international communications.

The telegraph was a major improvement upon the former communication technologies. It was therefore immediately adopted by private agents to conduct their business operations, as highlighted by Wenzlhuemer (2013, p.84): "from its very inception, the telegraph was intimately connected with the world of business, finance and trade". Nevertheless, its use was very expensive, and therefore reserved to wealthy individuals and firms, governments, and news agencies.

The rise of global news agencies The emergence of news agencies is intimately linked to the sharp rise in newspapers diffusion, that made the press the dominant channel of information at the time. The first news agency appeared in 1832 and was named after its creator, Charles-Louis Havas. The success of Havas, which quickly established a monopoly on its domestic market (France)[6] and started expanding its activity to foreign markets, triggered the creation of competing news agencies. Two former Havas employees, Bernhard Wolff and Paul Julius Reuter, started their own business, creating respectively the Wolffs Telegraphisches Bureau (1849) in Germany (henceforth referred to

[4]The first functioning line, by Cooke and Wheatstone, was built in 1839 between London and a suburbs town.

[5]For instance, in France, the length of the domestic telegraph network was 3540 km in 1852 and 42986 km in 1870.

[6]In 1860, 9 out of 10 Parisian newspapers were subscribers of the French international news agency Havas and for almost all of them it was the most important source of content (**M1997**).

as “Wolff”) and Reuters (1851) in the UK, which quickly gained contracts with newspapers outside of their domestic market. In the 1850s, the market for international news is therefore an oligopoly with three major players : Havas, Reuters and Wolff. These three incumbents quickly understood that more profit could be made by colluding, i.e. by avoiding duplicate costs of news production in some countries, and avoiding competition in some markets. This led to the birth of the international news cartel in 1859.

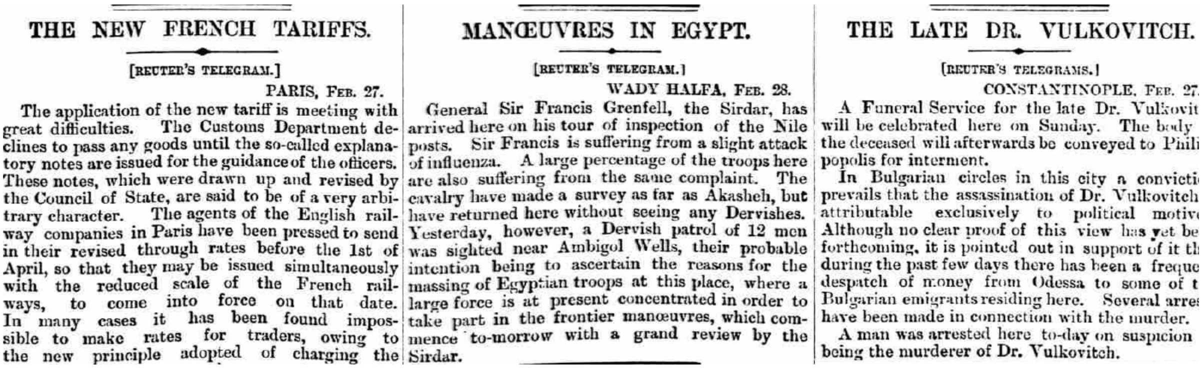
The main component of the 1859 agreement is that each agency was granted a monopoly position in some countries, meaning that no competitor could sell news to the press (or to local news agencies) in these countries. For instance, Havas was the only agency allowed to sell news in Spain, Reuters and Wolff agreed to voluntarily restrain from contracting in this country. However, a key dimension of the agreement is that the news agencies committed to share without fees information coming from their exclusivity zone. Coming back to the Spanish example, this means that Havas had to send the news coming from Spain to the two other cartel members. Moreover, to prevent the appearance of any serious competitor, the three colluding news agencies agreed to communicate only with each other, they could not sell news to another competing news agency. Finally, they pledged to develop the telegraphic infrastructure.

This first agreement mostly concerned only European territories, the rest of the world remained fair game. News agencies were free to collect and sell information in the neutral territories, but there was no systematic exchange of information between them for these territories. But over time, the agreement was reshaped to incorporate new markets that the news agencies deemed profitable. The main extensions occurred in 1867, 1876, 1889 and 1902. On each occasion, new countries were added to the cartel. Table 2.5 (in the appendix) indicates the countries that were added by each cartel agreement. Additionally, two news agencies joined the cartel: AP in 1876 and the Korrbureau in 1889. The cartel slowly dislocated in the aftermath of WWI. After the conflict, Wolff lost its territories, which were shared between the still highly cooperative duo Reuters-Havas. But AP and the other US news agencies (United Press and International News Service) put more pressure on the historic European duo. In the 1927 cartel agreement, many countries became shared territories. The early 1930s were the final blow to the cartel, which officially ceased to operate in 1934. Nowadays, the market is still an oligopoly, with Reuters and AP by far the biggest players and AFP (former Havas) as the third player.

With the news agency cartel, for the first time, a systematic collection and transmission of information across countries is in place. Even though the cartel reduced competition in each local market, it made the coverage of the country profitable (**A2013**; Bakker, 2014). The cartel organization also implied a level of information sharing across the global news agencies that arguably would not have been reached under a more competitive system. Indeed, news agencies really shared the news collected in their exclusive geographical areas. In some cases cooperation was even higher, Reuters and Havas entered into a joint-purse agreement in 1870. According to Bakker (2014) and **A2013**, this organization allowed for an increase in the coverage area, quality and quantity of information. The formal inclusion in the cartel matters since, even though a neutral territory could be served by a news agency, this situation was unlikely given that the market was less likely to be profitable.

It is important to keep in mind that the news agencies were private businesses, under no government mandate. Even though there may have been pressures from their domestic governments, especially during periods of conflict, profit maximization remained their primary objective. As none of the news agencies was State-owned, the decision to add a country in the cartel was not directly taken by governments. This does not mean that countries are added independently from diplomatic or economic considerations, but it implies that the costs and benefits for the news agency is likely to be a key driver of the decision. The case of South-America is a particularly good example: South-America, at the time more economically and politically linked to Great-Britain, was given to Havas to compensate for Reuters getting large territories in Asia.

Figure 2.1: Extract from *The Morning Post*, 02/29/1892



THE NEW FRENCH TARIFFS.

[REUTER'S TELEGRAM.]

PARIS, FEB. 27.

The application of the new tariff is meeting with great difficulties. The Customs Department declines to pass any goods until the so-called explanatory notes are issued for the guidance of the officers. These notes, which were drawn up and revised by the Council of State, are said to be of a very arbitrary character. The agents of the English railway companies in Paris have been pressed to send in their revised through rates before the 1st of April, so that they may be issued simultaneously with the reduced scale of the French railways, to come into force on that date. In many cases it has been found impossible to make rates for traders, owing to the new principle adopted of charging the

MANŒUVRES IN EGYPT.

[REUTER'S TELEGRAM.]

WADY HALFA, FEB. 28.

General Sir Francis Grenfell, the Sirdar, has arrived here on his tour of inspection of the Nile posts. Sir Francis is suffering from a slight attack of influenza. A large percentage of the troops here are also suffering from the same complaint. The cavalry have made a survey as far as Akasheh, but have returned here without seeing any Dervishes. Yesterday, however, a Dervish patrol of 12 men was sighted near Ambigol Wells, their probable intention being to ascertain the reasons for the massing of Egyptian troops at this place, where a large force is at present concentrated in order to take part in the frontier manœuvres, which commence to-morrow with a grand review by the Sirdar.

THE LATE DR. VULKOVITCH.

[REUTER'S TELEGRAMS.]

CONSTANTINOPLE, FEB. 27.

A Funeral Service for the late Dr. Vulkovit will be celebrated here on Sunday. The body the deceased will afterwards be conveyed to Phili popolis for interment.

In Bulgarian circles in this city a convicti prevails that the assassination of Dr. Vulkovitch attributable exclusively to political motiv Although no clear proof of this view has yet be forthcoming, it is pointed out in support of it th during the past few days there has been a freque despatch of money from Odessa to some of t Bulgarian emigrants residing here. Several arre have been made in connection with the murder.

A man was arrested here to-day on suspicion being the murderer of Dr. Vulkovitch.

News are available from all around the world through the "Reuter's telegrams", including from a country that is not part of Reuters' exclusive distribution area (France). Source: British Newspaper Archive.

As figure 2.1 illustrates, newspapers used the global news agencies as their main, and often sole, source of foreign information. Indeed, it was too costly for each newspaper to collect foreign news on its own.[7] They could either be direct subscribers to the international news agencies services, or clients of a national news agency which in turn relied on a global news agencies for foreign news. Government or private organizations could also subscribe to foreign news services from the global news agency serving their country.

News agencies and telegraph complement each other The reduction in information friction is fully realized when two countries are both covered by a news agency and linked by a telegraph. In the absence of a telegraph line, news agencies do not have the possibility to send news swiftly from one country to another. Conversely, in the absence of news agencies, information can flow among privileged users, but not systematically and with the large audience reached by the newspapers.

Because the telegraph was at the core of the news agency operations, the geographic extensions of the cartel coincided with the development of the telegraph network. The motto of Paul Julius Reuter was clear: "Follow the cable".[8] The cartel started by dividing Europe where the telegraph network was quite dense. Then South America and Australia were included in 1876, a few years after being linked to Europe by submarine cables (in 1874 for South America and 1872 for Australia). It does not however imply that all countries added to the cartel were already connected by a telegraph.[9]

[7] Even though a handful of newspapers had foreign correspondents in some specific locations for which they believed their readership would ask for very detailed reports, the cost was too high for this practice to be common. Only a few newspapers could afford foreign correspondents and these correspondents were often sent temporarily to cover big events (**A2010b**).

[8] Wenzlhuemer (2013, p. 90)

[9] Bolivia and Ecuador for instance were not connected to any international telegraph in 1876.

The telegraph provided the speed, reliability and privacy necessary to the news agencies. The link between news agencies an the telegraph is so tight that they were often referred to as "telegraphic news agencies", and sometimes even played an active role in the construction of new telegraph lines.[10]

Without the news agencies, a systematic and efficient transmission of information cannot take place, and the impact of the telegraph is restricted to its use as a communication device for those that can afford it. Indeed, sending telegrams was very expensive[11]and therefore *de facto* reserved to the most important and well established traders, while the press was relatively cheap and accessible to any potential trader. The role of newspapers as provider of public information is explicitly acknowledged by Ejrnæs and Persson (2010): "A flourishing commercial press turned what used to be exclusively or privately held knowledge into publicly accessible information. While the larger merchants' houses had access to telegraph transmission directly, others relied on 'cable news' reported in the press."

In fine, thanks to the telegraph, news agencies were able to transmit information considerably faster than before, especially for very distant countries. In figure 2.2, we plot the average delay between the date of an event and the date of publication of this event in the Time, a London newspaper. The dramatic drop in transmission times between London and South America coincides with the opening of the 1874 telegraph between Europe and South America. Australia was finally connected with Europe in 1872. Both South America and Australia entered the coverage of news agencies in 1876. From 1880 onwards, information is shared within days, compared with months a few decades ago.

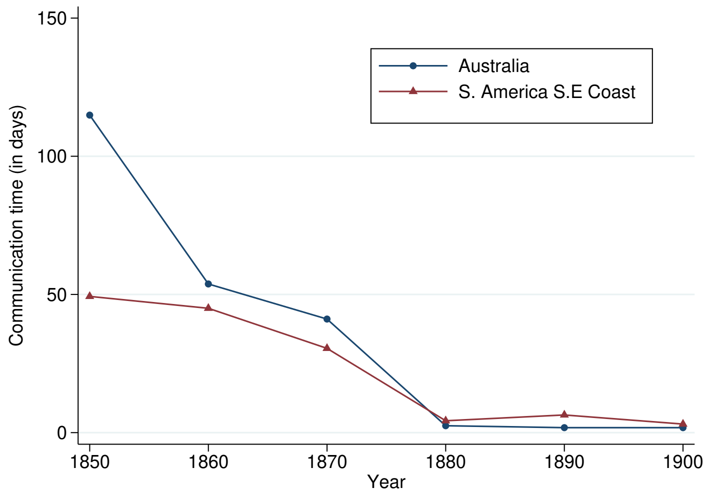
2.2 Data

Trade data: Information on historical bilateral trade flows comes from the TRADHIST database (Fouquin and Hugot, 2016), which combines data from previous databases with novel information extracted from primary sources (manuscripts from the customs archives). It also contains several bilateral variables linked to trade frictions, such as distance. On top of its extensive coverage of past trade flows, TRADHIST has the desirable feature that it gives preference to importer reported data, when available, which ensures higher accuracy. We nevertheless had to make two substantial modifications to the original data. First, because our analysis relies heavily on time variation, it was crucial to ensure consistency over time of the countries, so we grouped them according to the largest existing legal entity over the period. For instance, Sweden and Norway formed a single Kingdom until 1905, at which point Norway became an independent country, so we gathered them into a single entity, "Sweden-Norway". We obtained the trade flows of the so-formed entities by summing the trade flows of their components: trade flows between "Sweden-Norway" and Denmark after 1905 are the sum of trade flows between Sweden and Denmark, and trade flows between Norway and Denmark. Second, we re-coded the variables indicating bilateral colonial ties to ensure higher

[10]For instance, the transatlantic telegraph was built with the guarantee by Reuters to bring a "considerable volume of business" (Unesco, 1952, p.153.), and Bielsa (2008) explains that news agencies "have been instrumental in the creation of the material infrastructures for the production and circulation of information and in the development of worldwide networks, starting with the telegraph, which, in the second half of the nineteenth century became the first system for global communications"

[11]For instance sending a 10 word transatlantic telegram in 1866 corresponded one fifth of the annual wage of a US skilled worker.

Figure 2.2: Delay between the date of an event and its publication in the Time (London)



y-axis : average number of days between the date an event took place and the date it was reported in the Time (London). Data source : Wenzlhuemer (2013)

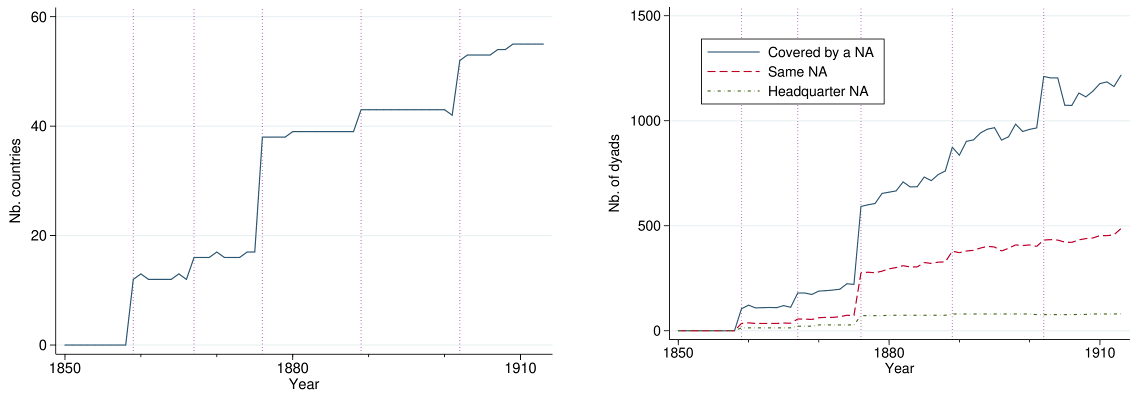
accuracy.

Our analysis starts in 1850, nine years prior to the first cartel agreement (1859) and fifteen years after the creation of the first news agency (Havas, 1835). It ends in 1914, since World War I is known to have significantly disrupted the trade patterns, and the cartel agreements were de facto less binding after 1918, even though they still formally existed. Over this period (1850-1914), five cartel agreements were struck : in 1859, 1867, 1876, 1889 and 1902. Our baseline unit of analysis is a directed pair of countries, i.e. a combination importer-exporter (France -> Uruguay for instance). Throughout the paper, we often refer to these pairs of countries as dyads. The country of origin is indexed by o, and the destination country by d.

News agency data: We hand-coded the geographic coverage of the news agency cartel based on information provided in Wolff (1991). More precisely, we defined a dummy variable $NewsAgency_{odt}$ taking value 1 when the importer and the exporter are covered by a news agency participating to the international news exchange agreements. This bilateral variable is not directed: if $NewsAgency_{odt} = 1$, then $NA_{dot} = 1$. Colonies are not systematically mentioned in the cartel agreements, so that it's hard to tell whether they were actually covered by a news agency or not. Therefore, in our baseline specification, we adopt the conservative approach of setting our NA dummy to 1 only for countries whose status is explicitly stated in Wolff (1991). In figure 2.3, we plot the number of countries mentioned in the cartel agreements, and the number of dyads in which both countries are covered by a news agency. As expected, the coverage increases over time, and more specifically rises sharply after each extension of the cartel agreements. We notice a big jump after the 1876 agreement, easily explained by the fact that at this date many countries were added to the cartel (South America and large parts of Asia). More descriptive statistics on news agency coverage are available in the appendix, figure 2.14.

Figure 2.3: Evolution over time of the news agency coverage

(a) Number of countries **(b)** Number of dyads



Notes: Each vertical purple dotted line corresponds to an agreement extending the geographic coverage of the cartel. In fig (b), the blue line counts the number of dyads that are covered by a news agency participating to the international news exchange agreements; the red dashed line counts the dyads that are covered by the same news agency and the yellow dotted line corresponds to the dyads in which the importer and the exporter are covered by the same news agency and one of them is the headquarter of the news agency.

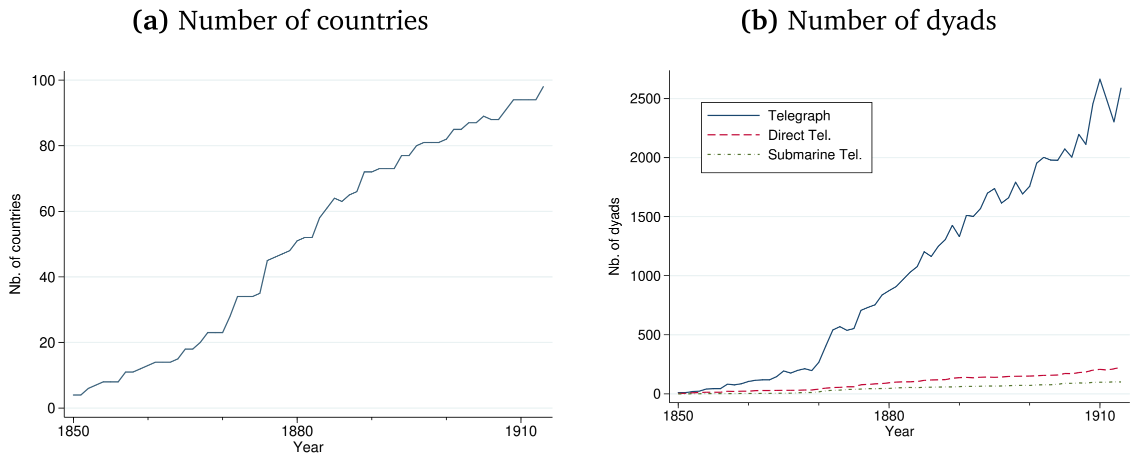
Telegraph data: Data on international telegraph links comes from two sources. For the submarine telegraph cables, we rely on the *Journal Télégraphique*, a monthly publication by the International Telegraph Union (ITU). In some issues, an appendix is available with an exhaustive list of the submarine cables in use and the date at which they started operating[12]. This data was collected by Roland Wenzlhuemer, who kindly accepted to share it with us. The last nomenclature dates back to 1903, before the end of our sample, but at this time the main submarine cables had already been laid down, and most of the countries were already telegraphically connected. Regarding the terrestrial telegraphic lines, no such nomenclature is available, so we had to rely on the visual analysis of maps from the "Bureau international des administrations télégraphiques", digitized and made available by the Bibliothèque nationale de France (BnF). The list of the maps we used is available in the appendix, along with one example of such maps (figure 2.9). The first map dates back to 1856 and the last one to 1912. Finally in order to start our sample in 1850, we use the list of the first international telegraphic lines from Wolff (1991).

We define a dummy variable $DirectTelegraph_{odt}$ taking value one if a telegraph links directly country o and country d at date t. For the submarine cables, we know for sure the first year in which the two countries are linked since the construction dates are given in the nomenclatures. For the terrestrial cables, we assume that $DirectTelegraph_{odt} = 0$ until the first time we see the telegraph on a map. To clarify, the first telegraph between Chile and Argentina was built in 1871.[13] We first see it on the 1875 map, so we code $DirectTelegraph_{ARG,CHL,t} = 0$ until $t = 1874$. Nevertheless, given the relatively small time span between each map, the imprecision should be small. Since telegraphs can always be used in both directions, the link is symmetric, so $DirectTelegraph_{odt} =$

[12] "Nomenclature des câbles formant le réseau sous-marin du globe dressée d'après des documents officiels par le Bureau international des administrations télégraphiques" in 1875, 1877, 1883, 1887, 1889, 1892, 1894, 1897, 1901 and 1903. See appendix for more details.

[13] Between Valparaiso (Chile) and Villa Nueva (Province of Córdoba, Argentina).

Figure 2.4: Evolution over time of the telegraph coverage



Notes: In fig (b), the blue line counts the number of dyads that are connected by a telegraph, directly or indirectly; the red dashed line counts the dyads that are directly connected and the green dotted line corresponds to the dyads that are connected by a submarine telegraph cable.

$DirectTelegraph_{dot}$. Based on the direct connections data, we build a dummy variable indicating whether any pair of countries is indirectly connected, meaning that the two countries are connected either directly or via one or many intermediary countries, i.e. they belong to the same connected sub-graph. For instance, if $DirectTelegraph_{ok} = 1$ and $DirectTelegraph_{kd} = 1$, then $Tel_{od} = 1$ even if $DirectTelegraph_{od} = 0$. In more formal terms, two countries o and d are considered as indirectly connected if there exists at least one path of any length n that connects o to d.

Figure 2.4 depicts the number of countries with at least one international telegraph connection (left-hand side) and the number of dyads connected by the telegraph (right-hand side). We distinguish between the dyads connected by a submarine cable, the dyads directly connected (by land or submarine cables) and the dyads that are directly or indirectly connected. In all cases, the number of connections increases, but the number of direct connections lags far behind the total number of connections. This is due to the network structure: adding a new direct link is likely to create more than one indirect link. Appendix figure 2.13 provides more descriptive statistics on the evolution over time of the number of telegraph connections.

This measure of telegraphic connection within a dyad has some drawbacks. First, it does not give any idea of the quality, and speed of the telegraphic connection between countries. It does not tell us about the speed at which information can be exchanged, we can calculate the shortest path but this not necessarily the fastest one as cables are not all of the same quality. Second, this measure tells us whether or not two countries are connected, but it does not tell us how much information is (and can be) exchanged between them. Nevertheless, we believe it is a valid proxy of the ability of two countries to communicate in a fast and reliable manner through this new technology. It differs considerably from the measure used in the first attempt to assess the effect of telegraph on trade (Lew and Cater, 2006), namely the product of the number of telegraphs sent by each country. This variable is not bilateral in essence, two countries may indeed send a lot of telegraph but are not connected and therefore cannot communicate with each other. Moreover it does not allow the inclusion of destination-year and origin-year fixed effects, that are now standard in gravity equations.

3 Main results

3.1 Estimation

To determine the effect of news agencies and telegraph on bilateral trade, we estimate the panel version of gravity equations. As recalled by Head and Mayer (2014b), a gravity equation for bilateral trade flows can be obtained from all the main models of international trade, so that our results do not require any assumption on the most appropriate way to model trade flows in the XIXth century. The gravity equation of trade flows refers to a multiplicative structure of the type:

$$Y_{odt} = O_{ot} D_{dt} \phi_{odt} \eta_{odt}$$

where Y_{odt} denotes the bilateral trade flow between the origin o and the destination d during year t. O_{ot} is a measure of the capability of country o to export, whatever the destination, and D_{dt} captures the general propensity of country d to import, whatever the origin of these imports. ϕ_{odt} is a bilateral resistance term incorporating the effect of all the trade frictions between the importer and the exporter. Finally, η_{odt} is an error term with mean zero. Among the factors determining the bilateral resistance term, ϕ_{odt}, some are fixed over time, others do vary over time :

$$\phi_{odt} = \underbrace{B_{od}}_{\text{Fixed bilateral frictions}} \times \underbrace{\exp(\boldsymbol{\beta}'\mathbf{X}_{odt})}_{\text{Time-varying bilateral frictions}}$$

Plugging this into the gravity equation, and taking the expected value of trade flows:

$$\begin{aligned}
\mathbb{E}(Y_{odt}) &= O_{ot} D_{dt} \phi_{odt} \\
\mathbb{E}(Y_{odt}) &= \exp(\ln(O_{ot} D_{dt} \phi_{odt})) \\
\mathbb{E}(Y_{odt}) &= \exp(\underbrace{\ln(O_{ot})}_{\text{FE}_{ot}} + \underbrace{\ln(D_{dt})}_{\text{FE}_{dt}} + \underbrace{\ln(B_{od})}_{\text{FE}_{od}} + \boldsymbol{\beta}'\mathbf{X}_{odt})
\end{aligned}$$

This leaves us with the following conditional expectation for the bilateral trade flow:

$$\mathbb{E}(Y_{odt}) = \exp(\text{FE}_{ot} + \text{FE}_{dt} + \text{FE}_{od} + \boldsymbol{\beta}'\mathbf{X}_{odt}) \tag{2.1}$$

Equation (2.1) shows that if we include sets of importer × year, destination × year and country pair fixed effects, and if the vector of time varying frictions $\mathbf{X}_{odt}$ is complete, then each component of the vector $\boldsymbol{\beta}$ can be recovered without bias. The variables we are interested in are part of $\mathbf{X}_{odt}$, since they are bilateral and time varying, so that obtaining unbiased estimates of $\boldsymbol{\beta}$ is what we desire. Estimating equation (2.1), any country specific variable, such as the quality of institutions, GDP, or productivity is fully captured by the origin-time or destination-time fixed effects. Similarly, the specific nature of a relationship between two trade partners, i.e. the fact that two countries may trade more together because of cultural or historical factors, or for any other idiosyncratic reason, is captured by the dyadic fixed effects as long as it is not time varying.

As explained in the previous section, we constructed two dummy variables proxying for the bilateral coverage by a global news agency ($NewsAgency_{odt}$) and the connection by a telegraph ($Telegraph_{odt}$). Because the bilateral information flow should be higher when these two dummy variables both take value 1, we are especially interested in their interaction term, $NewsAgency_{odt} \times Telegraph_{odt}$. If news agencies and telegraphs are complementary, because the telegraph allows for a more efficient communication among the network of global news agencies, while news agencies provide content that can be shared with a wider audience than private telegrams, the effect of $NewsAgency_{odt} \times Telegraph_{odt}$ is expected to be positive: the importer and the exporter get more information on each other, which is expected to increase trade between them. $NewsAgency_{odt} \times Telegraph_{odt}$ is not captured by the fixed effects since its value changes over time within a dyad and over trade partners within a country-year. The sign of the coefficients on $NewsAgency_{odt}$ and $Telegraph_{odt}$ could be either positive or null. It would be positive if the sole fact of being covered by a news agency (or benefiting from a telegraphic connection) increases the amount of information transmitted, or null if only the conjunction of the two matters.

News agency coverage and telegraphic links could be correlated with changes in the colonial ties linking both countries[14], which are in turn linked to trade flows, so that failing to account for those ties may result in an omitted variable bias. Therefore, in our baseline specification, the vector of time-varying variables, $\mathbf{X}_{odt}$, also includes three controls for colonial ties: $BothColonized_{odt}$ indicates whether both countries currently are colonies (not necessarily of the same empire), $SameColonizer_{odt}$ takes value 1 if the importer and the exporter belong to the same colonial empire, i.e. are currently colonized by the same country, and finally $MetropoleColony_{odt}$ captures the specific metropole-colony ties (the importer is the colonizer of the exporter, or vice-versa).

To summarize, the panel data gravity estimation allows us to get rid of potential confounding factors at the dyadic level and at the country-year level. The estimates are identified from time variations in trade flows, and therefore rely solely on dyads whose status changed over time. Concerning news agencies, the overwhelming majority of these changes consists in entering the cartel agreements. Very few countries are reassigned to another news agency, and no country is dropped out of the cartel agreements. For telegraphic connection, the changes comes from telegraphic links creations.

To estimate equation (2.1), two methods can be used. The first one consists in log-linearizing. Then, the trade friction coefficients $\boldsymbol{\beta}$ can be estimated through OLS:

$$Ln(Y_{odt}) = \text{FE}_{ot} + \text{FE}_{dt} + \text{FE}_{ot} + \boldsymbol{\beta}'\mathbf{X}_{odt} + \varepsilon_{odt}$$

Nevertheless, this OLS estimator is biased under heteroskedasticity (Silva and Tenreyro, 2006). To overcome this, one can use a pseudo maximum-likelihood estimator, assuming either a Poisson distribution for the density function (Poisson Pseudo Maximum Likelihood, henceforth PPML), or a Multinomial distribution (Multinomial Pseudo Maximum Likelihood). Both assumptions yield identical estimates as long as destination fixed effects are included (Sotelo, 2017), which is the case in all of our specifications. Besides the fact that it corrects for the heteroskedasticity driven bias affecting the OLS estimator, the PPML estimator essentially differs from its OLS counterpart on two accounts:

1. The weight put on large trade flows : PPML gives more importance to large trade flows than OLS (Head and Mayer, 2014b). To ensure that countries are treated more equally, Sotelo (2017) proposes to use market shares, i.e. trade flow between the two countries divided by total imports of the destination country.

[14] The colonial ties are not absorbed by the dyadic fixed effects since they vary over time during our period of interest.

2. The ability to handle zero trade flows : because it relies on a log-linearization, OLS requires to drop the zeros, while PPML does not, allowing for the incorporation of an extensive margin in the estimates (by extensive margin, we mean the switch from zero trade to strictly positive trade for a given dyad). Whether this is a desirable feature or not depends on the (unknown) true nature of the zero trade flows: are they true zero trade flows or unrecorded trade flows?

Because of its ability to correct for the heteroskedasticity driven bias, PPML is our preferred estimator. Nevertheless, we provide additional results based on OLS estimations, that are not qualitatively different from the ones we obtain from the PPML estimations. When estimating PPML models, we use as dependent variable trade shares instead of trade levels, to ensure that the weighting scheme is more comparable to the one from the the OLS estimations. These trade shares are denoted S_{odt} in the rest of the paper, and are defined as the share of imports of country d coming from country o:

$$S_{odt} = \frac{Y_{odt}}{\sum_{o \neq d} Y_{odt}}$$

In our dataset, some zero trade flows are recorded, which correspond to cases in which the creators of TRADHIST considered that there was indeed a null trade flow between the two countries, i.e. that the missing trade flows was not due to a lack of data. However, this is necessarily based on assumptions on a threshold above which the data is deemed sufficiently complete to confidently attribute the absence of a recorded flow to an absence of transaction. Moreover, we merge some geographical entities to ensure the consistency of each country included in our dataset over the whole period of study, and zero trade flows are not defined for these newly created country pairs. Our approach is therefore to use the information on zero trade flows in our preferred specification, but to provide as robustness checks estimates from specifications in which we either drop all the zero trade flows, or assume that all the non strictly positive trade flows between any existing pair of countries actually are zero trade flows, i.e. forcing the sample to be perfectly balanced.

