Title
Knowledge Spillovers through Networks of Scientists

Author
Zacchia, Paolo

Publication Date
2015

Peer reviewed|Thesis/dissertation

Knowledge Spillovers through Networks of Scientists

by

Paolo Zacchia

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Enrico Moretti, Chair
Patrick M. Kline
Benjamin R. Handel
Brian Wright

Summer 2015

Knowledge Spillovers through Networks of Scientists

Abstract

Knowledge Spillovers through Networks of Scientists

by

Paolo Zacchia

Doctor of Philosophy in Economics

University of California, Berkeley

Enrico Moretti, Chair

In this thematic dissertation, which is divided in three chapters, I discuss new evidence and insights about knowledge spillovers, that is the process by which new productive ideas benefit individuals and organizations that did not originally devise them. The general theme of the dissertation is the empirical search for one specific micro-level mechanism of information diffusion: that is, social interactions between inventors who are connected in a network of professional collaborations. In the first and second chapter I examine how inventor-level networks function as a channel of R&D spillovers across firms. The two chapters separately introduce two different and novel methodologies to estimate spillover effects in this setting. Both methods stem, however, from a shared theoretical framework, and the resulting econometric estimates are similar across the two approaches. In the third chapter I analyze a different but related issue: that is how geographical proximity facilitates knowledge exchange between inventors, in particular between "superstar" scientists and their coauthors. In the remainder of this abstract I provide a brief summary of each of the three chapters.

The first chapter illustrates a new methodology of quantitatively assessing R&D spillovers across firms in a reduced form fashion. As mentioned, I directly test the hypothesis that interactions between inventors and scientists of different organizations drive knowledge spillovers. To this end, I construct a network of publicly traded companies where each link is a function of the relative proportion of any two firms' inventors who have patent coauthors in both said organizations. I use this measure of connection to weigh the impact of R&D performed by each firm on the innovation and productivity outcomes of its neighbors in the network. An empirical concern is that the resulting estimates may reflect, rather than genuine spillover effects, some unobserved, simultaneous drivers of both R&D and firm performance, which are common to closely connected firms. By formally modeling the strategic dependence of R&D choices in the network I characterize the conditions under which specific instrumental variables, based on the R&D choices of more distantly related firms ("indirect friends"), solve the problem. Empirical results show that substantial effects of external R&D on firm performance and innovation are identified by this approach. I calculate the resulting marginal social return of R&D to be approximately 24% of the corresponding marginal private return.

In the second chapter I adopt a more structural approach to estimate the parameters of the model from the first chapter. Instead of quantifying spillovers in terms of their ultimate effects on firm-level productivity and innovation outcomes, I estimate the effect of connected firms' R&D in stimulating private R&D investment strategies. On the basis the equilibrium conditions of the R&D investment game, formulated under general assumptions, I establish a set of moment conditions of both the first and second order, relative to equilibrium R&D, that are exploited for estimation. The model separately identifies spillover effects and the variance of common shocks thanks to a zero conditional covariance restriction. Its intuition is summarized as follows: for any two firms that are only indirectly related in the network, their reciprocal strategic dependence is removed by holding constant the choices of their intermediate links. Thus, any residual cross-correlation in R&D must be driven by shared external factors, whose size can be estimated. The empirical results evidence that the R&D of connected firms positively stimulates private investment with an elasticity of about 0.2.

In the third and final chapter I analyze spillover effects on the production of patents following episodes in which superstar inventors relocate to a new city. In particular, in order to determine whether local externalities have a restricted network dimension or a wider spatial breadth, I estimate changes in patterns of patenting activity for two different groups of inventors: the restricted group of coauthors of the superstar, and all other inventors from one specific urban area. The analysis is performed for both the locality where the superstar moves and for the one that is left. I restrict the attention to outcome measures of patent output that exclude any joint work with the superstar, so to isolate spillovers from complementarity effects. The results from the event study evidence a large and persistent positive effect on the coauthors of the superstar who reside in the city of destination (averaging about 0.1 more patents per inventor each year), and a negative trend affecting those who live in the locality of departure. These effects increase with the relative ranking of the superstar in the patent distribution. Conversely, no city-wide spillover effect can be attested, offering little support to place-based policies aimed at generating a positive influx of highly skilled individuals.

Iustum et tenacem propositi virum
non civium ardor prava iubentium,
non voltus instantis tyranni
mente quatit solida neque Auster,
dux inquieti turbidus Hadriae,
nec fulminantis magna manus Iovis:
si fractus inlabatur orbis,
inpavidum ferient ruinae.

Horatius, Carmen III, 3, 1-8

Contents

List of Figures

List of Tables

Acknowledgments

I express my gratitude to Enrico Moretti for his continued and patient supervision of this project through its multiple stages, and the support he provided over the most critical parts of the Ph.D. program. I am no less indebted to Bryan Graham, Benjamin Handel, Patrick Kline and Brian Wright for all the advice and guidance given with their parallel supervision: each of them shared with me his own distinct ideas, all of which have proven to be uniquely valuable. Thanks are due to the entire faculty of the U.C. Berkeley Economics Department, in particular to David Card, Bronwyn Hall, Schachar Kariv, Edward Miguel, Jesse Rothstein, Gérard Roland and Christopher Walters for having fostered, often dialectically, my personal and professional growth since the coursework stage through the phase of transition to research until the writing of this dissertation. I remark or extend my appreciation for the suggestions given by Nicholas Bloom, Irina Denisova, David Domeij, Sergej Izmalkov, Petra Moser, John Van Reenen, Massimo Riccaboni, Martin Watzinger and all participants to seminars held at Berkeley, at the 2014 Munich Conference in Innovation and Competition, at the 2014 IWCEE workshop in Rome, at the Bank of Italy, the Russian School of Economics, the Stockholm School of Economics, and IMT Lucca where parts of this work have been presented.

I am thankful to all fellow students I have met at Berkeley during my graduate studies. Among them stands out Santiago Pereda Fernández, an extremely smart colleague as well as a unique friend. I recognize that the amount of academic and personal spillovers exchanged between the two of us over the years is not best described by a symmetric measure, and in particular I acknowledge his fundamental input for many of the results and intuitions discussed in the second chapter of the present dissertation; in those cases, the use of the first singular personal pronoun in the text is to be interpreted in a reversely emphatic sense, as some sort of *singulare minoritatis*. I am looking forward to transform that chapter into a unique, joint research work that best complements our respective ideas, intuitions and strengths. Special thanks are also due to Dorian Carloni, Francesco D'Acunto, Simon Galle, Tadeja Gračner, Daniel Gross, Hedvig Horvath, Jevgenija Jarmoš, Elena Manresa, Takeshi Murooka, Carl Nadler, Markus Pelger, Raffaele Saggio, David Silver, Sinaia Urrusti-Frenk, José Pablo Vásquez-Carvajal and Michael Weber for our reciprocal exchange of ideas and moral support over the years. I wish to all of them to achieve their aspired results in their professional future, to reflect their proven intellectual and personal stature.

A different type of gratitude shall be acknowledged to all the people who are dear to my heart, foremost my family: Silvia, Maria Grazia, Marco, Giovanna, Rosanna, Andrea and Silvia. Notwithstanding physical distances, their affection and care was never missed. Sadly, I cannot any longer personally extend a message of gratitude to my grandfather Pasquale, yet I will always hold dear his positive attitude and life guidance. I thank my friends Alessandro, Andrea, Carolina, Francesca, Francesco, Giuditta, Ivan, Leonardo, Marcello, Michele, Nicola, Simona and Ugo for all their encouragement and closeness manifested in crucial moments. Finally, I remark my appreciation for how Zsoka has reminded me what beauty means, in our profession as well as more generally in the world, at a time when I could barely see any. Thanks everyone.

Spillovers and Networks: Introduction

Thanks to the cumulative nature of research and, in general, of human knowledge, any invention or scientific discovery may bear consequences that go beyond the original intentions of those who originally devised it. This is no less true for firms and organizations that devote private resources to research and development. Knowledge about scientific discoveries, technical advances, or configurations of new products often spreads to other economic actors who may benefit from it in different ways. They can pursue promising avenues of research that have been opened up by others, implement specific improvements of production techniques that have been devised elsewhere, or revise their development plans in order to adapt to the innovations of competitors in rapidly changing industries.

The concept of *knowledge spillovers* has been central in economic analysis at least since Marshall (1890) posited its role to explain the apparent productivity advantages for firms to cluster near one another in manufacturing districts of 19th century England. Starting with him, knowledge spillovers have entered economic theories of industrial innovation, geographic agglomeration, economic growth, international trade and more. However, the exact mechanisms through which knowledge spills from one agent or organization to another are still unclear. Conjectures about human interaction and spatial proximity as drivers of information exchange are typically associated to methods of measuring spillovers that are unable to test their hypotheses directly, for these are typically based on aggregate R&D metrics.

Theories of knowledge spillovers are predicated on a variety of social mechanisms, which all explain the occurrence of information exchange through individual interaction. Foremost among them is the role played by professional connections and relationships that transcend the borders of single organizations. This corresponds to the hypothesis that those individuals that are involved in the process of knowledge creation within firms, chiefly inventors and scientists, diffuse their findings among colleagues in their entire industry. This mechanism is, however, a very difficult one to observe directly. The major obstacle has been thus far the lack of appropriate information about individual inventors working for different companies, and about their professional connections. Recent progress on the elaboration of patent-level and inventor-level data opens, however, new opportunities to test this crucial hypothesis.

Another often invoked mechanism, geographic proximity, plays an ambiguous position. It is intuitive to think that spreading the word about new discoveries is faster and easier if individuals are spatially close. In fact, geographic proximity can be simply thought as a factor that is merely complementary to the formerly described mechanism. Simply put, spillovers due to cross-firm professional connections are advantaged by spatial vicinity. According to a more general view, however, geographic proximity may have an additional and independent effect. This consists in favoring the diffusion of new ideas and knowledge within localities even beyond narrow circles, thereby resulting in agglomeration economies. Despite the relevance of this issue, economic research has not thus far clearly distinguished between the two different potential channels by which spatial proximity might drive knowledge exchange.

Regardless of which theoretical mechanism one endeavors to assess, there is usually an additional difficulty in empirically separating the spillover effect occurring within any group of economic agents on some ultimate outcome of interest (such as innovation or productivity), from the results of other external or internal factors that are common to the same entities. This problem, which is usually formulated in technical terms as the so-called "reflection" problem in the econometrics of social interactions, is primarily a question about the logical interpretation of real world observations about economically relevant issues. For example, one should be careful in calling the emergence of spillovers when observing faster patenting activity occurring between two connected firms or two neighboring inventors: the same fact may be the consequence of parallel technological trends in the former case, or of a common environment that is especially favorable to innovation in the latter.

In this dissertation, which is unified by common themes, I offer new insights and empirical evidence about knowledge spillovers by addressing the empirical question implied by the various hypotheses I have outlined above. Furthermore, I formulate and apply new empirical strategies that are aimed at identifying true spillovers without mistaking them for other economic forces. The first two chapters, in particular, are devoted to test the role of cross-firm professional connections in generating R&D spillovers. Both empirical analyses employ a newly combined dataset featuring measures of connections between firms. The two chapters differ in the new methodologies employed to address the same question. While the method described in the first chapter is more "reduced form" in spirit, the one in the second chapter is instead structural: remarkably, they yield comparable estimates of spillover effects. The third chapter, instead, focuses on the interaction between geography and knowledge spillovers. It employs an event study methodology in order to identify the effect of the location choices of "superstar" inventors on both the patent output of local collaborators and local innovation.

Another set of themes shared by all the chapters of this dissertation is constituted by the analysis and the empirics of networks. These are, in fact, ubiquitous in the economics of innovation: scientists are professionally associated to each other, firms might be related by collaboration and competition relationships, cities are linked through not only transportation and commercial, but also intellectual flows. All this is reflected in the characteristics of the data employed in the three chapters, in the respective analytical frameworks (all of them explicitly or implicitly featuring network relationships), and finally in the intuitions behind the identification strategies proposed in the first two chapters. Hence, this dissertation should be of interest not only for the advances in the understanding of knowledge spillovers it puts forward, but also for the analytical models it offers to explain recent, general stylized findings about the empirics of networks, as well as for the econometric methodologies introduced to estimate "network" effects or, more generally, peer effects occurring within networks.

The analyses and the results contained in this dissertation help answering some questions, but leave many others unaddressed. In particular, the important issue of the network formation process, and how this relates to spillover effects and information exchange in the context of innovation studies, calls for additional and separate analyses of similar depth and extent. Similarly, the search for the micro-level mechanisms behind localized knowledge spillovers is still at its beginnings; and much work, especially in terms of data collection and assembling, is needed for it to fully blossom. Ideas for future research work, or even full-fledged research programs, similarly stem from the premises to the conclusions contained in each chapter. I would like the reader to share with me a sense of an eternally advanced frontier of knowledge, as well as the enthusiasm for discoveries continually opening new paths.

Chapter 1

Network Effects on Productivity and Innovation

1.1 Introduction

The analysis developed in this chapter contributes to the empirical literature on the quantitative assessment of R&D spillovers by directly measuring the role of individual relationships in the diffusion of industrially valuable knowledge. I estimate the effect of R&D performed by different firms, that are linked through their scientists, on their reciprocal performance and innovation rates. In particular, I use coauthorship of past patents in order to identify inventors who are likely to maintain personal linkages across different organizations over time. For each pair of firms, I measure the degree of interaction between the two R&D teams and the potential for information exchange by the relative proportion of cross-connected coauthors. This metric changes over time, as scientists move across firms or acquire new coauthors.

By combining firm-level data with patent data that identify individual inventors over the course of their patenting history, I am able to construct a dynamic network of knowledge exchange. This network includes the largest, most innovative and R&D intensive U.S. firms, and it becomes tighter over time thanks to the increase in the total number of connections. The R&D of connected firms, weighted by the intensity of the links, is significantly and positively correlated with firm performance and innovation rates as measured by patent counts. This contrasts with well-established measures of spillover that rely, for instance, on technological similarities between firms (Jaffe, 1986, 1989). Within the network, these measures are not significantly and robustly correlated with relevant firm-level outcomes.

It is arduous, however, to assign a causal interpretation to these findings. As in the case of other types of studies on spillovers and externalities between economic agents, these correlations may simply reflect the existence of common unobserved confounders simultaneously driving R&D, innovation, and firm performance. For example, a sudden technological breakthrough in a specific technological niche where few connected firms operate may facilitate at the same time follow-up discoveries as well as cost savings. An alternative, no less likely scenario is one where a more rapid technological change is associated to a lower baseline profitability, conditional on firm size and input choices (including R&D). This corresponds to the classical case of industries at an earlier stage of their life cycle, which are simultaneously characterized by rapid product innovation and high costs. In both scenarios the unobservability of these confounders would bias, in either direction, standard estimates of R&D spillovers. This problem corresponds to the one of "correlated effects" per the classification by Manski (1993) of identification problems in the estimation of spillover effects.

Thanks to the characteristics of the network that I observe, I am able to formulate a novel empirical strategy that addresses the problem of common confounders. The basic intuition is straightforward. Unobserved factors that correlate across a pair of connected firms – call them i and j – may bias standard estimates of spillovers as long as their R&D expenditures also reflect those factors. Suppose further that a third firm k is connected to j but not to i. In addition, one can expect that its R&D is affected by some external circumstances that are shared with j but not with i. In this case, for firm i the R&D investment of k can be used as an exogenous predictor of that of j, and vice versa. As long as the network displays a sufficient number of similar "intransitive triads," appropriate instrumental variables can be derived from the choices of "indirect friends". These are independent from one firm's unobservables, but correlate with the R&D choices of direct connections.

I describe a model featuring a game of R&D investment played in a network of firms. R&D exerts reciprocal spillovers across connections; in addition, firms are hit by shocks that are exogenously correlated through the network. As a result, equilibrium R&D also co-varies across neighboring nodes, and the resulting correlation is endogenously amplified by the strategic anticipation of other firms' investment choices. However, under reasonable assumptions that allow for flexible patterns of cross-correlation in the shocks as well as for varying information structures of the game, the model predicts the existence of a degree of separation at which the R&D of different firms is independent. Since they are still correlated with the choices of direct friends, the choices of firms that are that distant would serve as valid instruments. Empirically, I find that that there is no significant cross-correlation in R&D choices at three degrees of separation. This motivates the use and comparison of instrumental variables based upon the R&D of indirect friends of second or third degree.

By applying this strategy, I still find substantial effects of connected firms' R&D on firm performance – expressed in terms of productivity and market value – as well as on patent production. However, when instrumenting peers' knowledge investment only with the R&D choices of indirect friends of third degree, I obtain substantially larger point estimates of spillover effects, for both the productivity and the patent production outcomes. That such difference is only apparent when applying the third-degree instrument in isolation is remarkably consistent with the proposed model. I interpret these findings as evidence that R&D is indeed driven by common factors across connected firms, consistently with the second scenario mentioned above in which these factors are also associated with lower conditional profitability. In light the results obtained by applying the proposed empirical strategy, I estimate the marginal social return of R&D to be about 24% of the private return.

This chapter builds on the traditional literature of industrial and innovation economics that tries to assess the determinants of productivity at the firm level, and in particular to measure the returns of R&D or patenting activity.[1] The quest for R&D spillovers in particular, initiated with the original intuitions of Griliches (1964, 1979, 1992), has been established in its current empirical framework by Jaffe in his cited contributions. In recent work Bloom, Schankerman and Van Reenen (2013) separate the positive effect of spillovers from the negative *business stealing* effect on firm performance, a longstanding issue in the literature. While they posit a role for inventors' personal or professional interaction in generating spillovers, in their empirical analysis they do not explicitly test this mechanism directly, nor do they control for the possibility of correlated confounders.

This work provides empirical evidence to support the hypothesis that spillovers are caused by the exchange of ideas between individuals. Thus, it is related to the research on the micro-level determinants of performance in the workplace. Moretti (2004) argues that productivity is related to how well-educated the workforce is in the city where a plant is located, suggesting that knowledge spillovers have a local scope.[2] Mas and Moretti (2009) demonstrate how "peer effects" apply at work, as coworkers intensify their efforts when they watch others increasingly doing so. Serafinelli (2013) shows that firm productivity is related to positive flows of workers with experience from companies at the top of the productivity distribution.

Given the structure of the interrelationships between firms that I uncover, this chapter is also related to the literature on social and economic networks. In particular, the dependencies between the strategic choices of firms and their outcomes are best understood within a game-theoretic framework. The model developed in this chapter is inspired by the ones of Calvó-Armengol, Patacchini and Zenou (2009) as well as Kranton, D'Amours and Bramoullé (2014).[3] The former in particular features network strategic complements, and provides an explicit empirical counterpart to the analytical framework. However, it does not allow for endogeneity of the "spatial" error term. This is the primary identification concern in this chapter, which I address through the strategy outlined above.[4]

---

[1] For surveys on each, see respectively Syverson (2011) and Hall, Mairesse and Mohnen (2010).

[2] Moretti (2011) lists knowledge spillovers as one of the micro-level determinants of agglomeration economies, although their relative quantitative contribution is uncertain. There are many complementarities between the literature that documents and looks for the causes of agglomeration economies, and the studies on R&D spillovers, especially when they place a focus on spatial proximity as a driver of knowledge exchange. Bloom et al. in a section of their paper attempt to distinguish a geographic component of spillovers by placing more weight on R&D performed by other firms in the same state. Lychagin et al. (2010) do a similar exercise by exploiting within-firm variation of their R&D activity at a finer geographic level. They document, essentially descriptively, a strong correlation of firm productivity with the density of R&D performed in the same narrow area where a firm's inventors reside.

[3] In particular, this model in this chapter includes an unobserved, spatially correlated component which affects the outcome of firms, whose realization is private information. In some ways, this resembles some features with the model in Conley and Udry (2010) who include an individual-specific production shifter of which neighbors in the network have limited information. The two models are, however, different. While their paper has an explicit focus on social learning (specifically, about different types of production technologies) this work focuses on R&D spillovers and strategic interaction. The spatially correlated firm-specific productivity shifter is to be interpreted here as the source of unobserved "common shocks".

[4] This strategy resembles in some ways the one featured in Bramoullé, Djebbari and Fortin (2009) and De

This chapter is organized as follows. Section 1 illustrates the game-theoretic framework that models R&D investment in a network. Section 2 describes the coauthorship-based measures of connections, and provides a description of the resulting dynamic network. Section 3 outlines the econometric framework and discusses the empirical strategy of the chapter. Section 4 presents the empirical results of the analysis and their implications. Finally, Section 5 indulges in some concluding remarks. Several Appendices of this dissertation complement both the theoretical and the empirical results from this chapter.

1.2 Analytical Framework

In this section I outline the analytical model that informs the empirical analysis of this chapter. The focus of the model is to explore the equilibrium relationship of firms' choice of R&D investment when they exert network externalities on each other and are subject to correlated shocks. Its main objective is to formalize and eventually generalize the intuition that motivates the identification strategy that is being proposed: when two indirectly connected firms i and k are not subject to the same confounders, they can reciprocally serve as predictors of the R&D choices made by their common link (call it firm j). This piece of intuition is useful in order to establish conditions for the identification of R&D spillovers when common confounding factors cannot be observed, and it can be extended to more complex structures of cross-correlation of the shocks through the network.

For illustrative purposes, I present first a stylized version of the model, which features only three firms in an "open triad" network, a simplified pattern of cross-correlation in the shocks and no additional production inputs; yet, it is sufficient to provide some relevant empirical predictions. I subsequently introduce the general version of the model, whose description is further expanded in Appendix A. This version allows to characterize the conditions for bounded spatial dependence in the network, which is necessary for identification. In a discussion at the end of this section I relate these results to features of the network topology, as well as to recent empirical findings about social networks. Within the entire framework, network formation is not explicitly modeled. Furthermore, the micro-level determinants of R&D spillovers are for now taken as given; they are appropriately treated in the next section.

1.2.1 Model's Setup

An economy consists of three firms, ordered as i, j and k, whose output depends on only one production factor: *knowledge* (Griliches, 1979). Knowledge is the result of R&D activity that is performed by teams of researchers – be they professional scientists, occasional inventors, academic collaborators of firms or other individuals – who are linked together in a network of professional relationships. These networks transcend the borders of the three firms. Thanks

Giorgi, Pellizzari and Redaelli (2010). However, these scholars are interested in addressing the simultaneity problem rather than regressors' endogeneity. Intuitively, they do not instrument peers' choices with those of indirect friends, but use the characteristics of "friends of friends" to predict peers' outcomes.

to the formal and informal exchange of information that happens in the network in the form of spillovers, a firm's knowledge depends not only on R&D that is performed in-house, but also on R&D from other firms that are connected in the network.

The knowledge input Z_i of firm i is a Cobb-Douglas function of the R&D investment of the three firms: S_i, S_j and S_k.

$$Z_i = S_i S_j^{\delta_{ij}} S_k^{\delta_{ik}} \tag{1.1}$$

Parameters δ_{ij}, δ_{ik} reflect the relative intensity of spillover effects between two firms.[5] I assume that $\delta_{ij}, \delta_{ik} \in [0, 1)$, which is to say that private R&D is intrinsically more valuable than spilled-over R&D. The latter is modeled as a strategic complement in this context, given that $\partial^2 Z_i / \partial S_i \partial S_j > 0$ and $\partial^2 Z_i / \partial S_i \partial S_k > 0$. I have chosen to characterize R&D as a strategic complement to stay consistent with the empirical framework adopted in this work. While the model would yield similar qualitative conclusions if a functional form involving strategic substitutes were chosen, some of the empirical implications would differ, because of the different strategic mechanism at play.[6]

The production function of firm i is, for $\gamma < 1$, given by:

$$Y_i\left(S_i, S_j, S_k, \omega_i\right) = Z_i^{\gamma} e^{\lambda \omega_i} = \left(S_i S_j^{\delta_{ij}} S_k^{\delta_{ik}}\right)^{\gamma} e^{\lambda \omega_i} \tag{1.2}$$

where ω_i is a stochastic shock that affects output. Such a shock, whose effect is parametrized by $\lambda \in \mathbb{R}$, reflects any technological factor that is specific of that firm and which affects in either direction its productivity or profitability. Examples may include temporal effects like the progression in the learning curve, or production inputs whose quality and quantity is difficult to observe, like the effort and the effectiveness of the firm management. In more abstract terms the shock ω_i can also be thought as a reduced-form representation, within a supply-side model, of demand-specific factors such as changes in the tastes of consumers for different varieties of goods.

Firm i's cost schedule for investing in R&D is also a function of ω_i:

$$C_i\left(S_i, \omega_i\right) = \kappa e^{-\omega_i} S_i \tag{1.3}$$

that is, the cost borne by one firm for each additional effective unit of R&D S_i decreases for larger values of ω_i. This represents circumstances such as radical discoveries that pave the way for successive, easier incremental innovations (which in practice increases the return of R&D) or the emergence of new opportunities for financing R&D. The reason to include the ω_i shock twice – both in the production and in the cost functions – is to establish a relationship between factors that affect the relative return of R&D and other factors that

[5] In line of principle, the model allows the intensity of spillovers to be asymmetric between two firms. For simplicity, I here treat spillovers as symmetric for each pair of firms, so that e.g. $\delta_{ij} = \delta_{ji}$.

[6] Whether R&D is in reality more of a strategic complement or a strategic substitute is a controversial matter: it is a notoriously hard dichotomy to test. It is usually thought to be both to some degree. As Jaffe (1986) put it, this is a ultimately a question on the assumed functional form, and standard econometric techniques are not the best means to assess curvature parameters beyond first derivatives.

have an effect on the profitability of the firm as a whole. Since λ can take also negative values, such a relationship is not necessarily unidirectional: with $\lambda < 0$, "cheaper" R&D is associated with lower total factor productivity.

This is best illustrated in a network setting, in which connected firms are allowed to share similar characteristics (similar values of ω_i). In particular, I assume that the shock is normally distributed, $\omega \sim \mathcal{N}(0, \sigma^2)$ and that:

$$\mathbb{C}\text{ov}(\omega_i, \omega_j) = \varsigma \delta_{ij}, \ \varsigma > 0 \tag{1.4}$$

firms that exert spillovers on each other are subject to similar unobserved circumstances the stronger their interdependence δ_{ij}. These circumstances may impact both the pace of R&D investment as well as on the overall profitability of the firms. A concrete, famous example in which the two effecs go together ($\lambda > 0$) is the ICT and computer industry. Close firms operating in that sector have been enjoying for decades parallel trends in the development of increasingly faster and cost-effective computers: the so-called "Moore's Law". Another well-known case that lends itself naturally to describe a situation in which the two effects might go in opposite directions ($\lambda < 0$) is the pharmaceutical sector. In that industry, the discovery of a new molecule typically stimulates parallel research on new drugs, which however require long and costly processes for testing and commercialization.[7] A classical stylized fact in the analysis of industry life cycles associates higher costs (and prices) to those early stages where product innovation is most rapid (Gort and Klepper, 1982; Klepper, 1996).

The idea that firms tend to share a more similar environment the more interdependent they are is intuitive, yet restrictive. In a static context and under certain conditions, it can be seen as the outcome of a model of network formation in presence of homophily, like the one proposed e.g. by Graham (2014). However, equation (1.4) rules out any correlation of the shock ω_i with ω_k if $\delta_{ik} = 0$, even if j and k are themselves very interconnected. This may not be a realistic assumption in many real-world innovative industries. In the context of the present model, (1.4) is a simplifying assumption. It allows me to illustrate how, in such a stylized setting, it can be possible to identify some exogenous variation for the R&D decisions of a firm's direct connections. In the general version of the game, played in a non-fully transitive network, this idea can be generalized to more complex assumptions that allow for cross-correlation in the shock even across indirect connections.

The firm objective function is its overall profit: revenues minus costs

$$\pi_i(S_i; S_j, S_k, \omega_i) = \left(S_i S_j^{\delta_{ij}} S_k^{\delta_{ik}}\right)^\gamma e^{\lambda \omega_i} - \kappa e^{-\omega_i} S_i \tag{1.5}$$

for any ordering of the three firms as i, j and k. The timing of the game is the following:

1. Nature draws ω_i, ω_j and ω_k, which are privately observed by each firm;
2. firms simultaneously make their R&D decisions;
3. payoffs (profits) are paid out.

[7] In the pharmaceutical industry, R&D is notoriously remunerative only in the long run. This has led to a separate debate on the optimal patent length for drugs.

This results in a game of private values in which information on individual types is not complete. The economic logic of this assumption is that managers are not omniscient about non-observable characteristics that drive innovation and market outcomes in other firms. However, they can make guesses (form beliefs) about them, and take into account the fact that the spilling-over firms are usually subjected to similar circumstances. That is, (1.4) is common knowledge. This makes explicit the positive feedback that occurs between "common shocks" and the joint choice of spillover strategic variables.[8] In Appendix A, I describe in more detail the implications of relaxing this assumption.

1.2.2 Equilibrium

The game is solved by invoking the Bayes-Nash equilibrium concept. That is, a profile of R&D investment strategies $\{S_i, S_j, S_k\}$ is an equilibrium if

$$\mathbb{E}\left[\pi_i\left(S_i^*; S_j^*, S_k^*\right)\middle|\,\omega_i\right] \geq \mathbb{E}\left[\pi_i\left(S_i; S_j^*, S_k^*\right)\middle|\,\omega_i\right] \quad \forall S_i \neq S_i^* \tag{1.6}$$

for firm i, and similarly for j and k. There are two strategic considerations at play. First, firms anticipate the R&D investment decisions of the others and respond accordingly in equilibrium. Because of strategic complementarities, they tend to raise their level of investment in reaction to increases observed in other firms. Second, they know that every firm attempts to raise their investment in R&D if subjected to favorable conditions (appropriate draws of ω). For that reason, they gauge the value of the shocks received by other firms using their limited information set: that is, their own realization of ω. Because of (1.4), the latter is only useful at predicting another firm's shock if a mutual spillover relationship exists.

The equilibrium is identified by solving separately for each firm's expected equilibrium: for each information set, the anticipated responses of other firms differs. The resulting equilibrium strategies are most conveniently expressed in logarithms. The Bayes-Nash equilibrium investment in R&D by firm i (j's and k's are symmetric) is given by

$$\log S_i^* = \frac{\left\{1-\tilde{\delta}_{jk}^2+\left[\left(\tilde{\delta}_{ij}+\tilde{\delta}_{ki}\tilde{\delta}_{kj}\right)\delta_{ij}+\left(\tilde{\delta}_{ki}+\tilde{\delta}_{ij}\tilde{\delta}_{kj}\right)\delta_{ik}\right]\frac{\varsigma}{\sigma^2}\right\}(1+\lambda)\,\omega_i+\Delta\log\left(\frac{\kappa}{\gamma}\right)}{(1-\gamma)\left(1-2\tilde{\delta}_{ij}\tilde{\delta}_{ik}\tilde{\delta}_{jk}+\tilde{\delta}_{ij}^2+\tilde{\delta}_{ik}^2+\tilde{\delta}_{jk}^2\right)} \tag{1.7}$$

where $\Delta \equiv 1-\tilde{\delta}_{jk}^2+\tilde{\delta}_{ij}+\tilde{\delta}_{ki}+\tilde{\delta}_{ki}\tilde{\delta}_{kj}+\tilde{\delta}_{ij}\tilde{\delta}_{kj}$ and $\tilde{\delta}=\delta\,(1-\gamma)^{-1}$. For the present work, three consequential considerations are most relevant.

[8] An alternative rationale for the whole set of assumptions is as follows. Firms do not know about other firms' shocks, and they are unaware of the underlying distribution from which ω is drawn (one can argue that it is indeed hard to predict ex ante the future distribution of a yet unobserved phenomenon). However, they can learn about *some* of the general conditions under which competitors and other organizations operate. This is thanks to the same mechanism that generates the spillovers: personal relationships crossing the borders of businesses. They are reflected by the parameters δ_{ij}, δ_{ik}, which are the analog of the connection metrics that are discussed in the next section. However realistic, such a set of alternative assumptions departs significantly from a typical game-theoretic framework. In a variant of this model it can be shown, however, to yield the same qualitative conclusions as the ones presented in this section.

Lemma 1. *The covariance between one firm's equilibrium strategy S_i^* and another firm's (say j) equilibrium strategy S_j^* reflects the underlying cross-correlation of their shocks, and is amplified by strategic dependence:*

$$\mathbb{C}\text{ov}\left(\log S_i^*, \log S_j^*\right) = \mathbb{C}\text{ov}\left(\omega_i, \omega_j\right)(1+\lambda)^2 \theta_i\theta_j = \theta_i\theta_j (1+\lambda)^2 \delta_{ij}\varsigma \tag{1.8}$$

for two constants θ_i (the factor multiplying $(1+\lambda)\,\omega_i$ in (1.7)) and θ_j.

The intuition is straightforward. Firms' R&D responds in equilibrium to their own realization of ω_i, not only because it affects the profitability of their own investment, but also because of the aforementioned mechanism of strategic interdependence. A high realization of ω_i signals a similar result for connected firms; if $\lambda > -1$ they would spend more in R&D. Because of complementarities, this would make the marginal investment in R&D by firm i even more valuable. As a result of this mutual feedback mechanism, the underlying correlation in the shock is reflected and amplified in the equilibrium outcomes.

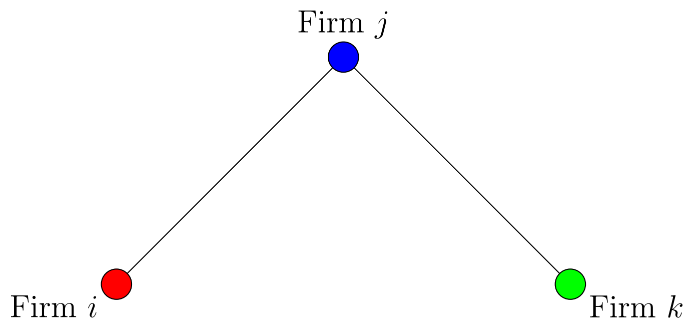
The next consideration is essentially a corollary of the previous one.

Lemma 2. *If two firms (say i and k) are disconnected to each other (so that $\delta_{ik} = 0$) but firm j is connected to both ($\delta_{ij}, \delta_{jk} \neq 0$), then the equilibrium strategies of firms i and k are uncorrelated to each other as well as to their respective shocks, but are both correlated to the equilibrium choice of R&D of firm j.*

This fact is a consequence of the assumption of incomplete information. In equilibrium, firms would respond – even if only marginally – to the choices of all the others, as long as they belong to a common network. The reason is that, even if unconnected, other firms may affect the choices of intermediate links. However, any firm can only make sensible guesses at the productivity shocks of others on the basis of their own signal. As long as the correlation between the shocks of two unconnected firms is zero, these are unable to predict their mutual realizations. Consequently, their equilibrium choices are independent.

This case is best represented with an example. Graph 1.1 displays an "intransitive triad", that is a specific (sub-)network of three nodes. Specifically, in an open triad out of three possible links only two exist. By framing the graph in the context of the model, both firms i and k receive spillovers from firm j, and their investment strategies respond endogenously to similar circumstances. Symmetrically, so does firm j relative to both firms i and k. However, as long as firms that are not connected do not share any correlated shock, either i and k are not in the position to anticipate how the other one in the pair *can affect the choices of firm* j. Since their choices are uncorrelated as a result, it is possible to "isolate" a piece of the variation of the choice of firm j that does not correlate to what i does, thanks to the choice of firm k (and vice versa). No similar type of variation can be isolated from firm j, since it is connected to both firm i and firm k.

The last consideration carries fundamental economic implications in light of the empirical analysis conducted in this work.