3.2 Results

Table 2.1 presents our main results. We estimate the panel gravity model from equation (2.1) using alternatively an OLS estimator, in columns (1) and (2), and a PPML estimator, in the other columns, (3) to (6). We also present results with or without controlling for colonial ties, and with different approaches in dealing with non strictly positive trade flows.

Column (4) is in our view the most appropriate specification: it corrects for the potential heteroskedasticity driven bias of the OLS estimates, while incorporating information from the most reasonable zero trade flows (the ones included in TRADHIST) and giving an equal weight to all countries thanks to the use of trade shares as dependent variable. With this specification, we find that the telegraph on its own increased trade by 38%, news agencies by 33% (although this estimate is not significant, a point we comment below) while the combination of the two resulted in a magnified increase, 30% additional trade. The variable identifying the improvement in information sharing between the two countries is the interaction term, and the fact that its effect is positive and significant confirms that improving the access to public information fosters trade.

In columns (1) and (3), we present results obtained when omitting the variables controlling for

Table 2.1: Effect of news agencies and telegraphs on trade flows, panel gravity estimates (1850-1913)

	OLS		PPML			
	$\text{Ln}(Y_{odt})$ (1)	$\text{Ln}(Y_{odt})$ (2)	S_{odt} (3)	S_{odt} (4)	$S_{odt} > 0$ (5)	S_{odt} (6)
News Ag. × Tel.	0.426^{b}	0.427^{b}	0.262^{c}	0.259^{c}	0.255^{b}	0.319
	[0.188]	[0.188]	[0.158]	[0.156]	[0.111]	[0.204]
Telegraph	0.319^{b}	0.323^{b}	0.323^{a}	0.334^{a}	0.181^{c}	0.383^{a}
	[0.147]	[0.147]	[0.116]	[0.115]	[0.096]	[0.146]
News Agency	0.286	0.287	0.232	0.222	0.178	0.075
	[0.210]	[0.210]	[0.184]	[0.183]	[0.137]	[0.237]
Observations	59910	59910	83373	83373	59910	140506
Sample	Complete	Complete	Complete	Complete	Complete	Balanced
Colony controls	×	✓	×	✓	✓	✓

Note: Data is aggregated at the country pair × year level. The dependent variable is the log of the bilateral trade flow in columns (1) and (2) and the share of imports of country d coming from country o (S_{odt}) in the remaining columns, (3) to (6). All estimations include destination × year, origin × year and country-pair fixed effects. $NewsAg. \times Tel.$ is a dummy indicating that both countries are covered by a news agency and linked by a telegraph. "Balanced" sample refers to the case in which we form all the possible combinations dyads × year and assign a zero trade flow if nothing was recorded in TRADHIST. In brackets are the standard errors, clustered by country-pair. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$.

colonial ties. These controls include three dummy variables: one indicating that both countries are colonized (potentially by a different country), the other one that they are colonized by the same country (they belong to the same colonial empire) and the last one that a country is a colony and the other one is its colonizer. When comparing these estimations to their counterpart with colonial ties controls (columns (2) and (4) respectively), we see that the estimates are remarkably close, suggesting that the correlation between our variables of interests and the colonial status is low.

The results are qualitatively unchanged when we use an OLS estimator. The presence of both a news agency and a telegraph in both countries still has a positive and significant effect on trade, which is even a bit larger than the one found with a PPML estimation (a 53% increase), but is also slightly less precisely estimated. For the telegraph, the effect as is remarkably close to the PPML estimate (38% more trade).

In columns (5) and (6), we experiment different assumptions on the appropriate way to treat zero trade flows. Column (5) presents our estimates after dropping all the zero trade flows. While the magnitude of the interaction term is left almost unchanged, the effect of the telegraph is reduced, suggesting that communication technologies *per se* have an important effect on the extensive margin of trade, *i.e.* on the probability that a dyad has a positive trade flow. In the last column (6), we conduct a different experiment, assuming that all the potential dyads for which we have no information in TRADHIST actually have a zero trade flow. With this balanced sample, the point estimates of the telegraph effect and of the interaction term grow, but the interaction term is less precisely estimated, and its effect is therefore not significant anymore.

The contribution of this paper is to establish the positive effect of a shock on information, identified by the joint presence of a news agency and a telegraphic link between the two countries. On top of being positive and significant, this pure effect of information is relatively large (a 30% increase in trade in our preferred specification). Interestingly, this magnitude is in line with the results of Dickstein and Morales (2018), who find that, for contemporary trade on chemicals, switching from minimal information to perfect foresight would result in aggregate exports rising by 25.1 to 33.5%. With a value of −5 for the trade elasticity of trade flows with respect to trade costs,[15] this 30% increase in trade value corresponds to a 5 percentage points decrease in the iceberg trade cost.[16]

The positive effect of telegraphs on trade is not a novel finding of this paper, but it confirms in a larger sample and over a longer time-horizon the results of Steinwender (2018). Her estimates imply an increase by 87%[17] of cotton exports from New-York to Liverpool after the opening of the transatlantic telegraph, an effect that incorporates both the reduction in communication costs and the more efficient provision of information, and that should therefore be compared to the one we find when summing the effects of the telegraph and of the interaction term.

The effect of news agencies *per se* (i.e., in the absence of telegraph) on trade is always positive, but never statistically significant. This suggests that in the absence of a telegraph connection, the news agencies were not able to efficiently share information. "Important news travelled along telegraph lines. And wherever there were such lines, there was also the latest news." (Wenzlhuemer, 2013, p.91). The corollary is that wherever there was no telegraph line, the news agencies were not able to operate in a fully satisfactory manner.

In appendix table 2.6, we introduce alternatively in the regression each of our variables of interest, i.e. we estimate one specification with news agencies as the sole explanatory variable of interest, one with telegraph only, and one with the interacted term only. The effects of both the telegraph and the news agencies appear a bit larger when the interaction is not properly accounted for, highlighting the necessity to consider simultaneously these two determinants of trade.

4 Robustness checks

All our specifications include time-varying origin and destination fixed effects, which rule out the hypothesis that our effect is driven by the general tendency of trade flows to rise over the period we study. These fixed effects also ensure that our estimates are isolated from channels not directly related to a decrease in bilateral trade frictions, such as an increase in GDP of the countries benefiting

[15]This value of the trade elasticity parameter comes from Fouquin and Hugot (2016), who estimate the elasticity of trade flows with respect to trade costs during the period we study and find it to be around −5, a value remarkably close to the median of the estimates obtained on more recent data (Head and Mayer, 2014b)

[16]The trade effects we estimate are the product of the semi-elasticity of trade costs with respect to our information shock and the elasticity of trade flows with respect to trade costs:

$$\frac{\partial \ln(Y_{odt})}{\partial X_{odt}} = \frac{\partial \ln(Y_{odt})}{\partial \ln(\tau_{odt})} \times \frac{\partial \ln(\tau_{odt})}{\partial X_{odt}}$$

where τ_{odt} denotes the iceberg trade cost. Plugging in our estimates: $0.26 = -5x$, hence $x = -0.052$. Hence the 5 percentage points decrease in iceberg trade costs.

[17]Table 8 p.676, column(9), the coefficient associated to the telegraph dummy with log exports as dependent variable is 0.63.

from news agency and telegraph coverage, or more openness to trade in general. Indeed these two factors depend solely on the origin or destination, and are therefore fully absorbed by the set of fixed effects.

We also include dyadic fixed effects which control for time-unvarying specific bilateral relationships. These dyadic fixed effects control for observable factors like distance or language proximity, but also for factors that would be harder to measure in a satisfactory manner, such as diplomatic or cultural proximity between the two countries. Therefore, the positive effect we find cannot be attributed to a cross-sectional positive correlation between our variable of interest and any omitted variable that would positively affect trade. In order for our identification to be biased, an unaccounted factor has to vary over time within a pair of countries and be correlated with news agency coverage and telegraphic connections.

We identify two such threats to our identification. The first one is that the date at which a telegraph is built between a country pair or at which the country pair is included in the syndication agreement may be driven either by anticipation of large trade flows or respond to observed past large trade flows. The second one is that it may be driven by diplomatic factors that would correlate with the bilateral trade policy, i.e. two countries may have a more favorable relationship that would be associated both with telegraph & news agencies link and lower trade barriers.

Our first robustness check is to isolate the sub-sample of the treated units for which the treatment date could be endogenous, in order to focus on the treated units for which it is exogenous. This is done in sub-section 4.1, where we present results estimated solely from indirect connections, and find that the positive trade effects we identified still do exist when relying on the most exogenous source of variation. To address more specifically the second threat, we check in sub-section 4.2 whether our variables of interest are correlated with tariffs on a subsample of our dataset (the dyads for which we do have tariff rates, that are unfortunately not available for our complete universe). We find no significant correlation, which suggests that trade policy is not closely tied with telegraphic and news agency linkages. Finally, in sub-section 4.3, we present the results of an event-study. They confirm the absence of pre-trend in trade for the dyads that will be connected in the future.

4.1 Separating direct and indirect connections

The timing of construction of a telegraphic line may be linked to anticipated trade flows, or to past trade flows between the two countries that would become connected. Even though the technical difficulties of the construction process made the precise timing of a successful telegraph opening hard to predict, especially for submarine cables, as argued by Steinwender (2018), our approach is relying on long-run time variations, which means that this short-run randomness in opening date may not be sufficient to ensure exogeneity of our explanatory variable. Our coefficient on the trade effect of telegraph could be either upward or downward biased, depending on the source of endogeneity. If telegraphic lines are built between countries for which it is expected that the trade relationship will deepen in the future, then we would overestimate the trade benefits of the telegraph. Conversely, if telegraphic lines are built between countries that already have a deep trade relationship and therefore a demand that is already very high for telegraphic services, then there would be less room for future trade growth and we would underestimate the trade effect of the telegraph.

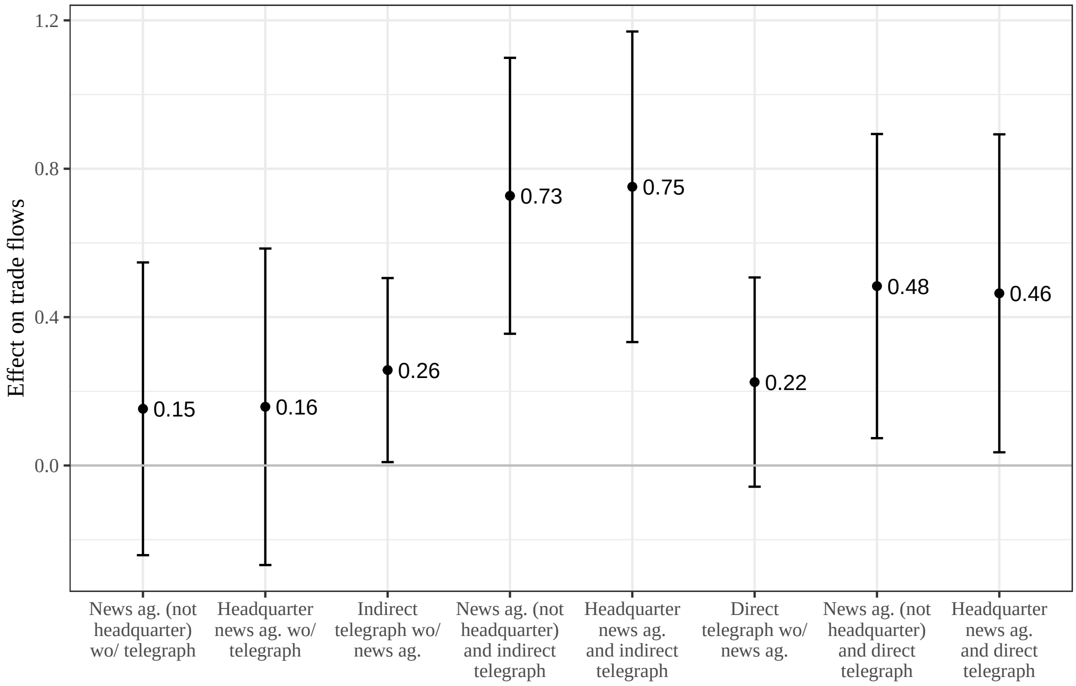
To alleviate these concerns, we focus on indirect connections. While countries may broadly control the date at which a direct telegraphic link is established between them, they have less power in deciding when the last missing segment to create a telegraphic path between them will be built. For instance, Brazil and China became indirectly linked in 1874, after the completion of a transatlantic telegraph between Portugal and Brazil (with relays in Madeiras and Cabo Verde). Arguably trade between China and Brazil had little influence on the timing of this connection. Therefore, we estimate separately the effect of direct and indirect connections, by adding a dummy variable $DirectTelegraph_{odt}$ that indicates a direct telegraph link. The coefficient on $Telegraph_{odt}$ then captures exclusively the effect of indirect links, and should provide a more accurate estimate of the causal effect.

Endogeneity concerns on the date of inclusion in the news agencies' syndication agreement are addressed with a similar strategy. The decision to add a country is linked to the expected profits from selling news in this country, which in tun may be linked to trade flows. Note however that the link has to be truly bilateral: the fact that countries are added based on their total economic size is not an issue for our identification since we include time varyin origin and destination fixed effects in our estimations. Also, we underline again that additions to the news agencies' syndication agreement occurred by waves: groups of countries were added in five staggered extensions, so the precise date of inclusion is unlikely to be at the hands of the concerned countries. If there was nevertheless a link between the date at which a pair of countries is included in the syndication agreement and bilateral trade flows, it could lead either to over- or underestimation of the causal effect. If two countries are included because of the prospect of large trade flows (and high demand for bilateral information), then we would overestimate the information effect. Conversely, if two countries are included because they faced an idiosyncratic positive shock on their past trade flows, we would underestimate the information effect.

We make use of the fact that the news agencies were not national but global operators. It is unlikely that a country would be able to exert pressure on a large foreign company to curb its choices towards covering certain areas. Lobbying efforts are easier when the firm is domestic. For instance, it would be harder for Spain than for France to influence the choices of Havas, the French news agency. Our strategy is therefore to decompose the country pairs between those in which one of the countries is the headquarter of the news agency, and the ones in which none of the countries is. The date of treatment is more likely to be endogenous in the former group, in which the news agency is a domestic firm for one of the countries than in the latter one, in which the news agency is a foreign operator for both countries. For instance, Argentina and Australia started being included in the news agency network in 1876, arguably an exogenous timing since the country pair became linked without any particular intent as a by-product of the agreement's extension. To isolate the dyads in which we could fear endogeneity, we add a dummy $HeadquarterNewsAgency_{odt}$, taking value 1 when both countries are covered by the same news agency and one of the countries is the headquarter of the news agency. Havas is considered to be based in France, Reuters in the UK, and Wolff in Germany.

Figure 2.5 presents the results of this estimation. We use our preferred specification: PPML estimator, with trade shares as dependent variables, on the complete sample (including the zero trade flows provided in TRADHIST). However, in this graph the reference situation is the absence of both a news agency and a telegraph, i.e. the interacted terms do not correspond to additional effects

Figure 2.5: Effect of telegraphs and news agencies on bilateral trade flows, distinguishing between direct and indirect links.



Notes: PPML estimates of the effect on trade flows, where the reference situation is the absence of news agency and telegraph. Interpretation: when two countries become indirectly linked by a telegraph but are not covered by a news agency, trade is expected to increase by 30% ($e^{0.26} - 1$). Bars indicates the 95% confidence interval, with standard errors clustered at the country-pair level.

as in table 2.1, but rather to total effects on trade. Additional effects are harder to interpret and are therefore presented later, in table 2.2.

We find that, in the absence of a telegraph, the trade effect of news agencies is not very different whether or not one of the countries is the headquarter of the news agency (a 17% vs a 16% increase in trade flows). This confirms that the endogenous timing of news agency coverage is not a big concern. Coefficients are a bit less similar when we compare direct and indirect telegraphic links: a 25% increase in trade for direct telegraphic lines vs a 30% increase for indirect telegraphic connections. This is consistent with the scenario in which direct telegraphic lines are built across countries that already had high trade levels before the construction, and for which there is less room for trade growth.

Most importantly, the trade effect of being indirectly connected both by a telegraphic line and a news agency largely exceeds the sum of the two effects taken separately (108% ($e^{0.73}$) vs 51% ($e^{0.15+0.26}$)), which confirms that, even when focusing on the dyads for which the connection date is more exogenous, we do find a positive and significant contribution of our information shock. However, the additional trade effect of being covered by a news agency is significantly lower for the country pairs that were directly connected by a telegraph than for those that were indirectly connected, suggesting that these country pairs already had abundant information on each other, and therefore benefited less from the arrival of a news agency. For these country pairs whose coverage date is more endogenous, there seems to be no magnified effect in presence of both a news agency and a telegraphic line.

Table 2.2: Panel gravity estimates, separating direct and indirect connection

	(1) S_{odt}	(2) S_{odt}	(3) S_{odt}	(4) $\text{Ln}(Y_{odt})$	(5) $S_{odt} > 0$
News Ag. × Tel.	0.259^c	0.317^b	0.316^b	0.564^a	0.310^a
	[0.156]	[0.158]	[0.161]	[0.193]	[0.116]
News Ag. × Direct Tel.		-0.211^c	-0.196^c	-0.607^a	-0.302^a
		[0.111]	[0.114]	[0.148]	[0.091]
Headquarter News Ag. × Tel.		0.019	0.017	-0.186	0.126
		[0.123]	[0.123]	[0.161]	[0.114]
Headquarter News Ag. × Direct Tel.		-0.044	-0.046	0.300	0.204
		[0.159]	[0.160]	[0.193]	[0.137]
Telegraph	0.334^a	0.257^b	0.251^b	0.219	0.118
	[0.115]	[0.117]	[0.118]	[0.147]	[0.099]
Direct Tel.		-0.032	-0.035	0.463^a	0.176^b
		[0.104]	[0.106]	[0.161]	[0.087]
News Agency	0.222	0.153	0.168	0.291	0.203
	[0.183]	[0.187]	[0.188]	[0.213]	[0.141]
Headquarter News Ag.		0.005	0.002	-0.101	-0.218^b
		[0.122]	[0.123]	[0.169]	[0.111]
Observations	83373	83373	83373	59910	59910
Estimator	PPML	PPML	PPML	OLS	PPML
Sample	Complete	Complete	Complete	Complete	Complete
Colony controls	✓	✓	✗	✓	✓

Note: Data is aggregated at the country-pair × year level. The dependent variable is the share of imports of destination d coming from origin o in the remaining columns. All specifications include destination × year, origin × year and country-pair fixed effects. $DirectTel.$ indicates a direct telegraph connection between o and d, while $HeadquarterNewsAg.$ indicates that both countries are covered by the same news agency and one of the two is the headquarter of the news agency. In brackets are the standard errors, clustered by country-pair. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$.

In table 2.2, we present results for a variety of specifications. The setup is slightly different from the one used in figure 2.5, since effects can now directly be interpreted as additional marginal effects, while they were total marginal effects in the graph. For instance, the coefficient on the interaction term "News Agency × Telegraph" denotes the additional effect of being both indirectly covered by a news agency and indirectly connected by a telegraph. This formulation allows to test directly our hypothesis of interest, i.e. whether the positive shock on information corresponding to this scenario increases trade.

In column (1), we remind the results of our preferred baseline specification, corresponding to column (4) of table 2.1. Column (2) corresponds to exactly the same specification as the one plotted in figure 2.5. The interaction term purged from direct connections remains positive and significant, with a magnitude close to our baseline estimate (trade increases by an additional 37% in presence of both a news agency and a telegraph). This also confirms what we were able to visualize from the graph: no difference in trade effect when one of the countries is the headquarter of the news agencies, a slightly lower (but not significantly lower) effect when the telegraph link is direct, and a significantly lower additional trade effects of news agencies when the telegraphic link is direct, suggesting that these country pairs started from high trade levels and therefore benefited less from the increase in information available.

Column (3) confirms that the colonial ties controls do not play a large role in our estimation, since the estimates are barely affected by their omission. In column (4), we use an OLS estimator and find that the results are essentially unchanged, except on two accounts: the additional effect of news agencies appear larger, and direct telegraphic links now appear to have a positive and significant effect on trade.[18] In column (5), we switch back to PPML but restrict the sample to strictly positive trade flows: the additional trade effect of news agencies and telegraphs is now very close to our baseline estimate from column (2), suggesting that the high magnitude observed in column (4) had more to do with the use of an OLS estimator than with the omission of the extensive margin.

The results in table 2.2 are consistent with the picture obtained by running a series of cross-sectional gravity estimations, one for each year of the sample. We define a dummy variable identifying the countries that are covered by a news agency for the first time in the 1876 cartel agreement, and check year after year whether, in cross-section, these countries trade more together. Similarly, we estimate the cross-sectional trade effect for the headquarter dyads that entered the cartel agreement in 1876. The evolution over time of these cross sectional estimates is plotted in figure 2.6.[19]

On the right hand side graph, we see that the headquarter dyads tend to trade more together than what would be predicted by a standard gravity equation, even before the news agency coverage actually starts for them, i.e. before 1876. This suggests the existence of a privileged relationship that is not improved by the additional information, and even seems to fade away in the long-run. The pattern is very different for the "indirect" news agency connections (left hand side of the graph): the coefficients are never significant before the start of the news agency coverage, suggesting that there is no special relationship between these pairs of countries before 1876. After this date however, their trade relationship starts improving and a decade later they trade significantly more than the benchmark from the gravity equation.

4.2 Proxy trade policy with tariffs

Another approach to circumvent the issue of potential endogeneity of coverage by telegraph and news agencies is to have a look at the most easily quantifiable measure of trade policy: tariffs. TRADHIST provides information on the average customs duties for 8000 dyads-year. This measure is far from exhaustive, and has some well-known limitations.[20] Nevertheless it is likely to remain a

[18] This suggests that direct telegraph lines did not induce dyads with zero trade flows to trade, but rather intensified trade between dyads that already used to trade (in other words, direct telegraphic lines mostly affected the intensive margin). This hypothesis is comforted by the fact that, when we switch back to PPML but restrict the sample to strictly positive trade flows the coefficient on direct telegraphic lines again appears positive and significant (see column (5)).

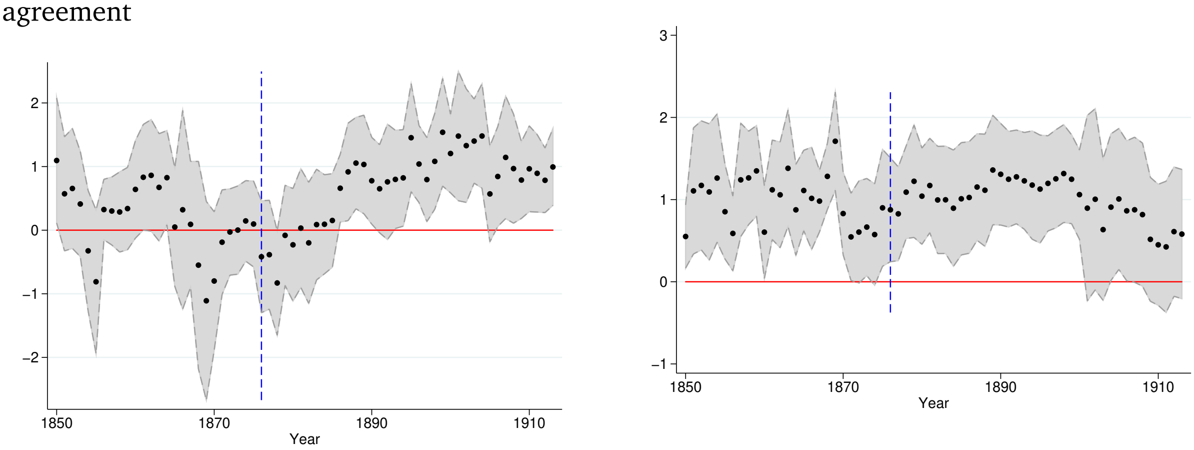
[19] For the curious reader, similar plots are available for the other variables included in our regressions, namely the colony control variables and the geographic distance (figure 2.17 in the appendix). Of particular interest is the fact that the so-called "distance puzzle" (i.e. the idea that the distance coefficient remains remarkably stable despite deep changes in the trade environment) is as vivid during the XIXth century as during the XXth century.

[20] If tariffs are very high on a given product, imports should decrease, therefore the weighted average of tariff is likely to be biased downwards.

Figure 2.6: Cross-sectional gravity estimates, 1876 cartel agreement

(a) Covered by a news agency in the 1876 cartel agreement

(b) Headquarter News Agency



Notes: Each dot corresponds to a PPML estimate from a cross-sectional gravity equation, with trade shares as dependent variable. Left hand side: dummy indicating that the dyad is covered first in the 1876 cartel agreement. Right hand side: dummy indicating that one of the country is the headquarter of the news agency, and that year of inclusion in the cartel agreements is 1876. Each regression includes origin and destination fixed effects, and controls for the log of distance.

good proxy for the bilateral openness between two countries.[21]

Ideally, we would include tariffs in the baseline regression as a robustness check. This is however not feasible given that we have tariffs only for a small subsamble of the dyads. The estimation would therefore be performed on a very different sample than in the baseline specification. However, if tariffs were correlated with our variables of interest in a way that may affect our estimates, then we would expect a within-dyad link between tariffs and news agency / telegraph coverage on the sample for which we have tariff data. To test this, we regress bilateral tariffs on our variables of interest and a set of dyadic and year fixed effects (table 2.3). We find no significant correlation between tariffs and our news agency / telegraph coverage dummies, which suggests that trade policy is not a major threat to identification.

4.3 Event study

The construction of the telegraph network and the extension of the global news agencies cartel were progressive over time, so that the first year in which the country is "treated" (i.e when both a telegraph link and a news agency coverage is available) differs across pairs of countries. There is a wide dispersion in treatment dates (plotted in figure 2.11, in the appendix). The fact that not all units in the panel receive treatment at the same time, and that some units are never treated allows estimating a dynamic model (event-study) to describe the evolution over time of the outcome before and after the treatment, yielding insights on the duration and the evolution of the treatment effect.

This is done by constructing a set of dummy variables, each corresponding to a certain number of years separating the dyad from its treatment date. Let K_{odt} denote the number of years to treatment

[21] We are especially reassured by the way this tariff measure correlates with our main variables of interest in a series of cross-sectional regressions (table 2.8, in the appendix). This confirms our priors, without threatening our identification strategy, which relies on time-variations.

Table 2.3: Correlation between bilateral tariff rates and our variables of interest.

	Average Tariff$_{odt}$			
	(1)	(2)	(3)	(4)
News Ag. × Tel.	-0.665	0.123	-1.567	-0.089
	[3.183]	[3.130]	[3.769]	[3.618]
News Ag. × Direct Tel.		-4.042		0.931
		[3.192]		[5.665]
Headquarter News Ag. × Tel.			4.528	5.672
			[5.223]	[5.232]
Headquarter News Ag. × Direct Tel.				-13.455^{b}
				[6.333]
Telegraph	-3.474	-4.341	-4.570	-6.517^{b}
	[2.662]	[2.650]	[2.777]	[2.766]
Direct Tel.		1.746		0.654
		[2.291]		[4.038]
News Agency	4.677	4.560	-0.171	-1.305
	[4.120]	[4.062]	[4.409]	[4.357]
Headquarter News Ag.			7.164	8.838
			[5.758]	[5.889]
Observations	4428	4428	4428	4428
Dyad FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Data is aggregated at the country pair × year level. The dependent variable is the average tariff rate. All specifications include country-pair and year fixed effects. In brackets are the standard errors, clustered by country-pair. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$.

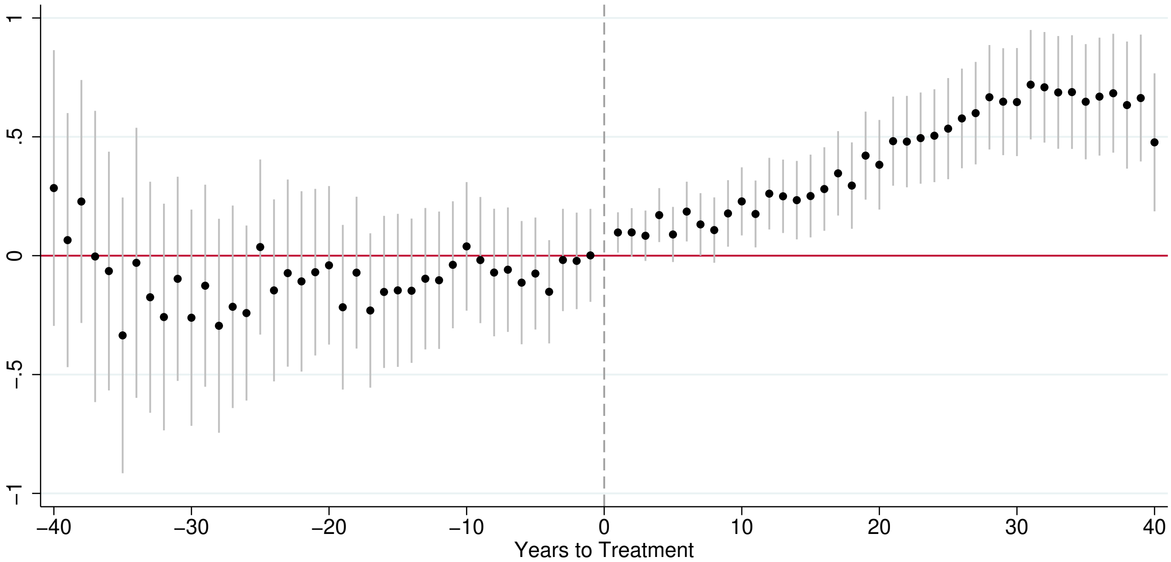
for dyad od at time t, so that, for instance, $K_{FRA,BRA,1870} = -6$, since the first year in which both France and Brazil are covered by a news agency and a telegraph is 1876. Therefore, for this observation, the dummy $\mathbb{1}\{K_{odt} = -6\}$ turns on, while all the other "years to treatment" dummy variables take value zero ($\mathbb{1}\{K_{odt} = k\} = 0 \ \forall k \neq -6$). Our aim is to estimate the marginal effects of each of these dummy variables, that we denote γ_k. If a dyad is never treated, as is the case for 77.84% of our sample, all the "years to treatment" dummy variables take value zero, i.e. $\mathbb{1}\{K_{odt} = k\} = 0 \ \forall k$

We regress the log of bilateral trade flows on dyadic and year fixed effects, the controls for colonial ties, the telegraph dummy and the news agency dummy, and on the above-described set of dummy variables indicating the number of years to treatment:

$$\ln(Y_{odt}) = \sum_{k=-40+}^{40+} \gamma_k \mathbb{1}\{K_{odt} = k\} + \mathrm{FE}_{od} + \mathrm{FE}_t + \boldsymbol{\beta}'\mathbf{X}_{odt} + \varepsilon_{odt} \tag{2.2}$$

where $k = 40+$ means "40 years or more after the treatment".

Figure 2.7: Evolution of bilateral trade before and after news agency and telegraph coverage (event-study)



Note: Each point on the graph corresponds to the coefficient on a dummy variable taking value 1 if the number of years to treatment is k (k negative before treatment and positive after). The treatment date is the first year in which both countries of the dyad are covered by a news agency and connected by the telegraph. The complete specification is provided in eq. (2.2)

Figure 2.7 shows how the treatment effect evolves before and after the treatment date. Before the treatment (left hand-side of the graph, with "years to treatment < 0"), we see that there is no particular trend, i.e. that pre-treatment trade never significantly differs from its treatment date level. In other words, dyads that are going to be connected by a news agency and a telegraph do not seem to be on an increasing or decreasing time trend before actually being connected. The picture is very different after the treatment (right hand-side of the graph): the dyads immediately start trading more, the increase is steady up to 30 years after the treatment, at which point the trade effect stabilizes at a rather high level.

The magnitude of the effect is comparable to the one found in our baseline estimations. The fact that it is long lasting is consistent with the permanent nature of the treatment: once a dyad is connected by a telegraph and a news agency, it never switches back to non coverage, so the additional flow of information never stops. Moreover, the long-run persistence of the effects of a shock on trade costs has been documented in other contexts (**X2021**). The slow increase suggests that, on top of direct knowledge of the foreign market conditions, our information shock affected trade through long-run channels, potentially an increase in Foreign Direct Investments, international migrations, or even a convergence in cultural tastes.

5 Testing the information channel

The previous section identifies a positive and significant impact of the telegraph and news agency coverage on trade. We attribute this effect to an improved access to information. However, our variables of interest are only proxies for the increase in the quantity and quality of information. To rule out other channels through which they may have affected trade, we test for the presence of effects that that are more specific to the information channel.

In the first sub-section, we document an increase in the volatility of trade flows after coverage by a news agency and a telegraph, consistent with a better ability of traders to adapt to varying market conditions. In a second sub-section, we provide quantitative evidence of an increase in the coverage of foreign countries in the press when they become connected by a telegraph and a news agency. This shows that the delivered information is indeed more abundant between the countries after they benefit from the positive shock we rely on.

5.1 Trade volatility

In this sub-section, we test a prediction of the Steinwender (2018) model linking trade and information: a reduction in information frictions increases trade variance. Indeed, exports and imports react to changes in expected demand from other countries. In the extreme case without any information, exporters ship every year the same amount, corresponding to the expected demand. With more information, exporters can respond faster and better to demand fluctuations. With perfect information the variance of exports should be equal to the variance of demand as the exporting country can perfectly adapt to the local demand.

We assess whether trade flows between two countries become more volatile after these two countries are covered by a news agency. To this end, we use several measures of volatility. The first one (column (1) in table 2.4 is the usual sample standard deviation. We compute for each dyad the standard deviation of trade flows before and after the dyad is covered by a news agency. We then regress this sample standard deviation on a dyadic fixed effect and a dummy taking value 1 if both countries are covered by a news agency:

$$\left(\widehat{Var_p}(Y_{od})\right)^{1/2} = \mathrm{FE}_{od} + \beta_1 Tel_{odp} + \beta_2 NA_{odp} + \beta_3 NA_{odp} \times Tel_{odp} + \varepsilon_{odp}$$

The second measure is the absolute value of the deviation of the trade flow at time t to the mean trade flow in each period (before and after news agency coverage): $|Y_{odt} - \bar{Y}_{odp(t)}|$. The advantage of this measure is that it allows having several observations within each dyad for the two time periods (before and after being covered by a news agency). Its drawback is that it can be very noisy, especially with historical datasets that may contain more errors than recent ones. We regress this absolute deviation on destination-year, origin-year and dyadic fixed effects, and a dummy taking value 1 if both countries are covered by a news agency:

$$|Y_{odt} - \bar{Y}_{odp(t)}| = \mathrm{FE}_{od} + \mathrm{FE}_{ot} + \mathrm{FE}_{dt} + \beta_1 Tel_{odt} + \beta_2 NA_{odt} + \beta_3 NA_{odt} \times Tel_{odt} + \varepsilon_{odt}$$

The level of our volatility measure does not have any meaningful interpretation, we are mostly interested in its change over time. For all possible versions of the outcome variable, we observe a significant and positive impact of the news agency coverage on the volatility of trade flows. This result is consistent with the idea that more information is available thanks to the news agencies, which allows trade partners to adapt to the demand shocks.

Table 2.4: Effect of news agency and telegraphs on trade volatility.

	(1)	(2)	(3)	(4)	(5)	(6)
	Standard Deviation	**Absolute Deviation**	**Absolute Deviation**	**Absolute Deviation**	**Log of Absolute Deviation**	**Log of Absolute Deviation**
$NA_{odt} \times Tel_{odt}$	742,259^{a}	543,376^{a}	600,525^{a}	702,133^{a}	0.672^{a}	0.700^{a}
	[94,746]	[125,333]	[130,750]	[199,400]	[0.140]	[0.122]
Tel_{odt}		183,238^{a}	156,144^{a}	306,314^{a}	0.872^{a}	0.726^{a}
		[43,094]	[39,933]	[81,863]	[0.0881]	[0.0755]
NA_{odt}		-130,861	-264,732^{b}	-381,128^{b}	0.0273	0.0113
		[102,553]	[105,574]	[162,226]	[0.124]	[0.110]
Observations	1194	117993	117993	62224	79715	62065
Sample	Complete	Complete	Complete	Flow>0	Complete	Flow>0

Note : All specifications include dyadic fixed effects. All specifications year fixed effects, except columns (1) and (2). Standard errors are clustered at the dyad level. In columns (4) and (6), the zero trade flows are omitted. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

5.2 Text analysis

In this section, we provide evidence of an increase in the coverage of foreign countries in French newspapers when these countries benefit from a telegraph connection with France or are included in the news agencies' syndication agreement. This analysis relies on a corpus of articles from the 16 main French newspapers, processed during the Europeana Newspapers project.[22]

For each newspaper and country, we count on a yearly basis the number of days in which the country appears at least once in the newspaper's articles. We define an occurrence as the presence in the text of either the country name or the capital name. We then create five distinct groups of countries, based on their date of accession to the news agencies' syndication agreement (1859, 1867, 1876, 1889 or 1902), and estimate separately in each of these groups the effect of news agencies and telegraph on the number of days of presence:

$$NbDaysPresence_{nct} = \exp(\mathrm{FE}_{nt} + \mathrm{FE}_{nc} + \sum_{group} [\beta_{1,group}\left(Tel_{ct} \times \mathbb{1}_{c\in group}\right) + \beta_{2,group}\left(NA_{ct} \times Tel_{ct} \times \mathbb{1}_{c\in group}\right)] + \varepsilon_{nct})$$

where c indexes the country, $group$ the group of countries to which the country belongs, based on its accession date to the syndication agreement, n the newspaper and t the year. The specification includes newspaper × year fixed effects that account for the fact that the total article lengths may vary differently over time depending on the newspaper, and newspaper × country, which means that we rely on variations over time of the country coverage within the newspaper. To summarize, we estimate whether the space devoted to a country within each newspaper increases when this country's telegraph and news agency status switches from "not included" to "included", allowing

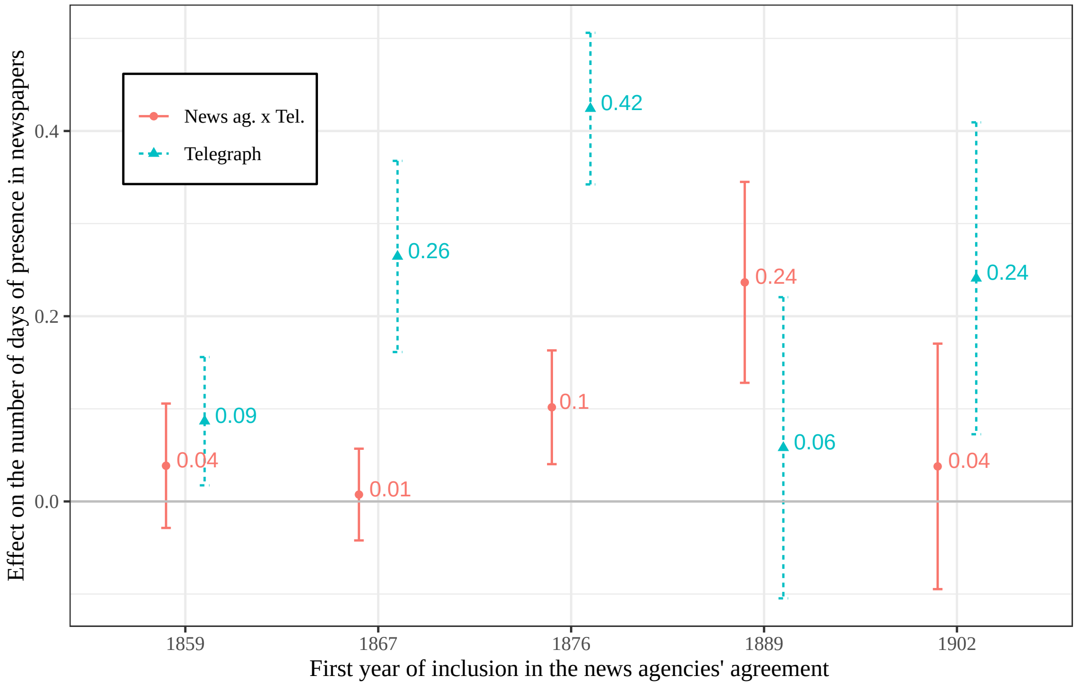
[22]More precisely, the dataset contains optical character recognition (OCR)'s transcription of the text of about 2 million pages, from L'Action française, Le Constitutionnel, La Croix, L'Echo de Paris, Le Figaro, Le Gaulois, L'Humanité, L'Intransigeant, Le Journal des débats politiques et littéraires, Le Matin, Le Petit Journal, Le Petit Parisien, La Presse, Le Siècle, Le Temps, L'Univers.

for heterogeneous effects depending on the wave of news agencies coverage extension to which the country belongs. Note that by construction the coefficient on news agencies alone is not identifiable since within a group all countries have the same accession date, so that the time fixed effects fully capture the news agencies effect.

An important limitation to this exercise is that the sample is entirely made of French newspapers, which only allows us to estimate the effect by relying on on the different coverage dates with respect to France, instead of the more diverse set of bilateral coverage dates used to obtain the estimates on trade flows in the two previous sections. In other words, the sample is much more limited in terms of geographical scope, which limits the sources of variations we can use to identify the potential effect.

Our results are plotted in figure 2.8. We find that the extensions that triggered the largest effect are the 1876 and 1889 agreements. This is probably due to the fact that these extensions included countries on which France likely had fewer information. On the contrary, earlier agreements covered either direct neighbors of France, or other European countries, for which it is likely that French newspapers already had more information. Moreover, we have a lower number of newspapers in our sample at the start of the period, leading to noisier estimates. This may explain why, even though they are always positive, the estimates are very low for the two first extensions, and not significant. Interestingly, the telegraph effect is always positive and significant in all but one group of countries. This implies that the telegraph was also a key determinant of the amount of public information available, but as argued in previous sections it also had a massive effect on private communication costs and can therefore not be used to isolate the pure effect of information on trade.

Figure 2.8: Effect of telegraphs and news agencies on the country coverage in French newspapers, by year of entry into the news agencies's syndication agreement.



Notes: Notes: PPML estimates, with newspaper × year and newspaper × country fixed effects. Bars indicates the 95% confidence interval, with standard errors clustered at the country × year level. We distinguish groups of countries based on their date of accession to the news agencies' syndication agreement, and allow for heterogeneous effects between those groups.

6 Conclusion

We use the joint expansion of the telegraph and the news agencies to disentangle the pure information effect from the effect of reduced communication costs. The positive trade effect of linking two countries through the telegraphic network is magnified when these two countries are also both part of a news agency syndication agreement facilitating the exchange of information. We estimate that the decrease in coordination costs allowed by the telegraph raised the value of trade by 40%, while the increase in the flow of information associated with the coverage by one of the global news agencies resulted in an additional 30% growth of trade.

The positive and significant effect of the telegraph and news agencies interaction subsists when focusing on indirect connections, which are less likely to suffer from the potential endogeneity bias linked to expectations of the operators. Additionally, we document patterns consistent with an increase in the flow of information: the variance of trade increases after the connection by a telegraph and a news agency, suggesting an improvement in the ability to adapt to demand shocks, and the presence of a foreign country in French newspapers increases after this country is linked to France by a telegraph and a news agency.

While estimated from a historical event, the results are relevant to understand contemporary trade flows, since exporters still may lack the necessary information, despite considerable improvement in communication technologies.

This paper does not take a stance on the precise mechanism through which better access to information on foreign countries affects trade. On top of the improved knowledge of foreign market conditions, news agencies and telegraphs may have affected other outcomes such as Foreign Direct Investment, human migration flows or even cultural tastes. We leave for future research to delve into these potential channels.

Bibliography

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt (2015). "Lessons from Schumpeterian growth theory." *American Economic Review* 105.5, pp. 94–99.

Aghion, Philippe and Xavier Jaravel (2015). "Knowledge Spillovers, Innovation and Growth." *Economic Journal* 583, pp. 533–573.

Agrawal, Ajay, Devesh Kapur, and John McHale (2008). "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics* 64.2, pp. 258–269.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi (2018). *Dancing with the stars: Innovation through interactions*. Tech. rep. National Bureau of Economic Research.

Akers, Noël J (2000). "The referencing of prior art documents in European patents and applications." *World Patent Information* 22.4, pp. 309–315.

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

Alcácer, Juan and Michelle Gittelman (2006). "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *Review of Economics and Statistics* 88.4, pp. 774–779.

Allen, Treb (2014). "Information Frictions in Trade." *Econometrica* 82.6, pp. 2041–2083.

Almeida, Paul and Bruce Kogut (1997). "The Exploration of Technological Diversity and the Geographic Localization of Innovation." *Small Business Economics* 9.1, pp. 21–31.

Almeida, Paul and Bruce Kogut (1999). "Localization of knowledge and the mobility of engineers in regional networks." *Management science* 45.7, pp. 905–917.

Axtell, Robert L (2001). "Zipf distribution of US firm sizes." *Science* 293.5536, p. 1818.

Bakker, Gerben (2014). "How they made news pay: news traders' quest for crisis-resistant business models."

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito (2019). "Production Networks, Geography, and Firm Performance." *Journal of Political Economy* 127.2, pp. 639–688.

Bielsa, Esperança (2008). "The pivotal role of news agencies in the context of globalization: a historical approach." *Global Networks* 8.3, pp. 347–366.

Bilir, L. Kamran and Eduardo Morales (2020). “Innovation in the Global Firm.” *Journal of Political Economy* 128.4, pp. 1566–1625.

Blake, Eric S., Todd B. Kimberlain, Robert J. Berg, John P. Cangialosi, and John L. Beven (2013). *Tropical Cyclone Report Hurricane Sandy*. National Hurricane Center.

Bloom, Nicholas, Mark Schankerman, and John van Reenen (2013). “Identifying Technology Spillovers and Product Market Rivalry.” *Econometrica* 81.4, pp. 1347–1393.

Breschi, Stefano and Francesco Lissoni (2009). “Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows.” *Journal of Economic Geography*, lbp008.

Buera, Francisco J. and Ezra Oberfield (2020). “The Global Diffusion of Ideas.” *Econometrica* 88.1, pp. 83–114.

Carayol, Nicolas, Laurent Bergé, Lorenzo Cassi, and Pascale Roux (2019). “Unintended triadic closure in social networks: The strategic formation of research collaborations between French inventors.” *Journal of Economic Behavior & Organization* 163.C, pp. 218–238.

Carvalho, Vasco M and Nico Voigtländer (2014). *Input diffusion and the evolution of production networks*. Tech. rep. National Bureau of Economic Research.

Catalini, Christian, Christian Fons-Rosen, and Patrick Gaulé (2018). *How Do Travel Costs Shape Collaboration?* Working Paper 24780. National Bureau of Economic Research.

Chaney, Thomas (2014). “The network structure of international trade.” *American Economic Review* 104.11, pp. 3600–3634.

Chaney, Thomas (2018a). “The Gravity Equation in International Trade: An Explanation.” *Journal of Political Economy* 126.1, pp. 150–177.

Chaney, Thomas (2018b). “The Gravity Equation in International Trade: An Explanation.” *Journal of Political Economy* 126.1, pp. 150–177.

Combes, Pierre-Philippe and Miren Lafourcade (2005). “Transport costs: measures, determinants, and regional policy implications for France.” *Journal of Economic Geography* 5.3, pp. 319–349.

Comin, Diego, Mikhail Dmitriev, and Esteban Rossi-Hansberg (2012). *The Spatial Diffusion of Technology*. NBER Working Papers 18534. National Bureau of Economic Research, Inc.

Comin, Diego and Bart Hobijn (2010). “An Exploration of Technology Diffusion.” *American Economic Review* 100.5, pp. 2031–59.

Dasgupta, Kunal and Jordi Mondria (2018). “Inattentive importers.” *Journal of International Economics* 112, pp. 150–165.

Dickstein, Michael J and Eduardo Morales (2018). “What do exporters know?” *The Quarterly Journal of Economics* 133.4, pp. 1753–1801.

Duguet, Emmanuel and Megan MacGarvie (2005). “How well do patent citations measure flows of technology? Evidence from French innovation surveys.” *Economics of Innovation and New Technology* 14.5, pp. 375–393.

Duranton, Gilles, Peter Morrow, and Matthew Turner (2014). "Roads and Trade: Evidence from the US." *The Review of Economic Studies* 81.2, pp. 681–724.

Ejrnæs, Mette and Karl Gunnar Persson (2010). "The gains from improved market efficiency: trade before and after the transatlantic telegraph." *European Review of Economic History* 14.3, pp. 361–381.

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

Estevadeordal, Antoni, Brian Frantz, and Alan M Taylor (2003). "The rise and fall of world trade, 1870–1939." *The Quarterly Journal of Economics* 118.2, pp. 359–407.

Feyrer, James (2011). *Distance, Trade, and Income – The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment*. 2011 Meeting Papers 1438. Society for Economic Dynamics.

Fouquin, Michel and Jules Hugot (2016). *Two Centuries of Bilateral Trade and Gravity Data: 1827-2014*. Working Papers 2016-14. CEPII.

Gabaix, Xavier (2016). "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30.1, pp. 185–206.

Ghiglino, Christian (2012). "Random walk to innovation: Why productivity follows a power law." *Journal of Economic Theory* 147.2, pp. 713–737.

Gibrat, Robert (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.

Gourieroux, C., A. Monfort, and E. Renault (1993). "Indirect inference." *Journal of Applied Econometrics* 8.S1, S85–S118.

Griffith, Rachel, Sokbae Lee, and John Van Reenen (2011). "Is distance dying at last? Falling home bias in fixed-effects models of patent citations." *Quantitative economics* 2.2, pp. 211–249.

Head, Keith, Yao Amber Li, and Asier Minondo (2019). "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics." *Review of Economics and Statistics* 101.4, pp. 713–727.

Head, Keith and Thierry Mayer (2013). "What separates us? Sources of resistance to globalization." *Canadian Journal of Economics* 46.4, pp. 1196–1231.

Head, Keith and Thierry Mayer (2014a). "Gravity Equations: Workhorse, Toolkit, and Cookbook." *the Handbook of International Economics*. Vol. 4. Elsevier. Chap. 3, pp. 131–195.

Head, Keith and Thierry Mayer (2014b). "Gravity Equations: Workhorse,Toolkit, and Cookbook." *Handbook of International Economics*. Vol. 4. Elsevier. Chap. Chapter 3, pp. 131–195.

Hortaçsu, Ali, F. Asís Martínez-Jerez, and Jason Douglas (2009). "The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre." *American Economic Journal: Microeconomics* 1.1, pp. 53–74.

Hugot, Jules and Camilo Umana Dajud (2016). “Trade Costs and the Suez and Panama Canals.” *CEPII Working Paper* 29.

Hummels, David and Georg Schaur (2013). “Time as a Trade Barrier.” *American Economic Review* 103.7, pp. 2935–59.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger (2018). “Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science.” *Quarterly Journal of Economics* 133.2, pp. 927–991.

Jackson, Matthew and Brian Rogers (2007). “Meeting Strangers and Friends of Friends: How Random Are Social Networks?” *American Economic Review* 97.3, pp. 890–915.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty (2000). “Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors.” *American Economic Review, Papers and Proceedings* 90.2, pp. 215–218.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993). “Geographic localization of knowledge spillovers as evidenced by patent citations.” *Quarterly Journal of Economics*, pp. 577–598.

Jensen, Robert (2007). “The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector.” *The Quarterly Journal of Economics* 122.3, pp. 879–924.

Juhász, Réka and Claudia Steinwender (2018). “Spinning the web: Codifiability, information frictions and trade.” *NBER working paper*.

Keller, Wolfgang and Stephen Ross Yeaple (2013). “The Gravity of Knowledge.” *American Economic Review* 103.4, pp. 1414–44.

Kerr, William R (2008). “Ethnic scientific communities and international technology diffusion.” *Review of Economics and Statistics* 90.3, pp. 518–537.

Lampe, Ryan (2012). “Strategic citation.” *Review of Economics and Statistics* 94.1, pp. 320–333.

Lanjouw, Jean, Ariel Pakes, and Jonathan Putnam (1998). “How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data.” *Journal of Industrial Economics* 46.4, pp. 405–32.

Lei, Zhen and Brian D. Wright (2017). “Why weak patents? Testing the examiner ignorance hypothesis.” *Journal of Public Economics* 148, pp. 43–56.

Lenoir, Clemence, Julien Martin, and Isabelle Mejean (2020). “Search Frictions in International Good Markets.”

Lew, Byron and Bruce Cater (2006). “The telegraph, coordination of tramp shipping, and growth in world trade, 1870-1910.” *European Review of Economic History* 10.2, pp. 147–173.

Li, Yao Amber (2014). “Borders and distance in knowledge spillovers: Dying over time or dying with age?—Evidence from patent citations.” *European Economic Review* 71, pp. 152–172.

Malgouyres, Clément, Thierry Mayer, Clément Mazet-Sonilhac, and Juan Carluccio (2020). "Technology-induced Trade Shocks? Evidence from Broadband Expansion in France." *mimeo*.

Maurseth, Per Botolf and Bart Verspagen (2002). "Knowledge Spillovers in Europe: A Patent Citations Analysis." *Scandinavian Journal of Economics* 104.4, pp. 531–45.

Mayer, Thierry and Soledad Zignago (2011). *Notes on CEPII's distances measures: The GeoDist database*. Working Papers. CEPII research center.

McFadden, Daniel (1973). "Conditional logit analysis of qualitative choice behavior." *Frontiers in Econometrics*. Ed. by P. Zarembka. Academic Press: New York. Chap. 4, pp. 105–142.

Mildenhall, Stephen, Steve Jakubowski, Siamak Daneshvaran, and Steve Bowen (2013). *Hurricane Sandy Event Recap Report*. AON Benfield.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura (2014). "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *Review of Economics and Statistics* 96.5, pp. 967–985.

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

Pascali, Luigi (2017). "The wind of change: Maritime technology, trade, and economic development." *American Economic Review* 107.9, pp. 2821–54.

Peri, Giovanni (2005). "Determinants of knowledge flows and their effect on innovation." *Review of Economics and Statistics* 87.2, pp. 308–322.

Serafinelli, Michel (2019). ""Good" Firms, Worker Flows, and Local Productivity." *Journal of Labor Economics* 37.3, pp. 747–792.

Silva, J. M. C. Santos and Silvana Tenreyro (2006). "The Log of Gravity." *Review of Economics and Statistics* 88.4, pp. 641–658.

Singh, Jasjit (2005). "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51.5, pp. 756–770.

Sotelo, Sebastian (2017). "Practical Aspects of Implementing the Multinomial PML Estimator."

Steinwender, Claudia (2018). "Real Effects of Information Frictions: When the States and the Kingdom Became United." *American Economic Review* 108.3, pp. 657–96.

Thompson, Peter (2006). "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations." *Review of Economics and Statistics* 88.2, pp. 383–388.

Thompson, Peter and Melanie Fox-Kean (2005). "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review* 95.1, pp. 450–460.

Unesco (1952). *News Agencies, Their Structure and Operation*. Greenwood Press.

Volpe Martincus, Christian and Juan Blyde (2013). “Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment.” *Journal of International Economics* 90.1, pp. 148–161.

Wenzlhuemer, Roland (2013). *Connecting the Nineteenth-Century World*. Cambridge University Press.

Wolff, Jacques (1991). “Structure, fonctionnement et évolution du marché international des nouvelles. Les agences de presse de 1835 à 1934.” *Revue Economique* 42.3, pp. 575–601.

Wrona, Jens (2018). “Border effects without borders: What divides Japan’s internal trade?” *International Economic Review* 59.3, pp. 1209–1262.

A Data

A.1 Construction of the database

News agency coverage:

Table 2.5: Countries added by each cartel agreement (source: Wolff (1991))

Agreement	Havas	Reuters	Wolff	AP	CorrBureau
1859	France Spain Italy Ottoman Emp.	Great Britain Ireland	Germany Russia Denmark Netherlands Sweden-Norway Finland Iceland	USA	
1867	Belgium Ottoman Emp. Egypt Portugal	Belgium Ottoman Emp. Egypt Netherlands			Austria-Hungary
1876	South America Switzerland Indochina	Far-East Australia New-Zealand Switzerland			
1889	Greece Bulgaria Romania Ottoman Emp.	Greece			Serbia Bulgaria Romania Ottoman Emp.
1902				Central America Puerto-Rico Philippines Cuba Hawaii	

Submarine telegraph cables:

Data shared by Roland Wenzlhuemer, built from the "Nomenclature des cables formant le réseau sous-marin du globe dressée d'après des documents officiels par le Bureau international des administrations télégraphiques", published in:

1. *Journal télégraphique* **3**, no 12 (1875)
2. *Journal télégraphique* **3**, no 29 (1877)
3. *Journal télégraphique* **7**, no 5 (1883)
4. *Journal télégraphique* **11**, no 4 (1887)
5. *Journal télégraphique* **13**, no 9 (1889)
6. *Journal télégraphique* **16**, no 4 (1892)
7. *Journal télégraphique* **18**, no 10 (1894)
8. *Journal télégraphique* **21**, no 11 (1897)

9. *Journal télégraphique* **25** (1901)

10. *Journal télégraphique* **27** (1903)

Terrestrial telegraph cables:

We constructed the database using a set of historical telegraph network maps whose references are provided below.

1. The Electric & International Telegraph Company's Map of the Telegraph Lines of Europe Published under the Authority of the Electric Telegraph Company by Day & Son, Lithographers to the Queen, 1856

2. Carte générale des grandes communications télégraphiques dans le Monde, dressée d'après des documents officiels par le Bureau international des administrations télégraphiques. C. v. Hoven, Imprimerie Lips (Berne), 1875

3. Carte générale des grandes communications télégraphiques dans le Monde, dressée d'après des documents officiels par le Bureau international des administrations télégraphiques. C. v. Hoven, (Berne), 1881

4. Carte des communications télégraphiques internationales et du régime extra-européen par le Bureau International des Administrations télégraphiques ; dressée et gravée par C.v. Hoven, (Berne), 1888

5. Carte des communications télégraphiques internationales et du régime extra-européen par le Bureau International des Administrations télégraphiques ; dressée et gravée par C.v. Hoven, (Berne), 1892

6. Carte des communications télégraphiques du régime européen dressée d'après des documents officiels par le Bureau international des administrations télégraphiques ; dessinée et gravée par C. v. Hover, (Berne), 1898

7. Carte générale des grandes communications télégraphiques dans le Monde, dressée d'après des documents officiels par le Bureau international des administrations télégraphiques. C. v. Hoven, (Berne), 1901

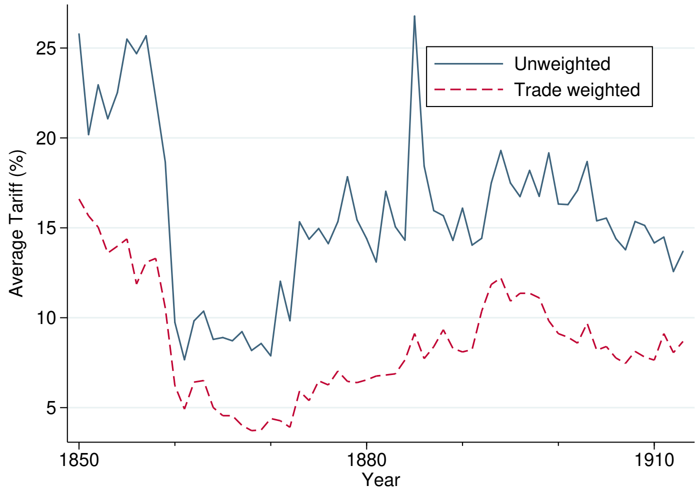
8. Carte générale des grandes communications télégraphiques dans le Monde, dressée d'après des documents officiels par le Bureau international des administrations télégraphiques. C. v. Hoven, (Berne), 1912

Figure 2.9: Map of the telegraph lines in 1875

Source: gallica.bnf.fr / BnF

A.2 Descriptive Statistics

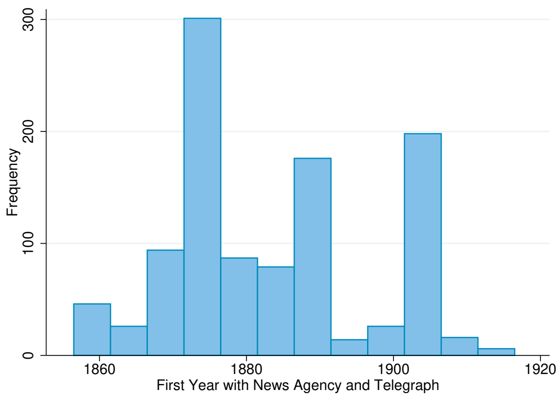
Figure 2.10: Average tariff rate on French imports



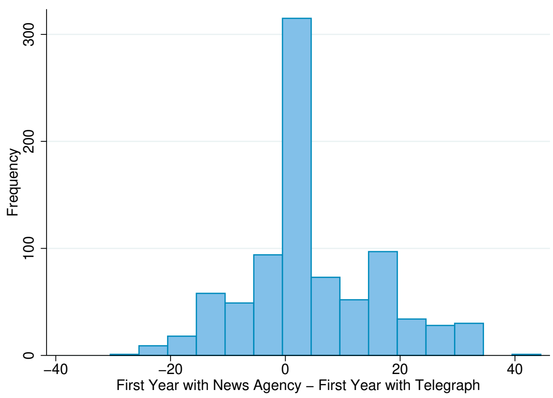
Notes: To compute the "unweighted" tariff, each origin is given the same weight, while for the "trade weighted" tariff, each origin is weighted by its trade flow with France.