Graph 1.1: An Intransitive Triad 3-nodes Network

Lemma 3. *Suppose that two firms (say i and j) are connected to each other (so that $\delta_{ij} \neq 0$). If $\theta_j > 0$ it holds that*

$$\lambda \mathbb{C}\text{ov}\left(\log S_j^*, \omega_i\right) = \lambda \left(1+\lambda\right) \theta_j \delta_{ij} \varsigma < 0 \quad \Leftrightarrow \quad \lambda \in (-1, 0) \tag{1.9}$$

that is, the covariance between the equilibrium R&D outcome of firm j and the shock received by firm i, multiplied by the effect of the shock itself on firm productivity, can only be negative for values of λ falling between -1 and 0.

The quantity $\lambda \mathbb{C}\text{ov}\left(\log S_j^*, \omega_i\right)$ looks familiar as the numerator of an omitted variable bias formula. Its sign reflects the bias of the spillover coefficient if one were to run a naive regression of firm productivity on the R&D of its connections. According to Lemma 3, as long as $\theta_j > 0$ this bias can be negative only for $-1 < \lambda < 0$. This describes a situation where there exist common drivers of R&D investment across group of firms, which also have simultaneous, negative yet milder effects on productivity and/or profitability. If $\lambda > 0$ the bias is positive, as friends' R&D would be correlated with factors that impact positively on the outcome. If $\lambda < -1$ the bias is also positive as R&D would be negatively correlated with ω, for firms respond to higher observations of the shock by reducing their R&D effort.

It is easy to show that under the assumptions of the model and for non-degenerate network structures, $\theta_i, \theta_j, \theta_k > 0$ (intuitively, this is due to complementarities). In such cases, the spatial cross-correlation of R&D across degrees of separation in a network is unequivocally non-negative for any value of λ, as shown by equation (1.8). As it is subsequently illustrated in this chapter, such a scenario corresponds to the empirical evidence, because the cross-correlation of R&D across degrees of separation is positive at one or two degrees, zero at further distances. Less intuitively though, the application of the identification strategy that is being proposed also indicates, after appropriately instrumenting for peer R&D investment, the presence of a negative bias in simplistic spillover estimates. Lemma 3 offers a precise economic rationale which is able to reconcile both facts: the presence of exogenous correlated shocks which positively drive R&D but negatively affect – albeit more weakly – productivity.

1.2.3 General Model

In the general version of the model, N firms maximize their profits by simultaneously choosing, in addition to R&D, a set of K conventional production inputs (e.g. capital and labor) $\{X_{ik}\}_k$. These input variables are characterized by linear costs $\{\xi_{ik}\}_k$. Any pair of firms i and j may exert mutual R&D spillovers, which are weighted by $g_{ij} \in [0,1]$. Firms have heterogeneous, input-independent components of productivity that can be divided into a deterministic part A_i and a stochastic component $e^{\lambda\omega_i}$. Their objective function reads as

$$\pi\left(X_{i1},\ldots,X_{iK};S_1,\ldots,S_N\right) = A_i\left(\prod_{k=1}^{K} X_{ik}^{\beta_k}\right) S_i^{\gamma}\left(\prod_{j=1}^{N} S_j^{g_{ij}}\right)^{\delta} e^{\lambda\omega_i} - \sum_{K=1}^{k}\xi_k X_{ik} - \kappa e^{-\omega_i} S_i \tag{1.10}$$

where γ is the elasticity of output (and profits) to private investment in R&D and δ is the elasticity of R&D by a firm j such that $g_{ij}=1$. The shocks ω have an arbitrary cross-correlation structure, and the information sets of each firm are not *a-priori* restricted.

If the non-zero connections $g_{ij} \neq 0$ are sufficiently sparse, this setup describes a network of firms. To characterize the general result of the model, it is helpful to adopt the terminology of social networks. In particular, define $d_{ij} \in \mathbb{N}$ as the *distance* or *minimum path length*[9] between firms i and j. The general result is expressed by the following two properties of equilibrium R&D choices $(\log S_1^*,\ldots,\log S_N^*)$, which are proven in Appendix A.

Proposition 1. *Suppose that there is no-cross correlation in the shocks ω if two firms have distance higher than C: $\mathbb{C}\text{ov}\left(\omega_i,\omega_j\right)=0$ if $d_{ij}>C$. Similarly, suppose that two firms that have distance higher than L do not observe their relative shocks ω. It follows that*

$$\mathbb{C}\text{ov}\left(\omega_i,\log S_j^*\right) = 0 \;\; \textit{if } d_{ij} > C+L \tag{1.11}$$

$$\mathbb{C}\text{ov}\left(\log S_i^*,\log S_j^*\right) = 0 \;\; \textit{if } d_{ij} > C+2L \tag{1.12}$$

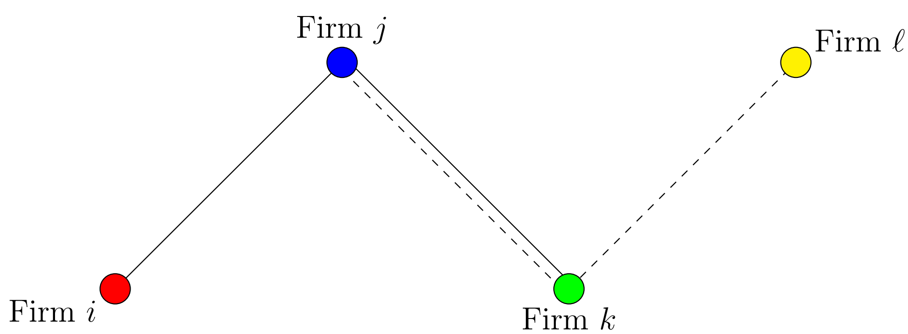
that is, the unobserved shock of one firm and the equilibrium strategy of another are independent as long as the two are distanced by a minimal path length higher than $C+L$; similarly the equilibrium strategies of any two firms at distance higher than $C+2L$ are also independent.

Proposition 1 places a bound in terms of "degrees of separation" on the equilibrium correlation across R&D choices and unobserved shocks in the network. The intuition is the following: even if in equilibrium firms endogenously internalize the shocks of other organizations that are "sufficiently close", and this in turn amplifies the exogenous cross-correlation, as long as both mechanisms are bounded their combined effect is as well. In other words,

[9] A path length is the total number of intermediate connections that indirectly connect two nodes (a *path*). Two nodes can be linked via more paths of different length, but usually only the shortest ones among them are of any interest. Two directly connected nodes have a minimum path length of 1; nodes j and k in Graph 1.1 have a minimum path length of 2. The minimum path length is popularly referred to as the "degree of separation" between two nodes.

the shocks other firms that are "very distant" in the network, whose R&D investment is of little relevance, are never internalized by individual firms. An implication of this result is that, for any firm i, the R&D choices of firms that are "sufficiently distant" in the network can be used as exogenous predictors of the R&D investment of i's direct links, which are located at distance 1. Such a relationship is due to the common dependence on some linear combination of shocks $\boldsymbol{\omega}$, a combination that excludes ω_i itself.

An example of this is provided in Graph 1.2, which displays a network of four firms (i, j, k, ℓ) constituted by two open triads with one edge in common: the link between j and k. If $C = 1$ and $L = 1$, firm i is able to fully observe ω_j, which provides predicting information about ω_k. Consequently, in equilibrium S_k^* is correlated to both ω_i and S_i^*. In a similar vein, the optimal R&D choice by firm ℓ, S_ℓ^*, correlates with S_j^*, yet not with ω_i which is "too distant" (firm ℓ cannot observe ω_j). Thus, the R&D of firm ℓ can function as an exogenous predictor of the R&D of firm i's link, and viceversa. The case in which $C = 2$ and $L = 0$, featuring two degrees of exogenous shock cross-correlation and an information set restricted to private values, is analogous. Firms only use their own realization of ω to optimize their R&D choice, inducing a cross-correlation in R&D as in the baseline case. This is bounded as long as C is finite, hence the R&D choices of firms i and ℓ can reciprocally work as peer predictors.



Graph 1.2: Two Semi-Overlapping Open Triads

A similar result can be established with respect to the cross-correlation of input variables and the R&D of firms that are sufficienty distant in the network.

Proposition 2. *Under the maintained hypotheses of Proposition 1, also the equilibrium conventional input choices of one firm are uncorrelated with the equilibrium R&D of firms located at distance higher than $C + 2L$.*

$$\mathbb{C}\text{ov}\left(\log X_{ik}^*, \log S_j^*\right) = 0 \text{ if } d_{ij} > C + 2L, \text{ for } k = 1, \dots, K \tag{1.13}$$

This result further supports the use of the R&D of firms that are distant enough as an instrument for the R&D of direct connections. Specifically, it motivates their exogeneity relative to other potentially endogenous control variables employed in the empirical analysis.

The intuition is that similarly to R&D, also conventional inputs are complementary to the shock ω. This generates some spatial correlation between the R&D of one firm and the conventional inputs of others, which is also bounded like the cross-correlation of R&D.

1.2.4 Discussion

What has been described is a conceptual framework that can help isolating some exogenous sources of variation in the characteristics of connected nodes in a network. However, its applicability to, say, the construction of instrumental variables depends on the network topology. Specifically, the network should be neither too tight nor too sparse, and display a sufficient number of intransitive triads like the one in graph 1.1. In some contexts, for specific nodes of a network it is not possible to isolate any genuine independent variation, as it the case of nodes j and k in graph 1.2. This might apply to nodes belonging to very tight clusters; for some of them, only few indirect connections of the desired distance can be isolated. As evidenced later though, given the characteristics of the data this does not appear to be a relevant concern for the empirical analysis performed in this work.

The main results, in any case, are not trivial. Recent consistent findings about the "three degrees of influence" in networks (Christakis and Fowler, 2013), an expression referring to the typical maximum extent of cross-correlation of nodes' characteristics,[10] currently lack a unified explanation. Economists have only recently started to consider the problem, and investigate their potential economic origins and empirical implications (Graham, 2014). However simple and stylized, this model offers a possible framework to explain these stylized facts through a combination of interacting exogenous factors and endogenous influences. In addition, a network approach incorporating spatially correlated heterogeneity can be informative for empirical analyses of peer effects, which face the challenge of how to properly account for the endogeneity problem induced by common confounders (Angrist, 2014).

In Section 3 the core result from this analysis, informed by some descriptive evidence from the data, is exploited to define an empirical strategy based on Instrumental Variables which addresses the chief endogeneity concern in this context: namely, the potential presence of common factors driving both R&D choices and the outcomes of connected firms.

1.3 Networks and Data

This section is divided in two parts. In the first part, I describe in abstract terms how I characterize the existence of spillover relationships between firms. In particular, I focus on linkages between their R&D-performing teams, on the basis of observable previous collabo-

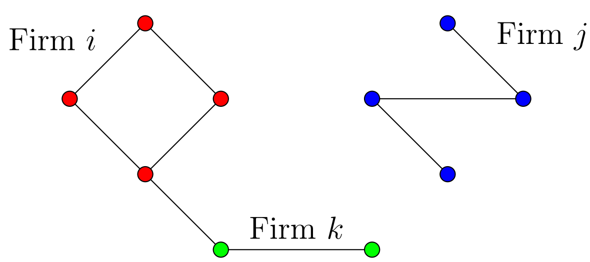
[10] In their work, Christakis and Fowler conduct a systematic investigation of the cross-correlation of adolescents' health habits across linkages of friendship networks. They document how most characteristics do not manifest any significant correlation across indirect friends of third degree or higher, with some characteristics registering some small third-degree correlation. For the most part, adolescent behavior displays "two degrees of influence", like – in the totally different context of this work – firm inverstment in R&D.

rations on patents. I formalize the metrics of *connection* that I empirically measure. In the second part, I describe the resulting dynamic network of R&D intensive firms selected from a specific panel of companies listed on the U.S. stock market. In addition, I provide some relevant descriptive statistics relative to the variables employed in the empirical analysis.

1.3.1 The Measures of Connection

Assume that there are three R&D intensive firms whose scientists are related to each others even beyond the borders of their respective organizations. Denote as M_i, M_j and M_k the sets of inventors belonging to each firm, with $M = M_i \cup M_j \cup M_k$. I define an existing *coauthorship relationship* between any two elements of M, be they m and n, with the notation $p^t_{mn} = 1$. This indicates that two individuals, at time t, share a past professional collaboration on a research project that resulted into a patent application listing both their names. Absent such a relationship, it is $p^t_{mn} = 0$. One could visualize the resulting network as a graph where each elements of M is a node, and nodes are linked by edges if $p = 1$.

Graph 1.3 depicts the first part of a stylized example (observed at some $t = 0$ point) of such a coauthorship network. The inventors of each firm (subsets of M) are nodes of the network displayed with different colors: red for i, blue for j, green for k. The coauthorship relationships p^0_{mn} are visualized as an edge connecting two nodes. The only existing cross-firm coauthorship relation is between an inventor of firm i and an inventor of firm k.



Graph 1.3: Inventors Network Example, $t = 0$

The central hypothesis of this analysis is that firms learn about other firms' R&D activities thanks to the inventors who are connected to scientists in other firms, because of continuing professional relationships or more informal channels. A natural implication of such an assumption is that the tighter is the connection between two R&D teams, the stronger are the spillovers occurring between two organizations. For this reason I define measures that quantitatively capture such a differential effect. A *measure of connection* $c^f_{(ij)t}$ between, say, firm i and firm j at time t is a function f of the fraction of inventors of either firm who are connected to inventors in the other firm, relative to the total size of both R&D teams:

$$c^f_{(ij)t} = f\left(\frac{\#\text{ inv.s of firm } i \text{ connected to } j \text{ at } t + \#\text{ inv.s of firm } j \text{ connected to } i \text{ at } t}{\#\text{ inv.s of firm } i \text{ at } t + \#\text{ inv.s of firm } j \text{ at } t}\right) \tag{1.14}$$

where $f : [0, 1] \to [0, 1]$. In the example of Figure 1.3, $c^f_{(ij)0} = c^f_{(jk)0} = 0$, while $c^f_{(ik)0} = f(1/3)$.

The facts that measures of connection are smaller than 1 and symmetric bear important qualitative implications. Because of the former, an extra unit of external R&D cannot be more valuable for a firm than internally performed R&D, which is a reasonable hypothesis.[11] As per the latter, the spillover relationship is assumed to be symmetric between two firms regardless of the relative size of their R&D departments.[12] In addition, it must be stressed that a connection measure essentially captures the relative number of personal professional relationships that have been established in the past, in terms of patent coauthorships. It is silent as regards distinguishing the relative importance of a single linkage.[13] In Appendix D I also examine alternative definitions of connections that are not symmetric.

Connection measures between two firms can change over time. Their dynamics are the result of conceptually different types of events that are in principle observable. They are: *i)* cross-firm R&D collaborations, such as joint ventures, resulting for example in collaborative patents; *ii)* the movement of inventors between firms. Both situations are usually thought of as drivers of knowledge transfer between firms, and they positively impact measures of connection. In addition, *iii)* entry and exit of inventors from the network also affect the calculated metrics. However, their net effect is ambiguous and depends on the specific circumstances of the inventors in question.[14]

Graph 1.4 extends the previous example by the advancing of one time period to $t = 1$, and examining the consequences of various changes in the underlying network. New linkages between inventors, due to newly appearing joint patents, are represented by dashed lines. In the example, some inventors of firm j have been observed to patent jointly with researchers from firm k, including an entrant inventor from the latter. There is also a new entrant in firm i, but he is not connected to anyone in other firms. Instead, among firm i's incumbents one inventor has now moved to j, while the one who used to maintain the connection with firm k has exited the network. As a result, $c^f_{(ij)1} = f(1/4)$, $c^f_{(jk)1} = f(1/2)$ and $c^f_{(ik)1} = 0$.

In the applied analysis I employ connection measures based on the square root function.

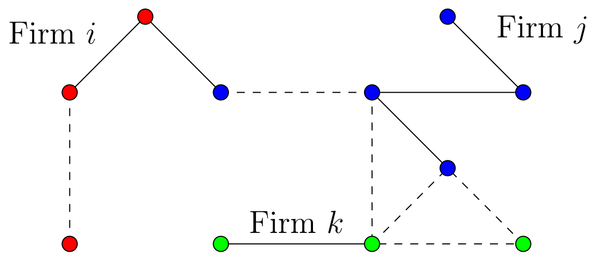
$$g_{(ij)t} = \left(\frac{\#\text{ inv.s of firm } i \text{ connected to } j \text{ at } t + \#\text{ inv.s of firm } j \text{ connected to } i \text{ at } t}{\#\text{ inv.s of firm } i \text{ at } t + \#\text{ inv.s of firm } j \text{ at } t} \right)^{\frac{1}{2}} \tag{1.15}$$

[11] This corresponds to the assumption that $\delta_{ij} < 1$ for any i, j in the analytical framework.

[12] This is apparent from the example in Figure 1.3 where the two connected firms have different size. This can have advantages: for example, it conveniently handles measurement errors in the assignment of individual inventors to firms. It may not be the most appropriate description of reality, however. One possibility is that just few "insiders" may be sufficient to grasp most of the secrets of one firm's R&D activity.

[13] A departure from this assumption is to consider that connections between two inventors that involve prolonged relationships over the years, relationships that result in many jointly filed patents, can be more relevant than others. Similarly, connections involving superstar inventors who issue many patents, of which some have been extremely well cited, can be of exceptional importance.

[14] New entrants increase the denominator of (1.14), but can potentially generate new cross-firm linkages, thus determining the tightening of connections. Similarly, the exit of scientists would decrease the denominator of (1.14), as well as the numerator if the exiting inventors were linking firms together.



Graph 1.4: Inventors Network Example, $t = 1$

This choice responds to a precise economic assumption. The typical anecdotal narrative on technological spillovers usually involves some solitary individual who transfers, perhaps by mistake, much of the knowledge internally developed by one firm to some of its partners or competitors. The very expression "spillovers" is verbally associated in such anecdotes to the "leakage" of few accumulating "drops" of knowledge. By applying the square root function to the ratio of connected inventors, I attribute more importance to the pairs of firms with relatively fewer connections. In the remainder of this chapter I use the expression "connection" to indicate the squared root metric. In Appendix D I present the empirical results from applying alternative definitions of connections; including ones based on the pure ratio of cross-connected inventors.

1.3.2 Networks Description

In the empirical analysis, I combine two different data sources. I construct the dynamic firm-level network on the basis of 707 firms from the unbalanced panel employed by Bloom et al. (henceforth BSV). Their sample consists of mostly manufacturing, R&D intensive firms listed in the U.S. stock market and belonging to the COMPUSTAT database, that are observed in the time interval of 1976-2001. This panel is representative of the bulk of private R&D performed in the US, which is concentrated among the largest and most productive firms. The dataset assembled by BSV includes information on accounting measures, various indicators of innovation performance, as well as the Jaffe-type measures that they use in their paper to disentangle different types of externalities.

I match the firm-level identifiers to the NBER patent data in order to obtain the list of patents assigned to each firm in the time interval under analysis. I subsequently match the official USPTO patent numbers to the Harvard Patent Network Dataverse (HNPD), a dataset that allows to identify the individual inventors signing each patent application. This is made possible by applying a specific disambiguation algorithm based upon the information contained in patents registered at the USPTO; specifically, some formulation of inventors' names and their ZIP codes of residence (Li et al., 2014). Ultimately, this results in the selection of 1,315,060 patents signed by 565,019 inventors.

To calculate the connection measures, I need to associate inventors to each other as well as to firms. The first task is accomplished by looking at jointly signed patents. Specifically, for two inventors m and n, I assign $p^t_{mn} = 1$ if at time $t+1$ the USPTO has received at least one patent *application* signed by both inventors. The implicit assumption is that the two inventors are involved in a professional relationship at least one year prior to the application.[15] Similarly, in order to assign inventors to firms one has to extrapolate facts on the basis of limited available information. I use the sequence of patents signed by inventor m and assigned to firm f in order to define a time interval in which one can reasonably presume that the individual was crucial for the R&D activity of that organization. The details of the assignment rule are provided in Appendix C.

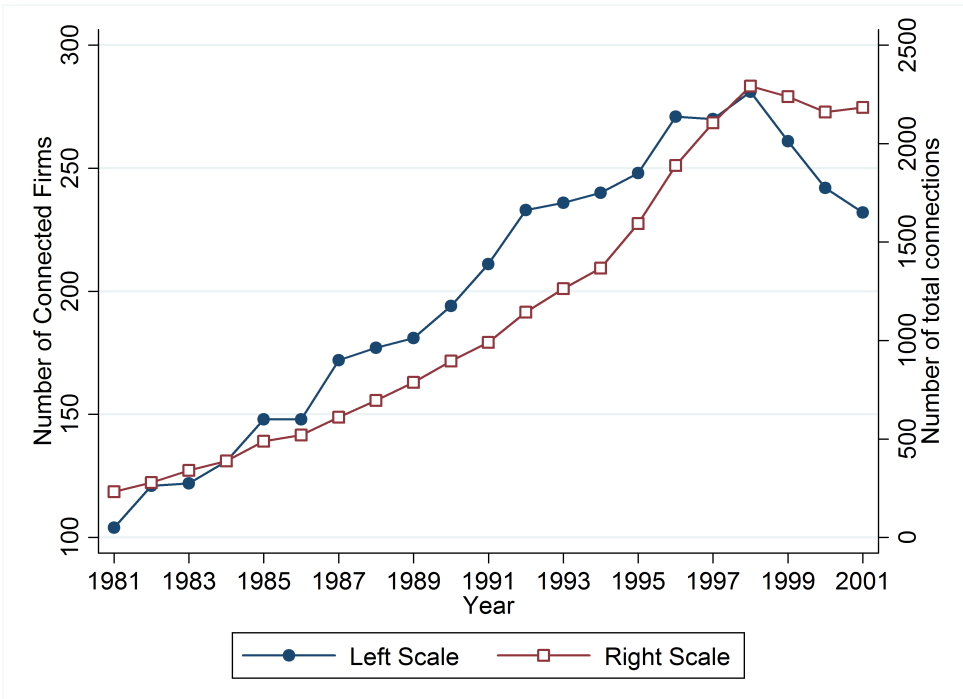


Figure 1.1: Connected Firms and Total Connections over time

I calculate the connection measure for each pair of firms and year. In total, 460 firms out of 707 display at least one positive connection with another firm in any year from 1981 to 2001.[16] The number of firms that are actually connected in any year varies with time: some of the initially unconnected firms would eventually develop bonds. Similarly, the firms that are already connected in 1981 may experience variations in the number of their connections, possibly resulting in the loss of all of them. Because of this, one never observes all the 460 firms of the dynamic network in each cross section. Figure 1.1 shows how many connected firms appear in each year, as well as the total number of yearly observed bilateral connections.

[15]Given the lag structure of R&D outcomes (patents) it is likely that this is an overly restrictive assumption. On the other hand, it is desirable to avoid assigning relationships that did not exist in reality.

[16]I calculate existing connections in 1981 thanks to information on patents that were issued since 1976.

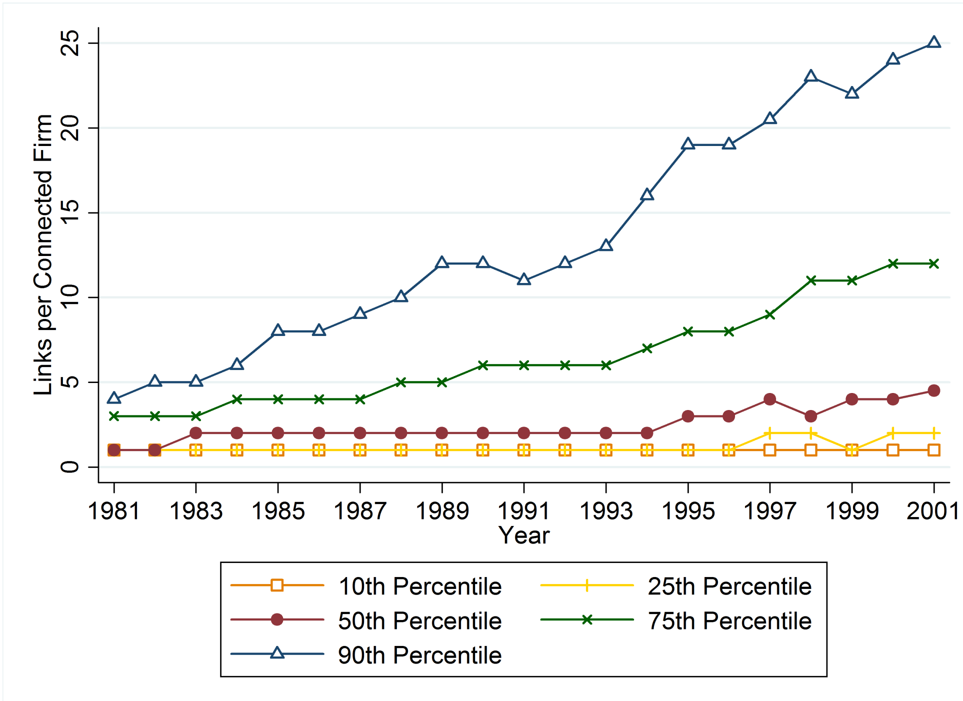


Figure 1.2: Degree distribution (binary connections) over time

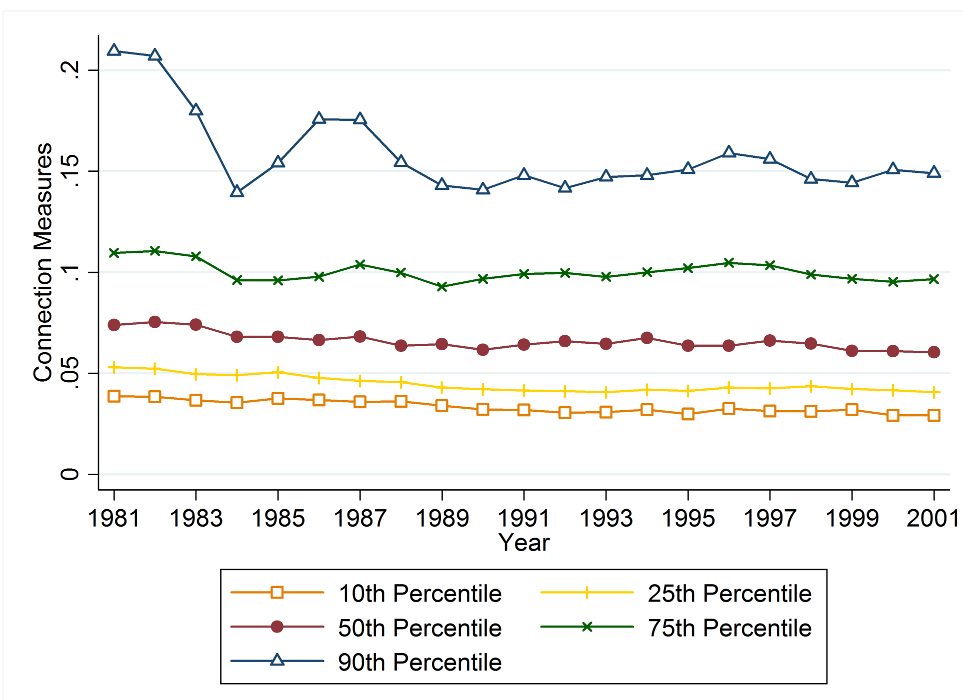


Figure 1.3: Distribution of the of connections ($g_{(ij)t}$) over time

Figure 1.1 displays a steady rise in the total number of connected firms between 1981 and 1998, to be followed by a drop from 1998 to 2001 because of losses of singleton connections by smaller firms. However, the total number of linkages, and thus the overall density of the network, remains quite stable during the final years of the sample. Another way to appreciate this temporal evolution is to visualize the actual network, in the form of graphs, as it looks like in different years. Selected graphs (for the years 1985, 1990, 1995 and 2000) are reported in Appendix E due to space limitations.

Figure 1.2 shows the yearly degree distributions. Like in many networks, it is very asymmetric and tends to widen over time. The most connected firms in the early '80s have less than 10 links, but several dozens of them around year 2000. Similarly, the average number of connections increases from about 1.5 to about 5. Each of them measures on average 0.083 (with a 0.066 standard deviation), but this average hides an asymmetric distribution which is displayed in figure 1.3 and which is quite stable over time. In order to interpret the empirical estimates, one may also want to consider the *total* amount of spillovers that firms receive from their connections. A measure that combines the variability in the degree distribution together with the variability in the strength of links is the *row sum* of connections, which is defined as $\bar{g}_{it} = \sum_{j \neq i} g_{(ij)t}$. Its yearly empirical distributions are displayed in figure 1.4. The aggregate mean and standard deviation of $\bar{g}_{it}$ are respectively 0.44 and 0.18. Apparently, the increase in its spread over time is due to the widening of the degree distribution.

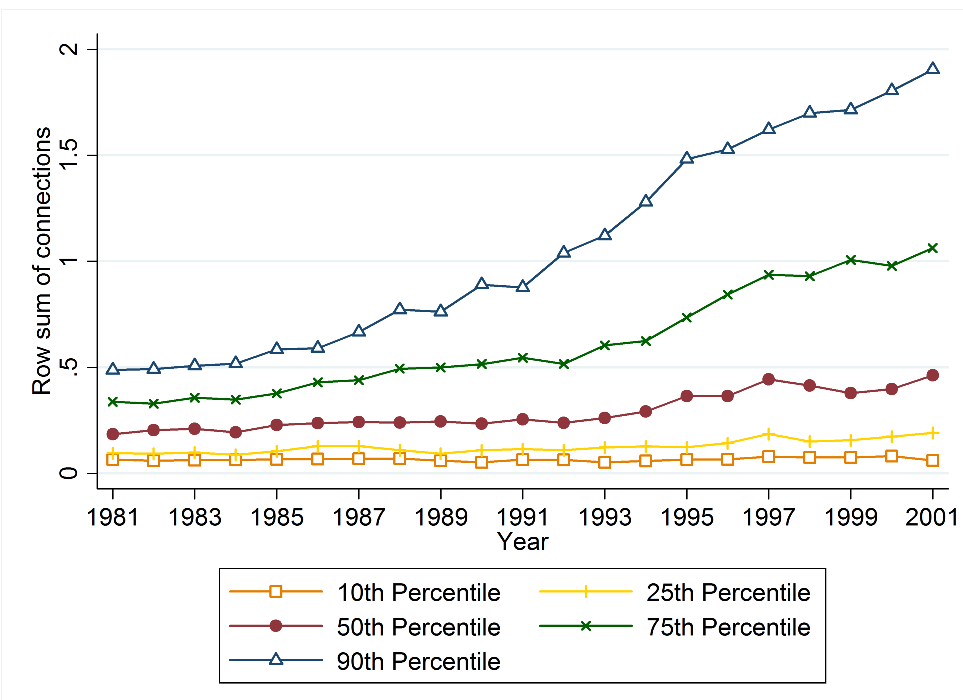


Figure 1.4: Distribution of the Row-sum of connections ($\bar{g}_{(it)} = \sum_{j \neq i} g_{(ij)t}$) over time

Table 1.1 provides some firm-level summary statistics. I divide the sample into five groups: the firms that do not belong to the network, and four groups for those that do. In particular, I calculate the *overall* sum of connections for each firm as $\bar{\bar{g}}_i = \sum_t \bar{g}_{it}$ and assign each firm to a group on the basis of its classification by quartile of $\bar{\bar{g}}_i$. Quartile 1 contains the least connected firms in the network over the time interval; quartile 4 contains the most connected ones.[17] For each group, I provide the mean and standard deviation of specific variables by pooling all the years in the sample. For this reason, in addition to real sales (Y_i) and firm size by number of employees (L_i), I report the ratio of Y_i to several input or spillover measures. Table 1.1 highlights the fact that the firms that belong to the network – in particular the most connected among them – are larger, more R&D intensive and more productive than loosely or non connected ones.

Table 1.1: Summary Statistics, 1981-2001

	No	Quartile of $\sum_t \bar{g}_{it}$			
	Network	1	2	3	4
Y_i: Sales (Millions 1996$)	540.6	1121.8	1473.2	2250.7	11099.8
	(1217.9)	(2407.7)	(2721.0)	(4629.8)	(20979.8)
L_i: Employees (Thousands)	3.418	6.987	9.524	12.39	56.36
	(6.454)	(15.57)	(16.88)	(22.29)	(96.20)
Y_i/L_i	142.7	142.8	169.3	169.1	220.0
	(91.47)	(116.6)	(110.5)	(124.9)	(178.6)
Y_i/K_i	7.189	5.668	5.619	5.461	4.762
	(6.474)	(3.618)	(4.513)	(3.855)	(3.919)
V_i/W_i: Tobin's Q	1.812	1.882	2.521	2.726	3.420
	(1.870)	(1.757)	(2.939)	(3.259)	(4.084)
P_i: Citation-weighted patents	4.099	15.60	22.76	70.35	647.8
	(12.45)	(43.28)	(44.74)	(136.2)	(1328.6)
$Y_i/R\&D_i$	47.30	20.96	61.50	11.71	4.505
	(155.2)	(78.92)	(581.2)	(35.40)	(3.977)
$Y_i/\prod_j R\&D_j^{g(ij)t}$		1004.5	900.6	579.6	197.5
		(2276.4)	(1855.6)	(1790.2)	(1198.5)
$Y_i/$ Jaffe Measure (i)	57.42	112.7	148.7	218.6	1016.9
	(129.0)	(242.8)	(285.4)	(443.1)	(1862.5)
No. of Observations	4210	1781	1748	1874	1940

[17]The four quartile-groups do not exactly contain the same number of observations because of attrition in the unbalanced panel.

1.4 Econometric Model

In this section, I outline the econometric methodology that I employ to estimate connection-based spillovers. This section is divided in three parts. In the first part, I introduce the workhorse model I use to evaluate the productivity effects of connections' R&D: an augmented production function. In the second part, I describe the Instrumental Variable strategy that I adopt in order to control for correlated effects, grounded on the theoretical framework developed in Section 1.2. I propose to aggregate the R&D choices of indirect connections of second or third degree in order to predict those of more direct links. In the third part I describe the two models that I employ to estimate spillovers on the market value and innovation rate of firms.

1.4.1 Production Function

I specify the firms' production function as a Cobb-Douglas:

$$Y_{it} = A_i Z_{it} \left(\prod_{k=1}^{K} X_{k_{it}}^{\beta_k} \right) e^{\tau_t + \upsilon_{it}} \tag{1.16}$$

with most variables being indexed by a firm-specific identifier $i = 1, \ldots, N$ and by time $t = 1, \ldots, T$. Specifically, Y_{it} represents output (measured as deflated sales), A_i is a firm-specific, time-invariant productivity shifter, Z_{it} is the knowledge capital, and $\{X_{k_{it}}\}_k$ is a set of conventional inputs or other observable factors affecting firm production. The unobservable factors, as well as any additional unpredictable determinant of production, are represented by the error term υ_{it}, while τ_t is a time-specific effect common to all firms.

The knowledge capital input is a function of the R&D performed in the network:[18]

$$Z_{it} = S_{it}^{\gamma} \left(\prod_{j=1}^{N} S_{jt}^{g_{(ij)t}} \right)^{\delta} \tag{1.17}$$

where S_{jt} denotes the R&D stock of firm j at time t, and $g_{(ij)t}$ is the connection measure between firms i and j at time t, with $g_{(ii)t} = 0$ for all i and for all t. The R&D stock S_{it} is constructed, following a customary approach in the literature, as the depreciated sum of past expenditures on R&D up to year $t-1$. To account for the known fact that the innovation and productivity effects of R&D materialize with a temporal lag, current expenditures in R&D are excluded from the calculation of the yearly stock.