Figure 2.11: Distribution of the "treatment dates" (event-study)

(a) Distribution of the first years in which dyads become connected by a telegraph and covered by a news agency



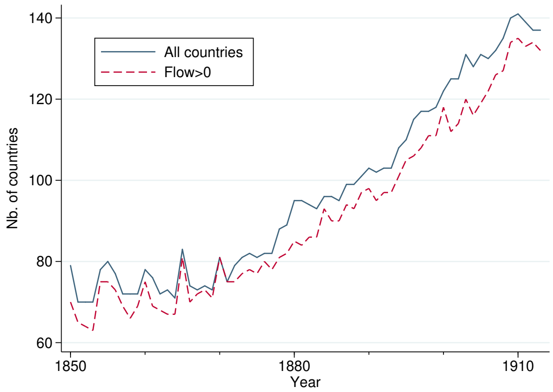
(b) Distribution of the difference between the first year of news agency coverage and the first year telegraph connection



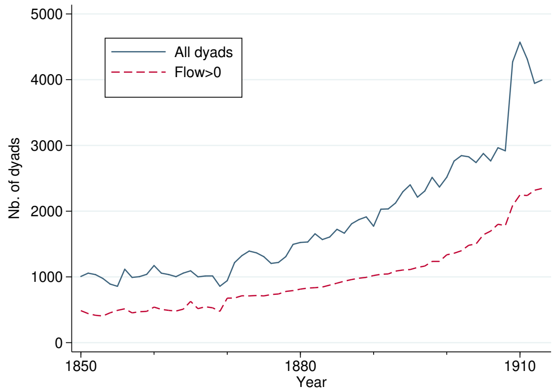
Notes: Among the dyads that end up being covered by a global news agencies and connected by a telegraph (the "treatment group"), we plot the distribution of the treatment year (a) and the distribution of the difference between the first year of news agency coverage and the first year telegraph connection (b).

Figure 2.12: Evolution over time of the number of exporters and dyads, of the average trade flow per dyad and country, and of total trade

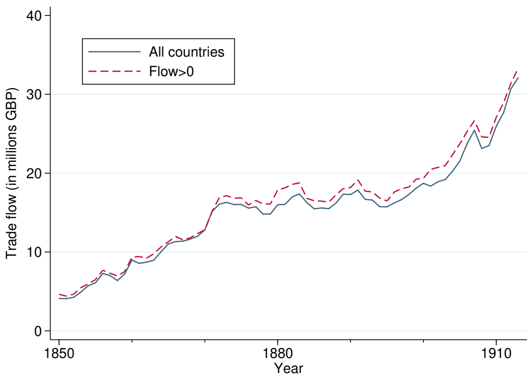
(a) Number of exporters



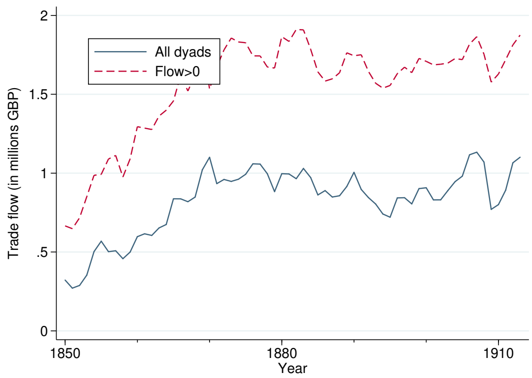
(b) Number of dyads



(c) Average trade flow per exporter (in millions GBP)



(d) Average trade flow per dyad (in millions GBP)



(e) Total trade (in billions GBP)

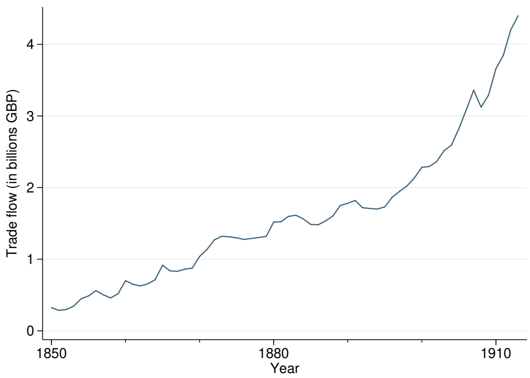
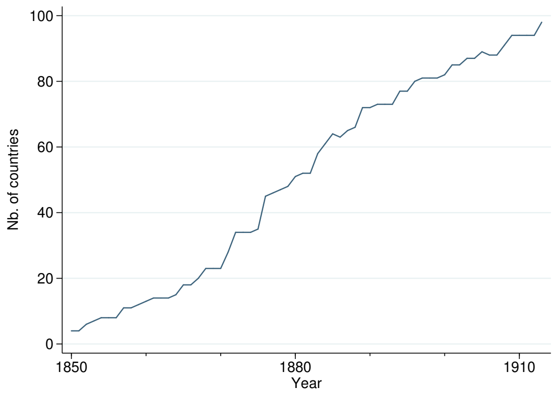
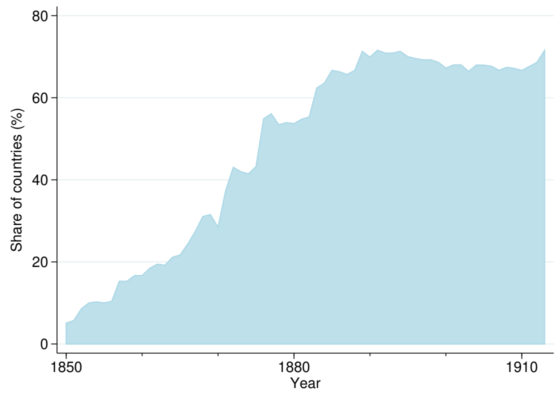


Figure 2.13: Evolution over time of the telegraph coverage

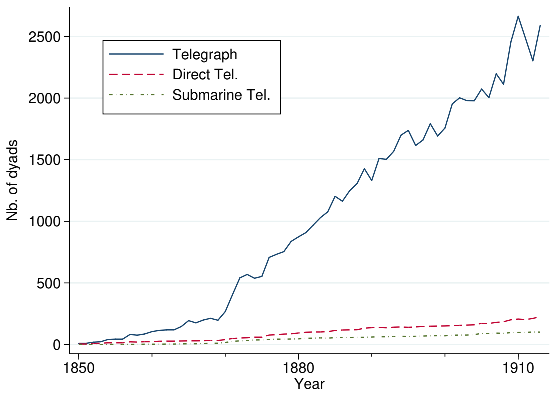
(a) Number of countries



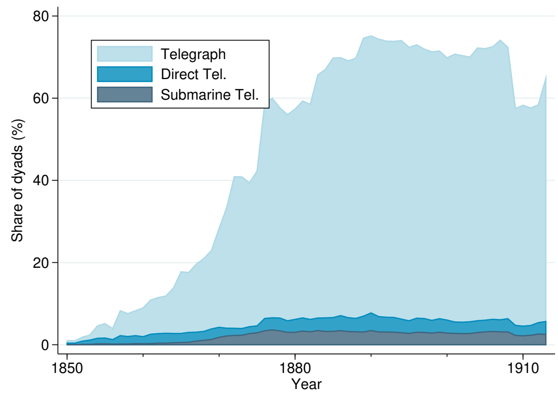
(b) Share of countries (in %)



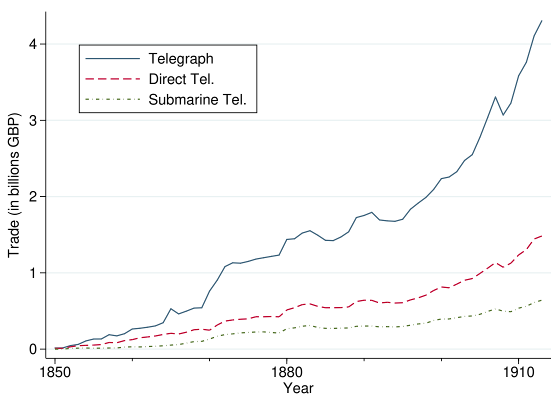
(c) Number of dyads



(d) Share of dyads (in %)



(e) Trade (in billions GBP)



(f) Share of trade (in %)

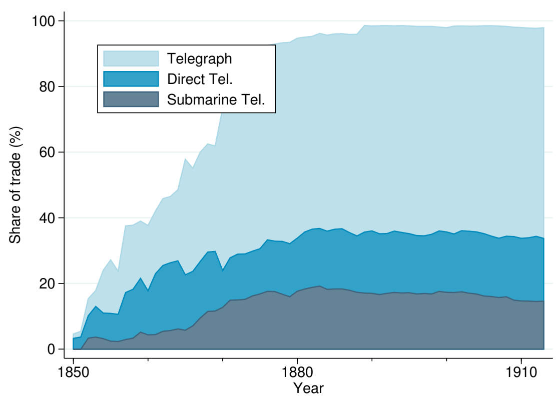
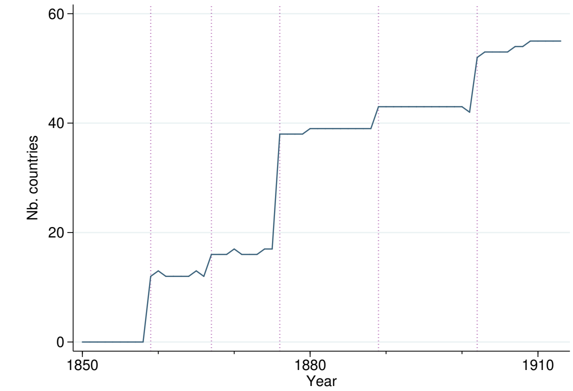
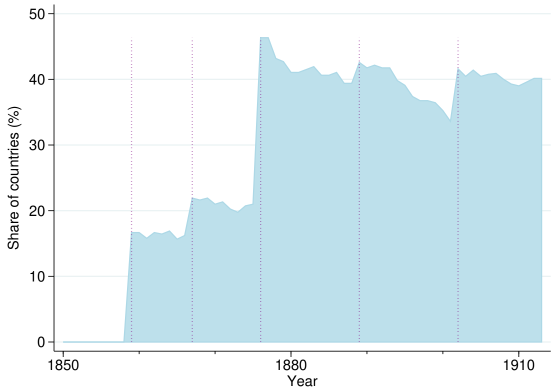


Figure 2.14: Evolution over time of news agency coverage

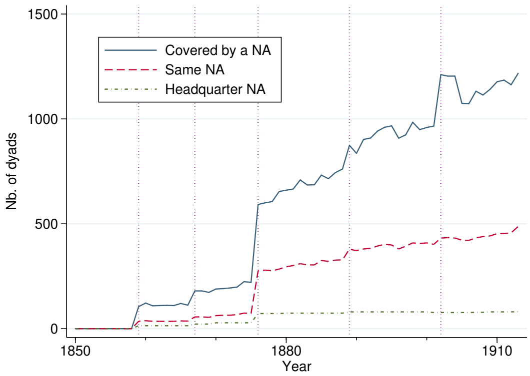
(a) Number of countries



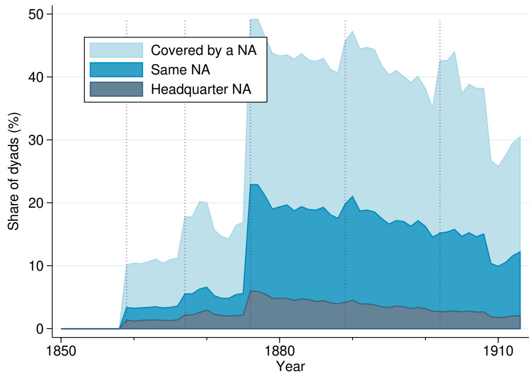
(b) Share of countries (in %)



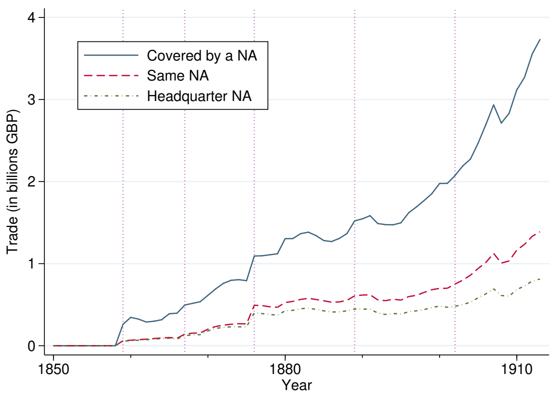
(c) Number of dyads



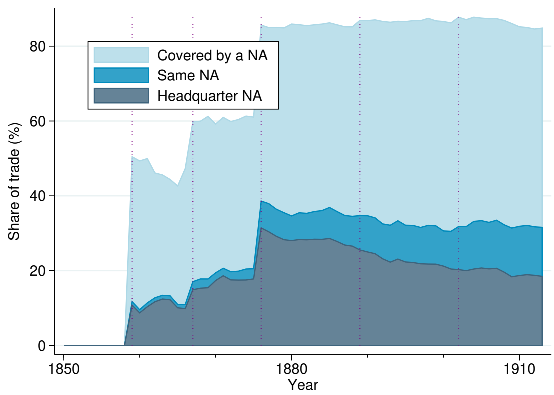
(d) Share of dyads (in %)



(e) Trade (in billions GBP)



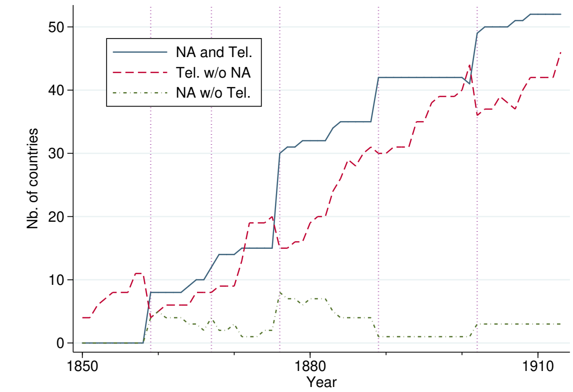
(f) Share of trade (in %)



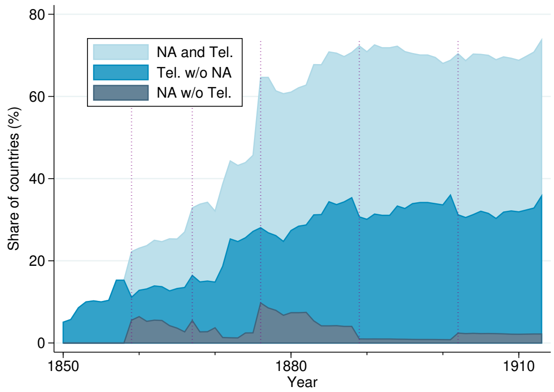
Note : Each vertical red line corresponds to an agreement extending the geographic coverage of the cartel. Only the countries (fig (a) and (b)) or the dyads (fig (c) and (d)) with non zero trade flows are counted.

Figure 2.15: Evolution over time of telegraph and news agency coverage

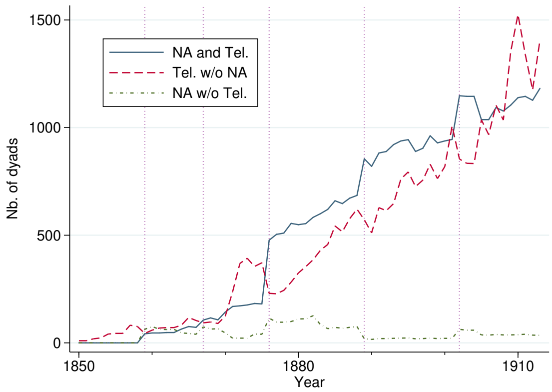
(a) Number of countries



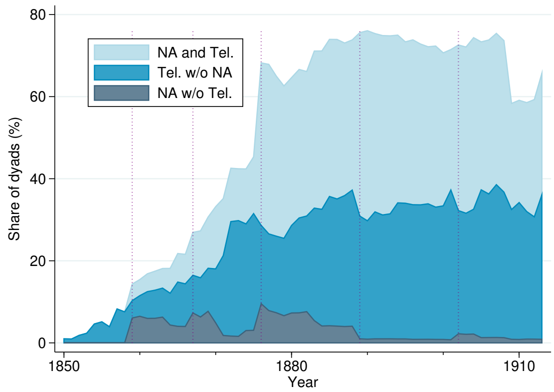
(b) Share of countries (in %)



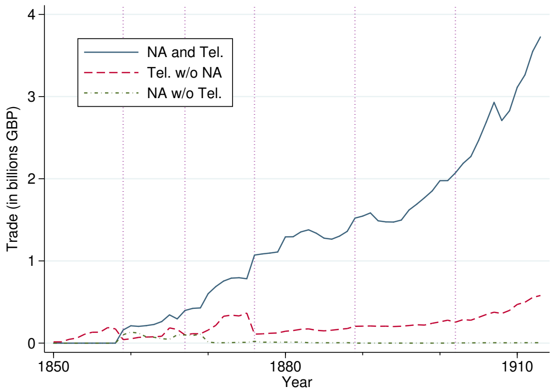
(c) Number of dyads



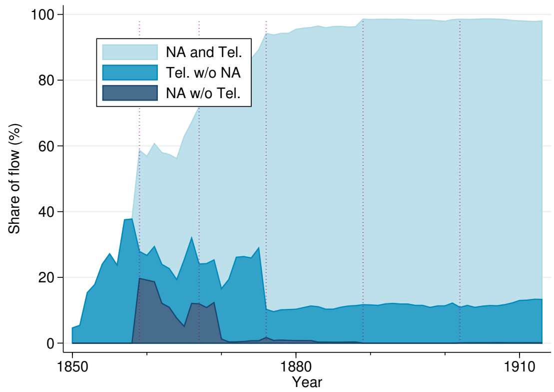
(d) Share of dyads (in %)



(e) Trade (in billions GBP)



(f) Share of trade (in %)



Note : "NA and Tel" refers to the observations that have a telegraph link and are covered by a global news agency. Each vertical red line corresponds to an agreement extending the geographic coverage of the cartel.

B Additional results

B.1 Panel Estimates

Table 2.6: Effect of news agencies and telegraphs on trade flows, introducing each variable of interest separately.

	(1)	(2)	(3)	(4)
	S_{odt}	S_{odt}	S_{odt}	S_{odt}
News Ag. × Tel.	0.556^a			0.259^c
	[0.120]			[0.156]
Telegraph		0.448^a		0.334^a
		[0.112]		[0.115]
News Agency			0.493^a	0.222
			[0.141]	[0.183]
Observations	83373	83373	83373	83373
Estimator	PPML	PPML	PPML	PPML
Sample	Complete	Complete	Complete	Complete
Colony controls	✓	✓	✓	✓

Note: Data is aggregated at the country pair × year level. The dependent variable is the share of imports of destination d coming from origin o. All specifications include destination × year, origin × year and country-pair fixed effects. All the dummy variables are mutually exclusive. In brackets are the standard errors, clustered by country-pair. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$.

Table 2.7: Effect of news agencies and telegraphs on trade flows, mutually exclusive dummy variables (1850-1913).

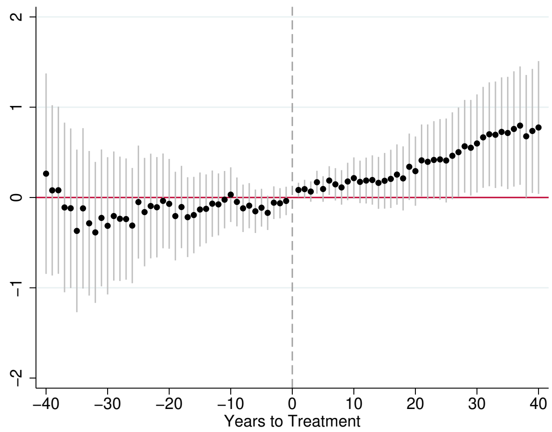
	(1)	(2)	(3)	(4)	(5)	(6)
	$\text{Ln}(Y_{odt})$	$\text{Ln}(Y_{odt})$	S_{odt}	S_{odt}	$S_{odt} > 0$	S_{odt}
News Ag. × Tel.	1.031^a	1.037^a	0.818^a	0.815^a	0.615^a	0.778^a
	[0.214]	[0.214]	[0.175]	[0.175]	[0.127]	[0.216]
Telegraph	0.319^b	0.323^b	0.323^a	0.334^a	0.181^c	0.383^a
	[0.147]	[0.147]	[0.116]	[0.115]	[0.096]	[0.146]
News Agency	0.286	0.287	0.232	0.222	0.178	0.075
	[0.210]	[0.210]	[0.184]	[0.183]	[0.137]	[0.237]
Observations	59910	59910	83373	83373	59910	140506
Estimator	OLS	OLS	PPML	PPML	PPML	PPML
Sample	Complete	Complete	Complete	Complete	Complete	Balanced
Colony controls	×	✓	×	✓	✓	✓

Note: Data is aggregated at the country pair × year level. The dependent variable is the log of the bilateral trade flow in columns (1) and (2), and the share of imports of destination d coming from origin o in the remaining columns, (3) to (6). All specifications include destination × year, origin × year and country-pair fixed effects. All the dummy variables are mutually exclusive. In brackets are the standard errors, clustered by country-pair. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$.

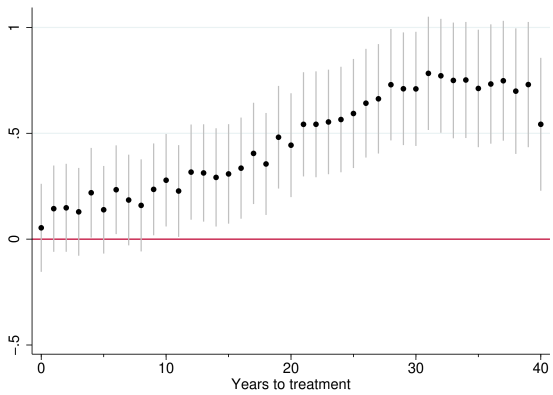
B.2 Event-Study

Figure 2.16: Additional Results of the Event-Study

(a) With dyad-specific time trends



(b) Without the pre-treatment dummies



Notes: Spikes indicate the 95% confidence interval.

B.3 Cross-sectional Estimates

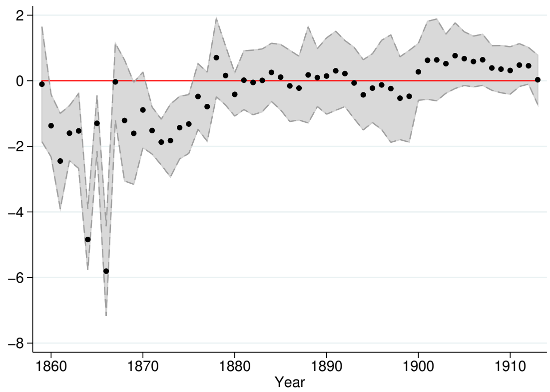
Table 2.8: Cross-dyads correlation between tariffs and our variables of interest

	Average Tariff$_{odt}$			
	(1)	(2)	(3)	(4)
News Ag. × Tel.	-3.775	-3.681	5.854	5.294
	[8.672]	[8.710]	[5.020]	[5.213]
News Ag. × Direct Tel.		2.864		6.216
		[4.400]		[4.299]
Headquarter News Ag. × Tel.			-18.235	-15.502
			[11.758]	[11.519]
Headquarter News Ag. × Direct Tel.				-9.565
				[5.864]
Telegraph	-11.829^{b}	-10.850^{b}	-11.937^{b}	-10.856^{b}
	[4.684]	[5.028]	[4.709]	[5.057]
Direct Tel.		-6.928		-7.100
		[4.299]		[4.313]
News Agency	-0.342	-0.146	-12.486^{b}	-12.256^{b}
	[9.270]	[9.304]	[4.868]	[4.874]
Headquarter News Ag.			25.410^{c}	25.452^{c}
			[13.162]	[13.271]
Observations	4436	4436	4436	4436
Dyad FE	×	×	×	×
Year FE	✓	✓	✓	✓

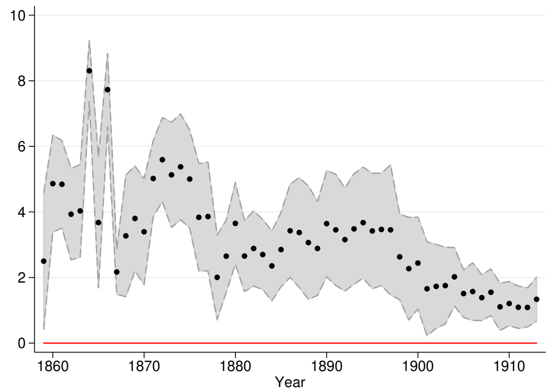
Note: Data is aggregated at the country pair × year level. The dependent variable is the average tariff rate. All specifications include year fixed effects. In brackets are robust standard-errors. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$.

Figure 2.17: Evolution over time of cross-sectional gravity estimates

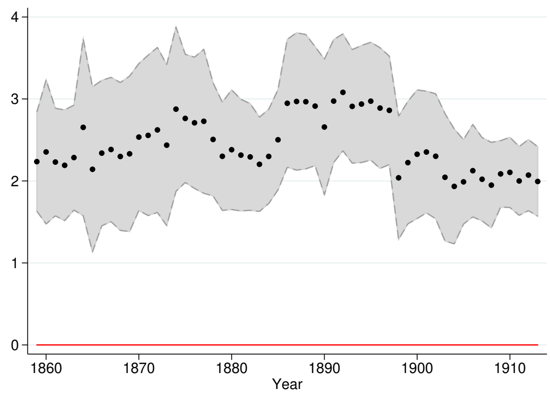
(a) Both countries colonized



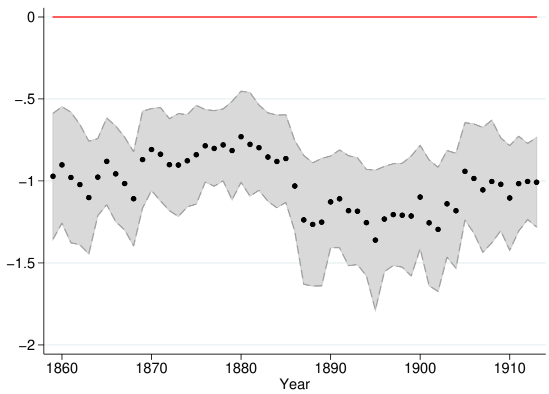
(b) Same colonizer



(c) Metropole-colony



(d) Log of geographic distance



Notes: Each dot corresponds to a PPML estimate from a cross-sectional gravity equation, with trade shares as dependent variable, and with origin and destination fixed effects.

Chapter 3

Trade and Transport Costs: Evidence from Hurricane Sandy

This chapter is co-authored with Emanuele Mazzini (OECD)

Abstract

The fact that international trade flows are approximately inversely proportional to distance and that this distance elasticity cannot be entirely explained by transport costs is well-established. This paper investigates the situation of intra-national trade costs, and reaches a similar conclusion: we find that the total distance elasticity of trade flows within the USA is around -0.84, while if transport costs were the only source of spatial frictions, this distance elasticity would be approximately 14 times lower (around -0.06). We establish this result by using hurricane Sandy as a natural experiment shifting upwards transport costs in some areas of the US. Pairs of origin and destination are heterogeneous in their exposure to the hurricane since the share of the usual route between them going through the affected areas varies. This provides us with a set of bilateral changes in transport costs, whose effect on trade can be estimated.

1 Introduction

The fact that distance impedes trade flows is one of the most robust findings in the empirical trade literature. The most recent meta-analysis (Head and Mayer, 2013), including 1835 estimates of the distance elasticity of trade flows, reveals that this elasticity hovers around -1. While transport costs appear as a natural explanation to rationalize this distance effect, their elasticity with respect to distance would have to be well above the current estimates to be consistent with a distance elasticity of trade flows of -1. Moreover, the distance elasticity did not decrease during the last decades, while transport costs experienced a dramatic fall over the same period. This points to the existence of some other types of spatial frictions, which Head and Mayer (2013) named "dark trade costs", in an analogy with the "dark matter" in cosmology.[1]

[1]Dark matter in astrophysics is an hypothetical type of matter, non observed but whose existence is inferred from its gravitational effects. The same applies to "dark trade costs": although they are hardly observable, they may be responsible for up to 85% of the distance effect on trade (Head and Mayer, 2013).

Potential sources for these frictions are diverse. They include, for example, differences in culture and tastes, a lack of mutual trust, and the spatial decay of information. Intuitively, we expect these mechanisms to affect less trade within countries than trade between countries. Indeed, culture and tastes are arguably more similar within a country than between countries, the spatial decay of information should be lower, and mutual trust should be higher. Additionally, tariffs and the "grey trade costs" of crossing borders (non-tariff barriers to trade) are absent. Therefore, we may wonder whether "dark trade costs" do exist within countries, or whether the only substantial source of spatial frictions is the transport cost. Our work sheds light on this question, focusing on the case of the USA. We provide an upper bound for the part of the distance elasticity of trade flows that is due to transport costs. We find that while the total distance elasticity of internal trade flows is -0.84, this distance elasticity would be significantly smaller, around -0.06, if there were no other trade costs than transport costs. This allows us to unambiguously reject the hypothesis that there are no "dark trade costs" within the US.

We obtain this result by making use of the massive disruptions on the transport infrastructure caused by hurricane Sandy in October 2012. These disruptions led to a sizable increase in transport costs in the affected area (the North-East of the US). Dyads for which a large share of the usual optimal route goes through the affected region are more affected than dyads for which the usual optimal path avoids the damaged area. This supplies us with a set of bilateral variations in transport costs. Our specification controls for the numerous potential side-effects of the hurricane that may have affected trade, such as the decrease in production or the operational issues faced by the firms.

Although the question of the nature of the "dark trade costs" is out of the scope of this paper, we see a promising explanation in the recent contribution of Chaney (2018b), who shows that the constant distance elasticity of trade can be rationalized by a model in which firms trade only through a network of contacts, and this network is spatially clustered. With this explanation, the distance elasticity of trade flows is not primarily linked to transport costs, and its magnitude depends on the structural parameters underlying the network formation dynamics. The scarcity of business links over large distances may be an important component of the "dark trade costs". This scarcity is confirmed by Bernard et al. (2019) who, studying buyer-supplier relationships among Japanese firms, find that distance strongly reduces the probability of forming a business link.

The second aspect of our contribution is methodological : we show how an indirect inference estimator can be applied to structural gravity models in order to estimate the local effect of Sandy on transport costs in the affected region. The intuition of this estimation method is that we find the value of the increase in transport costs for which the change in trade patterns we observe in real data matches the one we obtain in data simulated from a structural gravity model. More precisely, we use the fact that the bilateral changes in trade costs due to Sandy result in variations of the origin and destination specific multilateral resistance terms. We find the value of the local change in transport costs for which the variations of the multilateral resistance terms in simulated data are the closest to their empirical counterparts.

To the best of our knowledge, Feyrer (2011) was the first paper to use a natural experiment in order to isolate the role of transport costs in explaining the distance elasticity of trade from the role of other trade costs. He computed the change in bilateral sea distances that resulted from the closing of the Suez Canal in 1967 and its reopening in 1975 and found a distance elasticity comprised between

-0.5 and -0.2, much lower than the total distance elasticity (-1), which he interpreted as evidence that transport costs were not sufficient to explain the whole distance elasticity of international trade flows. His work was deepened by Hugot and Umana Dajud (2016), who added the Panama canal to the analysis and improved the estimation method. They found a distance elasticity of -0.15, again significantly below -1. These two papers however are not able to conclude on the existence of "dark trade costs" within countries, since they only consider maritime international trade flows. The idea that natural disasters may locally and temporarily increase transport costs in some areas of a given country was confirmed by Volpe Martincus and Blyde (2013), who showed that the 2010 earthquake in Chile had a negative effect on the exports of firms for which the optimal route was passing through affected regions. However, they did not attempt to relate their estimates to the distance elasticity of trade, unlike our work.