By combining (1.16) with (1.17) and taking logarithms on both sides of the resulting equation, one obtains the workhorse empirical model:

$$\log Y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \log X_{k_{it}} + \gamma \log S_{it} + \delta \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tau_t + \upsilon_{it} \tag{1.18}$$

[18] The functional form is analogous to the one specified in equation 1.1 in Section 1 and further expanded in Appendix A. For a similar multiplicative specification of R&D spillovers in the production function, see e.g. Lychagin et al. (2010) and Manresa (2014).

with $\alpha_i = \log A_i$. Parameter δ represents the overall strength of the R&D spillovers in the network. It is interpreted as the elasticity of a connection-weighted neighbor's R&D on one firm's productivity. It is useful for different kinds of thought experiments: for example, a firm i connected to a neighbor j with connection $g_{ij} = 0.4$ receives a 0.4δ percentage increase in productivity following a 1% increase in the R&D stock of firm j. Similarly, a firm with many connections of overall strength $\bar{g}_{it} = 4$ receives a 4δ percentage increase in productivity following a 1% rise in the research effort of *all* its neighbors.

In most specifications, I estimate the model using the same set of controls $\{X_{k_{it}}\}_k$ as in BSV. These include measures of the capital and labor inputs elaborated from accounting data, as well as synthetic controls for industry-level sales and price indicators. In addition, I include the main spillover variables employed in the study by BSV. The first one of them corresponds to the classical "Jaffe" measure of spillovers, which is based on the similarity in the technological classification of any two firms' set of patents. It is meant to capture the positive effect of knowledge spillovers. The second measure controls for the "business stealing" effect of competitors' R&D in downstream product markets. It weighs R&D on the basis of the overlap of two firms' sales across industries.[19] In order to mitigate concerns of endogeneity, in their study BSV substitute several variables in $\{X_{k_{it}}\}_k$ – including conventional inputs and measures of spillover – with their first lags. I conform to their choices so to facilitate the comparison and interpretation of the respective results.

I also include in many regressions a measure that accounts for the relative intensity of R&D performed in the metropolitan areas where a firm's inventors are mostly concentrated. This way, I attempt to control for the possibility that connections between firms would just capture their spatial proximity together with other parallel endogenous factors.[20] I call this measure "Geospills"; Appendix C provides additional details on its construction. Overall, I simultaneously include different measures of spillovers in the same estimation models. Thus, I am able to more convincingly restrict the interpretation of the estimates of δ to the sole effect of the R&D performed by firms that are linked through the coauthor-induced network.

1.4.2 Instrumental Variables

The main econometric concern relative to the reliability of any OLS estimate of model 1.18 is the presence of common confounders that drive *both* the choice of R&D and productivity for connected firms. If such confounders are not observed, the OLS estimate of δ incorporates their effect on the outcome, to the degree that they are correlated to R&D of connected firms. They correspond to the *correlated effects* as per the analysis by Manski (1993) of

[19] These are respectively called "Spilltech" and "Spillsic" by BSV. In their analysis, only "Spilltech" – the Jaffe measure, described in more detail in Appendix C – has a significant effect in the production function. The "Spillsic" measure has, though, a significant effect on other outcome measures in the analysis by BSV.

[20] I construct this measure in close analogy with how I construct the measure of connection, by weighing R&D of other firms with appropriate pair-specific and time-varying metrics. For the sake of simplicity, I call such metrics "measures of proximity". They are based on the relative number of inventors who, for every pair of firms, are observed to be resident in the same set of metropolitan areas, defined at the CBSA level.

spillovers in the classroom. They are the empirical counterpart of the shocks ω_i that have been introduced in the analytical framework. A well-known real world example of common confounders is "Moore's Law," which has characterized developments in the ICT industry in terms of both productivity increases and R&D patterns.

This corresponds to a situation in which the real population regression function is given by

$$\log Y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \log X_{k_{it}} + \gamma \log S_{it} + \delta \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tau_t + \lambda\omega_{it} + \varepsilon_{it} \tag{1.19}$$

where $\mathbb{E}[\omega_{it}] = 0$, and

$$\mathbb{E}[\omega_{it}\omega_{jt}] \propto g_{(ij)t} \tag{1.20}$$

$$\mathbb{E}[\omega_{it} \log S_{it} | \{\log X_{k_{it}}\}_k] \neq 0 \tag{1.21}$$

and finally, ε_{it} is a pure white noise term. In this case, by estimating (1.18) where $\upsilon_{it} = \lambda\omega_{it} + \varepsilon_{it}$, one is introducing a bias in the estimate of δ whose sign depends on the signs of the expectation in (1.21) as well as of λ. According to the analysis conducted in Section 1.2 and expanded in Appendix A, this bias is negative only for small negative values of λ.

The same analysis suggests a solution to this problem when open triads, such as the one from Graph 1.1, can be observed in the network. As long as the unobserved confounders are only correlated between connected firms, that is $\mathbb{E}[\omega_{it}\omega_{kt}] = 0$ if $i \neq k$ and $g_{(ik)t} = 0$, the R&D patterns of an indirectly connected firm k are uncorrelated with the unobservable shocks ω_i of firm i. The economic rationale is as follows. In equilibrium, the choice of R&D of a firm is a function of the R&D choices of all the members of the network. With imperfect information, however, a firm is unable to anticipate the shocks received by firms beyond its direct connections. Hence all the cross-variance in R&D choices is driven by (1.20). At the same time, the R&D trajectories of firm j – a direct connection of i – and k are correlated if $g_{(jk)t} \neq 0$, as a result of their common confounders as well as the amplification effect induced by complementarities.

In terms of the usual interpretation of IVs, the R&D of firm k – an *indirect connection* or *indirect friend* – can serve as an instrument for the R&D of firm j. Formally:

$$\mathbb{E}[\omega_{it} \log S_{kt} | \{\log X_{k_{it}}\}_k, \log S_{it}] = 0 \tag{1.22}$$

$$\mathbb{C}\text{ov}[\log S_{kt} \log S_{jt} | \{\log X_{k_{it}}\}_k, \log S_{it}] \neq 0 \tag{1.23}$$

where (1.22) can be thought of as the component of a more general moment condition and (1.23) motivates the power of the instrument. Since a firm generally has more than one connection, and each of them corresponds with more than one indirect linkage, in theory one can combine the entire resulting set of moments in several ways. Here I propose a straightforward way to aggregate all the indirect connections' R&D into a single instrumental variable. For every indirect connection k, define the indicator $\tilde{h}^i_{(jk)t} = g_{(jk)t}\mathbb{I}\left[g_{(ik)t} = 0\right]$. The "indirect spillovers" instrument reads as

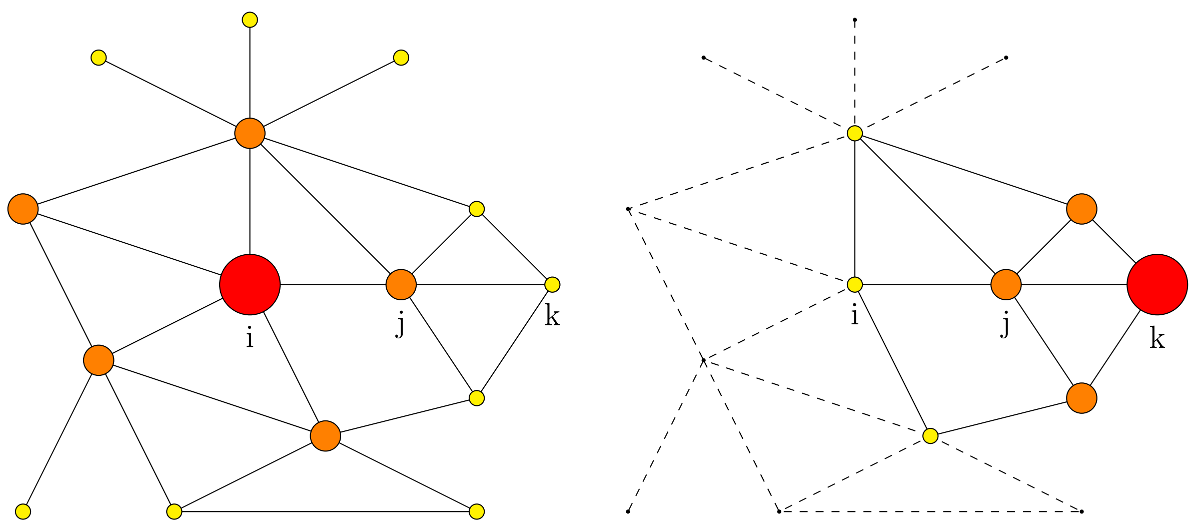
$$\sum_{k\neq i} h_{(ik)t}\log S_{kt} = \sum_{j\neq i} g_{(ij)t} \sum_{k\neq i,j} \tilde{h}^i_{(jk)t}\log S_{kt} \tag{1.24}$$

$$= \sum_{k\neq i,j}\left(\sum_{j\neq i}\left(g_{(ij)t}g_{(jk)t}\right)\mathbb{I}\left[g_{(ik)t} = 0\right]\right)\log S_{kt} \tag{1.25}$$

where the weights $h_{(ik)t}$ on the left hand side are implicitely defined by the expression on the right hand side. It is a direct consequence of (1.22) that

$$\mathbb{E}\left[\left(\sum_{k\neq i} h_{(ik)t}\log S_{kt}\right)\omega_i \middle| \left\{\log X_{k_{it}}\right\}_k, \log S_{it}\right] = 0 \tag{1.26}$$

holds, and that at the same time the instrument should retain some predictive power for the endogenous regressor $\sum_{j=1}^{N} g_{(ij)t}\log S_{jt}$ as long as (1.23) is true. As discussed in more detail in the appendices, an implication of the model is that the instrument is also uncorrelated with the set of estimated inputs, thereby allowing to estimate γ consistently.



Graph 1.5: Direct and indirect connections

Graph 1.5 provides a visual representation of the "indirect friends" that are captured by (1.24) in an artificial network. One of the many open triads in this network is indexed by i, j and k, the missing link being between i and k. On the left panel, the very central node i is shown as a large, black circle. Its direct connections are medium-sized grey nodes; the indirect connections are the small white ones. The right panel similarly focuses on the very peripheral node k. This example also illustrates how the number of indirect connections per direct ones can be similar for nodes that occupy different positions in a network. The identifying variation captured by the instrument is the variation in the log of R&D stock among indirect connections.

A possible concern with this strategy is that the exclusion restriction is violated, because the instrument is still correlated with the unobservable factors ω_{it}. According to the analytical framework, this can occur in two cases: *i*) equation (1.20) may not be a complete description of reality if, for example, the correlation in the unobserved shock extends up to two degrees; *ii*) the imperfect information assumption fails and firms may fully anticipate the shocks of their connections, in which case (1.22) is no longer valid. To evaluate the extent by which this may be a problem, I propose a diagnostic tool in the form of the spatial correlogram of R&D variables, where distance is defined in terms of degrees of separation in the network. By definition, the cross-correlation in the ω shocks, or the dependence of R&D from other firms' shocks, cannot be observed; however they are both reflected in the cross correlation of R&D as per the mechanisms outlined in Section 1.2.

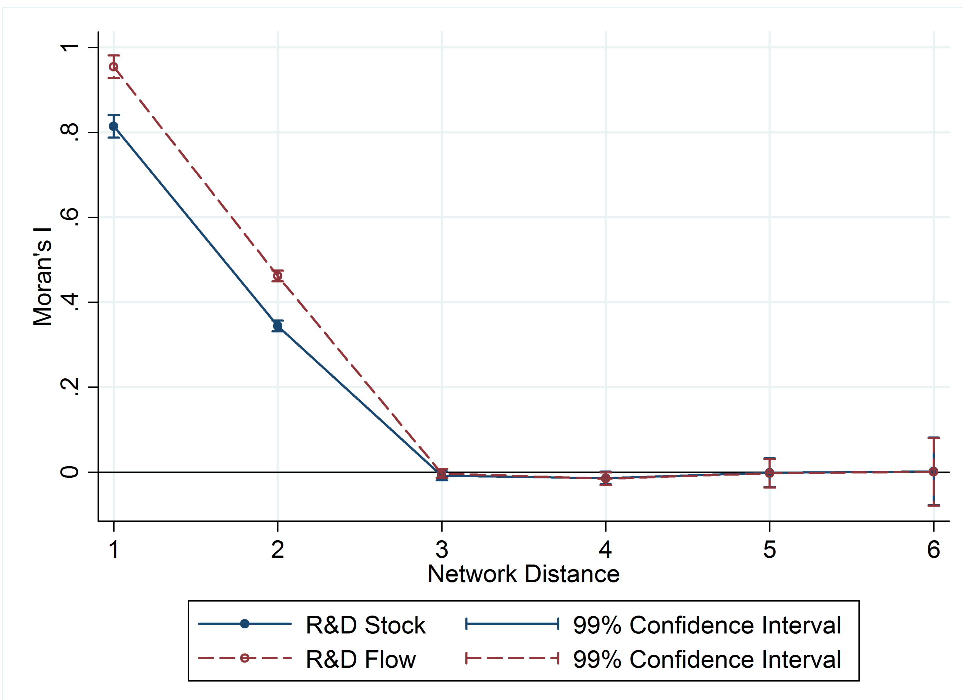


Figure 1.5: Spatial Correlogram of R&D Measures

Figure 1.5 displays the spatial correlation of both R&D flows and R&D stock variables in the network by pooling all years together. The correlation is measured by the Moran's I statistic, a standard tool in spatial analysis.[21] Figure 1.5 illustrates a strong correlation for

[21]Moran's I is a correlation measure calculated by weighting each pair of observations by the strength

direct connections (distance 1), a correlation of half strength for indirect links (distance 2) and zero correlation for all further distances; this is a typical pattern encountered in many other real-world networks. The correlation for R&D stock variables is mechanically weaker than the one of R&D flows, since it can account for past time periods when two firms were not connected. According to equations (1.11) and (1.12), this evidence is compatible either with a situation in which $(C, L) = (0, 1)$ or, more realistically, one where $(C, L) = (2, 0)$. While in the former case the simpler instrument proposed in (1.24-1.25) would be sufficient to eliminate the bias induced by ω, in the latter it would not. Only an instrument based on the R&D choices of firms located at distance 3 – uncorrelated with all zero-distance factors, but correlated with the R&D of firms located at distance 1 – would solve the problem.

It is thus necessary to define another instrument, which looks up to more distant nodes in the network. Specifically, I consider the "indirectly indirect" connections, that is the nodes with three degrees of separation that have no direct linkages with either direct "friends" or indirect connections of second degree. By aggregating all the R&D choices of those firms, I can define this additional instrument. In analogy with its two-degrees counterpart, let $\tilde{q}^i_{(k\ell)t} = h_{(k\ell)t}\mathbb{I}\left[g_{(i\ell)t} = 0\right]$ and aggregate over ℓ:

$$\sum_{\ell\neq i} q_{(i\ell)t}\log S_{\ell t} = \sum_{k\neq i} h_{(ik)t} \sum_{\ell\neq i,j,k} \tilde{q}^i_{(k\ell)t}\log S_{\ell t} \tag{1.27}$$

$$= \sum_{\ell\neq i,j,k}\sum_{k\neq i,j}\sum_{j\neq i}\left(g_{(ij)t}g_{(jk)t}g_{(k\ell)t}\right)\mathbb{I}\left[g_{(ik)t} = 0\right]\mathbb{I}\left[g_{(i\ell)t} = 0\right]\log S_{\ell t} \tag{1.28}$$

where the weights $q_{(i\ell)t}$, implicitly defined by the expression on the right-hand side, represent the strength of all indirect connection paths between i and ℓ, and are equal to zero if any firm ℓ has a direct connection with either direct or second-degree indirect links of firm i. In the empirical analysis I employ this instrument, which is based on indirect connections of third degree, either in isolation or in conjuction with its simpler second degree counterpart. The observation of differences in the point estimates for δ across these two cases can be an indication that the instrument defined in (1.24-1.25) is in fact (more strongly) correlated with the unobserved common factors.

Finally, it is important to consider that the cross-correlation in the unobserved shocks of the type expressed in (1.20) invalidates standard asymptotic properties of any 2SLS estimator. I address this problem by dividing the network into "communities", within which the network is particularly dense, and by which I cluster standard errors. Unlike other settings featuring clustered common variability, within a network it is less easy to separate groups characterized by a common dependence in the error terms, as the structure of binary relationships across nodes is non-trivial. This requires to construct "communities" that are as large as possible; in particular I cluster standard errors by 20 groups and apply a small sample correction. These communities are constructed by running the "Louvain algorithm"

---

of their relationship at any given degree of distance. In Figure 1.5 I report the calculated statistics along with a 99% confidence interval based on their sample standard deviations. The asymptotic distributional properties of the Moran's I statistic are discussed in Kelejian and Prucha (2001).

(Blondel et al., 2008)[22] on the "pooled" network resulting from the aggregation of all connections over time (see Appendix E for further description). For consistency, I employ the same clustering unit choice across all estimated models, from OLS to non-linear approaches. The results are robust to the choice of the clustering unit and of the number of communities.

1.4.3 Additional Outcomes

In empirical studies of R&D spillovers, it is customary to assess the effect of other firms' R&D not only on output or productivity, but also on other outcomes and indicators of firm performance and innovation rate. In his seminal study, Jaffe also measured the effect of spillovers on firms' market value and patent output. BSV follow in his legacy. Under their shared theoretical framework, especially under the maintained hypothesis of R&D as a strategic complement, spillovers stimulate R&D efforts and increase the number of inventions. The effects on productivity can be indirect (thanks to new or better patents/products) or direct (because of the immediate applicability of spilled knowledge in the production process). This ultimately results in better firm performance and increased market value.

I follow suit and measure the effect of the R&D performed by "connections" on outcomes other than output or productivity, largely following the empirical specifications by BSV. I begin from a market value specification: I regress the Tobin's q on a model of this sort:

$$\log\left(\frac{V_{it}}{W_{it}}\right) = \tilde{\alpha}_i + \sum_{k=1}^{\tilde{K}} \tilde{\beta}_k Q_{k_{it}} + \tilde{\delta} \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tilde{\tau}_t + \tilde{\upsilon}_{it} \tag{1.29}$$

where V_{it} is the market value of a firm measured at time t and W_{it} is the replacement value of its assets. Notice here that the set of controls is different than in model (1.18). In particular, $\{Q_{k_{it}}\}_k$ includes the Jaffe measure of spillovers as well as a polynomial of sixth degree of the ratio S_{it}/K_{it}, to control for differences in R&D intensity among firms.[23]

The estimation of spillovers effects on the innovation rate is based on a count model for citation-weighted patents P_{it}:

$$P_{it} = \exp\left(\breve{\alpha}_s + \sum_{k=1}^{\breve{K}} \breve{\beta}_k H_{k_{it}} + \breve{\gamma} \log S_{it} + \breve{\delta} \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \breve{\tau}_t + \breve{\upsilon}_{it}\right) \tag{1.30}$$

which, in order to account for values of $P_{it} = 0$, is typically estimated via maximum likelihood[24] (Hausman et al., 1984; Blundell et al., 1995). The set of controls $\{H_{k_{it}}\}_k$ includes a

[22] The Louvain algorithm, a popular tool in network analysis used to identify communities or "clusters", identifies a hierarchy of densely connected groups that display few interconnections across each other. This algorithm can be fine-tuned in order to return a different number of communities of different size, for different choices of the resolution parameter. Here, a resolution parameter of about 0.6 returns circa 20 groups.

[23] A polynomial of S_{it}/K_{it} is derived from a Taylor series approximation of the right-hand side term $\log(1 + \theta S_{it}/K_{it})$, which is customary in the specification of Tobin's q type of models. BSV show that results are similar whether one uses a six-degree polynomial or non-linear estimation methods instead.

[24] To guarantee convergence of the estimation algorithm, it is convenient not to include firm-specific fixed effects. I introduce four-digits industry fixed effects instead, that are indexed by s.

term for the lag the dependent variable ($\log P_{i(t-1)}$) as well as the Jaffe measure. In order to control for endogeneity, I adapt my IV strategy by employing a control function approach. Specifically, I take the residuals from first stage linear regressions of the endogenous outcome on the excluded instruments and the other controls, and I include them in the estimation of the non-linear model.

1.5 Empirical Results

In this section I present the empirical results of the analysis; it is divided in five parts. In the first part I present the baseline (OLS) results for the production function. In the second part, I address the endogeneity concerns of correlated confounders by applying the proposed IV strategy. In the third part I present the results for the firm value equation; in the fourth those for the patent count model. Finally, in the fifth and last part I offer some additional considerations, in particular about the economic relevance of the estimated effects.

1.5.1 Production Function, OLS

Table 1.2 displays the results from the estimation of equation (1.18). I take firm and year fixed effects and cluster the standard errors at the community level. Along with the estimate of γ and δ I report those for Capital and Labor. One can interpret the estimate of $\delta = 0.017$ from column (1) in several ways. It represents the average elasticity of output to the R&D performed by a fully connected firm at time t. Relative to non-completely connected firms, the corresponding estimated elasticity is equal to $\hat{\delta} g_{(ij)t}$. For example, a 1% increase in the R&D stock of an average connection translates into a 0.0015% rise in productivity. To the thought experiment of a 1% increase in the R&D stock of *all* of one firm's neighbors, it corresponds with a $\hat{\delta}\bar{g}_{it}$% rise (0.0075% on average). In comparison, the elasticity of privately undertaken R&D, $\hat{\gamma} = 0.044$ appears to be one order of magnitude larger.

Relative to column (1), in (2) I show the effect of controlling for the Jaffe measure of knowledge spillovers based on technological proximity, as well as for the geographic R&D intensity measure ("Geospills"). Their inclusion does not dramatically impact the point estimate $\hat{\delta}$, which falls to 0.0145 while remaining statistically significant. The geographic control, on the other hand, seems to have very little economic significance. In column (3) I restrict the sample only to those firms that enter the network at any point in time. This is an attempt to control for the possibility that the estimate $\hat{\delta}$ is driven by persistent productivity differences between firms that belong to the network and those that do not. This exercise has an interesting implication: while the estimate $\hat{\delta}$ is again not largely affected (0.012), the coefficient for the Jaffe measure of spillovers falls sharply and becomes statistically insignificant.[25] Since firms that do not belong to the network are the smallest and least R&D-intensive ones, this result implies that the positive correlation between real sales

[25]This holds true even if I do not estimate parameter δ on the subsample.

and the Jaffe measure is largely driven by small firms patenting in the most R&D-intensive technological fields.[26]

Table 1.2: Production Function, Ordinary Least Squares, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0365	0.0331	0.0436	0.0427	0.0414
	(0.0080)	(0.0080)	(0.0092)	(0.0111)	(0.0122)
Spillovers (δ)	0.0168	0.0145	0.0119	0.0136	0.0110
	(0.0034)	(0.0028)	(0.0027)	(0.0031)	(0.0029)
Geospills		0.0003	0.0003	0.0002	0.0002
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
Capital	0.1584	0.1568	0.1637	0.1604	0.1549
	(0.0113)	(0.0117)	(0.0202)	(0.0218)	(0.0212)
Labor	0.6477	0.6523	0.6461	0.6456	0.6433
	(0.0168)	(0.0170)	(0.0268)	(0.0297)	(0.0282)
Jaffe Tech. Proximity		0.2004	0.0300	0.0597	0.0281
		(0.0977)	(0.0821)	(0.0954)	(0.1075)
Fixed Effects	YES	YES	YES	YES	YES
Only Network	NO	NO	YES	YES	YES
No. of Communities					
(Community $\times$ Year Effects)	0	0	0	10	20
No. of Observations	11548	11548	6914	6914	6914

In columns (4) and (5) I also estimate an additional set of dummy variables, in a first attempt to control for the fact that connected firms may be subjected to similar shocks. Specifically, I absorb community-by-year effects, where communities are constructed by applying the Louvain algorithm. In particular, in column (4) I employ a network partition of 10 communities;[27] while in column (5) I construct the additional dummy variables from the same 20 communities that are also used for clustering standard errors. Increasing the number of clusters does not result in a dramatic variation of the point estimate $\hat{\delta}$ (in column (4), it actually increases); while the coefficient of the Jaffe Measure becomes negative. This suggests that the correlation between the connections-induced measure of spillovers and one firm's output is in fact driven by the variation in the R&D stock of that firm's linkages.

[26] This fact may also be interpreted as a censoring problem. COMPUSTAT only reports data for public firms. Small firms that go public are usually successful firms, and those that "make it into the news" are typically from fast developing high-tech sectors (and being in the news is itself endogenous). If a correlation exists between the Jaffe measure and the probability that small firms go public, this would be reflected in a positive bias in the estimate of the Jaffe measure when small public firms are included in the estimation sample. This issue certainly deserves further research.

[27] This results from the application of the Louvain algorithm with resolution 0.8.

1.5.2 Production Function, IV

I now illustrate the empirical results from the application of the IV strategy that addresses the problem of correlated confounders. I instrument the R&D stock of one firm's direct connections using those of its indirect "friends" of the second and third degree. The analytical framework suggests that instruments are to be taken at the degree of separation at which the R&D expenditures of firms that are that distant in the network, are uncorrelated. The spatial autocorrelation of R&D in the network evidenced by Figure 1.5 suggests that the best instrument is defined by aggregating indirect connections of third degree.

Table 1.3: Production Function, First Stage Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
2-degree Instrument	1.0912	1.1540			
	(0.0336)	(0.0446)			
3-degree Instrument		-0.5690	2.3408	2.2065	2.1658
		(0.1707)	(0.1536)	(0.1659)	(0.1746)
Private R&D	0.0946	0.1069	0.3677	0.3065	0.2670
	(0.0456)	(0.0446)	(0.1238)	(0.0990)	(0.1002)
Capital	0.0876	0.0982	0.4906	0.4807	0.4000
	(0.1022)	(0.0954)	(0.2925)	(0.2856)	(0.2985)
Labor	-0.1449	-0.1214	-0.7068	-0.6329	-0.6276
	(0.1217)	(0.1143)	(0.2523)	(0.2484)	(0.2668)
Jaffe Tech. Proximity	0.6669	0.6550	2.8423	3.0933	2.8289
	(0.3676)	(0.3468)	(1.2101)	(1.2268)	(1.1184)
Fixed Effects	YES	YES	YES	YES	YES
Only Network	YES	YES	YES	YES	YES
No. of Communities					
(Community $\times$ Year Effects)	0	0	0	10	20
F-statistic	791	843	131	114	64
No. of Observations	6914	6914	6914	6914	6914

I report various first stage equations from the 2SLS estimation of model (1.18) in Table 1.3. All estimates are restricted to the subsample formed by those firms that are part of the network. I regress the connections-induced spillovers variable on the aggregated log R&D stock of indirect connections of either second degree (column 1), second *and* third degree (2), third degree only (3). The estimates from column (4) and (5) are analogous to those in (3), but they additionally include two different sets of community-by-year fixed effects (based upon respectively 10 and 20 communities). In all cases I also include all the controls from equation (1.18). As expected, both instruments are strongly and positively correlated with the endogenous spillover variable.[28] The F-statistic from all first stage estimates is reassuringly strong; the lowest measured F-statistic – the one from column (5) – equals 64; and is associated to a t-Statistic of the third-degree instrument which is larger than 12. Conventional inputs seem to have little residual predictive power relative to the spillover variable, except for other R&D-based variables (private R&D and the Jaffe measure).

Table 1.4: Production Function, Two-Stages Least Squares, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0438	0.0439	0.0415	0.0397	0.0385
	(0.0090)	(0.0090)	(0.0102)	(0.0119)	(0.0130)
Spillovers (δ)	0.0116	0.0113	0.0160	0.0206	0.0184
	(0.0030)	(0.0029)	(0.0062)	(0.0064)	(0.0066)
Geospills	0.0003	0.0003	0.0002	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
Capital	0.1640	0.1641	0.1610	0.1560	0.1510
	(0.0202)	(0.0202)	(0.0202)	(0.0221)	(0.0212)
Labor	0.6458	0.6457	0.6491	0.6501	0.6480
	(0.0268)	(0.0268)	(0.0270)	(0.0298)	(0.0283)
Jaffe Tech. Proximity	0.0313	0.0321	0.0162	0.0341	0.0023
	(0.0837)	(0.0834)	(0.0902)	(0.1059)	(0.1173)
2nd degree IV	YES	YES	NO	NO	NO
3rd degree IV	NO	YES	YES	YES	YES
Hansen J-statistic		0.725			
(p-value)		(0.394)			
Fixed Effects	YES	YES	YES	YES	YES
Only Network	YES	YES	YES	YES	YES
No. of Communities					
(Community $\times$ Year Effects)	0	0	0	10	20
No. of Observations	6914	6914	6914	6914	6914

Table 1.4 shows the results from the 2SLS estimates corresponding column-to-column to the first stage regressions from Table 1.3. I focus the discussion on the estimate of parameter δ, as all other variables included in the model are estimated similarly to the OLS baseline.[29] By instrumenting the spillover variable with the R&D of indirect friends of second degree (column 1), δ is estimated around 0.0116, a figure substantially identical to the

[28]When the two instruments are included together, the coefficient for the third-degree instrument is negative and statistically different from zero. Nevertheless, in magnitude it is remarkably closer to zero, as it should be expected from conditioning on the R&D of "second-degree" firms.

[29]The exception is the Jaffe measure of technological proximity: its point estimate, apart from not being

one obtained from OLS estimates. When including both instruments (column 2) the result is similar: $\hat{\delta} = 0.0113$.[30] By only instrumenting for the third-degree indirect connections instead (column 3), the result is different: the point estimate of δ is substantially higher, hovering around 0.0160. Interestingly, the inclusion of community-by-year effects results in further increases of this point estimate: $\hat{\delta} = 0.0206$ with 10 communities (column 4) and $\hat{\delta} = 0.0184$ with 20 communities (column 5). All estimates of δ are statistically significant at the 1% level.[31]

These results are telling in two respects. First, they evidence a negative bias in simple OLS estimates. This bias is likely due to factors that are common to connected firms, which on the one hand are associated with higher R&D investment, but on the other hand have negative (albeit possibly temporary) effects on firm performance. Second, that a change in the point estimate of δ in only apparent when instrumenting for the third-degree instrument in isolation is remarkably consistent with the analytical framework outlined in this work, as well as with the descriptive evidence on the spatial autocorrelation of R&D from Figure 1.5. Both things point to the possibility that the second-degree instrument is actually correlated with the unobserved factors summarized by ω. This suspicion is corroborated by the comparison of the estimates of δ that are presented in Table 1.4.

1.5.3 Market Value

The results for the market value model (1.29) are displayed in table 1.5. In column (1) and (2) I estimate the model via OLS, respectively on the whole sample and on the network subsample. In columns (3), (4) and (5) I perform 2SLS on the subsample, using respectively the 2nd degree instrument, both instruments or the 3rd degree instrument in isolation. The estimates for the spillover parameter lie in an interval around 0.04; the corresponding elasticity of a symmetric increase in the R&D stock of all the connections of the average firm is about 0.02. Unlike the case of the production function, the application of the IV strategy does not clearly evidence the presence of a bias in the baseline OLS estimates. Noticeably, the 2SLS estimate of $\tilde{\delta}$ employing the third-degree instrument in isolation is closer to the OLS baseline (about 0.0350) than the two estimates obtained by using the second-degree instrument (both around 0.0425), while losing in precision. One possible interpretation of these results is as follows: those simultaneous factors that drive R&D, and which have synchronic effects on productivity, do not influence one firm's market value; presumably because the latter is a forward-looking variable and investors price in the future gains associated with current investment in R&D.

---

statistically significant, is less robust to different specifications; particularly, it is equal to zero in the estimate from column (5).

[30] The Hansen J overidentification test has a p-value of about 0.4, indicating that two instruments effectively capture different sources of variation. This is consistent with the hypotheses on the network structure of common dependence that have been outlined in Section 1.2.

[31] As expected, the standard errors are substantially larger – about double in magnitude – when instrumenting spillovers only for third-degree connections.

Table 1.5: Market Value, 1981-2001

	(1)	(2)	(3)	(4)	(5)
R&D Stock / Capital ($t-1$)	0.3877	0.4551	0.4487	0.4485	0.4529
	(0.2805)	(0.4010)	(0.3995)	(0.3997)	(0.3950)
Spillovers ($\tilde{\delta}$)	0.0391	0.0311	0.0422	0.0426	0.0349
	(0.0081)	(0.0069)	(0.0092)	(0.0093)	(0.0177)
Geospills	-0.0001	0.0000	-0.0002	-0.0002	-0.0001
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0006)
Jaffe Tech. Proximity	0.0129	-0.2430	-0.2885	-0.2899	-0.2585
	(0.1837)	(0.2617)	(0.2483)	(0.2477)	(0.2686)
BSV Business Stealing	-0.0338	0.1193	0.1161	0.1160	0.1182
	(0.0650)	(0.1053)	(0.1054)	(0.1055)	(0.1040)
Instrument(s)	OLS	OLS	2nd dist.	Both	3rd dist.
Fixed Effects	YES	YES	YES	YES	YES
Only Network	NO	YES	YES	YES	YES
No. of Observations	11357	6806	6806	6806	6806

1.5.4 Patent Count

The results for the patent count model (1.30) are reported in table 1.6, which is organized along the lines of table 1.5: column (1) reports the results from whole sample, column (2) those from the network subsample, while the results from the control function approach, based on the usual sequence of instrument combinations, are given in columns (3), (4) and (5). The coefficient for the spillover-connections parameter is estimated in an interval around 0.03 in columns from (1) to (4). The estimates from column (5) however, based on a control function approach using only the third-degree instrument, register a much larger point estimate for $\breve{\delta}$, equal to 0.0870. This is again interpreted as the elasticity of patent output relative to an increase of all connections' R&D, which is approximately equal to 0.0383 for the average firm in the network. It is worth noticing how the Jaffe measure again loses all its economic and statistical significance once the analysis is restricted to only firms included in the coauthorship-induced network.

The interpretation of the estimates from column (5) is non-trivial. The large increase in the point estimate of $\breve{\delta}$ is due, of course, to the presence of common factors that drive R&D for close firms. A scenario in which the R&D of individual firms is positively correlated to these factors, which by themselves would also have a negative effect on firm patent production, is one where firms react to periods of lower innovation rate by parallely increasing their R&D effort. An intriguing, alternative hypothesis is that simultaneous increases in R&D spending correlate to worse innovation on the quality margin. It should be reminded that the patent outcome on the left hand side of equation (1.30) weights patents by their citations. A typical narrative about some industries (e.g. the pharmaceutical sector) associates lower quality patents to an increase in the total number of USPTO registered inventions. Identifying the specific mechanism driving this result well deserves some separate investigation.

Table 1.6: Patent Count, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D ($\breve{\gamma}$)	0.0831	0.0761	0.0741	0.0769	0.0527
	(0.0313)	(0.0356)	(0.0361)	(0.0359)	(0.0391)
Spillovers ($\breve{\delta}$)	0.0295	0.0268	0.0377	0.0317	0.0870
	(0.0088)	(0.0063)	(0.0077)	(0.0073)	(0.0168)
Geospills	0.0008	0.0013	0.0012	0.0013	-0.0000
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0006)
Patents (t-1)	0.3956	0.4156	0.4143	0.4164	0.4103
	(0.0197)	(0.0176)	(0.0165)	(0.0167)	(0.0167)
Jaffe Tech. Proximity	0.3044	0.0026	0.0003	0.0003	-0.0014
	(0.0820)	(0.0552)	(0.0541)	(0.0541)	(0.0542)
Industry Dummies	YES	YES	YES	YES	YES
Only Network	NO	YES	YES	YES	YES
Control Function	NO	NO	YES	YES	YES
			2nd dist.	Both	3rd dist.
No. of Observations	11444	6587	6587	6587	6587

1.5.5 Discussion

The results from this section can be summarized as follows. The spillover variable constructed by aggregating the R&D of connected firms – where the definition of "connection" is based on the extent to which the inventors of two different firms are professionally related to each other – has a positive and sizeable effect on the productivity, market value and patent output of firms. This spillover variable is more robust to alternative model specifications and identification strategies, with respect to measures of spillovers that are more traditional in the literature. In addition, the application of the proposed IV strategy, when based on the R&D choices of indirect friends of third degree, induces an increase of the point estimates associated to the spillovers variable for both the productivity and the patent output models. This evidences the presence of negative biases, likely due to common confounders, in the estimates obtained from less sophisticated approaches. This fact is very much consistent with the analytical framework developed in this chapter, which identifies the relationship between the spatial autocorrelation of R&D in the network and estimation biases due to common shocks.