While some papers highlighted the existence of spatial frictions within countries, on top of transport costs, their approach yields more restrictive results than ours. For instance, Hortaçsu et al. (2009) studied eBay transaction within the 48 continental states of the US and found that distance had a negative effect on trade, even after controlling for shipping costs. Nevertheless their data covers only a small subsample of internal US trade flows, that may not be representative. Wrona (2018) identifies a border effect between East- and West-Japan, which is less general than our finding, in the sense that we show that internal trade flows are systematically more distance sensitive than what would be expected when taking only transport costs into account, not only when they cross a border.

The paper proceeds as follows. We start with a quick description of hurricane Sandy and its effects on transport infrastructure and trade flows (section 2). We provide some evidence of the disruptive effects of the hurricane, using data we obtained from states' Department of Transportation, and we explain how we computed the bilateral changes in transport costs induced by Sandy. Section 3 presents the theoretical framework we use, especially the way we model trade costs and their decomposition. It also details our identification strategy, which exploits the time variations in bilateral transport costs to estimate a gravity equation with dyadic fixed effects capturing the time invariant trade costs, and time varying origin and destination fixed effects capturing the direct economic impact and the other disruptions caused by the hurricane. Section 4 is dedicated to the presentation of our indirect inference estimator. Finally, we present our main result, establishing the existence of "dark trade costs" within the US, along with some robustness checks in section 5.

2 Hurricane Sandy

2.1 An exceptionnal and destructive storm

Sandy is a hurricane[2] that hit the Northeastern US at the end of October 2012. It can be considered as an exceptionally massive storm for the US, both because of its impact and characteristics. Sandy made landfall on October 29th, 2012 near Brigantine in New Jersey. At this date, it was not at the peak of its intensity (this peak was reached over Cuba), but it was still very intense and wide, with tropical storm-force winds extending 805 km from the center of circulation just prior to land-

[2]Note that when it reached the US territory, the storm was technically speaking no longer a hurricane, but rather a "post-tropical" cyclone, because it lacked the typical strong thunderstorm activity near its center and had lost its eye. Nevertheless, throughout this paper we will use the term hurricane to refer to it since this is the common practice.

fall.[3] The storm's angle of approach (a perpendicular one) and the fact that the landfall coincided with high tide and full moon also explain the exceptional effects of Sandy, since these two element led to record storm tide heights (the storm surge combined with astronomical tide) and the surge's large waves were driven directly into the coastal cities.

After landfall, Sandy slowed down and weakened, but its broad size nevertheless led to disruptions across the Eastern and Midwestern US, as well as Southeastern Canada. High winds, heavy rains and accumulating snows were recorded because of Sandy's remnants moving through southern Pennsylvania. In the central Appalachian Mountains, blizzard conditions occurred and strong winds spread into the Ohio Valley and the Great Lakes on October 31st. The hurricane completely dissipated in Eastern Canada in the next two days. The National Oceanic and Atmospheric Administration (NOAA) estimated that the entire area affected by the winds during the track extended over more than 5 million square kilometres and that more than 60 million people were directly affected across 24 states (Mildenhall et al. (2013), p. 13).

The exceptional intensity and width of this storm explains the high death toll: according to Blake et al. (2013), 72 people died in the Northeastern US as a direct consequence of the hurricane. Economic losses were also large: AON Benfield, a leading insurance intermediary, evaluates the total gross direct economic cost of hurricane Sandy as high as 68 billions USD (Mildenhall et al. (2013), p. 38), thus making it the second-costliest hurricane ever recorded in the USA (Katrina in 2005 being the first one) (Mildenhall et al. (2013), p.45).

Massive disruptions for road transport Hurricane Sandy caused a sizable temporary increase in transport costs for all the goods that had to transit in the Northeastern US because it severely damaged transport infrastructures in this area. Anecdotal evidence suggests that it was very difficult to circulate after the passage of the hurricane, as the words of James Hadden (project manager for the 511 Traffic and Travel Information Program at the New Jersey Department of Transportation) perfectly summarize: *"Roadway damage was beyond what anyone could have imagined"*.[4] Main roads and highways were closed or re-routed due to flooding, downed trees, downed wires or debris while trains and long-distance bus companies were forced to suspend operations across the Northeast for several days.[5]

This claim is supported by data we obtained from the New Jersey (NJ) and the New York (NY) Department of Transportation (henceforth DOT), which contains information about all the disruptions on the highway network recorded by the NJ DOT and the NY DOT during hurricane Sandy and its immediate aftermath.[6] The magnitude of the damages is confirmed by the figures given in table 3.1, where we classified the disruptions by categories and selected the categories that were unequivocally due to hurricane Sandy. Sandy caused overall more than 1000 disruptions on the highway network in the two considered states. Downed trees, signals, wires or poles represented the most important cause of disruptions (607 occurrences), followed by flooding (136 occurrences) and other generic weather related events (96 occurrences). Other major sources of disruptions include debris,

[3]Tropical storm-force winds correspond to a wind speed above 89 km/h, a speed at which, according to Wikipedia, "Trees are broken off or uprooted, structural damage likely". Hurricane-force winds (above 118 km/h) extended 280 kilometers. Source: `http://www.livescience.com/24380-hurricane-sandy-status-data.html`

[4]WRTM Workshop and Stakeholder Meeting, September 25, 2013

[5]See, for instance `http://wtop.com/news/2012/10/amtrak-bus-lines-cancel-service-for-sandy/`.

[6]To be precise, data extend until November 7th, 2012 for NJ but only until November 2nd, 2012 in the case of NY.

emergency interventions and overturned vehicles.

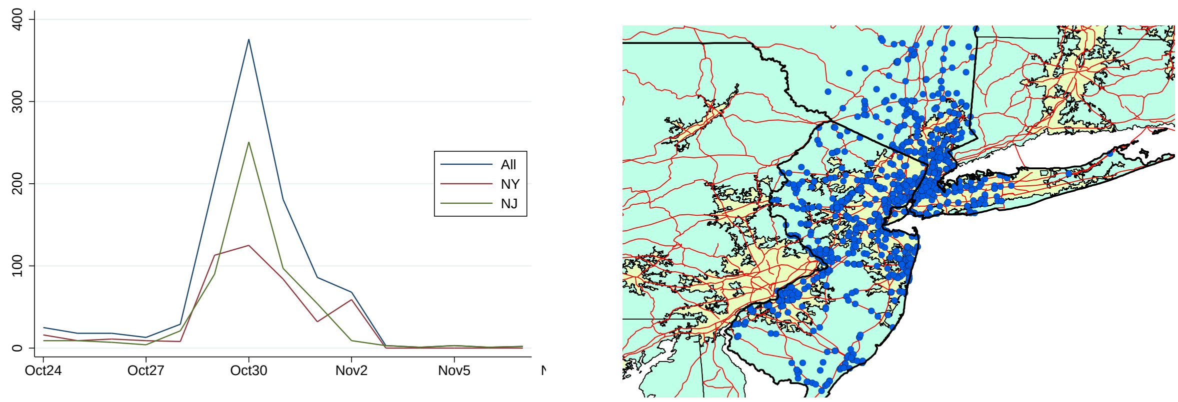
Table 3.1: Hurricane related disruptions in New Jersey and New York

Type	NJ	NY	Total
Downed pole/wire/signal/tree	345	262	607
Flooding	89	47	136
Weather related	39	57	96
Debris	6	46	52
Overturned vehicles	19	0	19
Emergency interventions	4	14	18
Totals	518	428	946

Note: Data sorted according to the total number of accidents in NY and NJ

Figure 3.1a shows that the number of hurricane related disruptions reached a remarkable peak on October 30th, with around 400 disruptive events recorded by the NJ and the NY DOT that day.[7] This suggests that the storm caused an unusual increase in the number of obstacles faced by potential road users. Figure 3.9, in appendix plots the number of non hurricane related disruptions, for NJ and NY separately. Consistently with our story, such disruptions did not experience any increase following the hurricane. Figure 3.1b indicates the location of the recorded disruptions. It shows that the disruptions affected the road network in a wide area, quite far inland, and were not limited to the coast.

Figure 3.1: Disruptions related to the hurricane



(a) **(b)**

(a) Evolution over time of the number of hurricane related disruptions in NJ and NY. (b) Map of the disruptions related to hurricane Sandy. Red lines show the highway network. Blue points represent disruptions recorded by NJ or NY DOT. Yellow surfaces correspond to Core-Based Statistical Areas (CBSA). Note that only NJ and NY DoT data included geographic coordinates, so that we are not able to plot disruptions occurring in other states.

The situation for motorists was aggravated by a shortage of gasoline, which led to rationing and price gouging in some instances.[8] Filling stations saw lines that were even miles long in some

[7]Note that these dates are the date at which the DOT record the disruptions, which might be a bit delayed compared to the date at which the disruption actually started because the DOT might not be able to instantaneously gather the whole information about the hurricane related damages.

[8]This shortage resulted from the combination of several elements on the production side: refineries in the affected

cases, forcing people to wait for several hours.[9] Table 3.15 reports that gasoline was available in only one third of gas stations on November, 2. Even on November, 9 the full supply of fuel was still not restored: 21% of gas stations had no gasoline supply. Finally, the EIA warns that the reported figures are "[...] not designed to reflect the specific experience of more severely affected areas", which suggests that the shortage could have been locally more severe.[10]

2.2 Effects on trade flows

The Commodity Flow Survey Data on trade flows comes from the Public Use Microdata (PUM) file of the 2012 Commodity Flow Survey (hereafter CFS). This survey is realized every five years by the Bureau of Transportation Statistics (hereafter BTS) and the Census Bureau. The 2012 CFS covers approximately 60 000 establishments in mining, manufacturing, wholesale, auxiliaries, and selected retail and services trade industries (a list of all the included industries can be found in the appendix, table 3.12). Once an establishment is selected, it receives a questionnaire every quarter. Reply is mandatory and the answers have to be exact, which ensures good quality of the data. In total, the CFS records 4,547,661 shipments in 2012.[11] The shipments included in the CFS do not necessarily correspond to trade transactions. However, it is common practice to use the CFS to proxy trade flows, as for instance in Duranton et al. (2014). For each shipment, we have information about twenty variables such as the origin and the destination areas, the transport mode, the NAICS code of the product as well as the value and the weight of the shipment.

The CFS distinguishes five single modes and five multiple modes of transport. Table 3.13 provides an exhaustive list of the modes of transport considered in the CFS and describe their respective importance in terms of trade flows. It shows that shipments carried by truck represent by far the large majority of commodity flows. This is the reason why our analysis of the effect of Sandy focuses on road trade flows.[12] As a robustness check, we nevertheless check that there was no substitution towards other modes of transport because of the hurricane.

Geographic information : the "CFS areas" For each shipment, we know the "CFS area" where it originated and arrived. There are 129 distinct CFS areas. Out of these, 83 correspond either to a Metropolitan Statistical Area (MSA) or a Combined Statistical Area (CSA), while the remaining 46 are "remainders", namely portions of states that are not within a MSA/CSA. We do not include in our sample the two CFS areas that are not within the continental USA, Hawaï and Alaska, because we focus on inland trade. After this selection, our sample gathers 3,143,535 observations in 45 distinct

areas were either shut down or run at reduced capacity, pipelines had to be closed for safety precautions, imports from the New York harbour were limited, and the disruptions on highways hampered the ability of trucks to deliver fuel when they were able to obtain it. The US Department of Energy additionally reported that about 8.5 million customers were without power during or after hurricane Sandy. Among them, many had to use electric generators requiring fuel, which led to a demand spike, while some were gas stations, unable to serve customers because of power outages. Cf http://www.eia.gov/todayinenergy/detail.cfm?id=8730

[9]See, for instance, http://money.cnn.com/2012/11/01/news/economy/gas-stations-supply-sandy/.

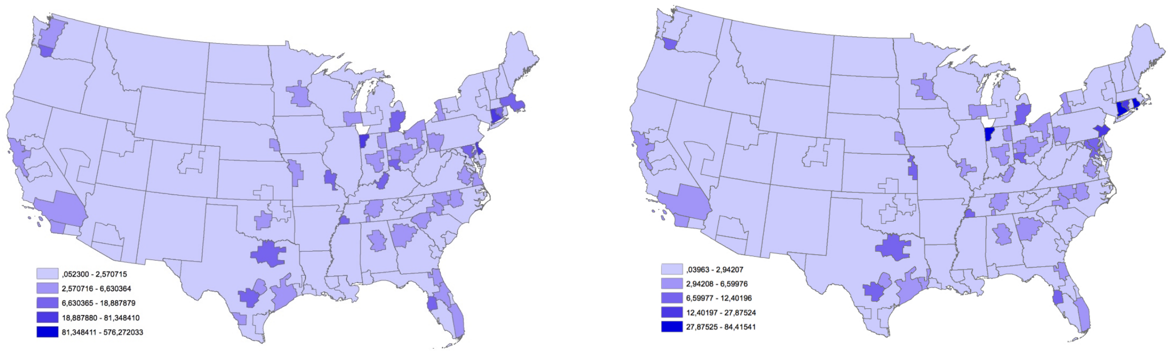
[10]The "New York City Metropolitan Area Retail Motor Gasoline Supply Report" is available at http://www.eia.gov/special/disruptions/hurricane/sandy/gasoline_updates.cfm.

[11]The term "shipment" is defined by the Census Bureau and the BTS as "a single movement of goods, commodities, or products from an establishment to a single customer or to another establishment owned or operated by the same company as the originating establishment (e.g., a warehouse, distribution center, or retail or wholesale outlet). Source: 2012 CFS Summary Report, p. ix

[12]We include in "road transport" all the goods shipped by truck, be this "for-hire truck" or "private truck".

NAICS. A list of all the CFS areas can be found in the appendix, in table 3.10 for "urban" CFS areas and table 3.11 for the "remainders".

Figure 3.2: Trade flows by CFS area



(a) **(b)**

(a) Total import value by CFS area (b) Total export value by CFS area

Figure 3.2 represents the total export value and the total import value of each CFS area, normalized by its surface. Unsurprisingly, we observe that urban CFS area trade more than rural CFS areas, which is consistent with economic activity being spatially concentrated in these areas. Given that the region affected by Sandy is highly urbanized, it accounts for a large share of US trade, which means that even a small increase in transport costs in this region may have a sizable effect on trade.

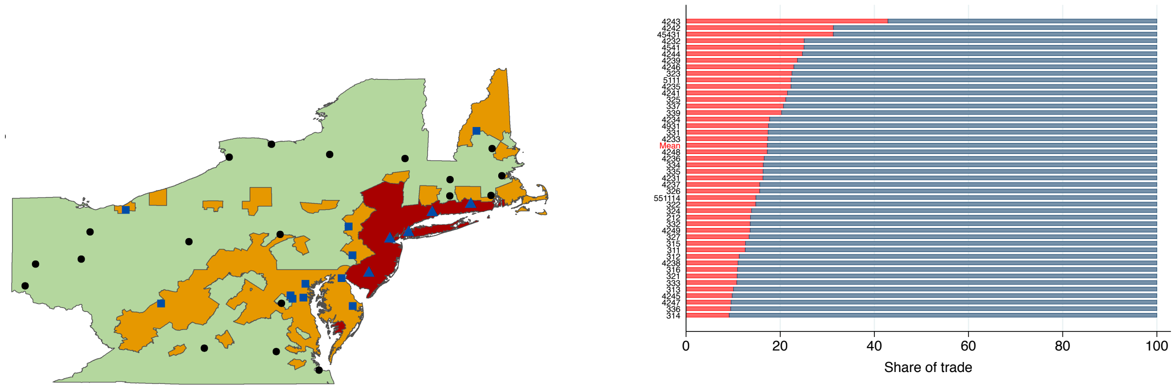
The affected areas: Individual Assistance and Public Assistance programs. In order to identify the areas that were affected by the Hurricane, we used geographic data from the Federal Emergency Management Agency (hereafter FEMA). We selected all the counties that benefited from Public Assistance (PA), or Individual Assistance (IA). Public Assistance is a program through which the FEMA provides a grant to "fund the repair, restoration, reconstruction or replacement of a public facility or infrastructure damaged or destroyed by a disaster", while IA provides a federal funding "to individuals and families who have sustained losses due to disasters"[13]. The counties that benefited from IA or PA are represented in blue in figure 3.3a. Note that sixteen CFS areas are at least partly in the devastated zone, and the zone is large enough for increases in transport costs to occur even for pairs of CFS areas that were not directly hit by the Hurricane, but for which the optimal itinerary goes through the affected area.

The idea that Sandy has hit a region that accounts for a large share of internal US trade is confirmed when we plot the share of trade flows for which the origin or the destination is in the region affected by the hurricane (figure 3.3b). This means that a large part of trade flows have been affected by Sandy, and that all industries were concerned, which suggests that the effects on trade will be large enough for us to measure them.

Total effect of Sandy on trade flows Secondly, we compare the evolution over time of trade flows in affected and unaffected areas. We aggregate all trade flows originating from or arriving to affected

[13]Both definitions are taken from the official FEMA website: http://www.fema.gov/news-release/2015/07/20/understanding-individual-assistance-and-public-assistance, consulted on January 20th, 2016

Figure 3.3: Geography and sectoral hetoregeneity of the disruptions

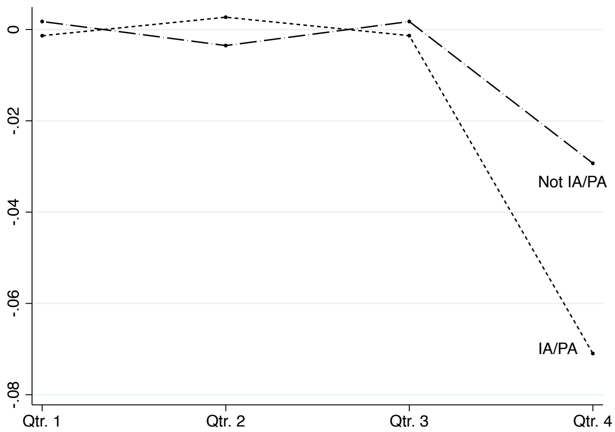


(a) **(b)**

(a) Map of the counties benefiting from Individual Assistance (red) or Public Assistance (orange). (b) Share of trade flows whose origin or destination is an affected area (IA or PA), by NAICS. We consider only trade flows during the 4th quarter. The signification of the NAICS codes is detailed in table 3.12.

areas and detrend this aggregate by regressing its values for the three first quarters on the time variable (quarter) and taking the difference between the predicted value and the observed value. We use a similar approach for all flows originating from or arriving to the unaffected areas. These detrended aggregate flows are plotted in figure 3.4. They fall in both groups during the 4th quarter, but the fall is more dramatic for flows for which the origin or the destination corresponds to an affected area. The fact that trade flows between non affected CFS areas also decrease could be explained by a global downward trend for the 4th quarter, but maybe also by the fact that the transport costs might have increased for them too, if the optimal path between them and their trade partners goes through the IA/PA counties. We insist on the fact that the effects highlighted by figure 3.4 embed not only the effects of transport costs, but also all the direct economic damages caused by the hurricane. Our identification strategy, that we will present in the next section, allows us to isolate the effect of transport costs.

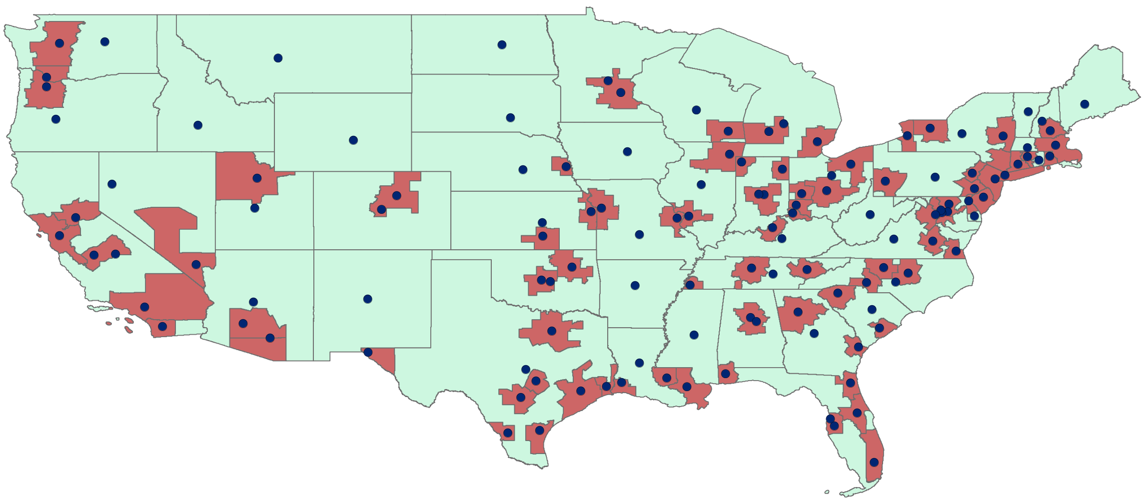
Figure 3.4: Detrended trade flows in affected and unaffected areas.



2.3 Determination of transport costs

The highway network Geographical data on the US road network come from Natural Earth[14]. The map includes only major roads, which is an advantage for our analysis because it corresponds approximately to the US National Truck Network, a network of approved state and interstate highways for commercial truck drivers[15]. We turn this map of the road network into a raster where each cell can take one of two values: either infinite value, if there is no road in the cell, or value 1 if there is a road in the cell. This value corresponds to the transport cost to go through the cell. Because we are only interested in the relative variation of transport costs, we could choose any value for the cost of going through one cell, and this would not affect our results. Hence we normalize this cost by setting its value to 1. A part of the road raster we use is visible in figure 3.6a (zoom on the North-East of the USA).

Figure 3.5: CFS areas and their weighted centroids



Notes: Red surfaces correspond to urban CFS areas, while green surfaces represent rural CFS areas. Dark blue points are the weighted centroids of the CFS areas.

Reduce each CFS area to its weighted centroid The computation of transport costs between any pair of CFS areas requires the conversion of each of these areas to a single point. We therefore determined the weighted centroid of each area, with weights based on population. More precisely, we first determine the centroid of each county within a given CFS area and assign a weight to each of these centroids, equal to the population of their county. The coordinates of the weighted centroid of the CFS area are then given by the weighted average of the coordinates of the centroids of each county within the considered CFS area. We use population weights because they can be considered as a proxy for economic activity, and it is more likely that a shipment leaves from (or arrives to) a place where economic activity is more intense. As a final step, we determine the point on a road which is the closest to the calculated weighted centroid and use this point in our transport costs

[14] Natural Earth is a website supported by the North American Cartographic Information Society (NACIS) proposing public domain maps at different scales on various themes: http://www.naturalearthdata.com

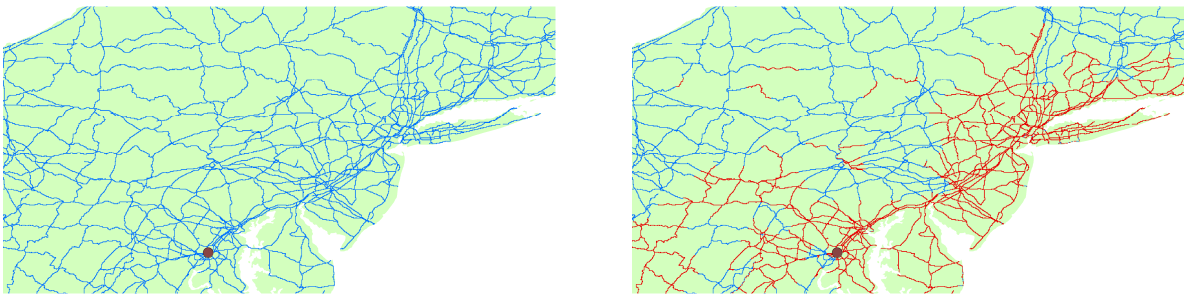
[15] The National Truck Network includes almost all of the Interstate Highway System and other specified non-Interstate highways.

computation.[16]

Bilateral transport costs without Sandy Throughout this paper, we will distinguish two "states of the world". The first one corresponds to the situation during hurricane Sandy and its immediate aftermath, where the road network is heavily affected. We will refer to it as the "Sandy state of the world" and use the superscript ^{S} to denote it. The second state of the world, which we will refer to as the "normal state of the world", corresponds to the rest of the year: no unusual disruption affects the road network. We will use the superscript ^{N} to denote this state of the world. Each state of the world corresponds to a different vector of bilateral transport costs: the transport costs in the "normal state of the world" are denoted T^N_{ni} and the transport costs in the "Sandy state of the world" are denoted T^S_{ni}.

T^N_{ni} is computed via a GIS software which has a built-in *least cost path* algorithm allowing to find the shortest path between all pairs of CFS areas and to return the associated cost, which is our measure of transport costs. The computation of transport costs during Sandy requires the creation of a new road raster, which differs from the previous one by the fact that we consider that transport costs have been multiplied by a certain factor, denoted κ in the area affected by the hurricane. Concretely, it means that road cells in the IA/PA counties now have a value of κ instead of 1 (see figure 3.6b). κ corresponds to the factor by which transport costs are multiplied in the affected area during the hurricane. We call this factor the "overcost parameter". For instance, if the overcost parameter $\kappa = 6$, it means that it is six times more costly to go through cells in the affected areas during the hurricane than in normal times. We explain in section 4 how we determined the value of the overcost parameter. With the new road raster, the procedure to determine the transport costs during Sandy is identical to the one described previously. For trade flows occurring within CFS areas, we cannot use our algorithm because the origin and the destination point would be the same, so we simply consider that transport costs have been multiplied by κ if the weighted centroid of the CFS area lies within an IA/PA county.

Figure 3.6: Zoom on the North-East region of our US road raster



(a) **(b)**

(a) Road raster in the "normal state of the world". Blue cells correspond to a cost of 1. Green cells correspond to an infinite cost (i.e., no road). (b) Road raster in the "Sandy state of the world". Red cells have a cost of κ.

We have two vectors of bilateral transport costs, T^N_{ni} and T^S_{ni}, each corresponding to a state of

[16]This final adjustment is necessary because the method we use for the computation of transport costs requires the origin and the destination points to be on a road.

the world. From these two vectors, we need to compute quarterly transport costs because our trade data is quarterly. During the three first quarters, we consider that there are no major disruptions, i.e. the state of the world is always normal, so that for $t = \{1,2,3\}$, $T_{nit} = T^N_{ni}$. In the 4th quarter, the state of the world can be Sandy or normal, so the transport costs are a weighted average of the transport costs in the "normal state of the world", T^N_{ni}, and the transport costs in the "Sandy state of the world", T^S_{ni}. More precisely, the weights depend on the duration of each state of the world.[17] Let χ denote the fraction of the fourth quarter spent in the "Sandy state of the world", i.e. the number of days in the "Sandy state of the world" divided by the total number of days in the fourth quarter (92 days). Then:

$$T_{ni,t=4} = \chi\, T^S_{ni} + (1-\chi)\, T^N_{ni}$$

In our baseline estimation, we consider that the "Sandy state of the world" lasts for ten days, hence $\chi = 10/92$. The choice of this duration is justified by anecdotal evidence and by the fact that a few days after Hurricane Sandy, a snow storm affected approximately the same area (the so-called "November 2012 nor'easter"). Nevertheless, we show that our results still hold for a wide range of values of χ, from $\chi = 1/92$ to $\chi = 20/92$.

3 Model

We assume that bilateral trade flows are given by a structural gravity equation. Structural gravity can be derived from most trade models and is therefore a fairly general specification. We do not need to specify a full trade model to carry our analysis, we just need the resulting trade flows to follow structural gravity. The trade flow for sector s from location i to location n at time t (denoted X_{nist}) takes the following form:

$$X_{nist} = \frac{Y_{ist}}{\Omega_{ist}} \frac{X_{nst}}{\Phi_{nst}} \phi_{nist} \tag{3.1}$$

where $Y_{ist} = \sum_n X_{nist}$ is the value of production in location i for sector s, while $X_{nst} = \sum_i X_{nist}$ is the value of expenditures in location n on goods from sector s. Our empirical counterparts are the total value of goods leaving from i (including those dispatched in the same CFS area) and the total value of goods arriving in n (including those from the same CFS area) respectively. Ω_{ist} and Φ_{nst} are multilateral resistance terms (henceforth MRT) defined as:

$$\Omega_{ist} = \sum_l \frac{\phi_{list} X_{lst}}{\Phi_{lst}} \tag{3.2a}$$

$$\Phi_{nst} = \sum_l \frac{\phi_{nlst} Y_{lst}}{\Omega_{lst}} \tag{3.2b}$$

The "bilateral resistance term" (henceforth BRT), ϕ_{nist}, is a power function of trade costs : $\phi_{nist} = {\tau_{nist}}^{\epsilon}$, where τ_{nist} is an iceberg trade cost and ϵ is the elasticity of trade flows with respect to trade costs. Our interest lies in the effect of distance on trade costs, and its decomposition between a part

[17] We simplify the problem by considering that the probability for a firm to be willing to send a shipment does not depend on the day, so that the probability that the firm is willing to send the shipment during the "Sandy state of the world" is given by the ratio of the disaster's duration over the quarter's duration (92 days)

related to transport costs and a part related to dark trade costs. For this purpose, we rewrite the trade cost τ_{nist} as the product of two components: transport costs, T_{nist}, and dark trade costs, C_{nist}.

$$\tau_{nist} = T_{nit} C_{nist} \tag{3.3}$$

Transport costs and dark trade costs are both positively related to distance, which makes it difficult to disentangle their respective contribution when regressing trade flows on distance. The idea that transport costs increase with distance is quite intuitive, and corresponds to the findings of Combes and Lafourcade (2005), who provide a comprehensive inventory of transport costs[18]. We model the effect of distance on transport costs in a simple and fairly general way, assuming that transport costs are a power function of road distance, with exponent α (the elasticity of transport costs with respect to distance):

$$T_{nit} = {d_{nit}}^{\alpha} \tag{3.4}$$

The current road distance may differ from the geographic distance. The current road distance corresponds to the geographic distance multiplied by a factor $\delta_{nit} \geq 1$, reflecting the fact that roads may not be straight lines and that some unusual detours may be imposed by the circumstances : $d_{nit} \equiv g_{ni}\, \delta_{nit}$, where g_{ni} is the geographic distance and δ_{nit} is the ratio between road distance and geographic distance. We normalize by setting $\delta_{nit} = 1$ for the three first quarters of 2012. As a consequence, we have $T_{nit} = {g_{ni}}^{\alpha}$ for $t = \{1, 2, 3\}$, and $T_{nit} = {d_{nit}}^{\alpha}$ for $t = 4$.

Most of the potential candidates for explaining dark trade costs are strongly correlated with distance. For instance, tastes should be more similar between closer regions, information should flow more easily over lower distances, and trust should be higher between neighbors. On top of this, Chaney (2018b) shows that the ability of firms to create the business links necessary to trade decreases with distance. We therefore model dark trade costs as a power function of distance, and incorporate this constant distance elasticity into a more general "dark trade costs" function, by allowing for a destination-industry specific component, c_{nst}, an origin-industry specific component, c_{ist}, and a dyad-industry specific component, c_{nis}:

$$C_{nist} = {g_{ni}}^{\gamma}\, c_{nst}\, c_{ist}\, c_{nis}$$

The dyad-industry component, c_{nis}, is time independent because the "truly bilateral" dark trade costs (e.g. proximity in culture and tastes , mutual trust, spatial decay of information) can reasonably be considered as constant over our period of analysis (the year 2012). The origin and destination specific components, c_{nst} and c_{ist}, are time dependent because they correspond to trade costs that Sandy may have increased, like the general level of mistrust or pessimism in the area, or operational issues. Nevertheless, these components are specific to the origin or the destination of the trade flow, so that they will be fully captured by the set of origin and destination fixed effects that we include.

[18] Among the different costs they consider, some are proportional to distance (fuel consumption, vehicle maintenance operating costs and tolls, others are proportional to the duration of the trip (driver's wage and accommodation), and because the duration of the trip increases with distance, such time-related costs are an increasing function of distance. We could even consider a broader range of time-related trade costs since, as emphasized in Hummels and Schaur (2013), time creates a delay between the moment when the production decision is made and the one when the product is sold, and market conditions might change during this interval. In the case of intermediary goods, such delay might even jeopardize the whole supply chain.

Plugging τ_{nit} into ϕ_{nit}:

$$
\begin{aligned}
\phi_{nist} &= (T_{nit}\, C_{nist})^{\epsilon} = ({d_{nit}}^{\alpha}\, {g_{ni}}^{\gamma})^{\epsilon}\, (c_{ist}\, c_{nst}\, c_{nis}\,)^{\epsilon} \\
\phi_{nist} &= ({\delta_{nit}}^{\alpha})^{\epsilon}\, ({g_{ni}}^{\alpha}\, {g_{ni}}^{\gamma})^{\epsilon}\, (c_{ist}\, c_{nst}\, c_{nis}\,)^{\epsilon}
\end{aligned}
$$

Defining $\rho \equiv \alpha + \gamma$:

$$
\begin{aligned}
\phi_{nist} &= \left({g_{ni}}^{\alpha+\gamma}\right)^{\epsilon}\, (c_{ist}\, c_{nst}\, c_{nis})^{\epsilon}\, ({\delta_{nit}}^{\alpha})^{\epsilon} \\
\phi_{nist} &= ({g_{ni}}^{\rho})^{\epsilon}\, (c_{ist}\, c_{nst}\, c_{nis})^{\epsilon}\, ({\delta_{nit}}^{\alpha})^{\epsilon}
\end{aligned}
$$

Plugging the BRT, ϕ_{nist}, into the structural gravity equation, we get :

$$
X_{nist} = \frac{Y_{ist}}{\Omega_{ist}} \frac{X_{nst}}{\Phi_{nst}}\, ({g_{ni}}^{\rho})^{\epsilon}\, (c_{ist}\, c_{nst}\, c_{nis}\,)^{\epsilon}\, ({\delta_{nit}}^{\alpha})^{\epsilon} \tag{3.5}
$$

From this equation, we observe that the distance elasticity of trade flows, $\epsilon\rho$, is the product of the elasticity of trade with respect to trade costs, ϵ, and the elasticity of trade costs with respect to distance, ρ:

$$
\begin{aligned}
\frac{\partial \ln(X_{nist})}{\partial \ln(g_{ni})} &= \underbrace{\frac{\partial \ln(X_{nist})}{\partial \ln(\tau_{nist})}}_{\epsilon}\, \underbrace{\frac{\partial \ln(\tau_{nist})}{\partial \ln(g_{ni})}}_{\rho} \\
\frac{\partial \ln(\text{Trade})}{\partial \ln(\text{Distance})} &= \frac{\partial \ln(\text{Trade})}{\partial \ln(\text{Trade costs})}\, \frac{\partial \ln(\text{Trade costs})}{\partial \ln(\text{Distance})}
\end{aligned}
$$

In turn, the distance elasticity of trade costs is the sum of the distance elasticity of transport costs and the distance elasticity of dark trade costs:

$$
\begin{aligned}
\underbrace{\frac{\partial \ln(\tau_{nist})}{\partial \ln(g_{ni})}}_{\rho} &= \underbrace{\frac{\partial \ln(T_{nit})}{\partial \ln(g_{ni})}}_{\alpha} + \underbrace{\frac{\partial \ln(C_{nist})}{\partial \ln(g_{ni})}}_{\rho} \\
\frac{\partial \ln(\text{Trade costs})}{\partial \ln(\text{Distance})} &= \frac{\partial \ln(\text{Transport costs})}{\partial \ln(\text{Distance})} + \frac{\partial \ln(\text{Dark trade costs})}{\partial \ln(\text{Distance})}
\end{aligned}
$$

Rewriting the distance elasticity of trade flows, we obtain:

$$
\begin{aligned}
\underbrace{\frac{\partial \ln(X_{nist})}{\partial \ln(g_{ni})}}_{\epsilon\rho} &= \underbrace{\frac{\partial \ln(X_{nist})}{\partial \ln(\tau_{nist})}}_{\epsilon} \left(\underbrace{\frac{\partial \ln(T_{nit})}{\partial \ln(g_{ni})}}_{\alpha} + \underbrace{\frac{\partial \ln(C_{nist})}{\partial \ln(g_{ni})}}_{\gamma} \right) \\
\epsilon\rho &= \epsilon\alpha + \epsilon\gamma
\end{aligned}
$$

This equation is key to our research question. It shows that the total distance elasticity of trade flows, $\epsilon\rho$, can be decomposed in two additive components: the transport cost effect, $\epsilon\alpha$ and the dark trade cost effect, $\epsilon\gamma$. Our purpose is to obtain an upper bound for $\epsilon\alpha$. If this upper bound is lower than $\epsilon\rho$, we can conclude that $\epsilon\gamma$ is strictly positive, which implies the existence of dark trade costs. Hurricane Sandy creates a variation in δ_{nit}, which we can use to estimate $\epsilon\alpha$. This appears clearly when we take the log of eq. (3.5):

$$\ln(X_{nist}) = \underbrace{\ln(Y_{ist}) - \ln(\Omega_{ist}) + \epsilon\ln(c_{ist})}_{\text{Origin-Industry-Quarter FE}} + \underbrace{\ln(X_{nst}) - \ln(\Phi_{nst}) + \epsilon\ln(c_{nst})}_{\text{Destination-Industry-Quarter FE}}$$
$$+ \underbrace{\epsilon\rho\ln(g_{ni}) + \epsilon\ln(c_{nis})}_{\text{Dyad-Industry FE}} + \epsilon\alpha\ln(\delta_{nit}) \tag{3.6}$$

$$\ln(X_{nist}) = O_{ist} + D_{nst} + B_{nis} + \epsilon\alpha\ln(\delta_{nit}) + \varepsilon_{nist} \tag{3.7}$$

This shows that $\epsilon\alpha$ can be estimated using the standard panel data structural gravity specification: regress the log of trade flows on the log of the change in distance induced by Sandy, with a time varying destination-industry fixed effect, D_{nst}, a time varying origin-industry fixed effect, O_{ist}, and a dyad-industry fixed effect, B_{nis}. The inclusion of dyad-industry fixed effects is crucial, because these fixed effects will capture all the time invariant trade costs that inflate the coefficient on distance in usual gravity estimations.

Looking more closely at the role of the origin fixed effects, , we see from eq. (3.6) that they include three components: the total production of the origin, Y_{ist}, the outward MRT of the origin, Ω_{ist}, and the origin-specific component of dark trade costs, c_{ist}. This means that the direct economic effect of Sandy on the origin region (loss of production capacity) is fully controlled for by the origin fixed effect: we expect Sandy to decrease Y_{ist} if i is in the North-East of the USA, but this will be absorbed by O_{ist}. Similarly, Sandy may have increased other types of trade frictions than the sole transport costs. For instance, the standard operating procedures were probably disrupted, or information did not flow as easily as in a normal period, or people lost confidence because of the damage they suffered. All these frictions to trade are not destination specific, in the sense that they affect trade towards all regions with the same magnitude. Hence, they correspond to an increase in the origin-specific component of dark trade costs, c_{ist}, which is again fully controlled for by the origin fixed effects, O_{ist}, so that these trade frictions do not affect our estimates of $\epsilon\alpha$. Finally the potential effects of Sandy on the outward MRT, Ω_{ist}, are also absorbed by the origin fixed effects.

Estimating the total distance elasticity of trade flows is fairly easy. Indeed, in the three first quarters of 2012, there are no unusual disruptions on the road network (no hurricane Sandy), so that we consider that $\delta_{nit} = 1$. Plugging again our decomposition of trade costs into the structural gravity equation, eq. (3.5), with $\delta_{nit} = 1$, we obtain:

$$\ln(X_{nist}) = \underbrace{\ln(Y_{ist}) - \ln(\Omega_{ist}) + \epsilon\ln(c_{ist})}_{\text{Origin-Industry FE}} + \underbrace{\ln(X_{nst}) - \ln(\Phi_{nst}) + \epsilon\ln(c_{nst})}_{\text{Destination-Industry FE}}$$
$$+ \epsilon\rho\ln(g_{ni}) + \epsilon\ln(c_{nis})$$

$$\ln(X_{nist}) = O_{ist} + D_{nst} + \epsilon\rho\ln(g_{ni}) + \varepsilon_{nist} \tag{3.8}$$

This shows that the total distance elasticity of trade flows, $\epsilon\rho$, can be estimated from the standard cross-section structural gravity specification, with origin-NAICS fixed effects, O_{ist} and destination-NAICS fixed effects, D_{nst}. We can either run one separate regression for each quarter, or pool data from the three first quarters.

From transport costs to distance: Our least path algorithm gives us a good approximation of the change in transport costs, T_{nit}, that we need to turn into a change in distance, d_{nit}. We rely on the functional form of the transport costs : $T_{nit} = {d_{nit}}^{\alpha}$, implies that $d_{nit} = {T_{nit}}^{\frac{1}{\alpha}}$. As a consequence, if we are able to find an upper bound for the value of the elasticity of transport costs with respect to distance, α, then we will also have a lower bound for the change in distance caused by Sandy. This lower bound is sufficient to answer our research question, since underestimating the change in distance will lead to an overestimation of the part of distance elasticity linked to transport costs, $\epsilon\alpha$, so that if we nevertheless find $\epsilon\alpha < \epsilon\rho$ it reinforces our claim.

A lower bound for α can be infered from the total distance elasticity of trade flows, $\epsilon\rho$. Indeed, $\rho = \alpha + \gamma$ and $\gamma \geq 0$ so that $\alpha \leq \rho$. We set the value of the elasticity of trade flows with respect to trade costs, ϵ, to -5, which corresponds to the mean value found in the literature when structural gravity is used, according to the meta-analysis performed by Head and Mayer (2014b). Our estimate of the total distance elasticity of trade flows, $\epsilon\rho$, is -0.84[19]. With $\epsilon = -5$, this implies $\rho = 0,17$. As a consequence $\alpha \leq 0.17$ so that $\frac{1}{\alpha} \geq 5.88$. Therefore, we compute $d_{nit} = {T_{nit}}^{5.88}$, and this gives us a lower bound for the distance equivalent of the change in transport costs. Finally, since $d_{nit} = g_{ni}\, \delta_{nit}$ and the geographic distance g_{ni} is time independent, the time variation in δ_{nit}, which we will use as a regressor to estimate the part of the distance elasticity linked to transport costs, $\epsilon\alpha$ (cf eq. (3.7)) is equal to the time variation in d_{nit}, for which we obtained a lower bound from ${T_{nit}}^{5.88}$.

4 Estimation of the overcost parameter (κ)

In order to compute the change in transport costs during Sandy between each CFS area, we used an “overcost parameter", denoted κ, defined as the factor by which transport costs are multiplied in the affected areas during the hurricane. The value of this parameter affects our final results: the higher we set κ, the higher will be the computed variation in transport costs, and therefore the higher the variation in our explanatory variable (for the same variation in the explained variable. Therefore setting a higher value for the “overcost parameter" results in a lower estimate of the part of distance elasticity linked to transport costs, $\epsilon\alpha$.

We overcome this issue by estimating a lower bound for the value of the “overcost parameter", using a method inspired from indirect inference (Gourieroux et al., 1993). The intuition behind our method is the following: we simulate trade flows with different values of κ and determine each time the implied changes in multilateral resistance terms (MRT), relying on the structural gravity equations. We can then compare the changes in MRT obtained from the simulated data to the ones estimated with real data, and find the value of κ that minimizes the distance between these two vectors. Our results suggest that a lower bound for κ is 6.

For feasibility reasons, we are not able to compute a different “overcost parameter” for each industry. Therefore the subscript s is not necessary in this section: we consider trade flows aggregated at the dyad level, instead of the dyad-NAICS level.

Effects of Sandy on trade costs The simulation of trade flows for any value of the “overcost parameter” requires us to distinguish again between two states of the world: the “Sandy state of the

[19]Results are presented in detail in table 3.3

world" and the "normal state of the world". We keep the same superscript convention, i.e. ^{N} for normal, and ^{S} for Sandy. Because we want to be able to easily generate trade flows for any value of κ, and trade flows depend ultimately on the bilateral trade costs, we need a simple formula expressing the bilateral trade costs during Sandy, τ_{ni}^S, as a function of the bilateral trade costs in the normal state of the world, τ_{ni}^N, and the "overcost parameter", κ. The derivation of such a function is straightforward under some simplifying assumptions that ensure that we will indeed estimate a lower bound of the "overcost parameter". We leave the details of this derivation in the appendix, and present only the resulting equation:

$$\tau_{ni}^S = (s_{ni}(\kappa - 1) + 1)\tau_{ni}^N \tag{3.9}$$

Simulate trade flows in the normal state of the world and during Sandy In the "normal state of the world", trade costs between any pair of CFS areas can be directly infered from the geographical distance between these CFS areas: $\tau_{ni}^N = {g_{ni}}^\rho$, where the elasticity of trade costs with respect to distance, ρ, is equal to 0.17, as shown in section 5. From these bilateral trade costs, we obtain the bilateral resistance terms (BRT) : $\phi_{ni}^N = (\tau_{ni}^N)^\epsilon$. Knowing all the BRT, we can compute the MRT (Ω_i^N and Φ_n^N) by solving equation (3.2). For this, we use the contraction mapping algorithm proposed by Head and Mayer.[20].The total value of exports of region i, Y_i, and the total value of imports of region n, X_n, are taken directly from the CFS data. Note that we slightly depart from structural gravity in the sense that we do not impose $Y_i = \sum_n X_{ni}$ and $X_n = \sum_i X_{ni}$. Our simulations are therefore revealing the modular trade impact (MTI) of Sandy, not its general equilibrium trade impact (GETI) (following the terminology used by Head and Mayer (2014b)).

Since we know $\phi_{ni}^N, \Omega_i^N, \Phi_n^N, Y_i$ and X_n, we have everything we need to obtain the simulated trade flows in normal times, denoted $\widetilde{X}_{ni}^N$, using the structural gravity equation, eq. (3.1). Note that we obtain a different simulated value for each quarter, because Y_i and X_n are time varying.

$$\widetilde{X}_{ni}^N = \frac{Y_i}{\Omega_i^N}\frac{X_n}{\Phi_n^N}\phi_{ni}^N$$

To simulate trade flows during Sandy, we compute a new vector of bilateral trade costs τ_{ni}^S. These trade costs are obtained from eq. (3.9). Basically, we take the trade costs in the "normal state of the world" τ_{ni}^N and multiply them by $(s_{ni}(\kappa - 1) + 1)$ to obtain the trade costs during Sandy (τ_{ni}^S). Once we have τ_{ni}^S, we apply the same procedure as above to obtain first ϕ_{ni}^S, then Ω_i^S and Φ_n^S, and finally the simulated trade flows during Sandy, $\widetilde{X}_{ni}^S$.

Given that our data is quarterly, we need to generate quarterly trade flows. In the first three quarters, there are no disruptions, hence the simulated quarterly trade flows are simply equal to the simulated flows for the"normal state of the world". For $t = \{1, 2, 3\}$: $\widetilde{X}_{nit} = \widetilde{X}_{nit}^N$. In the 4th quarter, the simulated trade flow is a weighted average of the simulated trade flows in the "normal state of the world" and the "Sandy state of the world": for $t = 4$, $\widetilde{X}_{ni,t=4} = \chi\widetilde{X}_{ni,t=4}^S + (1-\chi)\widetilde{X}_{ni,t=4}^N$, where χ is the ratio of the duration of Sandy's related disruptions over the duration of the 4th quarter. In our baseline estimation, we set $\chi = 10/92$, but we run alternative estimations with χ ranging from 1 to 20. Finally, given that we observe 11,73% of zero trade flows in our data, we select for each

[20] https://sites.google.com/site/hiegravity/stata-programs

quarter the 11,73% of flows that have the lowest values and set them to zero.

Indirect inference estimator Let us describe in one paragraph the intuition behind our estimator, before giving a more formal explanation. Firstly, we estimate the vectors of origin and destination fixed effects, $\mathbf{O}_{it}$ and $\mathbf{D}_{nt}$ in the real data, and we compute the difference between fixed effects in the 4th quarter and fixed effects in other quarters. This gives us a vector of parameters, which we denote $\hat{\theta}_0$. Then, we can simulate trade flows for any value of κ and estimate again the vectors of origin and destination fixed effects, this time in the simulated data. We compute the difference between fixed effects in the 4th quarter and fixed effects in other quarters, and we get a new vector of parameters whose values depend on κ, that we denote $\hat{\theta}(\kappa)$. Our estimate is the value of κ such that both vectors of parameters, $\hat{\theta}_0$ and $\hat{\theta}(\kappa)$, are "as close as possible"

More generally, the intuition of indirect inference estimators is that we estimate parameters from an "auxiliary model" and look for the value of the parameter(s) of interest such that the parameters estimated from the auxiliary model with simulated data match parameters estimated from the auxiliary model with real data. Here, our auxiliary model is the classical fixed effect specification of the gravity equation. And the parameters we want to match with simulated data are the time variations in the origin and destination fixed effects. The fixed effect specification of the gravity equation is given by the following equation:

$$\ln(X_{nit}) = B_{ni} + D_{nt} + O_{it} + \varepsilon_{nit} \tag{3.10}$$

For each CFS area j, we have 4 origin fixed effects O_{jt} (one for each quarter) and 4 destination fixed effects D_{jt} (again, one for each quarter). Let us denote $\mathbf{O}_t$ the column vector of the origin fixed effects in all CFS areas at time t and $\mathbf{D}_t$ the column vector of the destination fixed effects in all CFS areas at time t. We can define a vector $\boldsymbol{\theta}$ gathering the values of the time variation in origin and destination fixed effects, between the fourth quarter and the other quarters:

$$\boldsymbol{\theta} = \begin{bmatrix} \mathbf{O}_1 - \mathbf{O}_4 \\ \mathbf{O}_2 - \mathbf{O}_4 \\ \mathbf{O}_3 - \mathbf{O}_4 \\ \mathbf{D}_1 - \mathbf{D}_4 \\ \mathbf{D}_2 - \mathbf{D}_4 \\ \mathbf{D}_3 - \mathbf{D}_4 \end{bmatrix}$$

We first estimate the auxiliary model (eq. (3.10)) with real data. We obtain a set of origin and destination fixed effects, and we compute their time variation as explained above to obtain a vector $\hat{\boldsymbol{\theta}}_0$. This vector is the one we will try to match with our simulated data. Indeed, when we simulate trade flows, the values of the fixed effects in the 4th quarter depend on κ, because κ changes the BRT and as a consequence the MRT. Therefore we get a different vector of differences $\boldsymbol{\theta}$ for each value of κ, that we denote $\hat{\boldsymbol{\theta}}(\kappa)$. Our estimate of the “overcost parameter”, denoted $\hat{\kappa}$, is the value of κ for which $\hat{\boldsymbol{\theta}}(\kappa)$ is "as close as possible" to $\hat{\boldsymbol{\theta}}_0$. In other words, we seek κ such that the distance between $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}(\kappa)$ is minimized. This distance between $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}(\kappa)$ is measured following the Wald approach:

$$\hat{\kappa} = \arg\min_{\kappa} (\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}(\kappa))'(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}(\kappa))$$

Table 3.2: Value and standard-error of κ for different assumptions on the duration of the “Sandy state of the world”

χ	κ	s.e.
1	7.35	0.669
2	7.28	0.356
5	6.93	0.013
10	6.29	0.005
20	5.00	0.025

Standard errors are given by the following equation:

$$\sigma_{\hat{\kappa}} = \left[\left(\frac{\partial \hat{\boldsymbol{\theta}}(\kappa)}{\partial \kappa} \right)' \left(\frac{\partial \hat{\boldsymbol{\theta}}(\kappa)}{\partial \kappa} \right) \right]^{-1}$$

The estimate of κ that we obtain depends on the duration we assume for the “Sandy state of the world”, χ. For a duration of 10 days, the overcost parameter is 6.29, which we round down to 6. Remember that this is a lower bound for the true value of κ. Table 3.2 presents the results of our estimation for different values of χ

5 Results

Total distance elasticity of trade flows The estimation of the distance elasticity of trade flows, $\epsilon\rho$ is straightforward. We estimate cross-sectional gravity equations, one for each of the three first quarters of 2012. Let us remind the specification we use:

$$\ln(X_{nis}) = D_{ns} + O_{is} + \epsilon\rho \ln(g_{ni}) + \varepsilon_{nis}$$

where X_{nis} is the trade flow from i to n in NAICS s, D_{ns} is a destination-NAICS fixed effect, O_{is} is an origin-NAICS fixed effect and g_{ni} is the geographical distance between i and n. As explained previously, , we exclude all shipments exported outside of the US because we do not have information on their final destination, so that we are not able to determine the distance they cover.

Table 3.3 shows the results of our estimations of the distance elasticity of trade flows, when we consider each quarter separately (i.e. we run one distinct regression for each quarter). The coefficient does not exhibit much volatility over time, which suggests that the link between trade costs and distance and the elasticity of trade with respect to trade costs are quite stable over time. The average of the estimates over the three quarters is around -0.84, which lies in the lower part of the distribution of structural estimates found in Head and Mayer (2014b). This is consistent with the intuition that the spatial frictions are lower for flows within a country than for international flows.

Transport costs related part of the distance elasticity To estimates of the transport cost part of distance elasticity ($\epsilon\alpha$), we make use of the variation in transport costs created by Sandy. Let us remind the specification we use:

$$\ln(X_{nist}) = O_{ist} + D_{nst} + B_{nis} + \epsilon\alpha \ln(\delta_{nit}) + \varepsilon_{nist}$$

Table 3.3: Estimates of the total distance elasticity of trade flows, with one distinct regression per quarter

VARIABLES	(1) Flow T1	(2) Flow T2	(3) Flow T3
Distance	-0.827[a]	-0.850[a]	-0.849[a]
	[0.006]	[0.006]	[0.006]
Observations	122316	121065	117239
R^2	0.469	0.471	0.474

Robust standard errors in brackets. Significance levels: [a]: $p < 0.01$; [b]: $p < 0.05$; [c]: $p < 0.1$

.

Table 3.4: Baseline results with $\kappa = 6$

	Flow			
	(1)	(2)	(3)	(4)
Distance	-0.052[b]	-0.050[c]	-0.047	-0.058
	[0.025]	[0.030]	[0.033]	[0.040]
Observations	379228	358825	155173	142218
Rural excluded	×	×	✓	✓
Within CFS excl.	×	✓	×	✓
N. of clusters	126926	121538	52273	48734
R^2	0.777	0.753	0.790	0.756

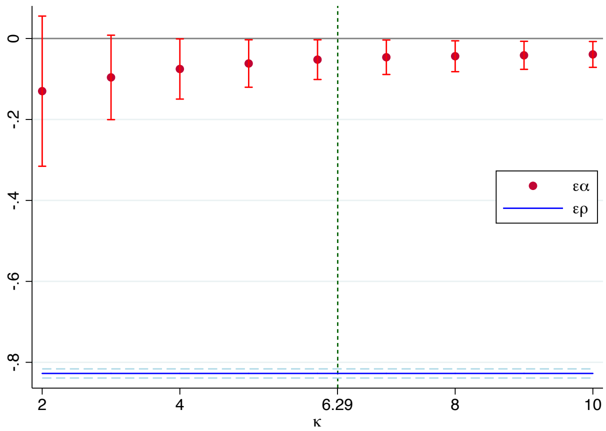
Standard errors in brackets, clustered at the dyad-industry level. Significance levels: [a]: $p < 0.01$; [b]: $p < 0.05$; [c]: $p < 0.1$

.

With the complete sample, we find an elasticity around -0.05 (table 3.4). This unambiguously leads us to conclude that the part of the distance elasticity of trade flows that can be explained by transport costs, $\epsilon\alpha$ is lower than the total distance elasticity, $\epsilon\rho$. Given that $\epsilon\rho = \epsilon\alpha + \epsilon\gamma$ (cf. section 3), it implies that $\epsilon\gamma > 0$, i.e. there are "dark trade costs" within the US. This conclusion would hold even if the value of the "overcost parameter", κ, that we estimated in section 4 was not the right one, as illustrated in figure 3.7, which plots the estimates of $\epsilon\alpha$ as a function of κ (red dots) and compares them with $\epsilon\rho$ (blue line). Note that we are not primarily interested in estimating the true value of $\epsilon\alpha$, an upper bound is sufficient for our main result to hold.

In the three last columns of table 3.4, we check that our results are robust to the omission of the dyads for which our measure of change in distance might be more noisy. Such a noisiness might occur for two reasons. The first reason is that we can not apply our least cost path algorithm to flows occuring within a given CFS area, as explained in section 2, so instead we simply consider that

Figure 3.7: Sensitivity of $\epsilon\alpha$ to κ for $\chi = 10/92$.



the transport costs are multiplied by κ in the "Sandy state of the world". The second reason is that the remainders ("rural CFS areas" are often very large; therefore their weighted centroid might be a poor proxy for the actual origin or destination point of the shipment, and this might ripple into our measure of the change in transport costs. As a consequence, we exclude alternatively the flows taking place within CFS areas, keeping only the flows that cross at least one CFS area boundary (column 2); the flows for which the origin or the destination is a rural CFS area, keeping only the flows between urban CFS areas (column 3); and both former types of flows together, keeping only the flows between urban CFS areas that cross at least one CFS area boundary (column 4). Although $\epsilon\alpha$ is estimated somewhat less precisely once these restrictions are imposed (see table 3.4), its magnitude is not very different, and the main conclusion still unambiguously holds: $\epsilon\alpha < \epsilon\rho$ so there must exists dark trade costs within the US.

Our result still holds for extreme assumptions on the duration of the "Sandy state of the world", as shown in figure 3.8. Figure 3.8a pictures the results we obtain if we assume that the disruptions linked to Sandy lasted one single day. In this case, the value of κ we estimate is 7.35, and, for this value, $\epsilon\alpha$ (red dot) is lower in absolute value than $\epsilon\rho$ (blue line). Similarly, figure 3.8b corresponds to the assumption that the disruptions caused by Sandy lasted twenty days. In this case, we clearly see that only part of the total distance elasticity can be explained by transport costs.

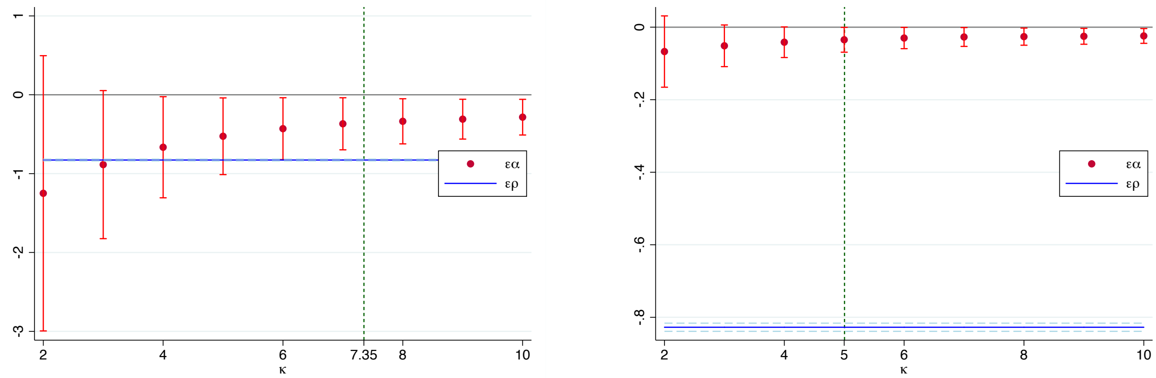
5.1 Robustness checks

Our baseline results lead to the clear conclusion that transport costs cannot account for the whole distance elasticity of trade flows within the US.

More restrictive definition of the affected areas

This result is robust to a more restrictive definition of the affected areas. Namely, we consider that only the counties that benefited from individual assistance (IA) are affected by the hurricane, instead of taking both the counties benefiting from individual assistance and public assistance (IA/PA). This amounts to reducing the area in which the transport costs increase because of Sandy.

Figure 3.8: Sensitivity of the estimates of $\epsilon\alpha$ to κ, for extreme assumptions on the duration of the "Sandy state of the world".



(a) $\chi = 1/92$ **(b)** $\chi = 20/92$

Table 3.5: Baseline results with $\kappa = 6$

	Flow			
	(1)	(2)	(3)	(4)
Distance	-0.085^b	-0.088^b	-0.081^c	-0.094^c
	[0.037]	[0.043]	[0.044]	[0.051]
Observations	379228	358825	155173	142218
Rural excluded	×	×	✓	✓
Within CFS excl.	×	✓	×	✓
N. of clusters	126926	121538	52273	48734
R^2	0.777	0.753	0.790	0.756

Standard errors in brackets, clustered at the dyad-industry level. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

.

Postponement or anticipation of some shipments

Nevertheless, there remains a last subject of concern that could threaten our claim: if shipments are postponed, the estimates of $\epsilon\alpha$ presented in table 3.4 exhibit a downward bias, because some of the trade destruction effect of the increase in transport costs during Sandy is offset by an increase in trade after the hurricane. The last part of this paper is therefore devoted to presenting evidence going against the hypothesis that shipments have been postponed.

Decomposition between intensive and extensive margin If shipments are postponed, then we expect an increase in the average shipment value for the affected dyads. Indeed, goods that could not be shipped because of the hurricane should be added to shipments taking place before or after the hurricane, which would increase their value. Hence, looking at the decomposition of the trade elasticity between an intensive margin (average value per shipment) and an extensive margin (number of shipments) can inform us about the presence, or the absence, of a postponement effect. More formally, let N_{nist} denote the number of shipments, then the intensive margin is defined as the elasticity of $\frac{X_{nist}}{N_{nist}}$ with respect to our explanatory variable, δ_{nit}, while the extensive margin is the elasticity of the number of shipments N_{nist} with respect to δ_{nit} . We therefore estimate the following equations:

$$\ln\left(\frac{X_{nist}}{N_{nist}}\right) = B_{nis} + D_{nst} + O_{ist} + \beta_1 \ln(\delta_{nit}) + \epsilon_{nist}$$
$$\ln(N_{nist}) = B_{nis} + D_{nst} + O_{ist} + \beta_2 \ln(\delta_{nit}) + \epsilon_{nist}$$

Table 3.6: Intensive margin, with $\kappa = 6$

	Value per shipment			
	(1)	(2)	(3)	(4)
Distance	0.021	0.030	0.035	0.036
	[0.022]	[0.027]	[0.029]	[0.036]
Observations	379228	358825	155173	142218
Rural excluded	×	×	✓	✓
Within CFS excl.	×	✓	×	✓
N. of clusters	126926	121538	52273	48734
R^2	0.728	0.720	0.735	0.726

Standard errors in brackets, clustered at the dyad-industry level. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

.