A way to quantify the economic relevance of this effect is to calculate the average Marginal Private Returns (MPR) and Marginal Social Returns (MSR) of R&D (see e.g. BSV). I define the MPR as the average increase in output relative to an increase in the R&D stock of the individual firm (MPR $= \frac{1}{N}\sum_{i=1}^{N} dY_i/dS_i$), while the MSR is the average increase in output relative to the average increase in the R&D stock of all the other firms (MPR $= \frac{1}{N}\sum_{i=1}^{N} dY_i/\overline{dS}$). These are easily calculated under the hypothesis of an homogeneous percentage increase in R&D by all firms ($dS_i/S_i = dS/S$ for all i). In this case, the average response of output dY/dS can be derived from (1.18) and decomposed as follows.

$$\frac{1}{N}\sum_{i=1}^{N}\frac{dY_i}{dS} = \underbrace{\hat{\gamma}\frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{S_i}}_{=\text{MPR}} + \underbrace{\hat{\delta}\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N} g_{ij}\frac{Y_i}{S_i}}_{=\text{MSR}} \tag{1.31}$$

To evaluate the MPR and the MSR, I use the estimates for $\hat{\gamma}$ and $\hat{\delta}$ from column (5) of Table 1.4, as well as the values $\frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{S_i} = 11.36$ and $\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} g_{ij}\frac{Y_i}{S_i} = 5.66$ calculated on the pooled panel. As a result, the MPR is approximately equal to 44% (= 11.36 × 3.85%) and the MSR to about 10.4% (= 5.66 × 1.84%):[32] the latter is approximately equal to 24% of the former. While not as large as the evaluations from other studies, these are realistic and economically significant values. Notice that these calculations do not take into account the "amplification effect" due to strategic response and complementarities, something that should be taken into account when evaluating, say, the effect of an R&D-stimulating policy.

The empirical strategy adopted in this chapter seeks to address the problem of common confounders. There are, however, some additional empirical concerns that remain unanswered. The first one is ultimately an issue of measurement error. Let us ignore for the moment the problem of measuring the innovation effort and the knowledge stock of connected firms with cumulated R&D expenditures, as well as the additional difficulty that the connection metrics may not fully capture the degree of interactions between R&D teams. Even in absence of these issues, there is a problem of *network sampling* that ultimately depends on the inability to observe nodes and links, even if their intensity were perfectly measurable (Chandrasekhar and Lewis, 2011). In the context of this chapter, given the type of sample selection implied by COMPUSTAT data (small and private firms are excluded), this would result in an underestimation of the effect of connections. A proper assessment of this problem requires the application of proper sampling strategies to high quality firm-level data that can be matched to patents: a non-trivial set of requirements.

The second problem is the endogeneity of the connections. In particular, if for any j such that $g_{(ij)t} \neq 0$, $g_{(ij)t}$ and the error term of firm i are correlated, any identifying assumption employed in this analysis is violated. This can happen if, for instance, the more productive

[32] Notice that these are the calculated returns from the R&D *stock*. To estimate the returns from annual R&D *expenditures*, one should divide these figures by the steady-state flow/stock ratio. Using the typical assumption of a 0.20 steady-state ratio, one gets an approximate 8.8% private return and an approximate 2.1% social return from yearly R&D expenditures.

firms are able to attract the most connected inventors, in which case the estimates of δ suffer from a positive bias. The reverse could also be true if, say, attracting connected inventors is a viable alternative to internally developing new knowledge for the least productive firms. To deal with this issue, I have already taken the step of narrowing the analysis down to only those firms that enter the network at any point in time. This generally results in a sizeable but not dramatic reduction of the point estimate for δ, while it greatly affects the estimate of the Jaffe measure. However, this does not account for the possibility that within firms, the dynamics of the rise and appearance of connections may be correlated with changes in unobservable characteristics of individual firms. The analysis of this problem is tightly connected to theories of network formation, and it very much deserves a separate study.

1.6 Conclusion

In this chapter I propose a new method of evaluating R&D spillovers. Thanks to information on patent coauthorship relationships between individuals that work for different organizations, I construct a network of firms that are connected through their R&D teams. I evaluate the dependence of firm productivity, market value and innovation rate from the R&D performed by firms connected in the network, weighted by the intensity of mutual links. Worried by the possibility of common confounders that simultaneously drive R&D choices and firm-level outcomes, I employ an identification strategy based on the network topology. In particular, I instrument the R&D choices of one firm's direct connections with those of sufficiently distant links. Under conditions specified by a formal model firms' interaction, appropriately constructed instrumental variables predict the intensity of spillovers received by one firm but are otherwise unrelated to its performance and innovation outcomes.

Estimates based on this definition of connections register sizeable spillovers of peers' R&D on the productivity, market value, and patent output of firms. These results, unlike those based on more traditional measures of R&D spillovers, are robust to different specifications, as well as to the restriction of the sample to the largest and most R&D intensive firms. In striking conformity with the prediction of the analytical model, the application of the identification strategy that I propose shows that, when instrumenting peers' R&D with the R&D of sufficiently distant firms, point estimates of spillover effects on productivity and patent output increase substantially. This suggests that common factors driving both R&D and firm outcomes might do so in opposite directions. Among other things, this finding may reflect the stylized fact that over the industry life cycle faster innovation is typically associated with higher costs. I use the estimates of spillovers obtained from the proposed methodology in order to evaluate the relative importance of the marginal social returns to R&D relative to the private returns, finding that the former are about 24% of the latter.

This work suggests several directions for further research. First, this analysis should be replicated in a setting (ideally, a specific high-tech industry) in which it is possible to observe all the possible connections of one firm, rather than a subset of them. This may allow to overcome the estimation problems that result from network sampling. Second, it would be interesting to evaluate the effects of spillovers in a context in which the identity and the individual professional histories of the connecting inventors can be actually observed. This would let researchers disentangle the relative importance of worker mobility, superstar inventors, academic collaborations, joint ventures, etc., in determining the extent of knowledge spillovers. Finally, on the methodological side, a fruitful line of research would be to integrate the analytical model on which the identification strategy of this chapter is grounded with explicit models of network formation, in order to better understand the determinants of cross-firm connections and their ultimate impact on R&D decisions and productivity.

Chapter 2

Identification of Spillovers in Network Games

2.1 Introduction

The main objective of the analysis conducted in the previous chapter is to uncover the result of R&D spillovers in terms of ultimate firm outcomes that are relevant for both stakeholders and policymakers: namely productivity, innovation, and measures of firm market value. In doing so, I have proposed and implemented an econometric methodology that accounts for the problem of correlated confounders, within a framework based on the direct estimation of R&D spillovers on those ultimate outcome measures. This strategy is motivated on some specific assumptions about the behavior of firms in response to the investment and market decisions of other businesses: namely, that firms do not respond to the choices of those who are "very distant" in the network. In this context the assumptions are supported by the data, but in other settings this approach may not be equally applicable.

In this chapter I propose an alternative method of estimating R&D spillovers occurring between firms. Instead of measuring the ultimate impact of other companies' R&D on some firm-level outcome, I estimate their elasticity on the R&D investment choice of the individual firm. This dependence is due to the strategic anticipation of network externalities, whether peers' R&D is a strategic complement or rather a substitute. In other words, I quantify R&D spillovers directly via the equilibrium relationships that characterize the solution of the firms' investment game. The same relationships allow to implement a solution to the reflection problem (Manski, 1993) and to the problem of correlated confounders, which is alternative to the one from the previous chapter. In particular, in order to disentangle different sources of cross-correlation in observed R&D, I exploit specific covariance restrictions that can be derived from the equilibria of the R&D investment game in presence of spillovers. The game itself is a more general version of the one proposed in the previous chapter. Specifically, it is characterized by fewer imposed assumptions on both the functional form, the stochastic properties of the unobserved shocks and the informational structure.

The method proposed here has several advantages: the obvious one is that, as mentioned, it requires fewer assumptions on the rules governing strategic interaction (the game), and consequently on the nature of the spatial cross-correlation of the unobserved shocks in the network. In addition, the approach being advanced is more robust to misspecifications of the shape of firms' objective function, and it is thus more flexible – potentially accomodating, for instance, the case of strategic substitutes as opposed to the one of strategic complements. As usual, what is gained in terms of generality is traded off with losses in terms of increased complexity. Moreover, this method is not generally applicable: as it is subsequently detailed, it is valid only in networks, such as the one under analysis, that are sufficiently sparse – that is, those that are characterized by a bounded node degree.

The identification in the model comes from the combination of the best response functions derived from the game and the covariance restrictions derived from the resulting equilibria. Intuitively, the covariance between the investment of two indirectly connected firms cannot depend on simultaneous strategic dependence from the R&D choices of their common peers, if these are held constant. Hence, any residual correlation between the R&D of these two firms can only depend – if anything – on the effect of common shocks. These covariance restrictions, together with the moment conditions relative to the variance of R&D, let also estimate the variance of common shocks. The empirical estimates of R&D spillovers, derived from the GMM implementation of the proposed approach, are quantitatively consistent with the results obtained in the previous chapter, supporting the validity of both methodologies. Specifically, the normalized elasticity of peers' R&D investment on one firm's knowledge capital is estimated to be about 0.2, which is consistent with the a normalized elasticity of peers' R&D on firm productivity of about 0.03.

The idea of exploiting covariance restrictions to measure spillovers is not novel and it has been previously employed in the estimation of peer effects in education (see e.g. Graham (2008), Pereda-Fernández (2015)). In this context instead, rather than modeling symmetrical interactions between individuals in groups (which is customary in the analysis of peer effects in the classroom) the focus is on non-overlapping, asymmetrical relationships that are typical of networks, where each firm interacts with a different set of other firms. Such heterogeneity of the individual reference groups implies a particularly defined set of restrictions in the covariances of their R&D investment levels, which are the basis for identification. Furthermore, in the present framework the covariance restrictions are combined with other moment conditions of the first and second order. To this end, I adapt the methods developed by Bonhomme and Robin (2009, 2010), which allow for a heteroschedastic variance of the error term modeled as a function of the covariates.

This chapter is organized as follows. Section 2.2 outlines the analytical framework, a more general version of the one proposed in Section 1.2. Section 2.3 discusses the identification and estimation method that is derived from the model. Section 2.4 provides further details on the data described in the previous chapter; such details are particularly relevant in light of the proposed estimation strategy. Section 2.5 presents the empirical results. Finally, Section 2.6 recapitulates the discussion, concluding the chapter. In addition, Appendix B provides mathematical proofs and further technical details relative to the proposed GMM model.

2.2 Analytical Framework

In this section I outline the analytical model on which the identification strategy proposed in this chapter is founded. In this model, I describe the strategic dependence of economic agents when they exert reciprocal positive externalities – "spillovers" – in a network. In particular, I characterize the stochastic properties of equilibrium outcomes when agents are subject to random exogenous shocks. I distinguish the resultant *endogenous* cross-correlation that arises because of agents' strategic choices in equilibrium, from the *exogenous* one – that follows from the stochastic dependence of the primitive shocks across connected agents. These two sources of cross-correlation have different economic interpretations, that are context-specific.

In its most general version, the model is flexible enough to be suitable for a wide range of economic settings of interest, and to allow for both strategic complementarities and substitutabilities. However, for both illustrative purposes and for consistency with the empirical application of our methodology, I contextualize the model in terms of a game of R&D investment played between firms that are linked in a network. Investment in R&D boosts firms' productivity; moreover, connected firms enjoy mutual R&D spillovers. In addition, firms are subject to exogenous shocks to their productivity, which are complementary to R&D. The degree of similarity of the shocks received by firms that are (possibly indirectly) connected in the network depends on the nature of specific hypotheses, which are separately discussed.

2.2.1 General Setup

Consider a network $\langle N, \mathbf{G} \rangle$, composed by N players connected to each other in a way that is summarized by the adjacency matrix $\mathbf{G}$ (of which each entry $g_{ij} \in [0,1]$, $g_{ii} = 0$ denotes the strength of the relationship between i and j). Every player is characterized by an objective function $\pi(\cdot)$ with the following general form:

$$\pi_i\left(S_i, \mathbf{S}_{-i}, \mathbf{g}_i; \omega_i\right) = b\left(S_i, \mathbf{S}_{-i}, \mathbf{g}_i\right) d\left(\omega_i\right) - \kappa S_i \tag{2.1}$$

where S is a strategic variable to provide which each agent bears a linear cost κ: to keep the exposition general I call it "effort". The payoff of each agent i also depends on the choices of all other players in a way that results from his position (row) in the network $\mathbf{g}_i$. By contrast, ω_i is a stochastic shock to the final outcome that is received by each agent. It is assumed that – with some liberty in the use of notation – $b'\left(S_i\right) > 0$, $b''\left(S_i\right) < 0$, $b'\left(S_j\right) > 0$, $b''\left(S_j\right) > 0$, $d'\left(\omega_i\right) > 0$ and $d''\left(\omega_i\right) < 0$: the objective function is increasing with decreasing returns in each input. The multiplicative separability of the objective function in the component that depends on strategic chices $b(\cdot)$ and the shock component $d(\cdot)$ ensures complementarity of the two $\left(\partial^2 \pi(\cdot) / \partial S \partial \omega > 0\right)$. That is, a positive shock on the outcome raises the marginal return of both private and social effort.

This setup allows for both strategic complements and substitutes. Whether the game features either is a matter of functional form – and economic – assumptions. In the case of strategic complements, an increase in effort by peers makes private effort even more worthwile

($\partial^2\pi(\cdot)/\partial S_i \partial S_j > 0$ for any j). A functional form for the objective function that entails strategic complements is as follows (see e.g. Manresa (2014) and Pereda-Fernández (2015)).

$$\pi_i(S_i, \mathbf{S}_{-i}, \mathbf{g}_i; \omega_i) = A_i S_i^{\gamma} \left(\prod_{j \neq i}^{N} S_j^{g_{ij}} \right)^{\delta} e^{\omega_i} - \kappa S_i \tag{2.2}$$

In the opposite case of strategic substitutes, access to peers' effort diminishes the marginal utility from private effort ($\partial^2\pi(\cdot)/\partial S_i \partial S_j < 0$ for any j). A functional form that is analogous to (2.2) but differs for featuring strategic substitutes, is the following

$$\pi_i(S_i, \mathbf{S}_{-i}, \mathbf{g}_i; \omega_i) = A_i \left(S_i + \delta \sum_{j \neq i}^{N} g_{ij} S_j \right)^{\gamma} \omega_i - \kappa S_i \tag{2.3}$$

which is a special case from the general class of functions examined by Kranton et al. (2014), with the addition of the exogenous shock component. In both functional forms, the parameter δ denotes the total strength of spillovers.

I frame the model as a game of R&D investment in the presence of spillovers between firms; where the objective function is profit, S is R&D, A is the residual component of the production function (including other inputs), and ω is an unobserved shock to productivity. This model differs from the one presented in the previous chapter in two respects. First, the present model is more general in terms of assumed functional forms. This makes the derived identification results very powerful, since they can be applied to a larger variety of settings (e.g. those featuring either strategic complementarities or substitutabilities). The "Bayesian" model from the first chapter is, however, more general in terms of informational assumptions; while the complete information setting featured in the present chapter is, there, only a special case. Incidentally, in this special case identification of spillovers is not possible; hence the importance of the identification results presented hereinafter.

2.2.2 Equilibrium

In a simultaneous game of complete information, agents (firms) make an optimal effort choice (R&D investment) taking into accounts those of their connections or peers. With strategic complements; they tend to raise their effort facing higher effort from their peers; with strategic substitutes the response is negative. The Nash Equilibrium is found as the fixed point of all agents' best reply functions. It turns out that the latter can be represented, for a wide class of functional forms, with the following expression:[1]

$$s_i - \vartheta \sum_{j \neq i}^{N} g_{ij} s_j = h(\bar{s}_i, \varpi_i) \tag{2.4}$$

where:

[1]To simplify the analysis, and following standard practice, I assume – here implicitly – regularity conditions ensuring that in equilibrium the optimal effort is always non-negative for each agent.

- $s_\ell = s(S_\ell)$ is a function of effort (R&D);
- $\vartheta(\delta)$ is a function of spillovers' strength δ;
- $\bar{s}_i$ is an agent-specific function of other parameters of the model such as A_i or κ;
- $\varpi_i(\omega_i)$ is a function of the shock ω_i;
- $h(\bar{s}_i, \varpi_i) \equiv h_i$ is a generic function increasing in both its arguments.

For instance, in the case of complementarities as per the objective function in (2.2), $s(S_\ell) = \log S_\ell$, $h(\cdot)$ is linear and $\vartheta = \delta(1-\gamma)^{-1}$. In the case of substitutabilities as per (2.3) instead, $s(S_\ell) = S_\ell$, $h(\cdot)$ is also linear and $\vartheta = -\delta$. In both cases, $\varpi_i = \lambda\omega_i$ where λ is a constant that can be normalized as $\lambda \equiv 1$ without loss of generality.[2] For illustrative purposes ω henceforth replaces ϖ in the notation, even if the two do not necessarily coincide.

The resulting equilibrium is most conveniently expressed in matrix form:[3]

$$\mathbf{s}^* = (\mathbf{I} - \vartheta\mathbf{G})^{-1}\,\mathbf{h}(\bar{\mathbf{s}}, \boldsymbol{\omega}) \tag{2.5}$$

where $\mathbf{h}(\bar{\mathbf{s}}, \boldsymbol{\omega}) = (h(\bar{s}_1, \omega_1), \ldots, h(\bar{s}_N, \omega_N))' \equiv \mathbf{h}$. The equilibrium is unique by applying realistic restrictions on the strength of spillovers (Calvó-Armengol et al. (2009); Kranton et al. (2014)).[4] In the equilibrium, the choices of each agent (firm) depend on the characteristics *and the shocks* of all others – not just those to which one is directly connected. By influencing the choices of its direct links, each agent (firm) also indirectly affects those who are more far away in the network, and internalizes this fact in equilibrium. In particular, each agent is particularly responsive to both choices and shocks of those agents who are very *central* in the network (that is, those who are well-connected to other well-connected agents). This is a known result from the analysis of games in networks and is not being explored further here.

2.2.3 Stochastic Properties: Endogenous Dependence

I now analyze the stochastic properties of the equilibrium under the hypothesis that the shocks ω are independent; specifically, the exogenous factors that affect the outcome (performance) of agents (firms) are uncorrelated, even across connected nodes in the network. Notwithstanding this hypothesis, it results that equilibrium effort (R&D) is cross-correlated throughout the entire network, since each agent responds strategically – even if marginally – to the shocks of any other agent it is remotely linked to. I define this a situation of *endogenous* dependence, so to distinguish it from the *exogenous* dependence, caused by some primitive cross-correlation in the shocks themselves. In order to properly describe this scenario, it is useful to formalize the independence across the ω shocks with the following hypothesis.

[2] Constant λ might be a function of other structural parameters, yet it includes a model-specific constant factor which, empirically, cannot be identified separately from the variance of ω (hence the normalization).

[3] Notice that the matrix $(\mathbf{I} - \vartheta\mathbf{G})$ is invertible almost surely.

[4] Uniqueness ensures that there are neither "explosive" equilibria in the case of complementarities, nor "asymmetric" equilibria – where few agents provide effort for many others – in the case of substitutabilities.

Hypothesis 1. Uncorrelated shocks: *The individual shocks are equal to the idiosyncratic shocks; specifically:* $\omega_i = \upsilon_i$. *Moreover, the components of the vector* $\boldsymbol{\upsilon} = \{\upsilon_i\}_{i=1}^N$ *are jointly independent and homoschedastic, and whose covariance matrix is as follows.*

$$\mathbf{\Omega}_{endo} \equiv \mathbb{V}\text{ar}\left(\boldsymbol{\omega}\right) = \sigma_\upsilon^2 \boldsymbol{I}$$

Under Hypothesis 1 and the additional assumption, made for illustrative purposes, that $h\left(\bar{s}_i, \omega_i\right) = \omega_i$, the cross-covariance of the equilibrium solution in 2.5 is:

$$\mathbb{V}\text{ar}\left(\mathbf{s}^*\right) = \sigma_\upsilon^2 \left(\mathbf{I} - \vartheta \mathbf{G}\right)^{-2} \tag{2.6}$$

this special case illustrates easily how, in equilibrium, the R&D choice of every firm is correlated to the R&D of every other firm connected to the same component of the network. The reason is that firms are strategically interdependent, even if indirectly: for any genereric firm i, the R&D choice of another, very distant firm ℓ matters for the calculation of the marginal return of R&D, since it directly or indirectly affects other firms in the network, including some of firm i's peers. This result follows from the assumption of complete information, and clearly invalidates the identification strategy proposed in the first chapter (which extensively illustrates – and exploits – the implications of incomplete information hypotheses).

This baseline scenario parallels some typical assumptions in the econometrics literature, according to which the only dependence across agents is attributed to observables. In this specific case, as eventually explained in more detail in this chapter, it is possible to isolate the spillover effect by appropriately conditioning on neighbors' choices. To get some intuition, consider an open triad like the one depicted in Graph 1.1 shown in the previous chapter: in that example, under the hypothesis of complete information (but no common shocks), the R&D investment choices of firms i and k are mutually interdependent since both affect – and are affected by – the choice of firm j. If firm j's R&D investment were held constant, such interdependence would be absent. In order to exploit this idea properly, one has to establish identification conditions on the basis of the equilibrium stochastic properties of the model.

2.2.4 Stochastic Properties: Adding Exogenous Dependence

I now abandon the hypothesis that firms are subject to purely independent shocks. In particular, I outline two alternative sets of assumptions describing different stochastic process that are dependent on the network topology. The two processes generate a primitive cross-correlation in the shock that extends to higher degrees of distance: in the first case (edge-specific shocks), the resulting cross-correlation in the shocks is limited to one degree of distance; in the second case (node-specific shocks), it extends further to the second degree. I do not formally model here the microeconomic origins (e.g. the specifics of the network formation process) that can rationalize either kind of stochastic process, but I provide some intuition; in fact, the main focus of this discussion is the description of the consequences that either assumption entails on the stochastic properties of the model's equilibrium.

Edge-specific shocks. This hypothesis corresponds to the idea that connected firms share some specific characteristics that are exclusive to them, that is, not shared by any other firm outside one link. In particular, I am introducing a set of additional shocks ψ_{ij} that are specific to each pair of firms. Potentially, there are $N(N-1)/2$ of these extra shocks; but only a fraction of this number, equaling the total number of edges of the network, actually enters the model. This can be thought as a way to model pair-specific characteristics. An economic rationale for this assumption is homophily, a type of network formation process where agents that are similar to each other develop reciprocal bonds. Formally, edge-specific shocks are expressed by the following hypothesis.

Hypothesis 2. Edge-specific shocks: *The individual shocks are equal to the idiosyncratic shocks plus the sum of the edge-specific shocks relative to the connected firms; specifically:* $\omega_i = \upsilon_i + \sum_{j=1}^{N} g_{ij}\psi_{ij}$. *Moreover, the edge-specific shock is symmetric, i.e.* $\psi_{ij} = \psi_{ji}$, *and the components of the vector*

$$(\boldsymbol{\upsilon}, \boldsymbol{\psi}) = \left(\{\upsilon_i\}_{i=1}^{N}, \left\{ \{\psi_{ij}\}_{i=1}^{N-1} \right\}_{j=i+1}^{N} \right)$$

are jointly independent and homoschedastic within type, with covariance matrix as follows.

$$\boldsymbol{\Omega}_{exoedge} \equiv \mathbb{V}\mathrm{ar}(\boldsymbol{\omega}) = \sigma_\upsilon^2 \boldsymbol{I} + \sigma_\psi^2 \left(diag\left(\boldsymbol{G}\iota\right) + \boldsymbol{G} \right)$$

Notice that the covariance structure described in (2) is a more general version of the one introduced in the analytical framework of the previous chapter; expressed in particular by equation (1.4) or (A.3).[5] Under Hypothesis 2 and for $h(\bar{s}_i, \omega_i) = \omega_i$ the equilibrium cross-correlation of R&D is

$$\mathbb{V}\mathrm{ar}(\mathbf{s}^*) = (\mathbf{I} - \vartheta\mathbf{G})^{-1} \left[\sigma_\upsilon^2 \mathbf{I} + \sigma_\psi^2 \left(diag\left(\mathbf{G}\iota\right) + \mathbf{G} \right) \right] (\mathbf{I} - \vartheta\mathbf{G})^{-1} \tag{2.7}$$

which, relative to the case of purely endogenous dependence, poses an additional problem of identification. In particular, with edge-specific shocks, the R&D of nodes located at distance two or higher are not mutually correlated when conditioning for intermediate connections (like j from Graph 1.1). Thus, it is not possible to identify the variance of ψ solely through covariance restrictions relative to unconnected firms; in order to obtain separate identification it is necessary – as illustrated later – to specify additional moment conditions.

Node-specific shocks. This hypothesis instead corresponds to the case in which some characteristic of each node are shared by all its peers, but without exclusivity. This is a more straightforward approach to model common shocks, as it allows some partial overlap of the shared factors across nodes. In particular, I am introducing a set of N extra shocks ϕ_i to which every firm is subject; these shocks are diffused amongst any firm i's peers with an intensity that is proportional to connections between pairs. This setup models best those circumstances in which some exogenous factors – such as specific technological opportunities – have a spatially limited diffusion in a network. It is formalized via the following assumption.

[5] See Appendix A for the latter.

Hypothesis 3. Node-specific shocks: *The individual shocks are equal to the idiosyncratic shocks plus the sum of the node-specific shocks relative to the connected firms; specifically:* $\omega_i = \upsilon_i + \phi_i + \sum_{j=1}^{N} g_{ij}\phi_j$. *Moreover, the components of the vector*

$$(\boldsymbol{\upsilon}, \boldsymbol{\phi}) = \left(\{\upsilon_i\}_{i=1}^{N}, \{\phi_i\}_{i=1}^{N} \right)$$

are jointly independent and homoschedastic within type, with covariance matrix as follows.

$$\boldsymbol{\Omega}_{exonode} \equiv \mathbb{V}\mathrm{ar}\left(\boldsymbol{\omega}\right) = \sigma_{\upsilon}^{2}\boldsymbol{I} + \sigma_{\phi}^{2}\left(\boldsymbol{I} + 2\boldsymbol{G} + \boldsymbol{G}^{2}\right)$$

The equilibrium cross-covariance in the shock extends up to two degrees of distances, like the exogenous cross-correlation with $C = 2$ that is hypothesized to rationalize the empirical evidence from Figure 1.5. In particular, in the special case when $h\left(\bar{s}_i, \omega_i\right) = \omega_i$ it reads as follows.

$$\mathbb{V}\mathrm{ar}\left(\mathbf{s}^{*}\right) = \left(\mathbf{I} - \vartheta\mathbf{G}\right)^{-1}\left[\sigma_{\upsilon}^{2}\mathbf{I} + \sigma_{\phi}^{2}\left(\mathbf{I} + 2\mathbf{G} + \mathbf{G}^{2}\right)\right]\left(\mathbf{I} - \vartheta\mathbf{G}\right)^{-1} \tag{2.8}$$

This poses less problems relative to the separate identification of the variance of ϕ: since the R&D of all nodes in any open triad are always correlated to each other in presence of node-specific shocks, identification is achieved with any covariance restriction.

2.3 Identification and Estimation

This section describes in more detail the identification strategy – and the resulting estimation method – proposed and implemented in this chapter. Such strategy, both in presence and in absence of exogenous dependence, is based on specific covariance restrictions relative to the R&D investment of different firms. The intuition that allows to identify spillover effects net of correlated confounders is that by conditioning on intermediate choices along a path, one isolates some cross-variability of different firms' R&D that does not depend on common confounders. The intuition that lets disentangle endogenous and exogenous dependence is that, after conditioning, the only residual correlated confounders that simultaneously affect both the variance and the covariance of two unrelated nodes are of the exogenous kind.

After presenting a functional form assumption specifying heteroschedasticity in the model as a function of covariates, I characterize the main identification conditions. I first present the simplest result, which is limited to only endogenous dependence; I eventually extend it to the case that includes exogenous factors. Subsequently, I discuss how to obtain accurate identification in presence of misspecification. Finally, after outlining a number of necessary conditions, I describe the GMM econometric model employed in the estimation.

2.3.1 Heteroschedasticity and Covariates

In order for the model to be realistic, it must include covariates and allow for heteroschedastic variance components. The following functional form assumption introduces them both, and models the former as a function of the latter. It encompasses a general class of functional forms that includes those most tipically employed in the applied analysis, like (2.2) or (2.3).

Assumption 1. Linearity of shocks in the best response function: *The equilibrium R&D of a firm depends on a vector of observable factors* $\mathbf{x}_i$ *and on that firm's unobservables* ω_i*. This dependence can be separated in two additive components as*

$$h(\bar{s}_i, \omega_i) = \tilde{s}(\mathbf{x}_i, \boldsymbol{\xi}) + f_\omega(\mathbf{x}_i, \boldsymbol{\zeta})\,\omega_i \tag{2.9}$$

where $\boldsymbol{\xi}$ *and* $\boldsymbol{\zeta}$ *are two finite sets of parameters.*

To avoid an overloading notation, let $\tilde{s}_i$ and $f_{\omega,i}$ respectively denote $\tilde{s}(\mathbf{x}_i, \boldsymbol{\xi})$ and $f_\omega(\mathbf{x}_i, \boldsymbol{\zeta})$.[6] Moreover, define $\mathbf{X}$ as the matrix collecting all horizontal vectors $\mathbf{x}'_1, \ldots, \mathbf{x}'_N$ and $\mathbf{F}_\omega(\mathbf{X})$ as the diagonal matrix whose i-th element equals $f_\omega(\mathbf{x}_i, \boldsymbol{\zeta})$. Under Assumption 1, it is possible to express the conditional covariance of $\mathbf{s}^*$ in general terms, to account for both the inclusion of covariates into the model and the presence of observables-dependent heteroschedasticity. Such a general expression is given by the following equation.

$$\mathbb{V}\text{ar}(\mathbf{s}^* | \mathbf{X}) = (\mathbf{I} - \vartheta\mathbf{G})^{-1}\mathbf{F}_\omega(\mathbf{X})\,\mathbb{V}\text{ar}(\boldsymbol{\omega})\,\mathbf{F}_\omega(\mathbf{X})(\mathbf{I} - \vartheta\mathbf{G})^{-1}$$

2.3.2 Identification: Endogenous Dependence

Consider the case in which there are no common shocks (Hypothesis 1): the only equilibrium dependence between nodes in a network is due to the endogenous response to individual shocks. The specific covariance matrix of $\mathbf{s}^*$ from equation (2.6), holding Assumption 1 and for $\mathbb{V}\text{ar}(\boldsymbol{\omega}) = \sigma_v^2\mathbf{I}$, takes the following form.

$$\mathbb{V}\text{ar}(\mathbf{s}^* | \mathbf{X}) = \sigma_v^2(\mathbf{I} - \vartheta\mathbf{G})^{-1}\mathbf{F}_\omega^2(\mathbf{X})(\mathbf{I} - \vartheta\mathbf{G})^{-1} \tag{2.10}$$

I now present the fundamental result that allows separate identification of spillover effects and unobserved shocks. This result is general: Assumption 1 is not necessary for it to hold.

Proposition 1. *Let the minimum path length between firms* i *and* k *be 2 or higher. Define the set of all firms that are directly connected to either* i *or* j*:* $\mathbf{H}_{ik} = \{h : \min\{g_{ij}, g_{kj}\} > 0\}$*. Under Hypothesis 1, conditional on* $\left(\mathbf{x}_i, \mathbf{x}_k, \left\{s_j^*\right\}_{j\in\mathbf{H}_{ik}}\right)$*, the elements in the pair* (s_i^*, s_k^*) *are mutually independent.*

Proof. By equation (2.4), $s_i^* = h_i + \vartheta\sum_{j=1}^N g_{ij}s_j^*$, and $s_k^* = h_k + \vartheta\sum_{j=1}^N g_{kj}s_j^*$. Define f_{ik}, $\tilde{f}_{ik}$ and $\tilde{f}$ respectively as the joint probability distribution functions of (s_i^*, s_k^*), (ω_i, ω_k), and ω_i.

$$\begin{aligned} f_{ik}\left(s_i^*, s_k^* \middle| \mathbf{x}_i, \mathbf{x}_k, \left\{s_j^*\right\}_{j\in\mathbf{H}_{ik}}\right) &= f_{ik}\left(h_i + \vartheta\sum_{j=1}^N g_{ij}s_j^*, h_k + \vartheta\sum_{j=1}^N g_{kj}s_j^* \middle| \mathbf{x}_i, \mathbf{x}_k, \left\{s_j^*\right\}_{j\in\mathbf{H}_{ik}}\right) \\ &\propto \tilde{f}_{ik}\left(\omega_i, \omega_k \middle| \mathbf{x}_i, \mathbf{x}_k, \left\{s_j^*\right\}_{j\in\mathbf{H}_{ik}}\right) \\ &= \tilde{f}\left(\omega_i \middle| \mathbf{x}_i, \left\{s_j^*\right\}_{j\in\mathbf{H}_{ik}}\right)\tilde{f}\left(\omega_k \middle| \mathbf{x}_k, \left\{s_j^*\right\}_{j\in\mathbf{H}_{ik}}\right) \end{aligned} \tag{2.11}$$

This proves the statement.

[6]Expression $\tilde{s}(\mathbf{x}_i, \boldsymbol{\xi})$ can be rationalized within a model that includes covariates like the one from Appendix A (where R&D is a function of other inputs), while $f_\omega(\mathbf{x}_i, \boldsymbol{\zeta})$ as an expression for heteroschedasticity.

This result expresses the fact that, when conditioning on the choices of intermediate nodes along the path between i and j, the covariance in R&D investment between the two disconnected firms that is induced by the unobserved factors ω equals zero. The intuition, as mentioned, is that by "holding constant" the R&D of all on intermediate firms located on all the paths linking i and k, one removes the mechanism of indirect strategic response between these two firms. By identifying a moment of the cross-correlation in R&D in which common confounders have no effect, it is possible to effectively recover the parameters of the model, including the variance of ω. As mentioned this result holds generally, regardless of the specific functional forms imposed on the best response function or the unobserved shock. If Assumption 1 held, the covariance restriction takes in particular the following form.

$$\begin{aligned}\mathbb{C}\text{ov}\left(s_i^*, s_k^* \middle| \mathbf{x}_i, \mathbf{x}_k, \left\{ s_j^* \right\}_{j \in \mathbf{H}_{ik}} \right) &= \mathbb{C}\text{ov}\left(f_{\omega,i} \upsilon_i, f_{\omega,k} \upsilon_k \middle| \mathbf{x}_i, \mathbf{x}_k \right) \\ &= \mathbb{I}\left(i = k \right) f_{\omega,i}^2 \sigma_\upsilon^2 \begin{cases} > 0 & \text{if } d_{ik} = 0 \\ = 0 & \text{if } d_{ik} > 0 \end{cases} \end{aligned} \tag{2.12}$$

Clearly, the conditional covariance is nonzero only if it involves the same two nodes: that is, it actually is the variance of the equilibrium R&D investment of one individual firm.