Table 3.7: Extensive margin, with $\kappa = 6$

	Value per shipment			
	(1)	(2)	(3)	(4)
Distance	-0.073^a	-0.080^a	-0.082^a	-0.093^a
	[0.018]	[0.022]	[0.024]	[0.029]
Observations	379228	358825	155173	142218
Rural excluded	×	×	✓	✓
Within CFS excl.	×	✓	×	✓
N. of clusters	126926	121538	52273	48734
R^2	0.862	0.829	0.877	0.836

Standard errors in brackets, clustered at the dyad-industry level. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

.

Results are given in tables 3.7 and 3.6. As expected, most of the effect of Sandy went through the extensive margin, a finding consistent with what Volpe Martincus and Blyde, 2013 observed after the 2010 Chilean earthquake. The intensive margin was not significantly affected. We interpret this as evidence against a postponement effect.

Restriction to most regular shipments Another way to test for the absence of postponement effect is to focus on shipments that need to be sent regularly, and can hardly be delayed. We do this using two different methods. The first one consists in selecting the industries that exhibit the most regular shipment patterns (with the implicit assumption that if they send shipments so regularly, it must mean that it is hard or costly for them to delay shipments). The second method uses the fact that some shipments are temperature controlled, and therefore cannot be delayed.

Select the most regular industries We select the industries for which the trade flows are the most stable over the three first quarters. A natural criterion to determine a low volatility is the coefficient of variation, i.e. the ratio between the standard-deviation and the mean of trade flows in this NAICS . However, given the large number of zero trade flows in our data, this coefficient would be very low for industries for which we have only little observations. Therefore we add another criterion to guide our decision: the share of zero trade flows. We combine these two criteria to form our regularity index, giving the same weight to each of them, and select the ten industries with the highest degree of regularity. A list of the selected industries is given in the appendix, table 3.8. The specification is the same as for the baseline regression, eq. (3.7). We find that the $\epsilon\alpha$ estimated using only the most regular industries (see table 3.8 below) is slightly higher (in absolute value) than the one estimated with all industries. Nevertheless, it is still far from $\epsilon\rho$, which confirms that transport costs are not sufficient to explain the whole distance elasticity of trade.

Table 3.8: Selected industries

	Flow			
	(1)	(2)	(3)	(4)
Distance	-0.073^b	-0.076^c	-0.067	-0.076
	[0.036]	[0.042]	[0.047]	[0.055]
Observations	154891	149870	61412	58114
Rural excluded	×	×	✓	✓
Within CFS excl.	×	✓	×	✓
N. of clusters	49661	48393	19753	18893
R^2	0.761	0.738	0.774	0.739

Standard errors in brackets, clustered at the dyad-industry level. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

.

Temperature controlled shipments Our CFS data includes a dummy variable indicating whether the shipment is "temperature controlled", i.e. whether it is carried in a vehicle designed to maintain the shipment at a certain temperature[21]. If a shipment is temperature controlled, it suggests that it tends to depreciate quickly over time and therefore that it is very costly to postpone it. As a consequence, if there is a postponement effect, then dyad-industries in which there are more temperature controlled shipments should be less affected by this postponement, and thus have a higher distance elasticity, because cancelled shipments cannot be postponed. We therefore compute the average share of temperature controlled shipments during the three first quarters for each dyad-industry and interact this share with distance. If there is a postponement effect, the coefficient on this interaction should be negative and significant. We estimate the following specification, where S_{nis} is the share of temperature controlled shipments during the three first quarters:

$$\ln(X_{nist}) = B_{nis} + D_{nst} + O_{ist} + \beta_1 \ln(\delta_{nit}) + \beta_2(\ln(\delta_{nit}) * S_{nis}) + \varepsilon_{nist}$$

Table 3.9: Temperature controlled goods

	Flow			
	(1)	(2)	(3)	(4)
Distance	-0.050^c	-0.047	-0.042	-0.055
	[0.026]	[0.031]	[0.034]	[0.042]
Distance × Sh. temp. contr.	-0.025	-0.028	-0.062	-0.024
	[0.074]	[0.081]	[0.104]	[0.114]
Observations	379228	358825	155173	142218
Rural excluded	×	×	✓	✓
Within CFS excl.	×	✓	×	✓
N. of clusters	126926	121538	52273	48734
R^2	0.777	0.753	0.790	0.756

Standard errors in brackets, clustered at the dyad-industry level. Significance levels: ^{a}: $p < 0.01$; ^{b}: $p < 0.05$; ^{c}: $p < 0.1$

.

Table 3.9 gives the results of this regression. As can be seen, the coefficient on the interaction term is not significant, which is an additional clue that there is no postponement effect.

6 Conclusion

Using hurricane Sandy as a natural experiment shifting upwards transport costs in some areas of the US, we show that transport costs cannot be the sole explanation behind the strong negative

[21] A temperature controlled shipment is defined as a shipment that is transported in a vehicle or container that regulates the temperature while en route (such as heating and refrigeration) or maintaining the temperature of the commodity at the time of loading (such as insulation). Source: http://www.rita.DOT.gov/bts/sites/rita.DOT.gov.bts/files/publications/commodity_flow_survey/html/def_terms.html, consulted on 27/05/2016

effect of distance on trade flows. More precisely, in our baseline estimation, we find that if transport costs were the only kind of trade costs correlated with distance, then the distance elasticity of trade flows within the US should be 14 times lower than what we actually observe. This result is robust to the exclusion of dyads for which the bilateral change in transport costs that we compute might have been less accurately determined. It also holds if we choose a more conservative perimeter for the areas affected by Sandy, or if we consider different durations for the disruptions caused by the hurricane. Finally, we provide a body of evidence that firms did not advance or postpone their shipments because of the hurricane, which would have resulted in a downward bias of our results.

The corollary of this finding is that other types of frictions must relate to distance, the so-called "dark trade costs". While a proper identification of the exact nature of these "dark trade costs" is out of the scope of this paper, we see a promising explanation in the recent developments of the business networks literature. Firms tend to form links with geographically close firms (Bernard et al., 2019), and trade flows occur through this supplier-customer network (Chaney, 2018b).

A Data description

Table 3.10: List of urban CFS areas ranked by export value

CFS Area	Exports (M$)	# destinations	Imports (M$)	# origins
Los Angeles	542,064	128	442,232	128
Dallas	410,957	128	375,901	128
Chicago (IL part)	383,354	128	316,610	128
New York (NJ part)	247,206	128	200,476	127
Houston	206,714	126	220,756	128
Atlanta	194,754	129	195,717	127
New York (NY part)	191,548	126	244,328	126
San Jose	176,665	127	170,256	125
San Antonio	159,823	114	173,115	123
Detroit	151,106	125	172,690	125
Boston (MA part)	132,273	127	161,687	125
Minneapolis	129,143	127	118,134	124
Columbus	116,392	125	116,407	121
Hartford	115,541	115	78,581	112
Cleveland	101,044	128	108,495	126
Seattle	94,600	119	117,294	125
Baltimore	91,526	118	72,175	125
Miami	90,327	124	132,269	126
Indianapolis	89,229	123	95,822	122
Philadelphia (PA part)	88,497	127	105,237	126
Milwaukee	88,006	127	74,314	121
New York (CT part)	87,837	125	77,141	114
Greensboro	84,637	124	64,548	118
Denver	80,080	119	81,927	125
Philadelphia (NJ part)	79,480	120	52,582	119
Phoenix	78,989	110	97,223	124
Pittsburgh	78,730	125	84,620	122
Nashville	69,271	123	65,462	122
Tampa	69,051	115	64,279	120
Portland (OR part)	68,264	120	65,871	122
Charlotte	67,046	126	64,012	122
Birmingham	63,276	124	69,594	118
Memphis	62,386	123	55,414	116
St-Louis (MO part)	61,215	126	55,693	123
Grand Rapids	61,041	124	49,631	116
Salt Lake City	59,108	122	74,972	123
Tulsa	58,589	118	43,244	116

Continued on next page

Table 3.10 – continued from previous page

CFS Area	Exports (M$)	#destinations	Imports (M$)	# origins
Austin	58,240	109	73,495	120
Louisville	56,714	124	71,496	119
Richmond	55,188	117	47,863	116
Sacramento	51,993	97	46,030	117
Raleigh	51,891	123	41,138	118
Cincinnati (OH part)	51,204	124	54,477	123
New York (PA part)	51,144	120	35,189	115
San Diego	49,932	120	64,954	120
Greenville	46,990	127	46,454	117
Kansas City (KS part)	45,468	122	33,861	119
Fort Wayne	45,301	120	31,329	110
Kansas City (MO part)	44,629	122	50,654	119
Orlando	44,517	120	52,626	123
Jacksonville	42,242	113	41,036	116
Buffalo	41,294	122	37,281	113
Albany	37,861	116	35,996	112
Boston (RI part)	36,842	118	30,254	101
Dayton	35,604	118	37,359	114
Rochester	34,752	121	33,941	110
Cincinnati (KY part)	34,130	115	21,156	106
Oklahoma City	33,578	116	47,405	120
Knoxville	32,800	115	30,593	115
Beaumont	31,019	88	22,055	78
New Orleans	30,408	104	41,616	116
Washington (VA part)	30,179	103	50,152	115
Chicago (IN part)	30,167	117	33,467	102
Omaha	28,040	117	29,275	111
Wichita	27,777	122	27,268	104
Fresno	26,625	98	21,152	99
Boston (NH part)	26,451	116	38,625	111
Philadelphia (DE part)	23,318	110	28,025	104
Baton Rouge	22,783	99	23,319	106
Virginia Beach	21,708	115	36,504	118
Las Vegas	21,064	94	32,562	116
St-Louis (IL part)	20,841	116	28,686	102
Washington (MD part)	19,111	90	42,790	114
Savannah	18,533	93	12,985	93
Tucson	16,832	99	16,620	105
El Paso	16,563	101	22,445	116

Continued on next page

Table 3.10 – continued from previous page

CFS Area	Exports (M$)	#destinations	Imports (M$)	# origins
Laredo	15,770	23	27,639	117
Charleston	15,246	112	17,085	100
Corpus Christi	13,367	46	19,831	81
Mobile	10,940	105	19,520	95
Portland (WA part)	9,178	100	9,322	76
Lake Charles	4,888	64	6,904	62
Washington (DC part)	2,386	18	8,812	78
Total	6,395,276	9,398	6,339,951	9,492

Table 3.11: List of remainder CFS areas ranked by export value

CFS Area	Exports (M$)	# destinations	Imports (M$)	# origins
Rem. of Texas	284,705	126	292,913	127
Rem. of Pennsylvania	199,180	129	185,466	127
Rem. of Illinois	165,996	126	149,311	125
Rem. of Wisconsin	153,842	127	143,950	125
Rem. of Iowa	144,175	127	143,787	125
Rem. of Ohio	130,466	127	112,213	123
Rem. of Indiana	126,331	128	128,343	124
Rem. of Mississipi	112,931	127	92,872	124
Rem. of N. Carolina	112,255	128	87,318	125
Rem. of Arkansas	95,135	126	103,843	124
Rem. of Kentucky	93,133	126	91,022	123
Rem. of Kansas	92,449	123	83,236	115
Rem. of Michigan	92,381	126	88,959	123
Rem. of New York	90,642	124	71,392	120
Rem. of Alabama	90,361	126	86,790	122
Rem. of Georgia	89,352	126	87,019	123
Rem. of Virginia	84,327	127	70,645	123
Rem. of California	80,732	113	96,672	123
Rem. of Tennessee	78,249	127	63,660	123
Rem. of Florida	70,242	120	103,297	125
Rem. of Missouri	69,771	125	82,783	122
Rem. of Minnesota	61,816	124	69,345	114
Rem. of S. Carolina	55,104	127	57,704	122
Rem. of Nebraska	50,515	119	50,077	112
Rem. of Louisiana	49,775	115	67,341	119
Rem. of S. Dakota	46,216	119	37,273	106

Continued on next page

Table 3.11 – continued from previous page

CFS Area	Exports (M$)	# destinations	Imports (M$)	# origins
Rem. of Oklahoma	42,004	118	59,348	110
Rem. of W. Virginia	37,157	124	46,109	118
Rem. of New Mexico	34,349	109	44,012	117
Rem. of Washington	34,067	111	40,768	113
Rem. of Colorado	31,795	117	42,781	120
Rem. of N. Dakota	29,991	106	42,592	107
Rem. of Maine	29,566	109	33,835	109
Rem. of Idaho	27,306	111	35,520	112
Rem. of Connecticut	26,169	98	29,575	82
Rem. of Nevada	25,986	106	23,849	111
Rem. of Maryland	24,497	114	22,859	104
Rem. of Oregon	23,644	109	34,439	109
Rem. of Massachussets	20,541	117	28,047	105
Rem. of Vermont	17,791	116	20,200	96
Rem. of Montana	16,781	86	26,097	108
Rem. of Wyoming	12,805	85	20,991	102
Rem. of Arizona	11,658	84	20,767	101
Rem. of Delaware	9,646	70	6,823	62
Rem. of Utah	9,336	69	11,480	87
Rem. of New Hampshire	4,900	100	8,071	69
Total	3,190,068	5,292	3,245,392	5,198

Table 3.12: Descriptive statistics by NAICS code

Sector	NAICS 2007	Value	Weight	# shipments	Value per shipment	# obs.	% zero flows	# dyads	# origins	# dest.	Avg. dist.
Manufacturing											
Mining	212	41,053	1,566,560	71.2	576.2	129,227	2.9	2,233	124	129	8.4
Food manufacturing	311	660,284	400,506	122.9	5,371.8	169,855	7.3	9,616	129	129	14.9
Beverage and tobacco	312	132,326	138,662	21.2	6,240.7	44,631	13.5	3,786	121	129	11.4
Textile mills	313	23,906	5,354	2.6	9,346.8	19,424	27.6	3,605	109	129	14.6
Textile product mills	314	17,666	4,319	6.5	2,716.5	20,056	31.3	4,081	122	129	14.7
Apparel	315	8,018	374	1.3	6,157.4	4,441	41.1	1,308	95	127	15.8
Leather and allied product	316	2,632	308	0.5	5,657.4	3,298	43.6	1,051	87	127	14.8
Wood product	321	66,313	167,786	17.2	3,853.3	92,055	9.1	5,48	126	129	11.8
Paper	322	144,812	103,557	18.0	8,027.0	88,85	11.6	7,176	121	129	12.7
Printing and related activities	323	57,292	18,407	51.8	1,105.9	75,005	14.6	6,875	129	129	13.3
Petroleum and coal products	324	178,434	397,882	27.7	6,433.4	50,68	13.0	3,853	121	129	10.5
Chemical	325	408,786	267,124	49.1	8,318.0	154,292	9.3	11,014	128	129	15.1
Plastics and rubber	326	186,738	48,383	30.6	6,111.1	115,772	11.8	10,292	127	129	14.8
Non-metallic mineral product	327	82,284	581,291	47.2	1,744.4	110,911	8.6	5,936	129	129	11.3
Primary metal	331	189,467	114,581	12.3	15,406.4	61,3	15.3	6,639	123	129	13.3
Fabricated metal product	332	266,445	80,452	67.4	3,954.8	139,98	10.2	10,129	128	129	14.9
Machinery	333	270,753	27,359	27.1	9,991.8	64,297	21.9	9,999	127	129	15.8
Computer and electronic product	334	105,257	2,948	8.9	11,826.7	18,428	37.5	4,831	126	129	17.0
Electrical equipment, appliances	335	87,805	13,273	10.7	8,233.2	36,229	29.1	7,487	120	129	16.2
Transportation equipment	336	453,886	55,024	26.8	16,945.2	51,146	18.2	6,107	123	129	14.2
Furniture and related	337	61,917	12,608	23.7	2,616.5	54,007	21.4	7,801	127	129	14.7
Miscellaneous	339	71,732	6,764	26.2	2,738.9	36,088	28.7	6,866	128	129	16.5
Wholesalers											
Motor vehicle and parts	4,231	415,85	53,151	680.8	610.8	81,055	9.7	4,452	127	129	10.6
Furniture and home furnishing	4,232	57,211	13,926	45.8	1,249.1	44,206	16.6	4,312	120	129	14.0
Lumber and other construction materials	4,233	108,051	286,664	96.1	1,124.3	112,067	4.8	3,251	128	129	7.6
Commercial equip.	4,234	214,771	15,653	160.5	1,338.5	36,554	19.8	4,206	125	129	14.1
Metal and mineral	4,235	184,693	125,692	77.7	2,377.6	97,199	8.4	5,023	126	129	10.1
Electrical and electronic goods	4,236	265,768	21,182	283.0	939.0	73,628	12.7	4,998	128	129	13.8
Hardware and plumbing	4,237	105,996	15,783	153.8	689.2	103,853	7.4	4,188	129	129	10.3
Machinery, equipment and supplies	4,238	271,964	54,421	226.3	1,201.7	96,765	11.2	6,002	129	129	12.8
Miscellaneous durable goods	4,239	110,164	145,324	50.9	2,162.5	56,882	14.9	4,943	129	129	12.4
Paper and paper products	4,241	80,828	33,27	100.5	803.9	64,898	8.5	3,142	127	129	8.9
Drugs and druggists' sundries	4,242	300,474	9,946	91.8	3,271.7	22,351	19.4	2,607	120	129	12.3
Apparel and piece goods	4,243	78,127	6,728	28.8	2,713.9	16,462	29.8	3,214	114	129	15.7
Grocery and related	4,244	618,21	295,168	388.5	1,591.5	164,889	4.4	4,921	129	129	11.5
Farm product raw material	4,245	115,921	271,231	18.4	6,300.3	35,241	7.5	1,436	107	129	10.1
Chemical and allied products	4,246	136,334	83,5	100.0	1,362.8	68,623	12.0	4,867	123	129	11.0
Petroleum and petroleum products	4,247	1,186,285	1,076,171	197.3	6,011.4	75,499	3.1	1,473	127	129	5.0
Beer, wine, and distilled alcoholic	4,248	118,976	49,759	128.2	927.7	84,463	1.9	990	128	129	8.6
Miscellaneous non-durable goods	4,249	239,482	147,828	113.9	2,103.0	98,095	8.1	4,594	127	129	11.5
Electronic shopping and mail-order houses	4,541	48,548	5,763	206.7	234.9	8,223	45.9	2,757	117	129	15.5
Warehousing and storage	4,931	1,086,774	247,565	128.2	8,479.8	79,23	10.5	5,576	124	129	11.5
Newspaper, periodical and book	5,111	36,929	10,078	352.2	104.8	29,518	12.0	1,817	121	129	11.3
Direct selling establishments	45,431	35,161	32,262	69.5	506.0	100,204	0.5	529	127	129	0.8
Corporate, subsidiary, and regional offices	551,114	251,02	78,162	42.2	5,943.9	19,981	22.0	2,67	108	129	11.5

Note: Shipment values are expressed in M$; weight is in thousand US tons (kt); # shipments is expressed in millions; shipment distance is the routed distance between shipment origin and destination (in hundreds of km) as computed by the US Census Bureau.

Table 3.13: Transport modes

Modes	Values			Percents from total		
	Value	Weight	Obs.	Value	Weight	Obs.
All modes	13,837,786	11,291,584	4,546,970	100.0	100.0	100.0
Single mode	11,869,117	10,864,393	3,354,106	85.8	96.2	73.8
Truck	10,123,625	8,048,389	3,231,969	73.2	71.3	71.1
For-hire truck	6,497,910	4,291,261	1,613,317	47.0	38.0	35.5
Private truck	3,624,027	3,754,398	1,618,282	26.2	33.2	35.6
Rail	430,203	1,536,294	38,458	3.1	13.6	0.8
Water	198,192	417,915	3,691	1.4	3.7	0.1
Air (incl truck & air)	438,009	4,542	68,809	3.2	0.0	1.5
Pipeline	433,481	507,032	3,673	3.1	4.5	0.1
Multiple mode	1,968,669	427,191	1,192,864	14.2	3.8	26.2
Parcel, USPS, or courier	1,687,586	28,514	1,165,297	12.2	0.3	25.6
Truck and rail	189,271	166,900	19,070	1.4	1.5	0.4
Truck and water	14,257	30,069	2,498	0.1	0.3	0.1
Rail and water	2,341	34,268	200	0.0	0.3	0.0
Other modes	21,352	78,357	963	0.2	0.7	0.0

Note: totals may not sum due to the censoring of some observations for confidentiality reasons. Shipment values are expressed in M$, weight in thousand US tons (kt). NB: 1 US ton corresponds to 907.185 kg. **Source:** own calculations based on 2012 CFS data.

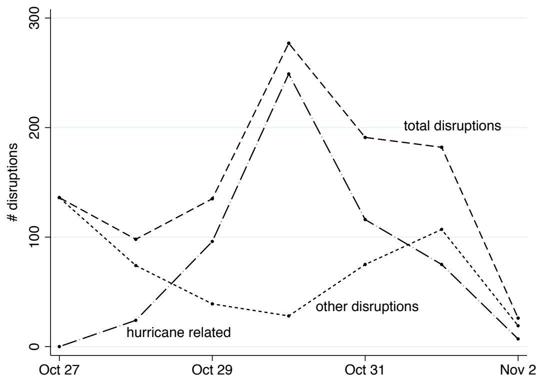
Table 3.14: Descriptive statistics - 129 origin and destination CFS areas; 218,133 dyad-industries

Variable	Mean	Std. Dev.	Median	Min	Max
Origin CFS Area					
Value	74,305.0	78,071.8	51,891.3	2,386.0	542,063.8
Weight	60,607.9	54,325.0	43,634.9	3,046.6	297,872.3
# shipments	34,201.7	40,002.4	23,329.5	1,521.3	316,378.7
Value per shipment	2.48	1.10	2.38	0.42	7.13
# obs.	24,107.4	14,047.5	20,993	2,072	88,195
Value per obs.	2.72	1.36	2.49	0.57	8.24
Share of zeros	12.26	2.89	12.20	5.71	20.89
shipment distance CFS	13.12	3.93	12.20	2.14	26.95
Destination CFS Area					
Value	74,305.0	70,187.4	52,581.6	6,822.9	442,231.5
Weight	60,607.9	54,012.4	42,287.9	4,054.5	295,574.5
# shipments	34,201.7	32,613.9	25,419.1	2,918.9	207,611.3
Value per shipment	2.34	0.88	2.28	0.46	5.15
# obs.	24,107.4	15,497.3	19,517	4,285	93,380
Value per obs.	2.82	1.05	2.66	1.18	9.09
Share of zeros	13.49	4.00	14.17	4.30	24.80
shipment distance CFS	12.89	4.89	11.01	6.26	29.45
Dyad - industry					
Value	43.9	543.9	3.1	0	131,371.1
Weight	35.8	686.2	0.5	0	136,257.5
# shipments	20.2	394.7	0.7	0	122,862.8
Value per shipment	16.47	226.44	4.16	0	51,160
# obs.	14.3	73.4	2.0	1	5,028
Value per obs.	4.16	24.21	0.95	0	4,821.9
Share of zeros	41.46	29.84	50.00	0	75
shipment distance CFS	15.18	11.70	12.34	0	56.55

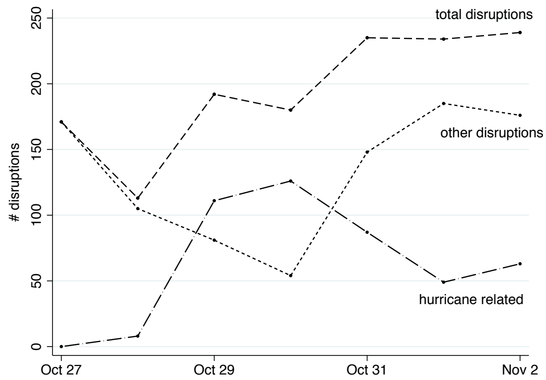
Note: shipment values and value per obs. are expressed in M$; weight is in thousand US tons (kt); # shipments and value per shipment are in thousands; distance is in hundreds of km.

B Disruptions for motorists after Sandy

Figure 3.9: Evolution over time of the number of disruptions recorded by the DoT.



(a) New-Jersey



(b) New-York

Table 3.15: Gasoline availability in the New-York City metropolitant area after Sandy

Station Response	Nov. 2	Nov. 3	Nov. 4	Nov. 5	Nov. 6	Nov. 7	Nov. 8	Nov. 9
No gasoline supply	10%	28%	24%	38%	28%	28%	21%	21%
Gasoline availability	33%	45%	59%	62%	66%	62%	72%	72%
No power at station	3%	3%	0%	0%	0%	0%	0%	0%
No contact w/station	53%	24%	17%	0%	7%	10%	7%	7%

Results of a survey implemented by the US Energy Information Administration (EIA) on gas stations in the New York City metropolitan area from the 2nd up to the 9th of November

C Derivation of equation (C)

We do not consider the variation in dark trade costs C_{ni}, so that the relative variation in trade costs τ_{ni} between the two states of the world is equal to the relative variation in transport costs T_{ni}:

$$\frac{\tau_{ni}^S}{\tau_{ni}^N} = \frac{T_{ni}^S}{T_{ni}^N}$$

Therefore, to find the change in trade costs caused by Sandy, we just need to find the change in transport costs. For this purpose, we rewrite the bilateral transport cost T_{ni} as the product of the average transport cost per km, denoted $\mathbb{E}(T_{ni})$, and the geographical distance between i and n , g_{ni}:

$$T_{ni} = \mathbb{E}(T_{ni})g_{ni}$$

During the "Sandy state of the world", transport costs increase in some areas. More precisely, they are multiplied by the "overcost parameter", κ. By definition, κ is the ratio between the transport cost per km during Sandy and the transport cost per km in the "normal state of the world":

$$\kappa = \frac{\mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 1)}{\mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 0)}$$

where $\mathbb{I}_{\text{D}}$ is a dummy variable equal to one when the concerned road segment is within an affected county (IA/PA county) and the state of the world is Sandy. The average transport cost per km in the "Sandy state of the world" is a weighted average of the transport cost per km in affected areas and the transport cost per km in non affected areas. The weights correspond to the share of the itinerary going through affected areas, s_{ni}[22] and the share of the itinerary going through non affected areas, $1 - s_{ni}$:

$$\mathbb{E}(T_{ni}^S) = s_{ni}\ \mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 1) + (1 - s_{ni})\ \mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 0)$$

which can be rewritten as:

$$\begin{aligned}
\mathbb{E}(T_{ni}^S) &= s_{ni}\ \kappa\ \mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 0) + (1 - s_{ni})\ \mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 0) \\
&= (s_{ni}(\kappa - 1) + 1)\ \mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 0)
\end{aligned}$$

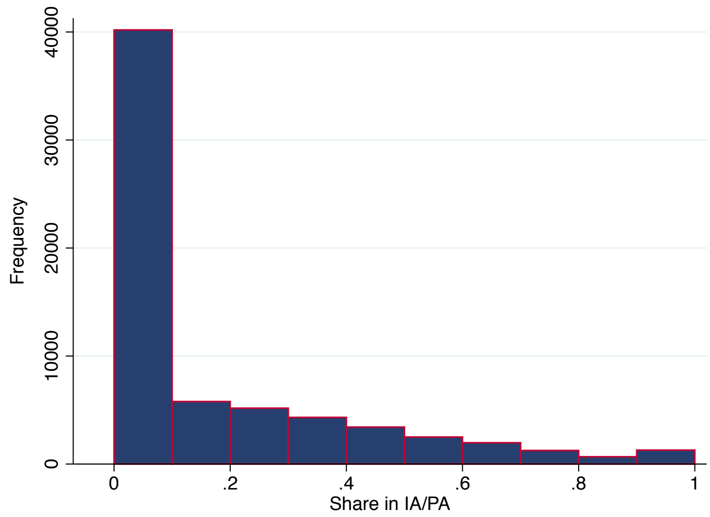
In the normal state of the world, $\mathbb{I}_{\text{D}} = 0$ everywhere, so the average transport cost is equal to the average transport cost in non affected areas: $\mathbb{E}(T_{ni}^N) = \mathbb{E}(T_{ni}|\mathbb{I}_{\text{D}} = 0)$. Therefore :

$$\begin{aligned}
\mathbb{E}(T_{ni}^S) &= (s_{ni}(\kappa - 1) + 1)\ \mathbb{E}(T_{ni}^N) \\
T_{ni}^S &= (s_{ni}(\kappa - 1) + 1)\ T_{ni}^N \\
\tau_{ni}^S &= (s_{ni}(\kappa - 1) + 1)\ \tau_{ni}^N
\end{aligned}$$

Ideally, we should redetermine the optimal path for each value of κ, because both g_{ni} and s_{ni} are affected by κ. However, for technical feasibility reasons, we have to disregard this path adjustment, so that both s_{ni} and g_{ni} are constant. In other words, for the estimation of κ (and only for this part of our work), we assume that agents do not change their path, whatever the value of κ. While not realistic, this hypothesis does not compromise our main results because we deliberately choose the path in such a way that κ will be underestimated. This downward bias on κ is obtained by using an upper bound for s_{ni} instead of the real value of s_{ni}. Indeed, overestimating the share of itinerary affected by the hurricane leads to overestimate the effect of κ on trade flows and as a consequence to underestimate κ. To find an upper bound for s_{ni}, we modify our road raster so that the cost of passing through a road cell in an IA/PA area is 10^{-6} whereas this cost for a road cell outside IA/PA areas is left unaltered at 1. As a consequence, the least cost path algorithm will choose the path that includes the largest possible share within IA/PA areas. We plot the distribution of s_{ni} in figure 3.10.

[22] i.e. road distance within IA/PA counties divided by total road distance of the itinerary

Figure 3.10: Distribution of s_{ni}



D List of the most regular industries

Table 3.16: List of the most regular industries

Sector	NAICS 2007
Food manufacturing	311
Wood product manufacturing	321
Paper manufacturing	322
Chemical manufacturing	325
Plastics and rubber products manufacturing	326
Nonmetallic mineral product manufacturing	327
Primary metal manufacturing	331
Fabricated metal product manufacturing	332
Grocery and related product merchant wholesalers	4244
Petroleum and petroleum products merchant wholesalers	4247

Bibliography

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt (2015). "Lessons from Schumpeterian growth theory." *American Economic Review* 105.5, pp. 94–99.

Aghion, Philippe and Xavier Jaravel (2015). "Knowledge Spillovers, Innovation and Growth." *Economic Journal* 583, pp. 533–573.

Agrawal, Ajay, Devesh Kapur, and John McHale (2008). "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics* 64.2, pp. 258–269.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi (2018). *Dancing with the stars: Innovation through interactions*. Tech. rep. National Bureau of Economic Research.