2.3.3 Identification: Adding Exogenous Dependence

With exogenous dependence, the covariance matrix $\mathbb{V}\text{ar}\left(\boldsymbol{\omega}\right) = \boldsymbol{\Omega}$ is no longer diagonal, reflecting the cross correlation of the unobservables. In particular, if Assumption 1 holds, the covariance matrix of $\mathbf{s}^*$ is expressed by

$$\mathbb{V}\text{ar}\left(\mathbf{s}^* \middle| \mathbf{X} \right) = \left(\mathbf{I} - \vartheta \mathbf{G} \right)^{-1} \mathbf{F}_\omega \left(\mathbf{X} \right) \left[\sigma_\upsilon^2 \mathbf{I} + \sigma_\psi^2 \left(diag\left(\mathbf{G} \iota \right) + \mathbf{G} \right) \right] \mathbf{F}_\omega \left(\mathbf{X} \right) \left(\mathbf{I} - \vartheta \mathbf{G} \right)^{-1} \tag{2.13}$$

with edge-specific shocks, and

$$\mathbb{V}\text{ar}\left(\mathbf{s}^* \middle| \mathbf{X} \right) = \left(\mathbf{I} - \vartheta \mathbf{G} \right)^{-1} \mathbf{F}_\omega \left(\mathbf{X} \right) \left[\sigma_\upsilon^2 \mathbf{I} + \sigma_\phi^2 \left(\mathbf{I} + 2\mathbf{G} + \mathbf{G}^2 \right) \right] \mathbf{F}_\omega \left(\mathbf{X} \right) \left(\mathbf{I} - \vartheta \mathbf{G} \right)^{-1} \tag{2.14}$$

with node-specific shocks.

Identification is based on two general results, analogous to the one from Proposition 1. The one for edge-specific shocks reads as follows.

Proposition 2. *Let the minimum path length between firms i and k be 2 or higher. Define the set of all firms that are directly connected to either i or j: $\mathbf{K}_{ik} = \{h : \min\{g_{ij}, g_{kj}\} > 0\}$. Under Hypothesis 2, conditional on $(\mathbf{x}_i, \mathbf{x}_k, \{s_j^*\}_{j \in \mathbf{K}_{ik}})$, the elements in the pair (s_i^*, s_k^*) are mutually independent.*

The proof is identical to the one of Proposition 1, taking advantage of the fact that exogenous correlation caused by edge-specific shocks does not extend beyond distance 1. Under Assumption 1, in particular, the covariance between firms i and k reads as

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}\left(s_i^*, s_k^* \middle| \mathbf{x}_i, \mathbf{x}_k, \left\{ s_j^* \right\}_{j \in \mathbf{K}_{ik}} \right) \\
&\qquad = \mathbb{C}\mathrm{ov}\left[f_{\omega,i}\left(\upsilon_i + \sum_{j=1}^{N} g_{ij}\psi_{ij} \right), f_{\omega,k}\left(\upsilon_k + \sum_{j=1}^{N} g_{ij}\psi_{ij} \right) \middle| \mathbf{x}_i, \mathbf{x}_k \right] \\
&\qquad = f_{\omega,i} f_{\omega,k} \left[\mathbb{I}\left(i=k\right)\sigma_\upsilon^2 + \left(\mathbb{I}\left(i=k\right)\bar{g}_i + g_{ik} \right)\sigma_\psi^2 \right] \begin{cases} > 0 & \text{if } d_{ik} \leq 1 \\ = 0 & \text{if } d_{ik} > 1 \end{cases}
\end{aligned} \tag{2.15}
$$

not extending beyond distance 1 – although it does not really matter, for nodes i and k from Proposition 2 must be at distance 2 or higher. In this case the identification of parameter σ_ψ^2 needs the inclusion, in the estimation, of variance moments from equation (2.13).

The result for the Node-specific shocks differs slightly from the preceding two.

Proposition 3. *Let the minimum path length between firms i and k be 2 or higher. Define the set of all firms that are directly connected to either i or j: $\mathbf{Q}_{ik} = \{h : \min\{g_{ij}, g_{kj}\} > 0\}$. Under Hypothesis 2, conditional on $(\mathbf{x}_i, \mathbf{x}_k, \left\{s_j^*\right\}_{j\in\mathbf{Q}_{ik}}, \{\phi_j\}_{j\in\mathbf{Q}_{ik}})$, the elements in the pair (s_i^*, s_k^*) are mutually independent.*

The difference is that in this case, in order to achieve independence, one needs to condition also on the sequence $\{\phi_j\}_{j\in\mathbf{Q}_{ik}}$, because the equilibrium R&D of any two nodes at distance 2 is still jointly dependent from the shocks of their immediate intermediate links: even if by conditioning on the R&D of intermediate links, endogenous dependence is effectively removed.[7] The proof of Proposition 3 is also analogous to that of Proposition 1, except that the last step is only possible by conditioning on $\{\phi_j\}_{j\in\mathbf{Q}_{ik}}$.

In any case, this extra requirement is not necessary for identification. To see why, let Assumption 1 hold; the resulting covariance between firms i and k is

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}\left(s_i^*, s_k^* \middle| \mathbf{x}_i, \mathbf{x}_k, \left\{ s_j^* \right\}_{j \in \mathbf{Q}_{ik}} \right) \\
&\qquad = \mathbb{C}\mathrm{ov}\left[f_{\omega,i}\left(\upsilon_i + \phi_i + \sum_{j=1}^{N} g_{ij}\phi_j \right), f_{\omega,k}\left(\upsilon_k + \phi_k + \sum_{j=1}^{N} g_{ij}\phi_j \right) \middle| \mathbf{x}_i, \mathbf{x}_k \right] \\
&= f_{\omega,i} f_{\omega,k} \left[\mathbb{I}\left(i=k\right)\left(\sigma_\upsilon^2 + \sigma_\phi^2\right) + \mathbb{I}\left(d_{ik}=1\right) 2 g_{ik}\sigma_\phi^2 + \sum_{j:d_{ij},d_{jk}>1}^{N} g_{ij}g_{kj}\sigma_\phi^2 \right] \begin{cases} > 0 & \text{if } d_{ik} \leq 2 \\ = 0 & \text{if } d_{ik} > 2 \end{cases}
\end{aligned} \tag{2.16}
$$

which extends up to distance 2 (but not beyond). Separate identification of parameter σ_ϕ^2 is very intuitive in this case: under Hypothesis 2, node-specific shocks are the only explanation of the excess covariance between any pair of disconnected nodes in an open triad.

[7] This explains why it is not necessary to, in fact, condition on the whole set $\{\phi_j\}_{j\in\mathbf{Q}_{ik}}$, but only on those elements of it such that $j \in \mathbf{Q}_{ik} \wedge (d_{ij} = 1 \vee d_{jk} = 1)$: these are the shocks received by direct connections of either i and j. I do not specify this into the expression of Proposition 3 so not to overcomplicate it. A weaker result, limited to pairs of nodes at distance 3 or higher, does not need this additional requirement.

2.3.4 Identification under Misspecification

Suppose that the functional form assumptions on the best response function and the unobservables are not correctly specified. In particular, assume that the R&D investment choice in isolation is given by $\tilde{s}_i(\mathbf{x}_i, \omega_i)$, and that this function does not depend on a finite number of parameters. Therefore, the identification results presented thus far do not apply. To show that the parameter that captures the spillovers ϑ can be identified under weaker conditions, first define $\mu(\mathbf{x}_i) \equiv \mathbb{E}[\tilde{s}_i|\mathbf{x}_i]$ and $\eta_i \equiv \tilde{s}_i - \mu(\mathbf{x}_i)$. Notice that η_i can depend in a nonlinear way on the unobservables, but the fundamental structure of the problem remains the same. Specifically, in the case of simple endogenous dependence, η_i depends only on υ_i; if the exogenous dependence is edge-specific, it also depends on $\{g_{ij}\psi_j\}_{j=1}^N$; it depends on $\{g_{ij}\phi_{ij}\}_{j=1}^N$ in addition to υ_i, finally, if the exogenous dependence is node-specific.

Write the optimal response function as:

$$s_i^* = \vartheta \sum_{h=1}^{N} g_{ih} s_h^* + \mu(\mathbf{x}_i) + \eta_i \tag{2.17}$$

then, by taking the appropriate conditional expectation, it is possible to obtain the following moment condition of the first order:

$$\mathbb{E}\left[s_i^* - \vartheta \sum_{h=1}^{N} g_{ih} s_h^* - \mu(\mathbf{x}_i) \middle| \mathbf{x}_i, \{s_k^*\}_{k:g_{ik}>0} \right] = 0 \tag{2.18}$$

hence for each given value of ϑ, and under standard regularity conditions, the nonparametric function $\mu(\cdot)$ is identified from equation (2.18). However, in order to identify ϑ it is necessary to include some additional moments in the estimation: variances and, crucially, covariances. The identification arguments are the same as in the exactly specified case, whatever the hypothesis one makes about the stochastic process generating ω_i. The general form of the covariance restriction involving two unrelated nodes i and k in an open triad is as follows.

$$\mathbb{C}\text{ov}\left(s_i^* - \vartheta \sum_{h=1}^{N} g_{ij} s_j^* - \mu(\mathbf{x}_i), s_k^* - \vartheta \sum_{j=1}^{N} g_{kj} s_j^* - \mu(\mathbf{x}_k) \middle| \mathbf{x}_i, \mathbf{x}_k, \left\{s_j^*\right\}_{j\in\mathbf{F}_{ik}} \right) = 0 \tag{2.19}$$

In this specific context it is possible to obtain separate identification of ϑ and $\mu(\cdot)$ by combining the equation in levels with the covariances. This is expressed by the following result, whose proof is given in Appendix B, that is based upon the hypothesis pure endogenous dependence (the cases with exogenous common shocks are analogous).

Proposition 4. *Let s_i^* be given by equation 2.17 and Hypothesis 3 hold. Then, the moments given in equations 2.18 and 2.19 identify $(\vartheta, \mu(\cdot))$.*

This result ensures that it is possible to obtain an accurate estimate of ϑ even if the model is incorrectly specified, as long as the true best response function is linear in ω_i and the empirical specification of $\mu(\cdot)$ is a reasonable approximation of $\tilde{s}(\mathbf{x}_i, \boldsymbol{\xi})$.

2.3.5 Estimation

The model is estimated by establishing a set of moment conditions defined at the triad level. This is the most straightforward method to match the main intuition behind identification: that is, to exploit the excess covariance restrictions between two unconnected nodes of an open triad. To see why, consider the i-th open triad, formed by the nodes $1(i)$, $2(i)$, and $3(i)$, such that $2(i)$ is the central node, and the other two are the external ones. Define

$$\epsilon_{mn} \equiv f_{\omega,m} f_{\omega,n} \left\{ \omega_m \omega_n - \mathbb{E}\left[\omega_m \omega_n\right] \right\} \tag{2.20}$$

then, the building block of the GMM problem is given by the following set of moment conditions $m_i(\boldsymbol{\theta})$:

$$m_i(\boldsymbol{\theta}) \equiv \begin{bmatrix} \mathbf{x}_{1(i)}\omega_{1(i)} \\ \mathbf{x}_{2(i)}\omega_{2(i)} \\ \mathbf{x}_{3(i)}\omega_{3(i)} \\ \mathbf{x}_{1(i)}\epsilon_{11(i)} \\ \mathbf{x}_{2(i)}\epsilon_{22(i)} \\ \mathbf{x}_{3(i)}\epsilon_{33(i)} \\ \left(\mathbf{x}_{1(i)} + \mathbf{x}_{2(i)}\right)\epsilon_{12(i)} \\ \left(\mathbf{x}_{2(i)} + \mathbf{x}_{3(i)}\right)\epsilon_{23(i)} \\ \left(\mathbf{x}_{1(i)} + \mathbf{x}_{3(i)}\right)\epsilon_{13(i)} \end{bmatrix} \tag{2.21}$$

where $\boldsymbol{\theta} = (\vartheta, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\sigma}^2)'$ is the vector that contains all the parameters of the model.[8] The dimension of $m_i(\boldsymbol{\theta})$ is $9K$, where $K = \dim(\mathbf{x}_i)$. However, the actual number of moments is smaller, since specific conditions in (2.21) refer to pairs of nodes that show up in multiple open triads. The fundamental moments in $m_i(\boldsymbol{\theta})$ are its $3K$ "covariance" elements

$$\left(\mathbf{x}_{m(i)} + \mathbf{x}_{n(i)}\right) f_{\omega,m(i)} f_{\omega,n(i)} \left\{ \omega_{m(i)}\omega_{n(i)} - \mathbb{E}\left[\omega_{m(i)}\omega_{n(i)}\right] \right\} \tag{2.22}$$

whose linear combinations, under Assumption 1, coincide with the covariance expressions – one per open triad – resulting from the equilibrium conditions.[9]

A problem with this approach is that the number of open triads and the number of nodes (firms) grow at different rates. For example, in an empty network (made of many disconnected nodes) the number of open triads is stuck as zero when the number of nodes grows. Conversely, in a star network (a network where a single node is connected to all others, which are directly unconnected to each other) the number of open triads grows at rate $\binom{N-1}{2}$, whenever the nodes grow at rate N. Scenarios like these might invalidate the asymptotic properties of the estimation. In order to overcome this problem I impose, as usual in the econometrics of networks, an appropriate "sparsity" assumption about the maximum density of the network.

[8] The horizontal vector $\boldsymbol{\sigma}^2$ contains the parameters relative to the variance of ω which vary with specific assumptions: $\boldsymbol{\sigma}^2 = (\sigma_\upsilon^2)$ under Hypothesis 1, $\boldsymbol{\sigma}^2 = (\sigma_\upsilon^2, \sigma_\psi^2)$ under Hypothesis 2, and $\boldsymbol{\sigma}^2 = (\sigma_\upsilon^2, \sigma_\phi^2)$ under Hypothesis 3.

[9] Compare with equations (2.12), (2.15) and (2.16) respectively under Hypotheses 1, 2 and 3.

Assumption 2. *The number of connections of each node i is bounded by a constant G^*.*

$$\lim_{N\to\infty}\sum_{j=1}^{N}\mathbb{I}\left[g_{ij}>0\right]<G^*\ \forall i=1,\dots,N \tag{2.23}$$

This sparsity assumption has an important implication, as it limits the growth rate of the number of open triads, T_N, to equal at most the growth rate of the number of nodes, N. Intuitively, Assumption 2 limits the centrality of any node, avoiding scenarios like the star network. This implies that the number of open triads in which a particular node is present is bounded, and therefore T_N is bounded by a multiple of N. The following Proposition formalizes this fact.

Proposition 5. *Let assumption 2 hold. Then, the growth rate of the number of open triads is at most the growth rate of the number of nodes, that is, $T_N = O(N)$.*

Proof. Under Assumption 2, a node i can be the central node of an open triad for at most $\binom{G^*}{2}$ triads. Since the total number of open triads is the sum over all nodes of the number of open triads in which one node is central, it follows that T_N is bounded by $\binom{G^*}{2}N$.

Let $\overline{K}\equiv\dim(\boldsymbol{\theta})$, $\overline{m}_n(\boldsymbol{\theta})\equiv\frac{1}{T_N}\sum_{i=1}^{T_N}m_i(\boldsymbol{\theta})$, and W_n be a $9K\times 9K$ weighting matrix. Then, the GMM estimator of this model is as follows.[10]

$$\hat{\boldsymbol{\theta}}\equiv\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\overline{m}_n(\boldsymbol{\theta})'\,W_n\overline{m}_n(\boldsymbol{\theta}) \tag{2.24}$$

Under Assumption 2 and typical regularity conditions, $\hat{\theta}$ is consistent and asymptotically normal. A discussion about the asymptotic properties of the estimator and the computation of its standard errors is provided in Appendix B.

Assumption 1 guarantees the identification of $\boldsymbol{\theta}$; however the two functions $\tilde{s}(\mathbf{x}_i,\boldsymbol{\xi})$ and $f_\omega(\mathbf{x}_i,\boldsymbol{\zeta})$ are left still undetermined. In order to complete the specification of the model, it is necessary to make some functional form assumptions. Here I posit the following.

Assumption 3. *The best response function is linear in the covariates $\mathbf{x}_i$, and the term that controls the heteroschedasticity, f_ω, is the exponential of a linear function in the covariates.*[11]

$$\tilde{s}(\mathbf{x}_i,\boldsymbol{\xi})=\mathbf{x}_i'\boldsymbol{\xi} \tag{2.25}$$

$$f_\omega(\mathbf{x}_i,\boldsymbol{\zeta})=\exp(\mathbf{x}_i'\boldsymbol{\zeta}) \tag{2.26}$$

This choice is flexible and computationally simple. A linear specification of $\tilde{s}(\mathbf{x}_i,\boldsymbol{\xi})$ can easily be enriched so to best approximate the true function; at the same time exponential expressions of the term for heteroschedasticity are among the most common and convenient in the applied practice.[12]

---

[10]Equivalently, $\hat{\boldsymbol{\theta}}$ solves $\frac{\partial\overline{m}_n(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}W_n\overline{m}_n\left(\hat{\boldsymbol{\theta}}\right)=0$.

[11]The set of observable covariates is specified as the same in the two cases, but it is not necessarily so, as one can impose restrictions. In the empirical analysis I restrict $f_\omega(\mathbf{x}_i,\boldsymbol{\zeta})$ to one covariate, the logarithm of the number of employees.

[12]Clearly, no parameter for a constant term can be identified in $\boldsymbol{\zeta}$ while simultaneously identifying σ_υ^2.

2.4 Triad-level Data

The estimation of the model is conducted on the same dataset employed in chapter 1: the COMPUSTAT firm-level data assembled by Bloom et al. (2013) matched to disambiguated patent data. Cross-firm linkages depend on the intensity of coauthorship relationship across their inventors. Using the same data allows a direct comparison of the empirical estimates of the spillovers parameter(s) across different procedures: the reduced-form approach of the previous chapter and the structural approach proposed here. Refer to Section 1.3 as well as to Appendix C for the full data description and network construction. In this section I comment on some characteristics of the dynamic network that are particularly relevant here.

2.4.1 Network Census

As motivated in the previous section, the validity of the estimation procedure proposed in this chapter rests on some conditions. First, that the network must be sufficiently *sparse* as per Assumption 2. As it is displayed on Figure 1.2, the sparsity requirement is well respected in the data, as the number of links per firm in any year rarely exceeds 20. Another, implicit condition is that the number of transitive or "closed" triads is small relative to the number of open triads. If open triads were too few relative to closed triads, identification would still be algebraically feasible, but at the same time it would be statistically weak.

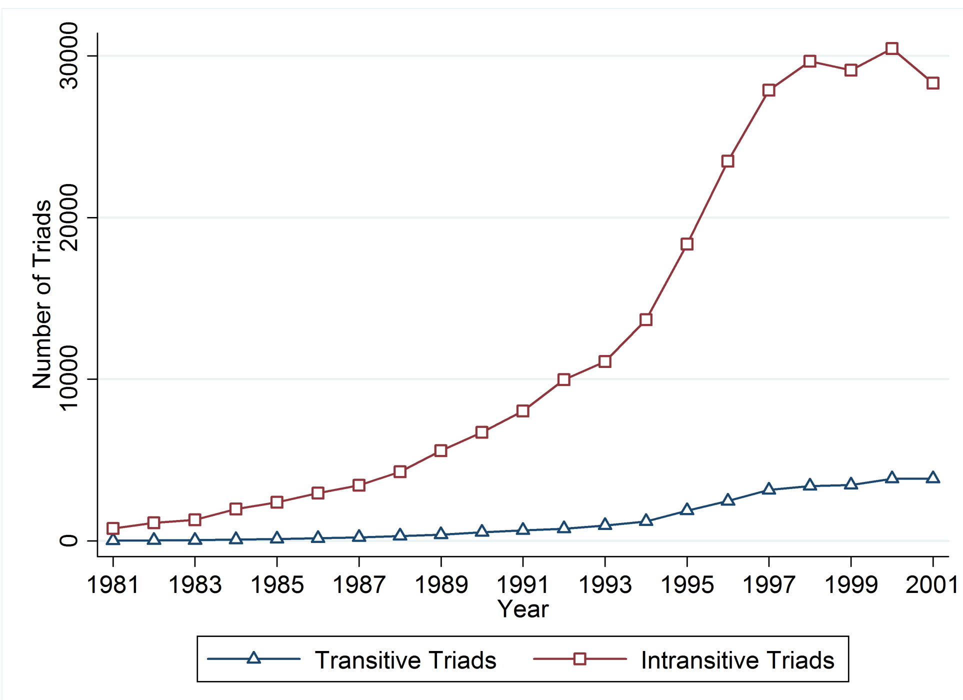


Figure 2.1: Network Census over time

The *Network Census*, that is the total count of both open and closed triads, is represented in its yearly evolution on figure 2.1. In the time interval 1981-2001, the same one analyzed in chapter 1, there are in total 160,365 open triads and 15,623 closed triads (these are about 9.1% of the total). Their number, particularly the one of open triads, grows over the years, to eventually stabilize at the end of the observation period. This makes the data at hand perfectly suitable for the estimation procedure that is being proposed. The final estimation sample is actually restricted, however, to 131,200 open triads, because of missing values for some variables of interest and panel attrition. The organization of the data at the triad level aggravates these problems, since it is sufficient that of three firms in a triad only one presents one problem of this kind for the entire triad to be eliminated from the estimation sample.

2.5 Empirical Results

In this section I present the empirical estimates of the parameters of model (2.21). Although the model is designed for a more general framework, throughout the section the assumed functional form of the objective function is analogous to the one in equation (2.2), featuring strategic complementarities. The estimated model is extended to include covariates as per Assumption 1; these enter the model logarithmically conforming to the Cobb-Douglas shape of the objective function. This choice is taken for consistency with the analysis conducted in chapter 1, and to facilitate comparisons of the estimated parameters. Throughout the section, the discussion is centered on the comparison of the estimates for the spillover parameter ϑ across estimates made under alternative hypotheses about the nature of common shocks.

The first part of this section is devoted to the results obtained under Hypothesis 1, where the spillovers parameter ϑ is estimated along the variance component σ_υ^2. A notable feature of the application of the model to the data is that the simultaneous estimation of σ_υ^2 together with a variance parameter associated to common shocks (σ_ψ^2 or σ_ϕ^2) usually returns a corner solution with $\hat{\sigma}_\upsilon^2 = 0$. Hence, I subsequently and separately present estimates of the model under Hypotheses 2 and 3; which include estimates of σ_ψ^2 or σ_ϕ^2 but not of σ_υ^2. Finally, in the last part of this section I provide a calibration analysis that relates the estimates of ϑ to the one of the more primitive parameter δ obtained in the previous chapter.

2.5.1 Results: No Common Shocks (Hypothesis 1)

In Figure 2.1 I present the results for the estimation of the model with no assumed exogenous common shocks. In order to account for constant unobservables, all variables employed in the estimation are ex-ante demeaned across both panel dimensions. Such a procedure is maintained throughout the analysis, and yields results that are similar to those one would obtain by estimating the GMM model with the direct inclusion of fixed effects. The main difference is that standard errors are slightly understated when employing pre-demeaned variables; however this approach has the non-negligible advantage of being noticeably more computationally convenient than the more rigorous approach.

Table 2.1: GMM Estimates under Hypothesis 1 and Homoschedasticity

	(1)	(2)	(3)	(4)
ϑ: Spillovers	0.0784	0.2062	0.2267	0.2282
	(0.0093)	(0.0668)	(0.0611)	(0.0573)
ξ: log(Capital)		0.2446	0.0931	0.1120
		(0.0280)	(0.0351)	(0.0313)
ξ: log(Labor)		0.2883	0.1045	0.1405
		(0.0884)	(0.0626)	(0.0682)
ξ: log(Sales) $(t-1)$			0.3207	0.2549
			(0.0360)	(0.0247)
ξ: log(Jaffe Tech. Proximity)				0.4509
				(0.1103)
ξ: log(BSV Business Stealing)				0.0166
				(0.0190)
σ_v^2	0.1144	0.0758	0.0748	0.0734
	(0.0174)	(0.0217)	(0.0223)	(0.0233)
Demeaned Variables:	ALL	ALL	ALL	ALL
Number of Open Triads	131200	131200	131024	131024

In column (1) I report the result for the simplest estimated model, which includes no covariates in the expression of the best response function as per (2.4). The spillover parameter ϑ is estimated very robustly at around 0.078, while σ_v^2 has a point estimate of about 0.114. In column (2) I include the two standard conventional inputs of the production function: Capital and Labor. The estimate of ϑ almost triples in size to about 0.206, while the coefficients for Capital and Labor are estimated at, respectively, 0.244 and 0.288. As expected, the estimated variance drops greatly, by about one third, down to 0.076. In column (3) I include a measure for lagged deflated sales, that can be thought as a proxy of future expected firm performance (which, if higher, would in turn encourage higher R&D spending). Coefficient ϑ is estimated even higher at around 0.225, while the coefficients for Capital and Labor are depressed as a result.[13] Finally, in column (4) I include Jaffe's (1986) measure of "Technological Spillovers" and the "Business Stealing" measure from Bloom et al. (2013). As expected, the former entails a positive elasticity on private R&D, while the coefficient associated to the latter is not statistically significant. The coefficient for ϑ is estimated similarly as in column (3). It is worth noticing that the inclusion of these additional covariates has a negligible impact on $\hat{\sigma}_v^2$. Most estimates, and all those of ϑ and σ_v^2, are statistically significant at the 1% level.

[13] Recall that the model estimates the equilibrium conditions of the game and not the production function; this result indicates that lagged sales is a strong sufficient statistic of other factors complementary to R&D.

That the ϑ coefficient is estimated with a larger value as covariates are included into the model, even if these covariates are themselves positively correlated with the R&D of connected firms, depends on the identification mechanism embedded into the model. Recall that the model includes moments of both the first (conditional mean) and second (variances and covariances) order, and that the parameter estimates are those values yielding the best aggregate fit of all these moments. If the expressions for the moments of the second order are incorrectly specified, and if, say, the covariances of R&D variables decrease with firm size (which is a realistic and reasonable hypothesis) then the omission of appropriate covariates would cause a downward bias on the estimated ϑ coefficient, as long as the spillovers pool determined by network connections is positively correlated with firm size.

Table 2.2: GMM Estimates under Hypothesis 1 and Heteroschedasticity

	(1)	(2)	(3)	(4)
ϑ: Spillovers	0.1047	0.2085	0.2290	0.2304
	(0.0091)	(0.0669)	(0.0614)	(0.0574)
ξ: log(Capital)		0.2412	0.0884	0.1073
		(0.0270)	(0.0341)	(0.0311)
ξ: log(Labor)		0.2981	0.1161	0.1520
		(0.0879)	(0.0609)	(0.0668)
ξ: log(Sales) $(t-1)$			0.3204	0.2548
			(0.0374)	(0.0235)
ξ: log(Jaffe Tech. Proximity)				0.4582
				(0.0964)
ξ: log(BSV Business Stealing)				0.0135
				(0.0198)
σ_v^2	0.0787	0.0772	0.0763	0.0750
	(0.0302)	(0.0221)	(0.0226)	(0.0236)
ζ: log(Labor)	1.7593	0.1095	0.1396	0.1349
	(0.2159)	(0.0354)	(0.0291)	(0.0282)
Demeaned Variables:	ALL	ALL	ALL	ALL
Number of Open Triads	131200	131200	131024	131024

In order provide a richer specification of the stochastic process driving the unobserved factors of the model, I allow their variance-covariance to be heteroschedastic but, for computational simplicity, to depend on one single variable. In particular, I estimate the ζ coefficient for Labor, (logarithm of the) number of employees of one firm, often used as a proxy of firm size. Table 2.2 presents, column-by-column, the results from the replication of the estimates from Table 2.1, extending them with the inclusion of said expression for heteroschedasticity.

The results evidence two facts. First, as one would expect from the identification analysis of the model, the estimates of ϑ and of the $\{\xi_k\}_k$ coefficients are virtually unchanged across Tables 2.2 and 2.1. Less intuitively, though, also the estimates of σ_v^2 are statistically identical in columns (2), (3) and (4). This contrasts with the positive, statistically significant and robust estimates of ζ_{Labor} from columns (2) to (4), all in the 0.11-0.14 range.[14] Since all firms in the sample have a size larger than e, the inclusion of an expression for heteroschedasticity results in a larger estimate of the overall model variance. The inclusion of further covariates in this expression yields similar qualitative results, that are not shown here for brevity.

2.5.2 Results: Edge-specific shocks (Hypothesis 2)

In Table 2.3 I show the results for the model with an Edge-Specific variance component, as per Hypothesis 2. As mentioned, the simultaneous estimation of σ_v^2 and σ_ψ^2 returns a zero corner solution for the former parameter, hence only the latter is estimated. From a scrutiny of the results two main considerations follow. First, the estimates for ϑ are slightly smaller, across all columns, than those from Table 2.1: e.g. in column (4) it is about equal to 0.201 (contrast with 0.230). The estimates of the $\{\xi_k\}_k$ parameters are, instead, virtually identical.

Table 2.3: GMM Estimates under Hypothesis 2 and Homoschedasticity

	(1)	(2)	(3)	(4)
ϑ: Spillovers	0.0786	0.1837	0.2033	0.2012
	(0.0073)	(0.0638)	(0.0582)	(0.0528)
ξ: log(Capital)		0.2484	0.0949	0.1041
		(0.0282)	(0.0377)	(0.0319)
ξ: log(Labor)		0.3222	0.1423	0.1848
		(0.0880)	(0.0641)	(0.0691)
ξ: log(Sales) $(t-1)$			0.3197	0.2614
			(0.0360)	(0.0288)
ξ: log(Jaffe Tech. Proximity)				0.3998
				(0.1201)
ξ: log(BSV Business Stealing)				0.0413
				(0.0186)
σ_ψ^2	0.0668	0.0401	0.0376	0.0342
	(0.0059)	(0.0071)	(0.0071)	(0.0060)
Demeaned Variables:	ALL	ALL	ALL	ALL
Number of Open Triads	131200	131200	131024	131024

[14]The exception are the results shown in column (1), an apparent case of misspecification.

The second consideration is that across all specification, the estimates for σ_ψ^2 are about one half of those of σ_υ^2 from both Tables 2.1 (e.g. 0.034 vs. 0.073 respectively). However, the estimates of σ_ψ^2 only capture a component of the variance: recall that the according to Hypothesis 2, $\mathbb{V}\text{ar}\left(\omega_i\right) = \sigma_\psi^2 \sum_{j \neq i} g_{ji}$ (firms with lots of connections are subject to multiple common shocks) if $\sigma_\upsilon^2 = 0$. Since the average row sum of connections is approximately 0.44 (see Section 1.3 and Figure 1.4) this means that the average firm in the network is subject to shocks with a $0.44 \times \hat{\sigma}_\psi^2$ variance, about 25% of the magnitude from the baseline case. Hence, the model with of edge-specific common shocks is seemingly capable of matching the empirical second moments with a much smaller estimated stochastic variability.

Table 2.4: GMM Estimates under Hypothesis 2 and Heteroschedasticity

	(1)	(2)	(3)	(4)
ϑ: Spillovers	0.1180	0.1868	0.2050	0.2021
	(0.0079)	(0.0643)	(0.0584)	(0.0531)
ξ: log(Capital)		0.2390	0.0848	0.0924
		(0.0267)	(0.0366)	(0.0314)
ξ: log(Labor)		0.3513	0.1743	0.2185
		(0.0885)	(0.0608)	(0.0645)
ξ: log(Sales) $(t-1)$			0.3157	0.2595
			(0.0383)	(0.0258)
ξ: log(Jaffe Tech. Proximity)				0.3976
				(0.1044)
ξ: log(BSV Business Stealing)				0.0374
				(0.0182)
σ_ψ^2	0.0623	0.0441	0.0408	0.0374
	(0.0065)	(0.0078)	(0.0077)	(0.0066)
ζ: log(Labor)	0.7142	0.2242	0.2286	0.2217
	(0.2483)	(0.0363)	(0.0274)	(0.0329)
Demeaned Variables:	ALL	ALL	ALL	ALL
Number of Open Triads	131200	131200	131024	131024

In Table 2.4 I present the same results, enriched by an expression for heteroschedasticity dependent on firm size (Labor). The comparison with the results from Table 2.2 is similar to the one made above: first, once again the model with edge-specific shocks results in smaller estimates of ϑ (down by about 10-15%). Second, the estimates of the variance component, which is now parametrized by σ_ψ^2, is much smaller. The estimates of ζ_{Labor} are instead about twice as large than those from Table 2.2, partly compensating for the fact that here they multiplicatively affect a much smaller variance factor than in the case of no common shocks.

2.5.3 Results: Node-specific shocks (Hypothesis 3)

In the last set of results I am presenting here, common shocks are assumed to be node-specific as per Hypothesis 3. Again, the GMM minimization problem returns a $\hat{\sigma}_v^2 = 0$ solution; thus only estimates of the variance parameter σ_ϕ^2 are reported. Table 2.5 displays the estimates of the model in the case with no specification for heteroschedasticity. The estimated results look even closer to the baseline case than those given in Table 2.3: the estimates of $\hat{\vartheta}$ are statistically identical, column-by-column, across Tables 2.1 and 2.5. Parameter σ_ϕ^2 is, in columns (2), (3) and (4) of Table 2.5, estimated in the range 0.065-0.67, which looks similar to the estimates of σ_v^2 from the baseline case. However, under Hypothesis 3 and with $\sigma_v^2 = 0$, the actual expression of the variance of ω is given by $\mathbb{V}\mathrm{ar}\,(\omega_i) = \sigma_\phi^2(1 + \sum_{j \neq i} g_{ij})$; hence for the average firm in the network the actual shock variance amounts to $1.50 \times \hat{\sigma}_\phi^2$, slightly less than 50% larger than in the baseline case. In terms of efficiently fitting the second order moments, the model with node-specific shocks seems to perform worse than the model with no common shocks and much worse than the model with edge-specific shocks.

Table 2.5: GMM Estimates under Hypothesis 3 and Homoschedasticity

	(1)	(2)	(3)	(4)
ϑ: Spillovers	0.0634	0.2051	0.2256	0.2261
	(0.0088)	(0.0655)	(0.0595)	(0.0553)
ξ: log(Capital)		0.2464	0.0938	0.1124
		(0.0280)	(0.0356)	(0.0315)
ξ: log(Labor)		0.2909	0.1077	0.1429
		(0.0871)	(0.0624)	(0.0669)
ξ: log(Sales) $(t-1)$			0.3215	0.2560
			(0.0352)	(0.0245)
ξ: log(Jaffe Tech. Proximity)				0.4500
				(0.1101)
ξ: log(BSV Business Stealing)				0.0190
				(0.0183)
σ_ϕ^2	0.1118	0.0677	0.0672	0.0651
	(0.0158)	(0.0181)	(0.0186)	(0.0186)
Demeaned Variables:	ALL	ALL	ALL	ALL
Number of Open Triads	131200	131200	131024	131024

Table 2.6 displays the results for the model with node-specific shocks and a variance heteroschedastically dependent on firm size. As in the previous case, the estimates of ϑ and of the $\{\xi_k\}_k$ parameters are very similar to the case with no common shocks, but the estimated overall variance of the model is substantially larger.