Akers, Noël J (2000). "The referencing of prior art documents in European patents and applications." *World Patent Information* 22.4, pp. 309–315.

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

Alcácer, Juan and Michelle Gittelman (2006). "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *Review of Economics and Statistics* 88.4, pp. 774–779.

Allen, Treb (2014). "Information Frictions in Trade." *Econometrica* 82.6, pp. 2041–2083.

Almeida, Paul and Bruce Kogut (1997). "The Exploration of Technological Diversity and the Geographic Localization of Innovation." *Small Business Economics* 9.1, pp. 21–31.

Almeida, Paul and Bruce Kogut (1999). "Localization of knowledge and the mobility of engineers in regional networks." *Management science* 45.7, pp. 905–917.

Axtell, Robert L (2001). "Zipf distribution of US firm sizes." *Science* 293.5536, p. 1818.

Bakker, Gerben (2014). "How they made news pay: news traders' quest for crisis-resistant business models."

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito (2019). "Production Networks, Geography, and Firm Performance." *Journal of Political Economy* 127.2, pp. 639–688.

Bielsa, Esperança (2008). "The pivotal role of news agencies in the context of globalization: a historical approach." *Global Networks* 8.3, pp. 347–366.

Bilir, L. Kamran and Eduardo Morales (2020). “Innovation in the Global Firm.” *Journal of Political Economy* 128.4, pp. 1566–1625.

Blake, Eric S., Todd B. Kimberlain, Robert J. Berg, John P. Cangialosi, and John L. Beven (2013). *Tropical Cyclone Report Hurricane Sandy*. National Hurricane Center.

Bloom, Nicholas, Mark Schankerman, and John van Reenen (2013). “Identifying Technology Spillovers and Product Market Rivalry.” *Econometrica* 81.4, pp. 1347–1393.

Breschi, Stefano and Francesco Lissoni (2009). “Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows.” *Journal of Economic Geography*, lbp008.

Buera, Francisco J. and Ezra Oberfield (2020). “The Global Diffusion of Ideas.” *Econometrica* 88.1, pp. 83–114.

Carayol, Nicolas, Laurent Bergé, Lorenzo Cassi, and Pascale Roux (2019). “Unintended triadic closure in social networks: The strategic formation of research collaborations between French inventors.” *Journal of Economic Behavior & Organization* 163.C, pp. 218–238.

Carvalho, Vasco M and Nico Voigtländer (2014). *Input diffusion and the evolution of production networks*. Tech. rep. National Bureau of Economic Research.

Catalini, Christian, Christian Fons-Rosen, and Patrick Gaulé (2018). *How Do Travel Costs Shape Collaboration?* Working Paper 24780. National Bureau of Economic Research.

Chaney, Thomas (2014). “The network structure of international trade.” *American Economic Review* 104.11, pp. 3600–3634.

Chaney, Thomas (2018a). “The Gravity Equation in International Trade: An Explanation.” *Journal of Political Economy* 126.1, pp. 150–177.

Chaney, Thomas (2018b). “The Gravity Equation in International Trade: An Explanation.” *Journal of Political Economy* 126.1, pp. 150–177.

Combes, Pierre-Philippe and Miren Lafourcade (2005). “Transport costs: measures, determinants, and regional policy implications for France.” *Journal of Economic Geography* 5.3, pp. 319–349.

Comin, Diego, Mikhail Dmitriev, and Esteban Rossi-Hansberg (2012). *The Spatial Diffusion of Technology*. NBER Working Papers 18534. National Bureau of Economic Research, Inc.

Comin, Diego and Bart Hobijn (2010). “An Exploration of Technology Diffusion.” *American Economic Review* 100.5, pp. 2031–59.

Dasgupta, Kunal and Jordi Mondria (2018). “Inattentive importers.” *Journal of International Economics* 112, pp. 150–165.

Dickstein, Michael J and Eduardo Morales (2018). “What do exporters know?” *The Quarterly Journal of Economics* 133.4, pp. 1753–1801.

Duguet, Emmanuel and Megan MacGarvie (2005). “How well do patent citations measure flows of technology? Evidence from French innovation surveys.” *Economics of Innovation and New Technology* 14.5, pp. 375–393.

Duranton, Gilles, Peter Morrow, and Matthew Turner (2014). "Roads and Trade: Evidence from the US." *The Review of Economic Studies* 81.2, pp. 681–724.

Ejrnæs, Mette and Karl Gunnar Persson (2010). "The gains from improved market efficiency: trade before and after the transatlantic telegraph." *European Review of Economic History* 14.3, pp. 361–381.

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

Estevadeordal, Antoni, Brian Frantz, and Alan M Taylor (2003). "The rise and fall of world trade, 1870–1939." *The Quarterly Journal of Economics* 118.2, pp. 359–407.

Feyrer, James (2011). *Distance, Trade, and Income – The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment*. 2011 Meeting Papers 1438. Society for Economic Dynamics.

Fouquin, Michel and Jules Hugot (2016). *Two Centuries of Bilateral Trade and Gravity Data: 1827-2014*. Working Papers 2016-14. CEPII.

Gabaix, Xavier (2016). "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30.1, pp. 185–206.

Ghiglino, Christian (2012). "Random walk to innovation: Why productivity follows a power law." *Journal of Economic Theory* 147.2, pp. 713–737.

Gibrat, Robert (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.

Gourieroux, C., A. Monfort, and E. Renault (1993). "Indirect inference." *Journal of Applied Econometrics* 8.S1, S85–S118.

Griffith, Rachel, Sokbae Lee, and John Van Reenen (2011). "Is distance dying at last? Falling home bias in fixed-effects models of patent citations." *Quantitative economics* 2.2, pp. 211–249.

Head, Keith, Yao Amber Li, and Asier Minondo (2019). "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics." *Review of Economics and Statistics* 101.4, pp. 713–727.

Head, Keith and Thierry Mayer (2013). "What separates us? Sources of resistance to globalization." *Canadian Journal of Economics* 46.4, pp. 1196–1231.

Head, Keith and Thierry Mayer (2014a). "Gravity Equations: Workhorse, Toolkit, and Cookbook." *the Handbook of International Economics*. Vol. 4. Elsevier. Chap. 3, pp. 131–195.

Head, Keith and Thierry Mayer (2014b). "Gravity Equations: Workhorse,Toolkit, and Cookbook." *Handbook of International Economics*. Vol. 4. Elsevier. Chap. Chapter 3, pp. 131–195.

Hortaçsu, Ali, F. Asís Martínez-Jerez, and Jason Douglas (2009). "The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre." *American Economic Journal: Microeconomics* 1.1, pp. 53–74.

Hugot, Jules and Camilo Umana Dajud (2016). “Trade Costs and the Suez and Panama Canals.” *CEPII Working Paper* 29.

Hummels, David and Georg Schaur (2013). “Time as a Trade Barrier.” *American Economic Review* 103.7, pp. 2935–59.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger (2018). “Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science.” *Quarterly Journal of Economics* 133.2, pp. 927–991.

Jackson, Matthew and Brian Rogers (2007). “Meeting Strangers and Friends of Friends: How Random Are Social Networks?” *American Economic Review* 97.3, pp. 890–915.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty (2000). “Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors.” *American Economic Review, Papers and Proceedings* 90.2, pp. 215–218.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993). “Geographic localization of knowledge spillovers as evidenced by patent citations.” *Quarterly Journal of Economics*, pp. 577–598.

Jensen, Robert (2007). “The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector.” *The Quarterly Journal of Economics* 122.3, pp. 879–924.

Juhász, Réka and Claudia Steinwender (2018). “Spinning the web: Codifiability, information frictions and trade.” *NBER working paper*.

Keller, Wolfgang and Stephen Ross Yeaple (2013). “The Gravity of Knowledge.” *American Economic Review* 103.4, pp. 1414–44.

Kerr, William R (2008). “Ethnic scientific communities and international technology diffusion.” *Review of Economics and Statistics* 90.3, pp. 518–537.

Lampe, Ryan (2012). “Strategic citation.” *Review of Economics and Statistics* 94.1, pp. 320–333.

Lanjouw, Jean, Ariel Pakes, and Jonathan Putnam (1998). “How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data.” *Journal of Industrial Economics* 46.4, pp. 405–32.

Lei, Zhen and Brian D. Wright (2017). “Why weak patents? Testing the examiner ignorance hypothesis.” *Journal of Public Economics* 148, pp. 43–56.

Lenoir, Clemence, Julien Martin, and Isabelle Mejean (2020). “Search Frictions in International Good Markets.”

Lew, Byron and Bruce Cater (2006). “The telegraph, coordination of tramp shipping, and growth in world trade, 1870-1910.” *European Review of Economic History* 10.2, pp. 147–173.

Li, Yao Amber (2014). “Borders and distance in knowledge spillovers: Dying over time or dying with age?—Evidence from patent citations.” *European Economic Review* 71, pp. 152–172.

Malgouyres, Clément, Thierry Mayer, Clément Mazet-Sonilhac, and Juan Carluccio (2020). "Technology-induced Trade Shocks? Evidence from Broadband Expansion in France." *mimeo*.

Maurseth, Per Botolf and Bart Verspagen (2002). "Knowledge Spillovers in Europe: A Patent Citations Analysis." *Scandinavian Journal of Economics* 104.4, pp. 531–45.

Mayer, Thierry and Soledad Zignago (2011). *Notes on CEPII's distances measures: The GeoDist database*. Working Papers. CEPII research center.

McFadden, Daniel (1973). "Conditional logit analysis of qualitative choice behavior." *Frontiers in Econometrics*. Ed. by P. Zarembka. Academic Press: New York. Chap. 4, pp. 105–142.

Mildenhall, Stephen, Steve Jakubowski, Siamak Daneshvaran, and Steve Bowen (2013). *Hurricane Sandy Event Recap Report*. AON Benfield.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura (2014). "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *Review of Economics and Statistics* 96.5, pp. 967–985.

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

Pascali, Luigi (2017). "The wind of change: Maritime technology, trade, and economic development." *American Economic Review* 107.9, pp. 2821–54.

Peri, Giovanni (2005). "Determinants of knowledge flows and their effect on innovation." *Review of Economics and Statistics* 87.2, pp. 308–322.

Serafinelli, Michel (2019). ""Good" Firms, Worker Flows, and Local Productivity." *Journal of Labor Economics* 37.3, pp. 747–792.

Silva, J. M. C. Santos and Silvana Tenreyro (2006). "The Log of Gravity." *Review of Economics and Statistics* 88.4, pp. 641–658.

Singh, Jasjit (2005). "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51.5, pp. 756–770.

Sotelo, Sebastian (2017). "Practical Aspects of Implementing the Multinomial PML Estimator."

Steinwender, Claudia (2018). "Real Effects of Information Frictions: When the States and the Kingdom Became United." *American Economic Review* 108.3, pp. 657–96.

Thompson, Peter (2006). "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations." *Review of Economics and Statistics* 88.2, pp. 383–388.

Thompson, Peter and Melanie Fox-Kean (2005). "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review* 95.1, pp. 450–460.

Unesco (1952). *News Agencies, Their Structure and Operation*. Greenwood Press.

Volpe Martincus, Christian and Juan Blyde (2013). “Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment.” *Journal of International Economics* 90.1, pp. 148–161.

Wenzlhuemer, Roland (2013). *Connecting the Nineteenth-Century World*. Cambridge University Press.

Wolff, Jacques (1991). “Structure, fonctionnement et évolution du marché international des nouvelles. Les agences de presse de 1835 à 1934.” *Revue Economique* 42.3, pp. 575–601.

Wrona, Jens (2018). “Border effects without borders: What divides Japan’s internal trade?” *International Economic Review* 59.3, pp. 1209–1262.

Institut d'études politiques de Paris
ÉCOLE DOCTORALE DE SCIENCES PO
Programme doctoral en économie
Département d'économie
Doctorat en sciences économiques

Trois essais sur les frictions spatiales

Pierre Cotterlaz

Thèse dirigée par
Thierry Mayer, Professeur des universités, IEP de Paris
soutenue le 3 juin 2021

Jury:
Maria Bas, Professeur des universités, CES, Université Paris 1 Panthéon-Sorbonne (Rapporteur)
Matthieu Crozet, Professeur des universités, Université Paris-Saclay (Rapporteur)
Anne-Célia Disdier, Professeur associé, Paris School of Economics, et Directrice de recherche, INRAE
Cécile Gaubert, Associate Professor, University of California, Berkeley
Emeric Henry, Professeur des universités, IEP de Paris
Thierry Mayer, Professeur des universités, IEP de Paris

Résumé

Trois Essais sur les Frictions Spatiales

Les frictions spatiales jouent un rôle crucial dans l'explication de nombreux phénomènes économiques. Dans cette thèse, nous étudions les origines, la prévalence et les conséquences de telles frictions à travers trois exemples.

Dans le premier chapitre, nous nous intéressons aux frictions spatiales pesant sur la diffusion de la connaissance. Nous expliquons l'effet négatif de la distance sur les flux de citations entre brevets par la structure des réseaux d'innovation. Nous montrons que la connaissance percole: les entreprises tendent à citer davantage les nouveaux brevets de leurs contacts existants, et à former de nouveaux liens avec des contacts de leurs contacts. Incorporer cette percolation dans un modèle de formation de réseau permet de rationaliser le lien négatif entre diffusion de la connaissance et distance.

Dans le second chapitre, nous explorons les liens entre frictions informationnelles et commerce international. Nous utilisons le contexte spécifique du XIXe siècle, au cours duquel émergent des agences de presse mondiales, facilitant l'acquisition par les acteurs économiques d'informations sur les marchés étrangers. Nous montrons que deux pays commercent davantage une fois qu'ils bénéficient de ce choc positif sur la capacité à obtenir de l'information. Les agences de presse s'insèrent donc parmi les nombreux facteurs explicatifs de la Première Mondialisation.

Le dernier chapitre cherche à déterminer si les coûts de transport constituent l'essentiel des obstacles au commerce à l'intérieur d'un pays. Alors qu'en matière de commerce international il est établi que les coûts du commerce ne se limitent pas aux coûts de transport, on dispose de moins d'éléments pour le commerce intra-national. Nous utilisons l'ouragan Sandy comme une expérience naturelle à l'origine d'une hausse des coûts de transport pour les flux transitant par certaines zones, et montrons que l'élasticité intra-USA des flux commerciaux à la distance serait bien plus faible si les coûts de transport étaient les seuls responsables de cette élasticité.

Chapitre 1: La Percolation de la Connaissance dans l'Espace

Malgré les améliorations considérables des technologies de l'information et de la communication au cours des trois dernières décennies, la distance géographique demeure un obstacle majeur à la diffusion de la connaissance. Nous estimons l'élasticité des flux internationaux de citations de brevets par rapport à la distance et trouvons qu'elle est restée stable entre 1980 et 2010, fluctuant autour de -0.3, ce qui signifie qu'une hausse de 10% de la distance entre deux pays est associée à une diminution de 3% des flux de citations entre ces deux pays. L'existence d'une élasticité négative et statistiquement significative est surprenante puisque les idées ne sont pas soumises aux frictions spatiales habituellement associées à la distance, telles que les coûts de transport où les droits de douane. La stabilité de cette élasticité interroge dans la mesure où la digitalisation et les technologies de communication, avec par exemple l'émergence d'outils de recherche en ligne des brevets, semblent n'avoir eu aucun effet sur les tendances agrégées de diffusion de la connaissance.

Ce chapitre montre que la dynamique de formation des réseaux au cours du cycle de vie de l'entreprise est cruciale pour la compréhension de l'effet agrégé de la distance: les entreprises jeunes de taille modeste ont des contacts spatialement proches, et leur réseau s'étend graduellement à mesure qu'elles croissent.

Notre contribution intervient en deux étapes. Dans un premier temps, nous étudions la manière dont se forment les liens entre innovateurs: nous mettons en évidence un phénomène appelé en économie des réseaux “fermeture triadique”, propriété selon laquelle les entreprises tendent à former davantage de liens avec des entreprises situées à deux nœuds de distance dans le réseau (en d’autres termes, avec des contacts de contacts). Pour dévoiler ce mécanisme, nous reconstituons le réseau à partir des citations de brevet, et évaluons l’influence de ce réseau sur la probabilité de formation d’un nouveau lien. Cela nous permet d’établir que les entreprises se réfèrent plus facilement à des connaissances générées par des entreprises qu’elles ont déjà citées dans le passé (leurs contacts), ainsi que par des entreprises citées par des entreprises qu’elles ont déjà citées (les contacts de leurs contacts). Ce processus de diffusion rappelle le phénomène de percolation en physique, puisque les idées apparaissent comme un fluide trouvant son chemin d’un innovateur à l’autre en suivant les liens d’un réseau.

Cette mise en évidence de la diffusion des idées entre entreprises via leur réseau proche repose sur une stratégie d’identification inédite. Nous utilisons le fait que certaines citations sont faites par le demandeur du brevet lui-même, tandis que d’autres sont ajoutées par l’examinateur du brevet. L’union de ces deux ensembles correspond aux citations qui auraient été réalisées dans un monde contre-factuel sans friction sur la diffusion de la connaissance. Nous estimons l’effet d’un lien direct sur la probabilité d’être cité par l’entreprise déposant le brevet elle-même, plutôt que par l’examinateur du brevet. Nos résultats montrent que les entreprises ont une probabilité 1.5 fois plus élevée que les examinateurs de citer des brevets détenus par leurs contacts, avec des effets hétérogènes selon la taille de l’entreprise. En outre, les entreprises ont une probabilité 35% plus élevée de citer des brevets qui ont été auparavant directement cités par leurs contacts. Ces effets persistent lors de la réalisation d’une large gamme de tests de robustesse.

Dans un deuxième temps, nous montrons les conséquences de ce mécanisme de formation des réseaux lorsque l’on adopte un point de vue plus agrégé, et en particulier comment il est suffisant pour expliquer l’effet de la distance sur les flux de connaissance. Pour ce faire, nous incorporons le processus de diffusion décrit ci-dessus dans un modèle. Les entreprises y croissent car leur réseau s’étend au fil du temps. Elles sont de moins en moins affectées par la distance à mesure que leur taille et leur âge augmentent, car elles ont eu davantage de temps pour étendre leur réseau. Ce modèle débouche sur deux prédictions, l’une sur la distribution des tailles des entreprises, l’autre sur la relation entre taille de l’entreprise et distance des citations de cette entreprise, qui combinées génèrent un effet agrégé de la distance. La première prédiction est que la distribution des tailles des innovateurs suit une loi de Pareto. La deuxième est qu’une fonction puissance lie la moyenne des carrés des distances auxquelles l’entreprise cite et la taille de l’entreprise.

Ces prédictions sont vérifiées dans les données. En plus d’être suffisantes pour générer une élasticité négative et constante, elles constituent en elles-mêmes des faits stylisés dignes d’intérêt. En effet, nous montrons que, au-delà d’une loi de Pareto, la distribution des tailles des innovateurs est empiriquement bien décrite par une loi de Zipf, ce qui la rattache aux nombreux objets économiques suivant cette loi. De la même manière, l’existence d’une relation systématique entre la taille d’un innovateur et la distance à laquelle se situe la connaissance qu’il utilise n’était pas documentée jusqu’à présent, et nous montrons en outre que cette relation est vérifiée dans des contextes variés, tant en cross-section qu’en panel.

Une conclusion importante de ce chapitre est que les petites entreprises sont les principales contributrices à l’effet agrégé de la distance. Au début de leur cycle de vie, les innovateurs mobilisent des connaissances produites par des contacts situés près d'eux, et au fur et à mesure de leur croissance ils tissent des liens avec des innovateurs plus lointains à travers leur réseau. Nous trouvons que malgré la stabilité de l'effet de la distance au cours du temps, le lien entre taille et distance s'est amoindri pendant la période que nous étudions, en grande partie parce que les petits innovateurs sont devenus capables d'accéder à des connaissances plus lointaines. Cela aurait dû induire une baisse de l'effet global de la distance, mais semble avoir été compensé par une hausse de la part des petits innovateurs au détriment des grands innovateurs.

Le mécanisme de formation du réseau que nous mettons en évidence est suffisamment général pour englober la plupart des explications habituellement avancées pour le caractère local des transferts de connaissance: il peut correspondre à des accords formels de collaboration R&D, mais également à des liens associés à une proximité culturelle ou ethnique (Agrawal et al., 2008; Kerr, 2008), à une mobilité inter-entreprises des ingénieurs (Almeida and Kogut, 1999; Breschi and Lissoni, 2009; Serafinelli, 2019) ou à des relations fournisseur-client (Carvalho and Voigtländer, 2014).

Chapitre 2: Information et Première Mondialisation: Agences de Presse et Commerce

Comme la connaissance, l'information ne se diffuse pas parfaitement d'un pays à l'autre. Ces frictions informationnelles sont susceptibles de constituer une entrave aux échanges internationaux, puisque la connaissance des caractéristiques des marchés étrangers (taille, prix, coûts du commerce et autres déterminants de la demande) est cruciale pour les exportateurs, et que pour les importateurs le choix du fournisseur dépend de l'information disponible sur les prix et la qualité des produits de différents marchés.

Dans le second chapitre, nous utilisons l'exemple historique de l'émergence des agences de presse mondiales pour quantifier les effets sur les flux commerciaux d'une facilitation de la circulation de l'information.

Les agences de presse collectent de l'information et la revendent aux médias (dans notre contexte, des journaux). Elles leur permettent d'enrichir leur contenu sur des pays qu'ils ne seraient pas capables de couvrir par leurs propres moyens. Au milieu du XIXe siècle, avec l'essor de la presse de masse, apparaissent les premières agences de presse mondiales. Le marché se structure rapidement sous la forme d'un oligopole où trois agences de presse dominantes s'entendent pour se partager les marchés nationaux. Dans ce cadre, elles s'accordent pour partager leurs informations et dissuader ainsi l'entrée de potentiels concurrents en s'assurant une information plus exhaustive que celle d'agences exclues de l'accord. Lorsque deux pays sont couverts par des agences de presse membres de cet accord, l'information circule donc plus facilement entre eux.

Le développement des agences de presse est intimement lié à la construction d'un réseau télégraphique international: les agences de presse utilisaient le télégraphe pour communiquer et ont fréquemment contribué à son expansion. Le télégraphe représentait une amélioration considérable par rapport aux précédentes technologies de communication, avec des délais de transmission plus courts et moins volatils. Cependant, bien qu'il rende les communications plus faciles, il ne donne pas accès en tant que tel à une source centralisée et fiable d'informations publiques. En effet, les messages télégraphiques sont privés, et il est facile d'en restreindre l'accès à quelques utilisateurs. A l'inverse, les agences de presse collectent et vendent de l'information qui peut ensuite être utilisée par chacun à un coût quasi nul. En d'autres termes, en l'absence d'agences de presse, les télégraphes réduisent les coûts de coordination et de communication, sans grand effet sur la quantité d'information dont dispose un large public. Nous utilisons cette distinction entre un usage du télégraphe facilement exclusif et la nature quasi publique de l'information fournie par les agences de presse pour séparer l'effet d'une baisse des coûts de coordination / communication de l'effet d'une amélioration de l'accès à l'information publique, une séparation que les précédentes études n'avaient pas la possibilité de faire.

Tous les pays n'ont pas été couverts simultanément par les télégraphes et les agences de presse. Bien que le succès du télégraphe ait été immédiat, le coût des infrastructures et des facteurs techniques ont rendu impossible une connexion rapide de l'ensemble des pays. De la même manière, les agences de presse mondiales n'ont pas immédiatement étendu leurs opérations au monde entier. Elles ont commencé par se partager l'Europe, et ont ensuite graduellement élargi le rayon de leur accord en cinq vagues successives, en 1859, 1867, 1876, 1889 et 1902. Cette entrée séquentielle des paires de pays dans le réseau des agences de presse et du télégraphe est clé pour notre stratégie d'identification, puisqu'elle nous permet d'estimer une équation de gravité en panel, avec en plus des habituels effets fixes "origine × année" et "destination × année" des effets fixes "paire de pays" qui contrôlent pour toutes les caractéristiques constantes au cours du temps des deux pays. Nos estimés reflètent donc l'augmentation des flux associée au choc positif d'information, purgée entre autres des variations agrégées de production de l'exportateur ou de dépenses de l'importateur, ainsi que de tout déterminant statique des flux commerciaux.

Notre approche pour capturer le pur effet de l'information est de de nous concentrer sur l'interaction entre télégraphe et agences de presse: alors que l'effet du télégraphe traduit la diminution des coûts de communication, l'interaction isole spécifiquement la contribution d'une amélioration de l'accès à l'information entre les deux pays. L'effet est substantiel: nous estimons que la valeur des flux commerciaux augmente de 30% supplémentaires lorsque deux pays sont inclus dans le réseau global de partage des nouvelles, en plus d'être reliés par le télégraphe. Nos résultats confirment également les estimés de précédentes études qui documentaient l'effet positif du télégraphe sur le commerce: nous trouvons que, en l'absence de couverture par une agence de presse, les flux commerciaux augmentent de 40% lorsque deux pays deviennent connectés par le télégraphe. Cependant, les agences de presse, en l'absence de télégraphe, ne sont pas associées à une hausse significative du commerce, ce qui suggère qu'elles étaient incapables d'opérer de manière satisfaisante lorsqu'elles étaient privées d'une technologie de communication adéquate.

Nous analysons ensuite la dynamique temporelle de l'effet à travers une "event-study", et trouvons que sa magnitude augmente progressivement, jusqu'à une trentaine d'années après la connexion de la dyade. Cette image est cohérente avec la lente constitution de réseaux commerciaux entre les pays qui ont bénéficié d'un accès amélioré à l'information. Enfin, nous mettons en évidence des résultats soutenant l'hypothèse que le surcroît de commerce est bien lié à une information plus abondante sur les pays étrangers concernés. En premier lieu, les flux bilatéraux deviennent plus volatils après que les deux pays sont connectés. Comme Steinwender, 2018 le montre, cette observation est cohérente avec un scénario où les partenaires s'adaptent davantage aux conditions du marché. Deuxièmement, en utilisant un corpus de textes de journaux français, nous mesurons une augmentation du nombre de mentions d'un pays étranger dans la presse française lorsque l'une des agences de presse mondiales commence à opérer dans ce pays et lorsqu'il devient relié à la France par une liaison télégraphique.

La diminution des frictions informationnelles est ainsi l'un des nombreux facteurs ayant contribué à la hausse soutenue des échanges internationaux pendant la seconde moitié du XIXe siècle (la "Première Globalisation"). Bien qu'obtenu à partir d'un évènement historique distant, ce résultat reste pertinent pour analyser le commerce contemporain, puisque l'information n'est toujours pas complète en dépit des améliorations considérables apportées aux technologies de communication. Ce chapitre ne tranche pas sur le mécanisme précis par lequel une hausse de la quantité d'information disponible affecte le commerce. Cependant, le fait que l'effet continue à augmenter progressivement pendant une longue période suggère que l'amélioration de la circulation de l'information a pu actionner des mécanismes agissant sur une durée longue, tels que les Investissements Directs à l'Etranger, les migrations humaines internationales ou même une convergence des goûts culturels.

Chapitre 3: Commerce et Coûts de Transport: l'Exemple de l'Ouragan Sandy

Les flux de commerce internationaux décroissent fortement lorsque la distance augmente, et seule une partie de cette baisse peut être attribuée aux coûts de transport. Cela témoigne de la présence d'autres coûts du commerce, "noirs" car non observables mais nécessaires pour expliquer la structure gravitaire des flux commerciaux (Head and Mayer, 2013). Les sources potentielles de ces frictions sont multiples. Elles incluent par exemples des différences de goût ou de culture, un manque de confiance mutuelle, et l'imparfaite diffusion spatiale de l'information (évoquée dans les deux premiers chapitres). On s'attend à ce que ces ces coûts noirs du commerce soient moins élevés au sein d'un pays qu'entre les pays: la culture et les goûts y sont plus similaires, l'information s'y diffuse plus facilement, et la confiance mutuelle y est plus importante. En outre, les droits de douane et autres coûts "gris" (barrières non tarifaires) associés au franchissement d'une frontière nationale y sont absents. Néanmoins, ce chapitre montre qu'une partie seulement de l'élasticité à la distance des flux commerciaux internes aux USA peut être attribuée aux coûts de transport, impliquant l'existence de sources additionnelles de frictions spatiales à l'intérieur même des pays.

Plus précisément, nous trouvons que si l'élasticité totale des flux intra-USA à la distance est de -0.84, cette élasticité serait bien plus faible, autour de -0.06, si les coûts de transport étaient les seuls obstacles au commerce. Ce résultat est établi à l'aide d'une expérience naturelle: l'ouragan Sandy, qui a frappé le Nord-Est des Etats-Unis fin octobre 2012. Cet ouragan a causé des dommages majeurs sur les infrastructures routières, à l'origine d'une hausse des coûts de transport dans les secteurs touchés. En fonction du chemin optimal les reliant, certaines paires de villes (dyades) sont plus affectées que d'autres: les dyades pour lesquelles une part importante du trajet habituel traverse les zones dévastées par l'ouragan subissent une hausse de coût de transport plus importante que les dyades pour lesquelles ce trajet évite les zones touchées. Par exemple, les coûts de transport entre Los Angeles et Seattle ne sont pas affectés, alors que les coûts de transport entre Boston et Miami le sont. Nous calculons une borne inférieure pour l'équivalent en termes de distance routière de ce changement de coûts de transport et régressons dans une équation de gravité en panel les flux commerciaux intra-USA sur cette mesure de distance qui varie au cours du temps. L'effet de la distance obtenu ainsi est plus faible que l'effet de la distance obtenu par l'estimation d'une équation de gravité en cross-section, ce qui confirme que l'élasticité des flux à la distance en cross-section incorpore des coûts du commerce distincts des coûts de transport.

Le changement de coûts de transport induit par Sandy est calculé à partir d'un algorithme de plus court chemin. Nous décomposons le réseau routier américain en une grille, où le franchissement de chaque cellule est associée à un coût, et cherchons le chemin entre deux points pour lequel le coût est minimisé. Un paramètre clé dont nous avons besoin pour évaluer les effets de Sandy est le "paramètre de surcoût", qui indique dans quelle mesure le coût de franchissement augmente dans les zones affectées par Sandy. Ce paramètre est estimé par une méthode d'inférence indirecte: nous minimisons la différence entre des moments observés et des moments prédits par un modèle de gravité structurelle.

Notre résultat sur l'incapacité des coûts de transport seuls à expliquer l'ensemble de l'élasticité des flux à la distance à l'intérieur des USA reste valide lorsque l'on exclut les dyades pour lesquelles le changement de coût de transport bilatéral que nous calculons pourrait être moins précis. Il subsiste également si nous optons pour un périmètre plus restrictif pour définir les zones affectées par Sandy, ou sous différentes hypothèses sur la durée des dommages causés par l'ouragan. En outre, nous montrons que les entreprises n'ont ni avancé ni reculé dans le temps leurs envois en raison de l'ouragan, ce qui aurait été à l'origine d'un biais à la baisse de nos résultats. En revanche, nous laissons ouverte pour des recherches ultérieures la question de l'identification précise des mécanismes par lesquels ces coûts noirs du commerce opèrent.