Table 2.6: GMM Estimates under Hypothesis 3 and Heteroschedasticity

	(1)	(2)	(3)	(4)
ϑ: Spillovers	0.0780	0.2071	0.2275	0.2278
	(0.0394)	(0.0655)	(0.0595)	(0.0554)
ξ: log(Capital)		0.2423	0.0882	0.1067
		(0.0268)	(0.0346)	(0.0313)
ξ: log(Labor)		0.3049	0.1218	0.1579
		(0.0863)	(0.0608)	(0.0648)
ξ: log(Sales) $(t-1)$			0.3214	0.2561
			(0.0364)	(0.0231)
ξ: log(Jaffe Tech. Proximity)				0.4602
				(0.0938)
ξ: log(BSV Business Stealing)				0.0152
				(0.0185)
σ^2_ϕ	0.1070	0.0693	0.0686	0.0665
	(0.0275)	(0.0183)	(0.0187)	(0.0188)
ζ: log(Labor)	0.2646	0.1439	0.1664	0.1602
	(0.4489)	(0.0330)	(0.0260)	(0.0283)
Demeaned Variables:	ALL	ALL	ALL	ALL
Number of Open Triads	131200	131200	131024	131024

2.5.4 Comparison of Spillovers Estimates

The estimate of the spillover coefficient ϑ is substantially larger than the estimate of parameter δ from the previous chapter. For example, my favorite estimate $\hat{\vartheta}$, from column (4) of Table 2.3 is about 0.202; while the 2SLS estimate $\hat{\delta}$ of the model with two instruments from column (4) of Table 1.4 is about 0.020: one order of magnitude smaller. It is worth reminding that the two coefficients do not represent the same structural parameter, but have a reciprocal relation that depends on the specific assumptions made on the model's functional form. Throughout both chapters, the functional form that is brought to the data entails strategic complementarities, which in its simplest version is akin to the expression in equation (2.2). In that model, $\vartheta = \delta\,(1-\gamma)^{-1}$. In the richer model that includes covariates and that is discussed at length in Appendix A, $\vartheta = \delta\,(1-\gamma-\sum_k \beta_k)^{-1}$, where β_k is the parameter associated to the k-th input in the production function.[15]

[15]The model in Appendix A assumes different degrees of incomplete information, but the equilibrium solution shares these characteristics with a model with complete information and a Cobb-Douglas production function that is multiplicative in peers' R&D.

In the empirical model presented in this chapter, ϑ and δ are not separately identified; intuitively, this is because parameter γ, the parameter that measures the elasticity of private investment in R&D on total output, is not estimated at all. Parameter ϑ, however, has an independent interpretation as the overall elasticity of connections' R&D on private R&D investment. In addition, a specific value for δ can be obtained from the present estimates through a calibration exercise. This requires to choose value for γ and the set of conventional inputs $\{\beta_k\}_k$; for the purpose of this analysis it is appropriate to pick typical estimates of the production function parameters from the estimates of the first chapter. For example, by setting $\gamma^* \simeq 0.04$, $\beta^*_{Capital} \simeq 0.16$, $\beta^*_{Labor} \simeq 0.65$ and with $\hat{\vartheta} \simeq 0.20$, one obtains:

$$\delta = \hat{\vartheta}\left(1 - \gamma^* - \sum_k \beta^*_k\right) \simeq 0.20 \times 0.15 = 0.03$$

a figure much closer to the 2SLS estimates of δ obtained in the previous chapter.

2.6 Conclusion

In this chapter I provide a different method to address the same research question from the previous one: how to estimate R&D spillovers occurring between firms whose inventors and scientists are professionally connected. In particular, instead of assessing the direct effect of connected firms' R&D on specific firm-level outcomes of interest in a reduced form framework, I structurally estimate their impact on the firm-level patterns of R&D investment, itself an intermediate input. Conceptually, this amounts to directly quantify the endogenous response of firm behavior to peer choices in presence of spillovers, one of the channels through which peer effects operate according to Manski's (1993) treatment of the subject. This is something seldom attempted in empirical research, mostly due to absence of information about agent effort across typical settings. In the context of R&D spillovers instead, effort can be directly observed in the form of R&D investment.

The estimation of the model is based on the equilibrium conditions of the R&D investment game, and identification in particular rests upon a set of conditional covariance relative to R&D choices of two firms related in any open triad of the network. The associated intuition is that any residual covariance between two firms that are indirectly connected in the network, if one appropriately conditions for the characteristics and choices of all other intermediate nodes between the two (which induce endogenous dependence), can be only driven by potentially common exogenous factors, if these exist. This allows to separately identify, in open triads, spillover effects and the potentially exogenous common shocks. The empirical results corroborate the main findings from the previous chapter. In particular, the overall normalized elasticity of connected firms' R&D on private R&D investment is estimated to be about 0.20 when accounting for common shocks. Under the chosen functional form assumption and in light of the quantitative estimates of other model parameters obtained in the previous chapter, this is consistent with R&D spillovers on productivity close to a 3% normalized elasticity, similar to the 2SLS estimates presented in the previous chapter.

This work shall be expanded in several directions. First, different specifications of the variance and covariance functional forms should be estimated, tested and compared with the baseline results, so to further validate (or possibly challenge) the main results. Second, it would be fruitful to estimate the model under radically different functional form assumptions, such as the one featuring strategic substitutes as per equation (2.3). This may greatly help to answer one long-standing empirical question in innovation economics: whether R&D is more a strategic complements or a strategic substitute.[16] Finally, it would be useful to develop and implement a methodology to test different assumptions on the stochastic dependence of R&D choices in the network (endogenous, edge-specific exogenous, or node-specific exogenous) which may be useful also in other contexts in order to more convincingly isolate genuine spillover effects from common shocks correlated with individual characteristics.

[16] A test on the estimated parameters of the best reply function is effectively a test on the curvature of the knowledge production function – what Jaffe (1986) complained OLS and other linear models are incapable of appropriately doing.

Chapter 3

The Local Spillovers of Superstar Inventors

3.1 Introduction

In this chapter I explore potential micro-level mechanisms driving the diffusion and exchange of knowledge in space. In particular, I focus on the interaction between spatial proximity and the existence of individual and professional connections between inventors. To this end, I analyze episodes in which "superstar" scientists – those in the top 5% of the patent distribution – relocate across cities. In particular, I examine the impact over time of these events on the "residual" patent output (patents not coauthored with the moving superstar) of four groups of inventors. They are the individuals who belong to the network of patent collaboration of the superstar and those who do not, for both localities of departure and destination. In addition, I analyze how these effects vary with the relative position of the moving inventor in the distribution of patents per individual.

The main findings of this event analysis can be summarized as follows. When a superstar moves there is a favorable and long-lasting effect on the number of patents independently realized by inventors who, at any point in time, have collaborated with the migrating superstar and are also based in his locality of destination. Quantitatively, this effect amounts on average to about 0.1 more patents per inventor each year, and it increases with the relative position of the migrating star in the distribution of patent production. However, when measuring the effect of patent quality by weighting for patent citations, this effect seems to be less persistent over time and to fade out after few years. Conversely, one superstar's collaborators who reside in his abandoned metropolitan area suffer from a negative effect on their count of residual patent application, albeit with a lag of about two years from the migration event – likely reflecting the natural lag of the R&D process.

However, there seems to be no strong evidence in favor of similar effects on the patent output of all other inventors in both cities involved. The only exception is when the superstar belongs to the top 0.1% of the patent distribution: in this case one can observe simultaneously an increase in the total patent production and in the number of active inventors in the city of destination, as well as a decrease in the average residual patent count per inventor. Taken together, these results reflect the existence of localized knowledge spillovers, which are confined however to the network of one inventor's coauthors. By the terminology introduced in this chapter, local knowledge spillovers happen on the *intensive* margin but not on the *extensive* one. As a consequence, the evidence in favor of "big push" type of place-based innovation policies – realized on the expectation that attracting few highly creative individuals may result into a wider equilibrium shift in local innovativeness – is mixed.

This chapter builds on the tradition of economic studies searching for geographically localized externalities associated with R&D activity. Jaffe, Trajtenberg and Henderson (1993, JTH) in their seminal work show how patent citations – seen as the "paper trail" of knowledge spillovers[1] – tend to come from the same urban area as the cited patent. Amongst subsequent research, notable is the critique of JTH by Thompson and Fox-Kean (2005, TFK), who show that improving on the method to control for the geographic distribution of economic activity would yield less robust results. In support of this claim, Agrawal et al. (2010) show that in "company towns" patent citations are concentrated within the same firm. Albeit different in spirit, other recent attempts to estimate localized human capital or R&D externalities are provided by Moretti (2004), Lychagin et al. (2010), Bloom et al. (2013).

The central identification problem faced by all these researchers is how to separate genuine geographic externalities from other common factors that are shared within the localities of interest. This issue is not specific to the analysis of knowledge spillovers, but is common to all urban and trade economists searching for general agglomeration economies (of which knowledge spillovers is usually thought to be one determinant, along with labor market pooling and scale economies). A solution, employed by Greenstone, Hornbeck and Moretti (2010) in their analysis of the effect of large plants on local productivity, is to exploit tail events affecting both "winning" and "losing" localities, whose long-run outcomes are compared. This chapter is based on a similar approach: the analysis of patenting outcomes by selected groups of inventors across places that receive and places that lose superstar inventors.[2]

This chapter is also related to all those studies examining the economics of top-end or "superstar" professionals, with a focus on very successful inventors and academics. Perhaps most famously among these studies, the work by Azoulay et al. (2010) provides evidence about the role of superstars in stimulating the intellectual production of their collaborators. The identification strategy employed in their paper, however, is not capable to provide an unambiguous test on the effect of the interaction between collaboration with a superstar and

[1] This definition is based upon one famous critical statement by Krugman (1991), according to whom "knowledge flows [...] are invisible; they leave no paper trail by which they may be measured and tracked."

[2] The two approaches are not identical, however. Specifically, Greenstone et al. compare places that benefited from the arrival of "Million Dollar Plants" with those that were almost selected for the opening of such facilities, but were ultimately discarded in favor of the "winning" localities. This can be thought as a source of semi-exogenous variation. In the present chapter this element, at present, is absent. Nevertheless, the event study methodology is still useful at comparing outcomes when it identifies sharp changes in the variables of interest associated with the event in question.

common spatial location.[3] There is also a growing interest in the migration of superstar inventors, considered drivers of local innovation. In particular, recent work has assessed the role of marginal tax rates in determining the location choices of superstar inventors, both across U.S. cities (Moretti and Wilson, 2014) and countries (Akcigit et al., 2015).

The remainder of this chapter is organized as follows. Section 3.2 provides a general conceptual framework, relating it to the empirical analysis. Section 3.3 introduces the data and provides a description of the superstars' "relocation events". Section 3.4 presents and comments the main empirical results from the event studies. Finally, Section 3.5 is dedicated to conclusive remarks. Additional estimates are separately reported in Appendix F.

3.2 Conceptual Framework

This section is divided in two parts. In the first one I provide a critical overview of theories of knowledge spillovers, highlighting how the idea of "geographical spillovers" conflates distinct aspects that ought to be analyzed separately. In the second part I outline a simple analytical model that distinguishes said aspects, and derive some welfare implications.

3.2.1 Knowledge Spillovers and Geography

In his classical analysis, Griliches (1979) distinguishes three sources of knowledge spillovers. They are: the *horizontal spillovers*, induced by learning about new product or processes of other firms competing in downstream product markets; the *vertical spillovers*, which are the result of economies of scale and scope derived from synergic relationships through supply chains[4]; and the *technological spillovers*, caused by the consequences of broad technological discoveries beyond narrow production sectors (for example, findings in the semiconductors industry may benefit ICT, those in biotech the pharmaceutical sector and so on).

This classification is useful because it allows one to conceptually separate how knowledge spillovers may affect the economic analysis of single supply chains, specific product markets and the economy as a whole; but it says nothing about the social mechanism that allows information and knowledge to spread. It can be argued that at root, however, all three sources of spillovers from Griliches' list are implicitely based on the same process: the exchange of ideas through social interaction between individuals who operate across different firms and, possibly, industries. A quest for knowledge spillovers should therefore try to identify this mechanism and its determinants.

This argument motivates why it can be erroneous to amend Griliches' list by adding *geographic spillovers*. The simple idea behind assuming a role for geographical closeness is that

[3]See the discussion at pg. 577 of the article.

[4]It may be questioned that innovation and/or productivity benefits caused by vertical linkages are a case of spillovers, instead of technological advantages induced by specific vertical configurations of production. If such advantages cannot be subject to enforcible contracts, strong vertical relationships between firms can be seen as an endogenous consequence (Williamson, 1981) rather than an exogenous cause.

intuitively, it makes it easier for individuals – in particular creative and innovative ones – to meet and communicate. In all evidence, this factor reinforces the main mechanism that is common to all Griliches' types of spillover: individual interaction. However, communication mediated via spatial closeness can benefit individuals working on joint projects, those employed in different organizations but of similar technological content, as well as individuals who are loosely professionally associated. In each circumstance it is important to distinguish whether the result of facilitated interaction (thanks to geographic proximity) is actually a case of spillovers – or, more generally, an externality.

I propose the following distinction between three means by which geographical proximity stimulates intellectual productivity and innovativeness. The first one is *complementarity* between joint intellectual effort and spatial closeness. For example, inventors may be *jointly* more prolific while working and talking everyday in the same laboratory, rather than collaborating on a project via email or Skype. The technological specificities of any production process are not, by definition, externalities. The second case is the one of *facilitated search*: individuals looking for partners and collaborators for a specific research venture or project might find their "ideal type" more easily if it lives close by.[5] Again, this is not the case of an externality, but rather of reduced search frictions.

Lastly, I list pure *local knowledge spillovers*. They are the result of stimulation of new thoughts and ideas thanks to communication and personal interaction. Their defining feature is that they are unrelated to any professional/production relationship: be it present or future, contractually defined or not. It is worth to further subdivide the effect of local spillovers on two margins, the *intensive* and the *extensive* one. This attribution depends on whether the individuals involved in the process have been professionally connected or not at any point in time. Suppose that one highly reputed inventor migrates to the Silicon Valley to join a specific research project at a local company. If the new collaborators of the superstar benefit from this relationship and acquire better skills and knowledge that eventually result in new inventions *at other firms*, the effect is on the *intensive* margin. If the superstar in question meets other professionals at the local golf club on Saturday evenings, and positively influences them with his ideas and insights, the effect is on the *extensive* margin.

Local knowledge spillovers constitute an externality, for they affect the reciprocal outcomes of agents in a way that is independent of individual decisions. As such, they open the door for policy interventions aimed at addressing inefficiencies. However, in order to motivate *place-based* policies on the basis of local knowledge spillovers, it is necessary that the extensive margin of the externality is large enough relative to intensive margin. If a

[5]"Facilitated search" bears some resemblance with the second Marshallian agglomeration force, the concept of tight labor markets, which is also grounded on efficiency-enhancing reduced search frictions. While thight labor markets are by definition framed in the context of the employer-employee search process, here I am discussing the more general process of sorting in "research teams", whether these are employed by specific firms or not. The stress is on the fact that the sorting process is driven by the interested agents – the researchers – themselves, as opposed to some hierarchic organizations. An example is this difference is the case of academic coauthors from different universities, as opposed to a senior academic looking for local research assistants.

policy aimed at attracting high-skilled researchers in order to stimulate local innovation disproportionately benefits a small well-connected network, it is perhaps not the most appropriate allocation of public funds. The empirical analysis conducted in this chapter aims, *inter alia*, at evaluating the relative importance of the intensive vs. the extensive margin of local spillovers.

3.2.2 The Model

I formalize these concepts in a very stylized model, which frames the setting of the empirical analysis. The model is loosely inspired by the version of the Rosen-Roback model[6] featured in Moretti (2011), although the following is a welfare analysis of individual choice in presence of externalities and not a general equilibrium framework.

Consider two cities A and B, indexed by $c = A, B$, with a fixed population of inventors I_c each. There is a superstar inventor who may only reside in one of the two cities c ($T_c \in \{0, 1\}$, $T_A + T_B = 1$). In both cities, there are some inventors N_c who are direct collaborators of the superstar, and would remain so regardless of which city the superstar elects for his residence. In addition, there are some other individuals M_c with whom the superstar would only collaborate if he or she lived in their same city. Finally, there are L_c inventors whose research is unrelated to that of the star. For both cities, $I_c \equiv L_c + M_c + N_c$. Define p_i as the number of patents realized by a generic individual i. Assume the following identities to hold for each city c:

$$\mathbb{E}\left[p_{it}\middle| i \in L_c, T_c = 1\right] - \mathbb{E}\left[p_{it}\middle| i \in L_c, T_c = 0\right] = \ell_L \tag{3.1}$$

$$\mathbb{E}\left[p_{it}\middle| i \in M_c, T_c = 1\right] - \mathbb{E}\left[p_{it}\middle| i \in M_c, T_c = 0\right] = \ell_M + f_M \tag{3.2}$$

$$\mathbb{E}\left[p_{it}\middle| i \in N_c, T_c = 1\right] - \mathbb{E}\left[p_{it}\middle| i \in N_c, T_c = 0\right] = \ell_N + f_N \tag{3.3}$$

where $f \geq 0$ denote patents coauthored by the superstar and $\ell \geq 0$ patents *not* coauthored by the superstar. According to the previous classification of the channels through which spatial proximity benefits intellectual production, it follows that:

- f_M is the result of *complementarties* between intellectual labor and spatial proximity;
- f_N denotes the effect of a *facilitated search* of coauthors;
- ℓ represent pure *local knowledge spillovers*, on both the intensive ($\ell_M + \ell_N$) and the extensive (ℓ_L) margins.

The superstar inventor chooses in which city to live. His choice is based on: a) the difference in number of patents he can realize by residing in a given locality – which is thought to affect the superstar's income and is ultimately a function of N_A, N_B, M_A, M_B; b) other factors affecting his or her preference relative for either locality (fiscal advantages,

[6]See Rosen (1979); Roback (1982, 1988).

local amenities, personal idiosyncratic preferences). In particular, I denote by E the patent-equivalent intensity of the superstar's preference for city A. Thus, he or she will chose city A according to the rule

$$T_A = \mathbb{I}\left[f_N\left(N_A - N_B\right) - f_M\left(M_A - M_B\right) + E \geq 0\right] \tag{3.4}$$

and symmetrically for T_B.

Assuming that all patents are of the same quality (not a realistic statement, but certainly an innocuous one for the aim of the present analysis) the problem of a social planner can be thought as the one of allocating a superstar in one city so to maximize the overall patent count and, thus, innovation. The social planner also takes into account the preferences of the superstar inventor[7] but, unlike him or her, would also internalize local knowledge spillovers. Define as $D_A = 1$ the situation where social welfare is maximized if the superstar lives in city A. In analogy with (3.4), it follows that

$$D_A = \mathbb{I}\left[\left(\ell_N + f_N\right)\left(N_A - N_B\right) - \left(\ell_M + f_M\right)\left(M_A - M_B\right) + \ell_L\left(L_A - L_B\right) + E \geq 0\right] \tag{3.5}$$

and symmetrically for D_B.

The following two statements summarize the welfare analysis of this simple model.

Lemma 1. If $\ell_L = 0$ or if $L_A = L_B$, then the locational choice of the superstar inventor coincides with the social optimum ($T_c = D_c$).

This follows straightforwardly from the fact that $\ell_M, \ell_N \geq 0$. The intuition is that even if the superstar inventor does not internalize the spillovers generated on his or her coauthors, the location decision would be based on the number of potential collaborations in either city. If there are no extensive-margin spillovers or they affect both cities equally, these are not marginally relevant for the evaluation of the social optimum. The social utility function would be in that case monotonic in the number of superstar's coauthors in either city.

Lemma 2. The choice of the superstar and the social optimum do not coincide ($T_c \neq D_c$) if

$$\ell_L\left(L_d - L_c\right) \geq \left(\ell_N + f_N\right)\left(N_c - N_d\right) - \left(\ell_M + f_M\right)\left(M_c - M_d\right) + E \tag{3.6}$$

for $c = A, B$ and $d = B, A$.

In other words, a policy designed to attract a superstar from one place to another (say via a compensating manipulation of E, for example through tax incentives) can be justified only if there is a sufficiently stronger potential for agglomeration economies in the locality of destination. This is given by the combination of extensive-margin local spillovers (ℓ_L) and the difference in the population of inventors unrelated to the superstar ($L_d - L_c$).

[7] This modeling choice, besides being theoretically sound, slightly simplifies the welfare analysis. The main qualitative insights are unaffacted.

3.3 Data and Relocation Events

This section is divided in four parts. In the first part, I provide a general introduction to the data. In the second one, I characterize the "relocation events" of superstars inventors in time and space. In the third I describe of the thought experiment on which the event study is based. In the fourth and last part I define the outcomes of interest. Summary statistics are reported throughout the section.

3.3.1 Inventor Data

In this study I employ the patent dataset elaborated by Li et al. (2014), which includes disambiguated identities of individual authors of patents that have been registered at the USPTO. The dataset is organized at the inventor-patent level, and it contains the ZIP code of the residence address for individual inventor at the time of applying for a patent. I match the ZIP codes to the Core Based Statistical Area (CBSA) as per the 2006 definition, and exclude all inventors who cannot be matched to any urban area (usually, they are individuals residing either in rural areas or abroad). This results in the selection of 1,758,580 patents with application dates spanning from 1975 to 2008, authored by 987,807 inventors.

It is a common instance that patents, especially those with several multiple inventors, are signed by individuals residing in differente CBSAs. Conversely, more than 85% of inventors only file patents from a single urban area, and less than 5% realize patents from more than two different cities over the course of their life. The typical pattern of patent-level information for individuals who have been relocating across cities is sharply discontinuous, denoting well-defined moving episodes. An example of such pattern, fabricated for illustratory purpose, is represented in table 3.1.

Table 3.1: Relocation Example 1

Inventor No.	Patent No.	Name on Patent	App. Year	CBSA ID	Sequence
17380242	1369841	JOHN SMITH	1987	114	1
17380242	6432207	J. SMITH	1989	114	1
17380242	5578017	JOHN SMITH	1990	114	1
17380242	6047949	JOHN SMITH	1990	114	1
17380242	6841760	JOHN SMITH	1991	114	1
17380242	1086679	SMITH, J. F.	1991	876	2
17380242	6184582	JOHN SMITH	1992	876	2
17380242	5219248	JOHN SMITH	1994	876	2

The residence address of this very hypothetical individual (John Smith[8]), as indicated on patent applications, is associated to two different urban areas corresponding to two con-

[8]Notice how the name reported on patents purposely differs in this constructed example. This illustrates the importance of disambiguating individual identities in patent data by exploiting available information (such as uniformity in the residence address, patent class, names of collaborators and so on).

secutive "sequences": one for patent applications received by the USPTO until 1991, and the other one for later patent applications (including one from 1991). In such a case, the data would unambiguously provide the information that inventor John Smith has moved to a different CBSA in 1991, which defines the relocation *event year* for this specific individual.

In a minority of cases, however, the data are less unequivoc, because urban areas associated to ZIP Codes of residence overlap over time. This can happen for different reasons – an inventor collaborates across multiple laboratories at the same time, he or she visits some institutions for a short period of time, or perhaps really has multiple residences (but can only report one at a time). In these circumstances I assign a new "sequence", defining one inventor's residential history, only if more than half of the patents come from a new urban area in the time interval defined by potential event years (ties are resolved conservatively). To clarify the procedure, I provide another constructed example in Table 3.2.

Table 3.2: Relocation Example 2

Inventor No.	Patent No.	Name on Patent	App. Year	CBSA ID	Sequence
04857953	8414094	PAUL FOGEL	1989	417	1
04857953	2110077	PAUL FOGEL	1992	417	1
04857953	5644092	PAUL FOGEL	1992	417	1
04857953	2648021	PAUL FOGEL	1993	559	2
04857953	9477426	PAUL FOGEL	1994	559	2
04857953	2769154	P. FOGEL	1994	417	2
04857953	1180159	P. FOGEL	1995	559	2
04857953	4079703	P. W. FOGEL	1995	271	2
04857953	7219492	P. W. FOGEL	1995	559	2
04857953	8716910	P. W. FOGEL	1996	559	2
04857953	5948404	P. W. FOGEL	1998	271	2

In the case of Paul Fogel, another imaginary inventor, three CBSAs appear on record: one from his first patent in 1989 (corresponding to the first sequence), a second city-sequence whose event year is 1993, and finally a third city associated to later patents. Moreover, the CBSA from the first sequence also appears once in a patent from 1994. If this piece of data were real, they would seem to indicate that inventor Paul Fogel has relocated in 1993 but has kept ties at other places. Since patents from the third locality look sporadic, they do not constitute a third sequence, and Paul Fogel is associated to CBSA No. 559 from 1993 onwards for the entirety of his second sequence.

3.3.2 Moving Superstars

I restrict the analysis of inventor relocation to "superstars": they are defined – following standard practice – as the individuals in the top 5% of the patents per inventor distribution. Inventors with more than twelve patents fall in this category. However, individuals in this group are highly heterogeneously prolific, as Figure 3.1 indicates. This graph, showing the right tail of the empirical distribution of patents per inventors, has been truncated at 10 patents on the left and 100 patents on the right. In fact, there are 703 inventors not shown in the graph who have authored more than 100 patents. The three individuals with the highest patent count have realized 2,410, 2,064 and 995 inventions respectively in the time interval under analysis.

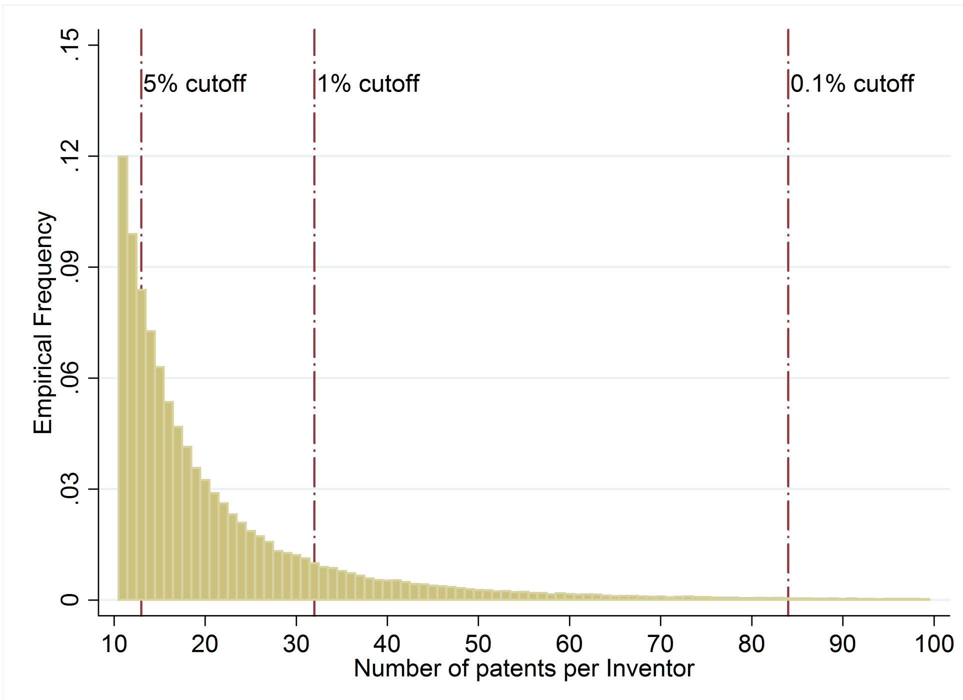


Figure 3.1: Patents per Inventor Distribution (Truncated)

Superstar inventors are considerably more mobile than other individuals. A staggering percentage of 23.7% is characterized by one relocation event as defined in the previous part of this section, and about 8% by more than one event. Superstars who are more prolific seem to be, as expected, slightly more mobile than those who are less so. Throughout the rest of this study I subdivide superstar inventors in three subgroups by their position in the patents distribution: the "Low" group, including inventors between the 5% and the 1% percentiles; the "Medium" group, encompassing those between the 1% and the 0.1% quantiles; and the "High" group, to which belong the individuals in the top 0.1% of the distribution (the 1% and 0.1% cutoffs are respectively 32 and 84 patents, as reported in Figure 3.1). The number of individuals in each of these three groups, broken down by number of observed moves, is reported in Table 3.3.

I restrict the analysis to the relocation events of individuals who move only once, whose event year falls between 1981 and 2001. The reason of these choices is to perform a correct

Table 3.3: Number of Moves by Superstar Group

Group	Low	(%)	Med.	(%)	High	(%)	All	(%)
0 Moves	26,331	(70.2%)	5,912	(65.4%)	701	(64.9%)	32,944	(69.2%)
1 Move	8,732	(23.3%)	2,298	(25.4%)	248	(23.0%)	11,278	(23.7%)
2 Moves	1,985	(5.3%)	623	(6.9%)	96	(8.9%)	2,704	(5.7%)
3+ Moves	450	(1.2%)	209	(2.3%)	35	(3.2%)	694	(1.4%)
Total	37,498	(100%)	9,042	(100%)	1,080	(100%)	47,620	(100%)

event analysis and to attribute to each event a sufficient number of observations for the outcome variables both before and after the event date. By selecting the 1981 and 2001 window, I am able to use observations as early as 1975 for 1981 events and as late as 2008 for 2001 events. By excluding superstars with more than one move, I eliminate instances where the time window to analyze certain events is truncated by close previous or subsequent relocations of the same superstar. The number of events in each year is displayed for each group of superstars in Figure 3.2, and it increases over time for all groups.

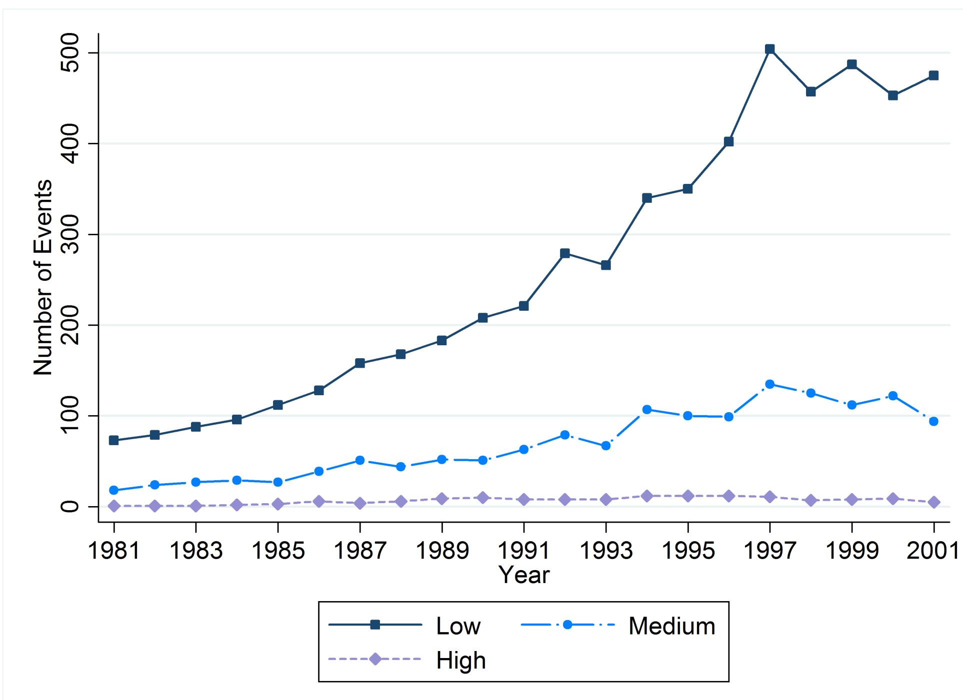


Figure 3.2: Number of Relocation Events in Each Year

The geography of the relocation events presents quite recognizable patterns. The two maps in Figures 3.3 and 3.4 display a classification of CBSAs by, respectively, the total number of "In-Migration" (where a superstar moves) and "Out-Migration" (whence a superstar leaves) events from 1981 to 2001. California is, by far, the state involved in most events: the San Jose urban area has witnessed 680 superstars arriving and 609 leaving (unsurprisingly the top figure in both categories), San Francisco has respectively 568 and 451, while Los Angeles has 255 and 399. By contrast, the New York urban area has experienced the departure of 325 superstars against only 155 arriving. These patterns look very similar if one breaks them down by more restricted time windows or by superstar group.

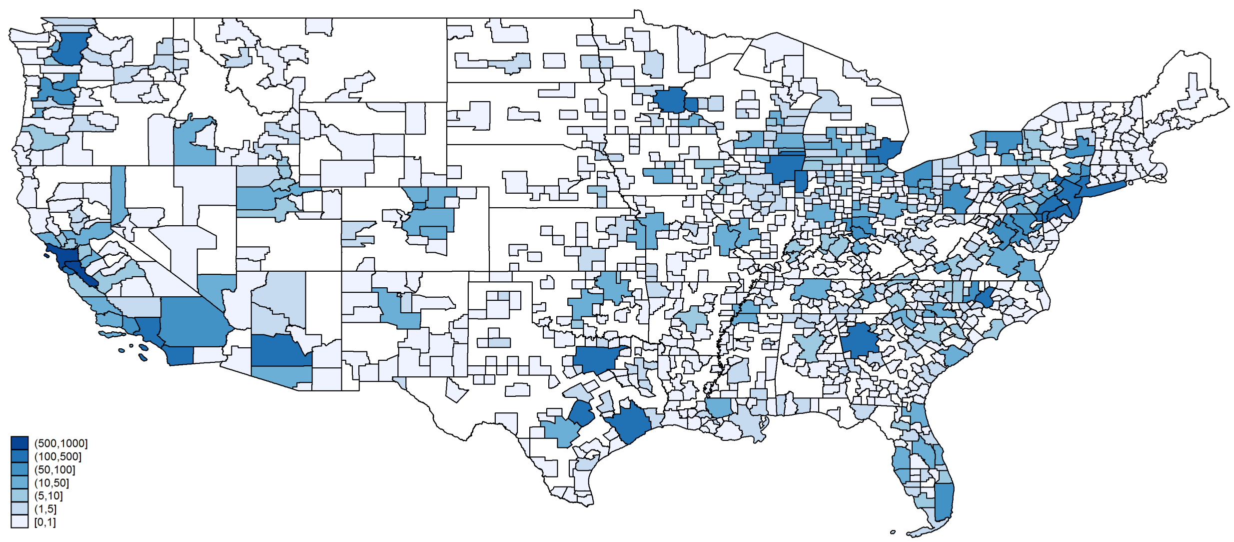


Figure 3.3: Map of All Moving-In Events

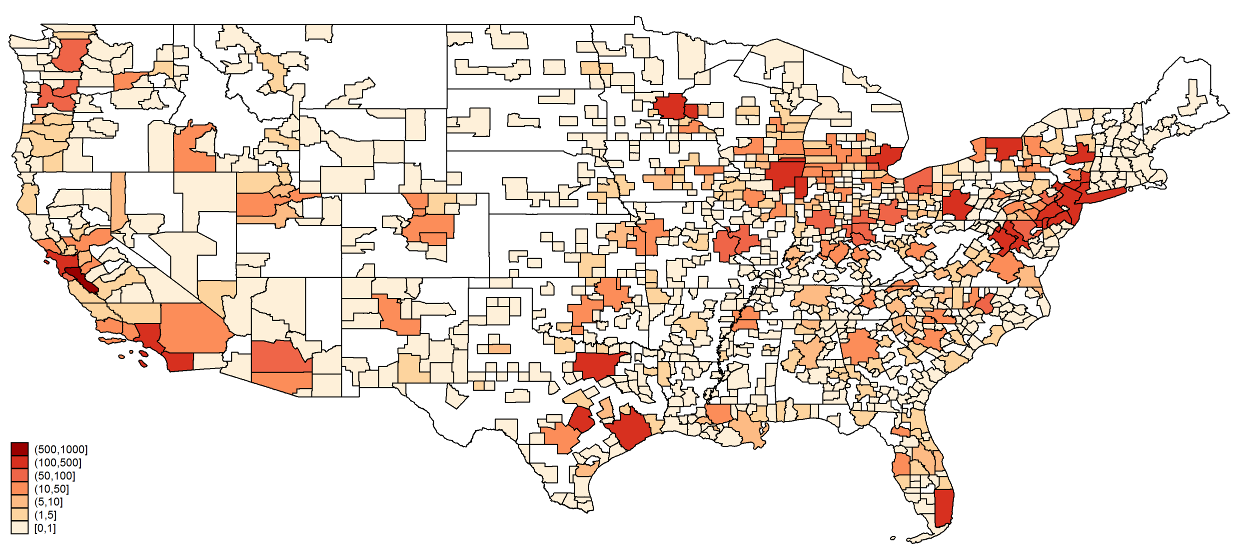
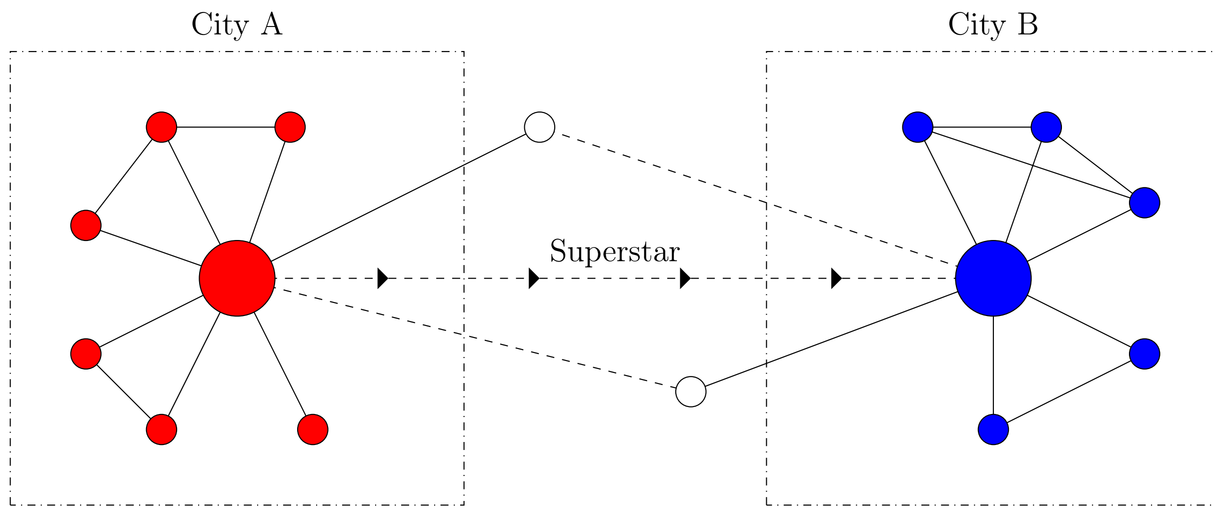


Figure 3.4: Map of All Moving-Out Events

3.3.3 Local Networks

This analysis has two aims: first, to identify genuine local knowledge spillovers net of complementarity effects and other technological characteristics of the intellectual production process; second, to disentangle the intensive and the extensive margin of localized spillovers. The former objective is addressed by estimating the impact of a superstar's relocation event on productivity measures of the superstar's local coauthorship networks, in both localities of destination and departure. These productivity measures are *residual patent counts* – that is, they exclude patents realized together with the superstar himself. The resulting estimates allow to evaluate the intensive margin of local knowledge spillovers: the exclusion of patents common with the superstar are meant to remove complementarity and "search" effects. The second objective is realized by estimating the impact of the superstar's arrival or departure on the residual patent count of an entire CBSA. This allows to assess the extensive margin of local knowledge spillovers against the intensive margin.



Graph 3.1: A Moving Superstar and Local Coauthorship Networks

A stylized example can easily illustrate the thought experiment on which the analysis is based. Graph 3.1 represents the relocation event of a superstar and his or her coauthorship network. The superstar in question (the larger node) moves from City A to City B, and has collaborators in both places – as well as some that do not reside in either city. The event study featured in this chapter evaluates the impact of the relocation event on the residual patent outcomes of two groups: the blue nodes in the graph – the coauthors in the locality of destination – and the red nodes – those in the city of departure. Under the hypothesis that the physical presence of the superstar generates local spillovers, one should expect a positive effect on the former and a negative effect on the latter. In addition, the impact of the event is evaluated against the residual patent production of all other inventors (excluding the superstar's direct network) in either City A or City B.

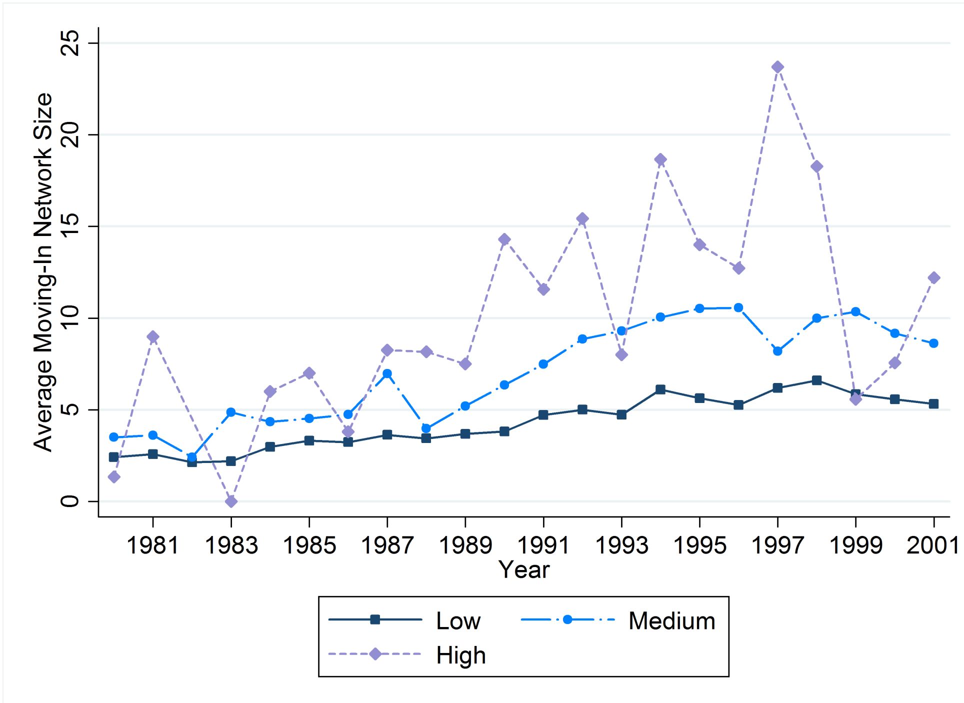


Figure 3.5: Average Network Size by group, cities of Destination

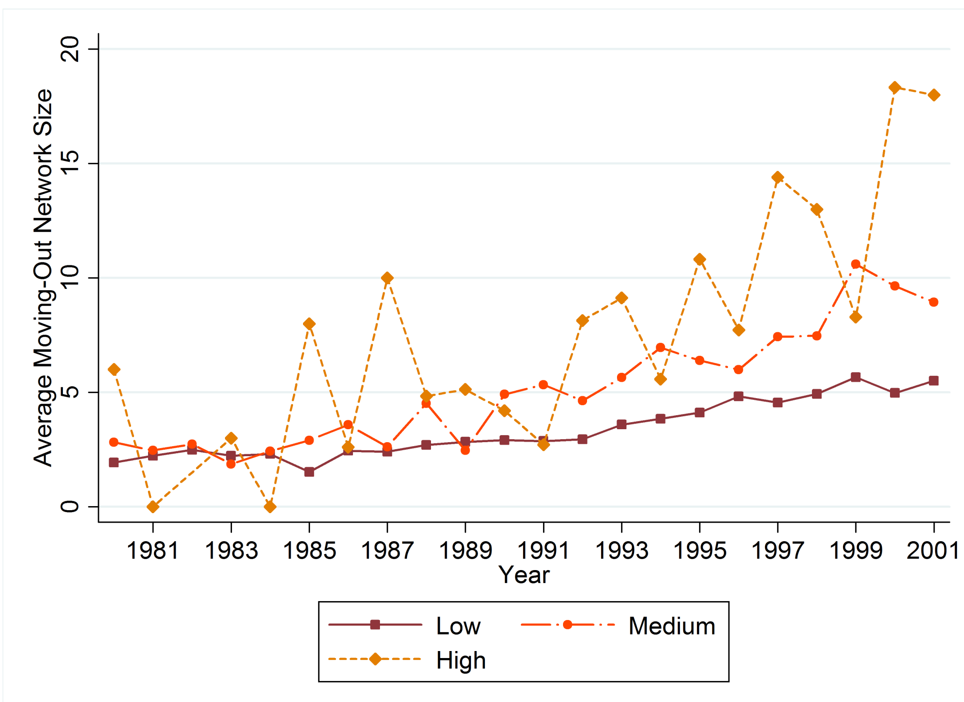


Figure 3.6: Average Network Size by group, cities of Departure

In the analysis, a coauthor of a superstar is defined as one inventor who has signed a patent application jointly with the superstar in an interval of six years before or after the relocation event. Every coauthor is assigned to a metropolitan area on the basis of the available information on his location of residence at the time when the superstar moves, which can be extracted from the data as described earlier in this section. For some superstar-events, no network can be reconstructed in either involved locality, leaving a final number of 6,044 events subject to the analysis. Interestingly, the average size of local networks does not increase wildly with the relative position of the superstar in the patent distribution.[9] Summary statistics relative to the network sizes, broken down by year and superstar group,[10] are reported in Figures 3.5 and 3.6 respectively for the cities of destination and of departure.

3.3.4 Patent Outcomes

I estimate the effect of superstar relocation on four network-level patent counts, two of them being citation-weighted and two of them not. I start defining them from the simplest case, the one of the pure patent count. For the superstar-event i, city $c \in \{I, O\}$ (where I is the "In" destination city and "O" is the "Out" departure city) and year t the pure patent count P_{ict} of the corresponding network is given by

$$P_{ict} = \sum_{n \in \mathcal{N}\{i,c\}} \hat{p}_{nit} \tag{3.7}$$

where $\mathcal{N}\{i,c\}$ is the set of coauthors of superstar i located in city c, and $\hat{p}_{nit}$ is the *residual* number of patent *applications* of inventor n in year t. "Residual," as a property of the patent count, signifies that patents of an inventor that are joint with superstar i are excluded from the count, and this is denoted by the "hat" symbol in $\hat{p}$.

While all empirical estimates control for network size, the pure patent count might not be a meaningful measure given the large variation in the dimension of the local networks of superstars. To account for this one may consider the *average* patent count, dividing (3.7) by network size.

$$\bar{P}_{ict} = \frac{1}{\dim\left(\mathcal{N}\{n,c\}\right)} \sum_{n \in \mathcal{N}\{n,c\}} \hat{p}_{nit} \tag{3.8}$$

Another issue is that some technological and research areas are characterized, relative to others, by larger team sizes – typically resulting in patents with many authors. This may bias the analysis either way depending on the frequency of the events by technological field. To address this, one could use the residual patent *shares* count $\hat{s}_{nit}$, that is the sum of all patent shares of inventor i (as opposed to $\hat{p}_{nit}$). A patent share is the inverse of the number of authors of a patent; in the case of a patent with three authors for instance, the share of

[9] Premilinary analysis shows that superstars in higher quantiles have a disproportionate number of coauthors dispersed in localities where they have never resided.

[10] An average of zero for the "High" group in a given year indicates that no local network can be reconstructed for the very top superstars on that year.

each author is 1/3. Shares of patents co-authored with the superstar are removed from the count. One may also look at the average share count in a network.

$$\tilde{P}_{ict} = \sum_{n \in \mathcal{N}\{i,c\}} \hat{s}_{nit} \tag{3.9}$$

$$\tilde{\bar{P}}_{ict} = \frac{1}{\dim\left(\mathcal{N}\left\{i,c\right\}\right)} \sum_{n \in \mathcal{N}\{n,c\}} \hat{s}_{nit} \tag{3.10}$$

Still, all these alternative measures do not account for patent heterogeneity – the fact that some patents are more valuable than others. A frequently used proxy of patent quality is the number of citations; thus I also estimate the effect of the events on appropriate citations-weighted patent outcomes. The citation-weighted analogues of (3.7), (3.8), (3.9) and (3.10) are:

$$B_{ict} = \sum_{n \in \mathcal{N}\{i,c\}} \hat{b}_{nit} \tag{3.11}$$

$$\bar{B}_{ict} = \frac{1}{\dim\left(\mathcal{N}\left\{i,c\right\}\right)} \sum_{n \in \mathcal{N}\{n,c\}} \hat{b}_{nit} \tag{3.12}$$

$$\tilde{B}_{ict} = \sum_{n \in \mathcal{N}\{i,c\}} \hat{z}_{nit} \tag{3.13}$$

$$\tilde{\bar{B}}_{ict} = \frac{1}{\dim\left(\mathcal{N}\left\{i,c\right\}\right)} \sum_{n \in \mathcal{N}\{n,c\}} \hat{z}_{nit} \tag{3.14}$$

where b_{nit} is the *citation-weighted* patent count of inventor n and z_{nit} is the count of citation-weighted shares.[11] In order to account for the truncation problem – the fact that more recent

[11] I provide more accurate definitions of $\hat{p}_{nit}$, $\hat{b}_{nit}$, $\hat{s}_{nit}$ and $\hat{z}_{nit}$, not given in the text in order not to overload it. Define $\mathbb{K}$ as the set of all patents, $\dim(\mathbb{K}) = K$, and index patents by $k = 1, \ldots, K$. Notation $\mathcal{P}(n,t)$ denotes the set of patents for which inventor n filed an application in year t. Define $S(k)$ as the number of authors of patent k. Finally, $C(k)$ is the truncation-adjusted citation weight. The definitions are as follows.

$$\hat{p}_{nit} = \sum_{k=1}^{K} \mathbb{I}\left[k \in \mathcal{P}(n,t) \wedge k \notin \mathcal{P}(i,t)\right] \tag{3.15}$$

$$\hat{s}_{nit} = \sum_{k=1}^{K} \mathbb{I}\left[k \in \mathcal{P}(n,t) \wedge k \notin \mathcal{P}(i,t)\right] \cdot \frac{1}{S(k)} \tag{3.16}$$

$$\hat{b}_{nit} = \sum_{k=1}^{K} \mathbb{I}\left[k \in \mathcal{P}(n,t) \wedge k \notin \mathcal{P}(i,t)\right] \cdot C(k) \tag{3.17}$$

$$\hat{z}_{nit} = \sum_{k=1}^{K} \mathbb{I}\left[k \in \mathcal{P}(n,t) \wedge k \notin \mathcal{P}(i,t)\right] \cdot \frac{C(k)}{S(k)} \tag{3.18}$$

patents have yet to receive citations at the time they are observed – I apply the correction method employed in Hall et al. (2001). Still, the method yields very imprecise inferences about the quality of later patents (such as those filed in the 2000's or in the late '90s), a problem that affects the precision of the analysis of even earlier events. For this reason, I restrict the analysis to relocation events prior to 1997 when using these measures.

For the estimation of extensive margin spillovers I use a set of alternative measures that is analogous to the one above. The main difference is that, in this case, each count is based on all inventors in city c excluding those who are part of the superstar's coauthorship network – a set that I denote as $\mathcal{M}\{i,c\}$ (with $\mathcal{M}\{i,c\} \cap \mathcal{N}\{i,c\} = \oslash$, that is, the two sets are complementary). The city-wide patent count, the analogue of P_{ict}, is defined as

$$C_{ict} = \sum_{n \in \mathcal{M}\{i,c\}} \hat{p}_{nit} \tag{3.19}$$

and similarly are $\bar{C}_{ict}$, $\tilde{C}_{ict}$, $\tilde{\tilde{C}}_{ict}$, G_{ict}, $\bar{G_{ict}}$, $\tilde{G}_{ict}$ and $\tilde{\tilde{G}}_{ict}$; where G denotes citation-weighting like B does in the case of network-level counts.

Figure 3.7 shows the averages of P_{int} and $\tilde{P}_{int}$ at the event-year level from 1981 to 2001, for both the localities of destination and those of departure. As one would expect, in the early years the networks "abandoned" by the superstar tend to be more productive than those that are "reached", while the reverse is true at the end of the time period. Figure 3.8 reports analogous summary statistics for C_{int} and $\tilde{C}_{int}$[12]. The trend of each variable looks remarkably similar for both "In" and "Out" localities, showing little systematic association of superstars' moving choices with changes in local innovativeness. The summary statistics for the average and the citation-weighted measures, not shown here for brevity, follow patterns that are very similar to those displayed in figures 3.7 and 3.8 for both network-level and city-level counts.

3.4 Event Study: Empirical Results

This section discusses the empirical results from the implementation of the event study, and is divided in five parts. In the first part, I outline the workhorse empirical model that is common to each estimation model. In the second part, I describe the estimates relative to the baseline outcomes of interest, the pure residual patent counts (both at the network and at the city level). In the third part I briefly comment on the results from the other outcome measures, excluding the citation-weighted ones. The latter are the subject of the fourth part of the section. Finally, the fifth and last part of the section discusses results for the baseline measure divided by superstar group. In Appendix F I report additional event analyses, in the form of graphs, that for reasons of conciseness and relative relevance are not included in the main body of this chapter.

[12] These averages can be approximated as a weighted average of patent counts at the city level, where each weights is the number of events associated with one city

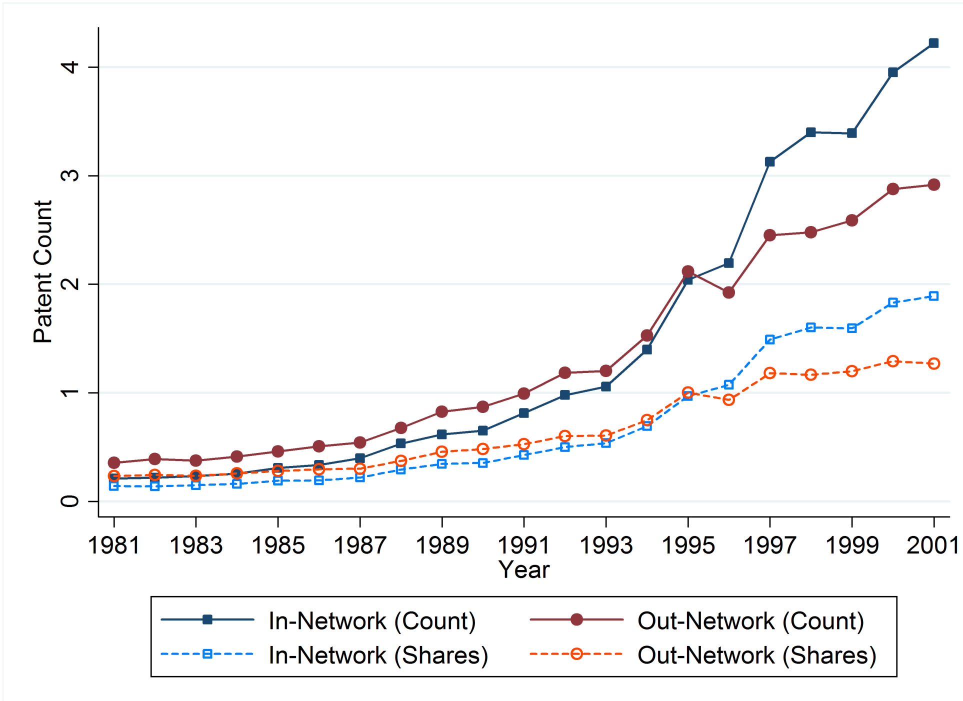


Figure 3.7: Network-level Patent Count averages, by year

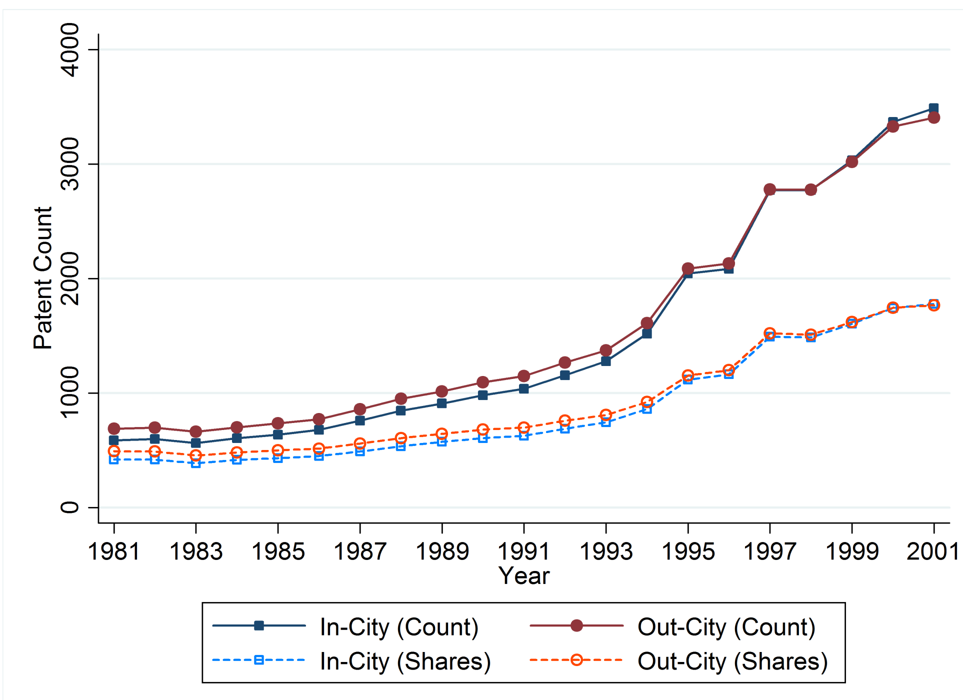


Figure 3.8: CBSA-level Patent Count averages, by year

3.4.1 Empirical Model

The empirical model common to all estimates presented hereinafter is standard in the event studies methodology. For each event i denote its event date as $s(i)$, a set of dummy variables $\left\{D^{-\bar{K}}, \ldots, D^{k}, \ldots, D^{\bar{K}}\right\}$ defined as $D_{ict}^{k} = \mathbb{I}\left[t = s(i) + k\right]$ for $-\bar{K} \geq k \geq \bar{K}$, and two additional "bound" dummies[13] $\left(D_{ict}^{\ell}, D_{ict}^{u}\right)$, that are similarly given as $D_{ict}^{\ell} = \mathbb{I}\left[t < s(i) - \bar{K}\right]$ and $D_{ict}^{u} = \mathbb{I}\left[t > s(i) + \bar{K}\right]$. The model reads as:

$$Y_{ict} = \alpha_{ic} + \sum_{k=-\bar{K}}^{\bar{K}} \beta_k D_{ict}^{k} + \gamma_\ell D_{ict}^{\ell} + \gamma_u D_{ict}^{u} + \delta_{ct} + \epsilon_{ict} \tag{3.20}$$

$$\beta_{-1} = 0 \tag{3.21}$$

where Y_{ict} is a patent count outcome of choice for either $c = I$ or $c = O$. In this study I set $\bar{K} = 6$, so that the model can be estimated on the 1975-2007 sample for all events occurred between 1981 and 2001 and the resulting panel that is internal to the bound dates is balanced. The estimates of $\{\beta^{-6}, \ldots, \beta^{0}, \ldots, \beta^{6}\}$ allow to identify the relationship between the event and the outcome of interest over time. Notice that event-city fixed effects α_{ic} control for all time-invariant characteristics, including network size for the estimates at the network level. The model is estimated via OLS and standard errors are clustered at the "event" i level, that is the local network or the CBSA (depending on the left-hand side variable of choice).

3.4.2 Baseline Results: Simple Patent Counts

Table 3.4 reports the results of the event study for the baseline patent measures: the simple counts P_{ict} and C_{ict}. Column (1) displays the estimates relative to the network geographically joined by the superstar (P_{iIt}). The average patent count increases slowly before the event, to be followed by a large increase after it. The post-event difference in patent count at the network level stabilizes at around 0.9-1 extra residual patent applications per year, which is a sizeable effect. Column (2) displays the estimates relative to the network that the superstar has left (P_{iOt}). Also in this case a slow increase in the patent count can be noticed before the event date, to be followed by a rapid decrease. Six years after the event, the "abandoned" network generates on average one residual patent less relative to the year prior to the event, with a resulting difference with the "In" Networks of about 1.75 patents. These results are better appreciated via graphical visualization; they are reported in Figure 3.9.

Column (3) in Table 3.4 shows the estimates relative to the entire patent count of the urban area where the superstar has moved (C_{iIt}). Again, the total number of residual patent applications rises before the event, to be followed by a decline. The results for the urban area left by the superstar (C_{iOt}), reported in column (4), follow a parellel trend, as Figure 3.10 displays. These estimates indicate that superstar inventors tend to move in periods of higher-than-usual innovative activity, for both localities involved in the event. While it is not

[13] Upperscripts ℓ and u stay respectively for "lower" and "upper" bound.

Table 3.4: Event Analysis: Simple Patent Count

	(1) IN-Network	(2) EX-Network	(3) In-City	(4) EX-City
–6	-0.7003	-0.2807	-83.2476	-52.1744
	(0.0558)	(0.0580)	(12.4917)	(12.2570)
–5	-0.6270	-0.1863	-71.9672	-39.9814
	(0.0536)	(0.0636)	(10.4017)	(10.2108)
–4	-0.5415	-0.0983	-51.7083	-21.7772
	(0.0506)	(0.0557)	(8.5916)	(8.3035)
–3	-0.4331	-0.0325	-41.0032	-13.1039
	(0.0471)	(0.0584)	(6.1669)	(5.9931)
–2	-0.2526	-0.0098	-18.8553	-6.2776
	(0.0452)	(0.0499)	(4.3807)	(4.1525)
–1	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)
0	0.7553	0.3203	6.9865	2.6266
	(0.1023)	(0.0940)	(4.7339)	(4.5236)
+1	1.1174	0.1352	12.7360	0.3500
	(0.1353)	(0.1017)	(7.5889)	(7.2810)
+2	1.2898	0.0415	7.0707	-4.3009
	(0.1390)	(0.1229)	(12.3952)	(12.0918)
+3	1.1808	-0.2394	-2.3283	-26.1491
	(0.1408)	(0.1482)	(17.8957)	(17.4420)
+4	1.0778	-0.5683	-34.3247	-58.5404
	(0.1535)	(0.1585)	(24.5080)	(23.7968)
+5	0.8506	-0.8300	-79.1706	-93.8742
	(0.1653)	(0.1745)	(32.0839)	(31.1991)
+6	0.7499	-1.0151	-103.5372	-121.0449
	(0.2027)	(0.1806)	(37.1872)	(36.5288)
N	187364	187364	187364	187364

immediate to deduce the exact reason why this happens, what is interesting is that no clear trend distinguishes the two types of localities. If these results seem to exclude any effect of the superstar itself on the productivity of all local inventors, on the other hand they also seem to exclude that the superstar's decision relates in either way to differentials in urban innovativeness.

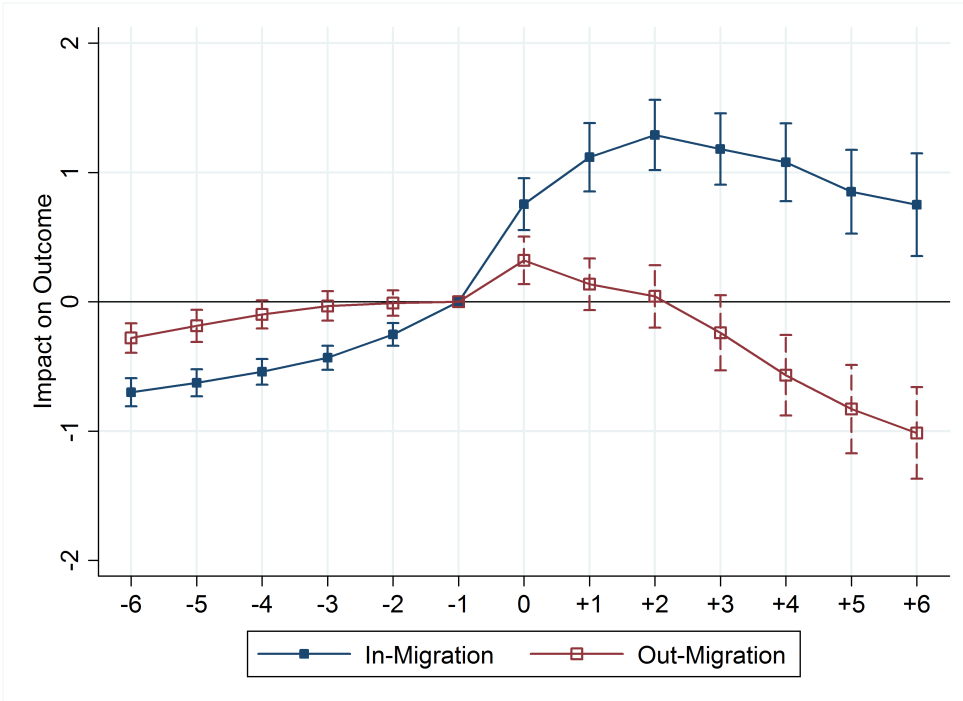


Figure 3.9: Network-level Estimates, Patent Counts

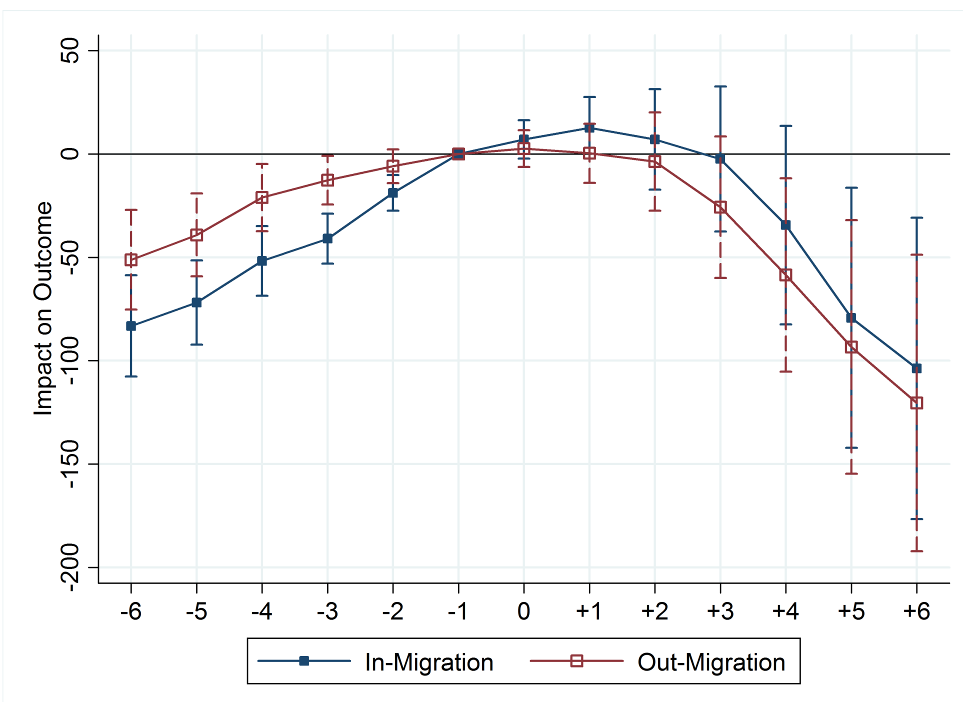


Figure 3.10: City-level Estimates, Patent Counts

Overall, these results suggest that there are important localized knowledge spillovers, because inventors who are related to the superstar tend to independently produce more patents if the superstar lives close to them. However, such spillovers seem to be confined to what I have defined the "intensive margin", that is the effect is limited to the set of close coauthors of the moving inventor. There is no *prima facie* evidence of spillover effects on other individuals in the city of destination, seemingly ruling out the existence of what I call the "extensive margin" of local spillovers. In accordance with the discussion conducted in the first section of this chapter, such a scenario does not provide any solid motivation for "big push" type of place-based innovation policies, especially those that are aimed at attracting highly skilled inventors in order to boost the growth of local high-tech industries.

3.4.3 Other Patent Count Measures

The simple patent count P_{ict} (or C_{ict} for estimates at the urban area level) is useful to make an immediate assessment of the absolute magnitude of the effects. As it was discussed earlier, however, this may not be the most appropriate measure if local networks – or cities – differ in the number of inventors or in the propensity of inventors to patent in large teams (which may vary across technological fields). For this reason the analysis is also conducted on other outcome measures, like the average patent count $\bar{P}_{ict}$ and the share count $\tilde{P}_{ict}$. The corresponding results are displayed in two separate graphs, respectively in Figures 3.11 and 3.12. The results look very similar to those from Figure 3.9; only the order of magnitude differs. They indicate that the effect is driven neither by the possibility that superstars move from places with few inventors to places with more, nor by the one that highly mobile superstars are more frequently found in more collaborative and prolific sectors. The results for the combined measure of average patent shares $\bar{\tilde{P}}_{ict}$, shown in Appendix F, also look alike. The results for alternative outcome measures at the CBSA-level are similarly deferred; neither do they evidence any kind of extensive margin local spillovers.

3.4.4 Citation-Weighted Patents

Different is the case of outcome measures for which patents are weighted by citations. Figure 3.13 reports the results from the event study estimation for the Network-level citation-weighted count B_{ict}, while Figure 3.14 displays the corresponding results for the CBSA-level measure G_{ict}. They are to be compared respectively with Figures 3.9 and 3.10. The results at the network level speak of an overall less pronounced effect of superstar relocation on the quality of his or her coauthors' residual patents. This effect is not only less distinguishable from the pre-event trend, but also fades away four years after the event. Furthermore, the estimated coefficients for the count five or six years following the event are statistically indistinguishable across "In" and "Out" localities. The results at the urban area level from Figure 3.14 present, instead, the same pattern of Figure 3.10. Studies on other B and G measures yield qualitatively identical results; most of them are shown in Appendix F.

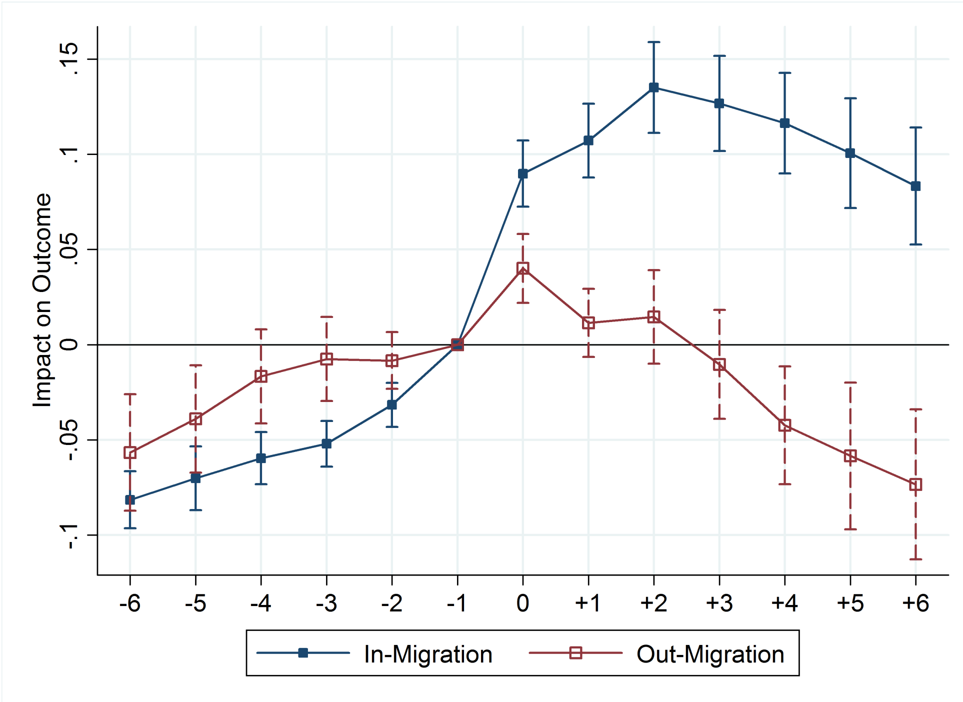


Figure 3.11: Network-level Estimates, Average Patent Counts

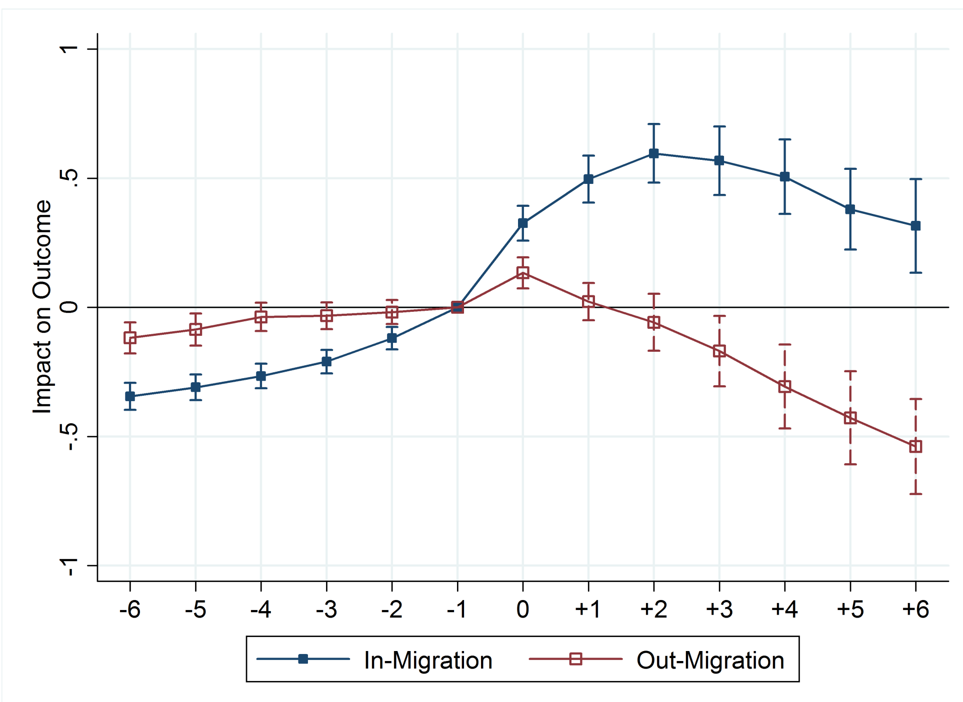


Figure 3.12: Network-level Estimates, Patent Shares

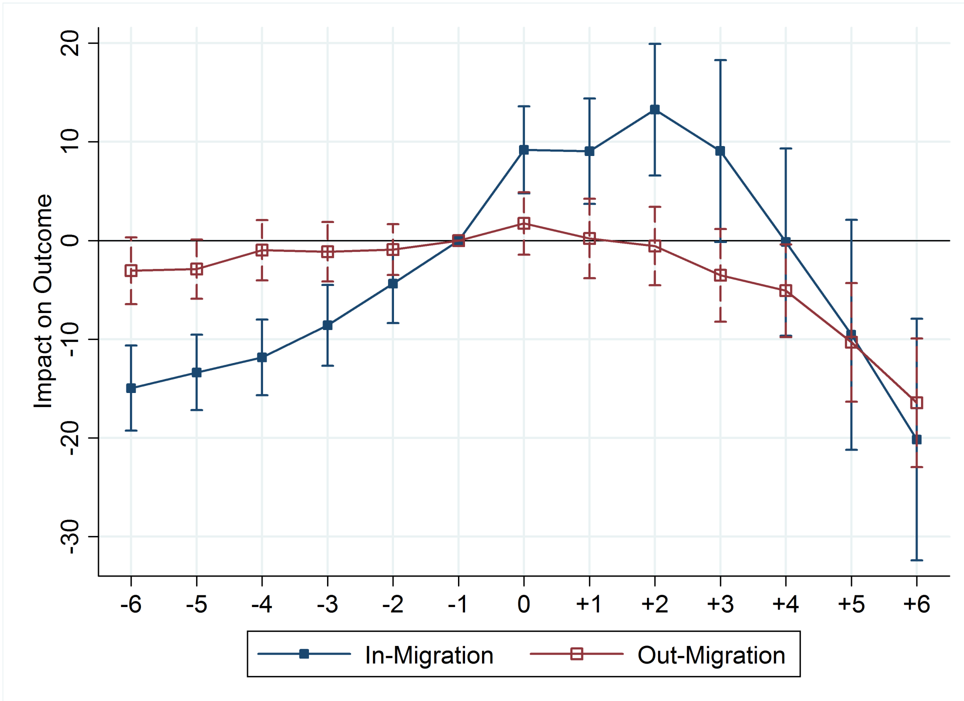


Figure 3.13: Network-level Estimates, Cit. Weighted Patent Counts

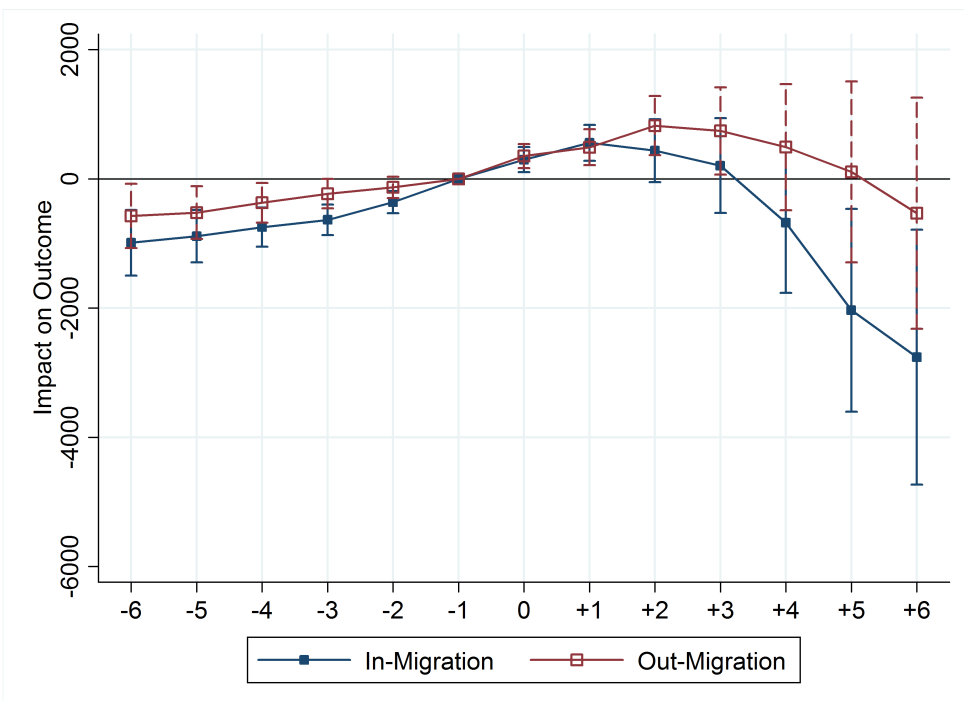


Figure 3.14: City-level Estimates, Cit. Weighted Patent Counts

These results are to be intepreted with caution, as one should reconcile a persistent effect of superstar relocation on the pure patent count of the "In" network with a temporary effect on the citation-weighted count for the same group. A possibility is that the spillover effect is indeed limited in time but is also associated with a "reputation" effect – that is, individuals who manifest an association with a local superstar become able to realize more low-quality patents, potentially crowding out the "good" ones. If this were true, there would also be a negative network effect associated with the migration of superstars. However, such a hypothesis is hardly testable within the present analysis and with the available information. In any case, it is worth to stress once again that citation-weighted estimates are conducted on a much reduced sample (excluding all events occurred after 1996) in order to mitigate the truncation problem. However, while the statistical power is certainly reduced by a great deal, the truncation problem might still be present – and this would disproportionately affect the estimates of later post-event years, differentially exacerbating measurement error issues. This can only be addressed with more data and an extended longitudinal dimension.

3.4.5 Superstar Heterogeneity

It may be argued that the effects of superstars' migrations are not the same if the inventor in question has 15 patents or instead 150 – or for that matter, 1,500. It is therefore interesting to estimate the model separately by group of superstars, distinguished by the position in the patent distribution. Specifically, I discuss results relative to the "Low", the "Middle" and the "High" groups of inventors defined in Section 2. Inventors in the "Low" group lie between the 5% and the 1% of the patent per inventor distribution; those in the "Middle" between the 1% and the 0.1% quantiles, and the "High" group includes the top 0.1%. While this partition may not be the best way to capture the heterogeneous effect of local spillovers as it varies with the superstar number of patents, the evidence of any difference across the three groups can still be indicative.

Figures 3.15 and 3.16 report the results for the estimates of P_{ict} and $\bar{P}_{ict}$ divided by group, and for $c = \{I, O\}$.[14] For both "In" and "Out" localities, the on-impact effects associated with the superstars in the Middle and High groups – especially with the latter – are remarkably stronger than those relative to the Low grup, in whose case the very presence of the effects is dubious. These differences do not depend on variations in network size, because they are consistent when using the "average" measure $\bar{P}_{ict}$. In the long run, geographical proximity with a superstar inventor in the top 0.1% of the distribution is associated with 6-10 extra residual patents in the typical group; on average these are 0.2-0.4 patents per inventor. The corresponding figure for the Medium group is stable around 0.2 extra patents. Conversely, the coauthors of a top 0.1% superstar left who have been left in the old city produce 0.15 less residual patents six years after departure of their more prolific colleague; 0.10 less for the Medium group. Standard erros for the top group's estimates are shown in Appendix F.

[14] In these figures, confidence intervals are omitted for the sake of clarity; most of the relevant effects are however significant, and despite the small statistical power associated with the "High" group.

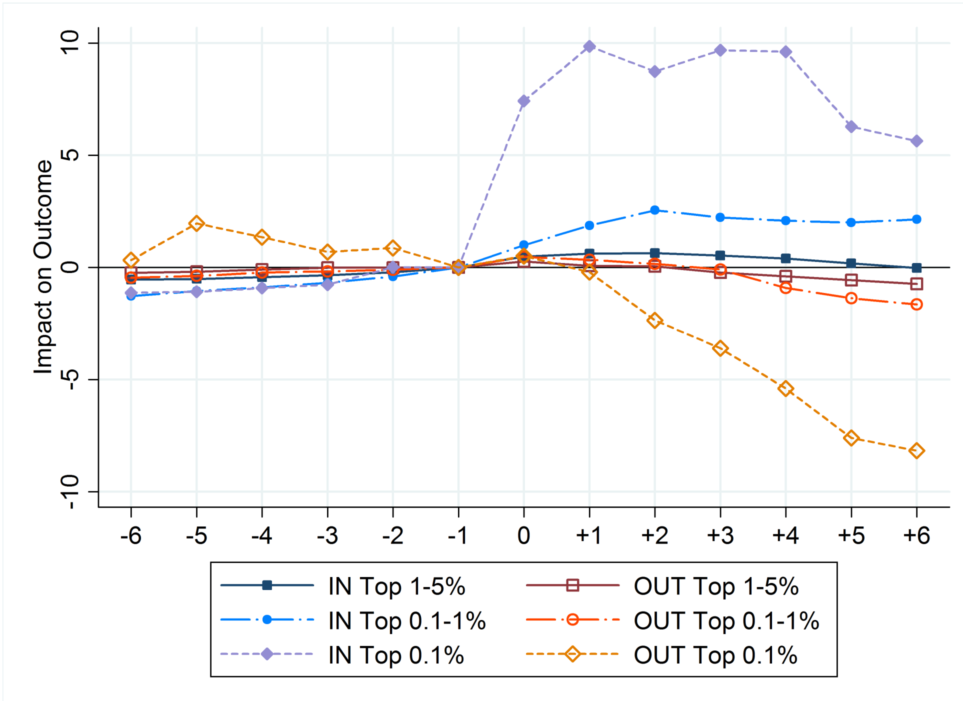


Figure 3.15: Network-level Estimates by Group, Patent Counts

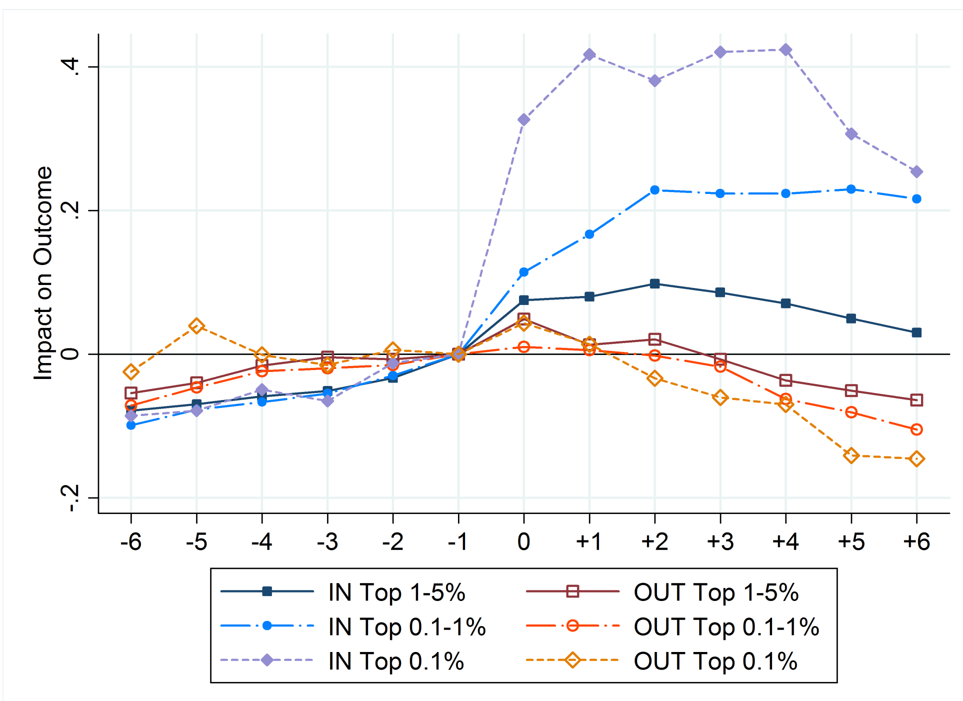


Figure 3.16: Network-level Estimates by Group, Average Patent Counts

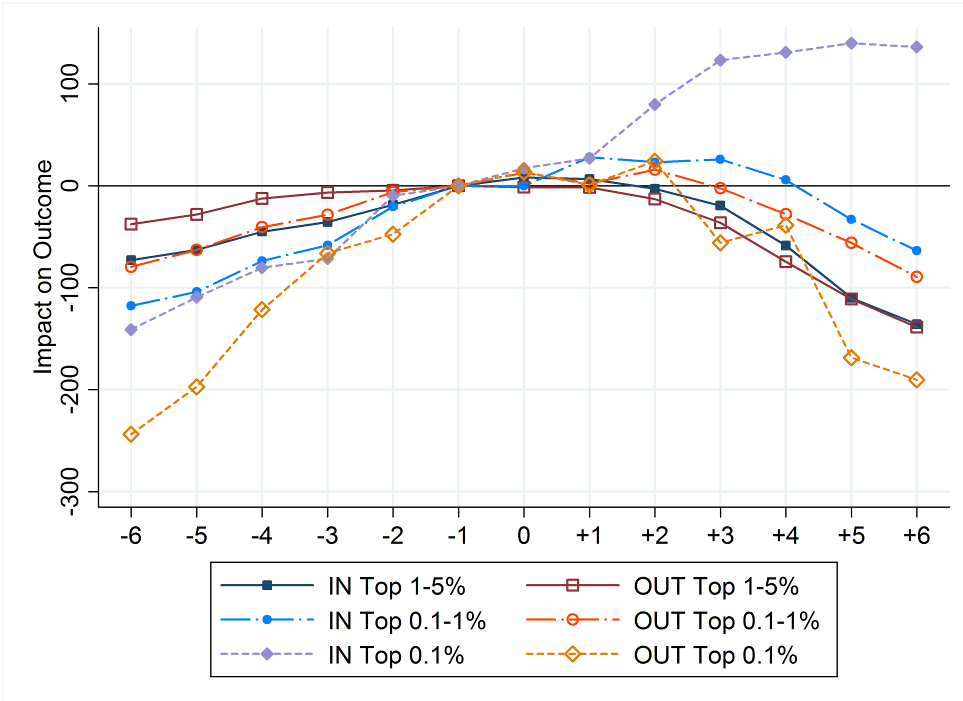


Figure 3.17: City-level Estimates by Group, Patent Counts

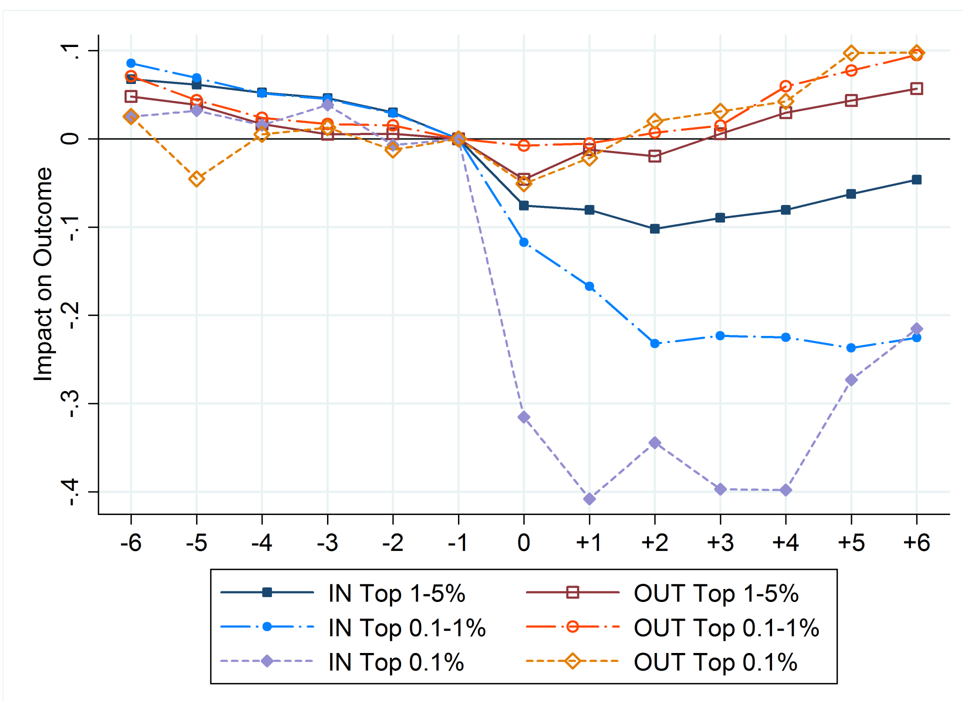


Figure 3.18: City-level Estimates by Group, Average Patent Counts

The results at the city level are more interesting when separated by superstar group. Figures 3.17 and 3.18 show the relative estimates respectively for C_{ict} and $\bar{C}_{ict}$. Unlike all other cases, there seems to be an increase in the overall (residual) patent count in the locality joined by a top 0.1% superstar – although this lags the event by two or three years. In addition (and remarkably) this is associated with a huge and on-impact *decrease* in the average count, which amounts to 0.4 *less* patents per inventor. This coincidence can only be explained by a simultaneous increase in the number of inventors in the same city, which precedes the unfolding of the effect. It would be interesting to explore whether this phenomenon is exogenous or endogenous to the relocation of first-class superstars. It certainly matches models of self-reinforcing agglomeration economies and local externalities, although it does not correspond to the definition of "extensive margin local spillovers" – in terms of additional individual residual patents – that has been given in this work.

To summarize, breaking down the analysis by superstar group effectively shows (as someone would expect) that the spillover effects associated with inventors who rank even higher in the patent distribution are indeed larger. In addition, the relocation of superstars in the "High" group – the top 0.1% – is associated with an increased patent output in the city of destination, which is driven by an increase in the number of low-productivity inventors. Results using share counts instead of unitary patent counts would yield the same set of conclusions. It is unfortunately not appropriate to make a similar analysis for citation-weighted measures while keeping the same, convenient definitions of superstar groups. In fact, the loss of statistical power associated to such an exercise would yield very noisy and hard to interpret estimates for the Middle and High groups.

3.5 Conclusion

This study analyzes patterns in patent production that precede and follow the relocation of a superstar inventor from one city to another. The objective of the analysis is twofold: first, it aims at separating complementarities between knowledge production and spatial proximity from pure knowledge spillovers; this is addressed by focusing the attention on the residual patents, those that are not co-authored with the moving superstar. The second objective is to distinguish two dimensions of local knowledge spillovers: the intensive margin, primarily benefitting the more direct colleagues of the superstar; and the extensive margin, a more generalized effect on all other inventors within one city. Such groups are distinguished in the empirical analysis by using past and future patent coauthorship relationships in order to identify the colleagues of a superstar.

The empirical results show that the location decision of a superstar inventor has a very strong effect on the residual productivity of his or her group of more direct collaborators. Those among them who reside in the location where a superstar moves tend to produce on average 0.1 more patents per year – a number that increases to 0.2 if the superstar belongs to the top 1% of the patent distribution and to 0.4% for the top 0.1%. Conversely, inventors who keep living close to the old home of the superstar experience a negative trend of their residual productivity shortly after the event. While clear intensive margin type of localized spillovers can be identified, there seems to be little to no evidence in favor of exensive margin spillovers. Consequently, this study offers little support to policies aimed at attracting big names in a technological field, on the expectation that this would generate wider agglomeration effects.

I would suggest three directions worth pursuing for future work to improve on the evidence presented in this study. First, the analysis should be enriched by including more control variables in the empirical model – primarily variables that are relative to the localities involved in the events, but also to the moving superstar. The latter requirement is particularly challenging, because efforts to link inventor-patent data to information on their labor market and personal characteristics have just begun. Second, the analysis should identify and incorporate some exogenous determinants that lead to the movement of superstars; this would help to remove potential biases induced by unobserved productivity determinants of groups of inventors. Finally, it may be interesting to distinguish more than two margins of local spillovers: perhaps the presence of a superstar is unable to influence most inventors of one city, but can still affect individuals who are more distantly related to the superstar in their extended coauthorship network.

Appendix A

Extended Analysis of the Bayesian Game

In this appendix I discuss in more detail the general version of the model outlined in section 1.2. This will let me prove Propositions 1 and 2 from the text. I also discuss the implications of introducing a temporal dimension into the model.

A.1 Setup

A game is played in a network of firms $\langle N, \mathbf{G} \rangle$: by adopting standard notation, N is the number of players (firms) and $\mathbf{G}$ is the *adjacency matrix* that describes their linkages. In particular, g_{ij} represents the strength of the connection between i and j and $g_{ii} = 0$ for all i. For simplicity, assume that the matrix is symmetric ($g_{ij} = g_{ji}$, the network is *undirected*) but this assumption does not qualitatively affect the analysis. Firms are characterized by a Cobb-Douglas production function, and the externality that defines the game is the fact that firm i's output depends not only on K standard inputs $\{X_{i1}, \dots, X_{iK}\}$ as well as its own R&D S_i, but also of the R&D performed by other firms in the network: $\mathbf{S}_{-i} = (S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_N)$. Specifically, the production function reads as:

$$Y_i = A_i \left(\prod_{k=1}^{K} X_{ik}^{\beta_k} \right) S_i^{\gamma} \left(\prod_{j=1}^{N} S_j^{g_{ij}} \right)^{\delta} \tag{A.1}$$

with $\sum_k \beta_k + \gamma + \delta \leq 1$ (constant or decreasing returns to scale). It follows from this functional form assumption that R&D is a strategic complement in this game: one can easily verify that $\partial^2 Y_i / \partial S_i \partial S_j \geq 0$ for a given j, with strict equality if $g_{ij} \neq 0$ and $\delta > 0$.

In equation (A.1), A_i denotes a firm-specific productivity shifter (total factor productivity), further decomposed as

$$A_i = \tilde{A}_i e^{\lambda \omega_i + \varepsilon_i} \tag{A.2}$$

where $\tilde{A}_i$ is a purely deterministic firm-specific component of productivity which is known by all firms in the network, whereas ω_i and ε_i are two unobservable shocks that also affect the outcome. In particular, ω_i is a shock that reflects institutional characteristics that may be common to firms that are connected in the network, with $\lambda \in \mathbb{R}$. In addition, it is assumed that firms are price-takers in their respective perfectly competitive markets (with price/numeraire normalized to one). Denote as $\{\xi_1, \ldots, \xi_K\}$ the vector that collects the linear unitary costs for each of the K traditional inputs $\{X_{i1}, \ldots, X_{iK}\}$. Also define $\chi_i \equiv \kappa e^{-\omega_i} > 0$ as the cost to invest in an additional unit of R&D. In words, the cost of R&D can depend for a firm from a fixed component κ as well as – negatively – on the unobserved shock ω_i. In practice, the model is introducing the presence of factors, that are common to connected firms, which have a simultaneous, but potentially opposite effect on both the convenience to invest in R&D and firm productivity or profitability.

I characterize the common "shocks" by making distributional assumptions on ω_i and ε_i:

$$\boldsymbol{\omega} \sim \mathcal{N}\left(0, \varsigma^2 \mathbf{I} + \psi \mathbf{G}\right) \tag{A.3}$$

$$\boldsymbol{\varepsilon} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}\right) \tag{A.4}$$

where, without loss of generality, the ratio ς^2/ψ (both are scalars) is large enough that the matrix $\varsigma^2 \mathbf{I} + \psi \mathbf{G}$ is semidefinite positive. In words, condition (A.3) says that if two firms share a connection, it is likely that they also share unobserved advantages (or disadvantages) from performing R&D and possibly on their productivity. If two firms are not connected, this correlation is absent. Notice in what follows that the joint normality assumptions facilitate the analysis but do not bear any qualitative implication. I represent the firm's objective function as:

$$\pi\left(X_{i1}, \ldots, X_{iK}; \mathbf{S}\right) = \tilde{A}_i \left(\prod_{k=1}^{K} X_{ik}^{\beta_k}\right) S_i^{\gamma} \left(\prod_{j=1}^{N} S_j^{g_{ij}}\right)^{\delta} e^{\lambda \omega_i + \varepsilon_i} - \sum_{K=1}^{k} \xi_k X_{ik} - \kappa e^{-\omega_i} S_i \tag{A.5}$$

where $\mathbf{S} = S_i \times \cdots \times S_N$. In what follows, I define and solve the game that results from the joint maximization of inputs and R&D decisions by all firms. The resulting equilibrium can be thought as the partial equilibrium outcome of this stylized economy. The game is described as one of incomplete information, where types are defined over the set $\Omega = \omega_i \times \cdots \times \omega_N$. While the overall distribution function from which types are drawn is common knowledge, a firm i only knows the realization of its own draw i. The sequence of the game is the following:

1. Nature draws $\boldsymbol{\omega}$ from (A.3);
2. each firm observes its own realization ω_i, but not those of other firms;
3. then, each firm makes a choice of the strategic variable S_i;
4. Nature draws $\boldsymbol{\varepsilon}$ from (A.4);
5. payoffs are paid out.

A.2 Solution

I characterize the game's solution by invoking the Bayes-Nash equilibrium concept. Therefore, an equilibrium profile of strategies $(S_1^*, \ldots, S_N^*)$ is such that, given optimal choices of the inputs $\{X^*_{ik}\}$

$$\mathbb{E}_{\Omega_{-i}}\pi_i\left[S^*_i,\mathbf{S}^*_{-i};\{X^*_{ik}\}\middle|\,\omega_i\right]\geq\mathbb{E}_{\Omega_{-i}}\pi_i\left[S_i,\mathbf{S}^*_{-i};\{X^*_{ik}\}\middle|\,\omega_i\right]\ \ \forall S_i\neq S^*_i$$

In order to determine this solution, it is convenient to calculate the Nash Equilibrium $\left(\bar{S}_1,\ldots,\bar{S}_N\right)$ that would occur under the case of complete information, that is every firm observes $\boldsymbol{\omega}$ entirely. Since the vector of equilibrium logarithmic R&D choices is, under complete information, a linear function of $\boldsymbol{\omega}$, and since non-private values ω_j affect the playoffs of each firm i only via their indirect effect on firm j's equilibrium R&D S_j, then the combination of conditional expected functions – one for every firm i – of this hypothetical Nash solution does indeed constitute the Bayes-Nash equilibrium of the game.

Pretend then that the game has complete information. Every firm maximizes the expected value, relative to ε_i, of (A.5); conditioning to the CNE profile of other players' strategies $\bar{\mathbf{S}}_{-i}$. The First Order Conditions relative to S_i and $\{X_{i1},\ldots,X_{iK}\}$ are:

$$\frac{\partial\pi_i}{\partial S_i}=\gamma\tilde{A}_i\left(\prod_{k=1}^{K}\bar{X}_{ik}^{\beta_k}\right)\left(\bar{S}_i\right)^{\gamma-1}\left(\prod_{j=1}^{N}\bar{S}_j^{g_{ij}}\right)^{\delta}e^{\lambda\omega_i+\frac{\sigma^2}{2}}-\kappa e^{-\omega_i}=0 \tag{A.6}$$

$$\frac{\partial\pi_i}{\partial X_{iq}}=\beta_q\tilde{A}_i\left(\prod_{k\neq q}^{K}\bar{X}_{ik}^{\beta_k}\right)\bar{X}_{iq}^{\beta_q-1}\left(\bar{S}_i\right)^{\gamma}\left(\prod_{j=1}^{N}\bar{S}_j^{g_{ij}}\right)^{\delta}e^{\lambda\omega_i+\frac{\sigma^2}{2}}-\xi_q=0 \tag{A.7}$$

with (A.7) taken for $q=1,\ldots,K$. The two conditions imply that for any q, $\bar{X}_{qi}=\frac{\beta_q\kappa}{\gamma\xi_q}\bar{S}_ie^{-\omega_i}$: finding the optimal profile of R&D decisions allows to easily calculate every input variable's value in equilibrium. Substitute these expressions into (A.6) and rearrange terms:

$$\bar{S}_i^{\left(1-\sum_{k=1}^{K}\beta_k-\gamma\right)}\left(\prod_{j=1}^{N}\bar{S}_j^{g_{ij}}\right)^{-\delta}=\tilde{A}_i\frac{\gamma}{\kappa}\left[\prod_{k=1}^{K}\left(\frac{\beta_k\kappa}{\gamma\xi_k}\right)^{\beta_k}\right]e^{\left(1-\sum_{k=1}^{K}\beta_k+\lambda\right)\omega_i+\frac{\sigma^2}{2}} \tag{A.8}$$

which, by taking logarithms, becomes

$$\log\bar{S}_i-\tilde{\delta}\sum_{j=1}^{N}g_{ij}\log\bar{S}_j=\tilde{\alpha}_i+\sum_{k=1}^{K}\tilde{\beta}_k\log\left(\frac{\beta_k\kappa}{\gamma\xi_k}\right)+\phi\left(\log\frac{\gamma}{\kappa}+\frac{\sigma^2}{2}\right)+\left(\tilde{\rho}+\tilde{\lambda}\right)\omega_i \tag{A.9}$$

with the following reparametrizations being applied.

$$\begin{aligned}
\phi &\equiv \frac{1}{1-\sum_{k=1}^{K}\beta_k-\gamma}>1\\
\tilde{\alpha}_i &\equiv \phi\log\tilde{A}_i\\
\tilde{\beta}_k &\equiv \phi\beta_k\\
\tilde{\delta} &\equiv \phi\delta\\
\tilde{\rho} &\equiv \phi\left(1-\sum_{k=1}^{K}\beta_k\right)>0\\
\tilde{\lambda} &\equiv \phi\lambda
\end{aligned}$$

Notice that the two combined parameters $\tilde{\rho}$ and $\tilde{\lambda}$ express two different kinds of dependence of the R&D choice from the shock ω_i. While $\tilde{\rho}$ measures the semi-elasticity of R&D that is due to more favorable conditions from doing R&D (what I have modeled as a shock to the cost component), $\tilde{\lambda}$ is the component of that semi-elasticity that is due to the shock entering directly into the production function. If more favorable technological opportunities improve productivity even in absence of R&D, the return of R&D itself is higher and firms would want to invest more conditional on costs.

Assume without loss of generality that $\tilde{\alpha}_i + \sum_{k=1}^{K} \tilde{\beta}_k \log\left(\frac{\beta_k \kappa}{\gamma \xi_k}\right) + \phi\left(\log \frac{\gamma}{\kappa} + \frac{\sigma^2}{2}\right) = 1$ for every firm i. One can represent the N conditions expressed by (A.9) in matrix form:

$$\left(\mathbf{I} - \tilde{\delta}\mathbf{G}\right)\bar{\mathbf{s}} = \iota + \left(\tilde{\rho} + \tilde{\lambda}\right)\boldsymbol{\omega} \tag{A.10}$$

where $\bar{\mathbf{s}} = \left(\log \bar{S}_1, \ldots, \log \bar{S}_N\right)'$. As long as $\det\left|\mathbf{I} - \tilde{\delta}\mathbf{G}\right| \neq 0$ (which happens almost surely), one can express in compact form the corresponding Nash Equilibrium of the game as a function of the inverse matrix $\mathbf{F} \equiv \left(\mathbf{I} - \tilde{\delta}\mathbf{G}\right)^{-1}$.

$$\bar{\mathbf{s}} = \mathbf{F}\left[\iota + \left(\tilde{\rho} + \tilde{\lambda}\right)\boldsymbol{\omega}\right] \tag{A.11}$$

In words, the CNE strategy of each firm i is a function of all the shocks ω_j received by all the other firms in the network. Each of these shocks is weighted by f_{ij}, itself a function of $\tilde{\delta}$, the ij-th element of matrix $\mathbf{F}$. This is a measure of the relative importance, for i, of j's R&D in terms of strategic dependence, and it is related to the concept of *eigenvector centrality* in networks. Intuitively, even if i and j are not directly connected, the choices of j may affect its neighbors – who are direct or indirect neighbors of i – and thus cause an equilibrium response of i's R&D because of strategic complementarities.

However, firms have generally limited information about the shocks received by other partners or competitors. Under the maintained assumption that they can observe only their own ω_i, the Bayes-Nash equilibrium of the incomplete information game (expressed in logarithms) $\mathbf{s}^* = (\log S_1^*, \ldots, \log S_N^*)'$ is found by taking the expected value of every row of (A.11) conditioning on each firm's information set.

$$\begin{aligned}
\log S_i^* &= \mathbb{E}_{\Omega_{-i}}\left[\log \bar{S}_i \middle| \omega_i\right] \\
&= \sum_{j=1}^{N} f_{ij} + \left(\tilde{\rho} + \tilde{\lambda}\right) \sum_{j=1}^{N} f_{ij} \mathbb{E}\left[\omega_j \middle| \omega_i\right] \\
&= \sum_{j=1}^{N} f_{ij} + \left(\tilde{\rho} + \tilde{\lambda}\right) \sum_{j=1}^{N} f_{ij} \frac{\mathbb{C}\text{ov}\left(\omega_i, \omega_j\right)}{\mathbb{V}\text{ar}\left(\omega_i\right)} \omega_i \\
&= \sum_{j=1}^{N} f_{ij} + \left(\tilde{\rho} + \tilde{\lambda}\right) \sum_{j \neq i}^{N} f_{ij} g_{ij} \frac{\psi}{\varsigma^2} \omega_i + \left(\tilde{\rho} + \tilde{\lambda}\right) f_{ii} \omega_i
\end{aligned} \tag{A.12}$$

Since firms cannot predict the shock received by other firms to which they are not connected, these do not affect their R&D choices. Firms anticipate, however, that their direct neighbors may be subjected to similar conditions as their own observation of ω_i, with R&D accordingly responding in equilibrium. Thanks to strategic complementarities, this results in an amplification of the response of firm i to their own shock ω_i. Expression (A.12) can be decomposed into an "idiosyncratic response" to ω_i and an effect of "amplification" of the shock.

$$\log S_i^* = \sum_{j=1}^{N} f_{ij} + \underbrace{\left(\tilde{\rho}+\tilde{\lambda}\right)\sum_{j\neq i}^{N} f_{ij}g_{ij}\frac{\psi}{\varsigma^2}\omega_i}_{\text{amplification}} + \underbrace{\left(\tilde{\rho}+\tilde{\lambda}\right) f_{ii}\omega_i}_{\text{idiosyncratic response}}$$

By defining $d_i \equiv \left(\tilde{\rho}+\tilde{\lambda}\right)\left(\sum_{j\neq i}^{N} f_{ij}g_{ij}\psi\varsigma^{-2} + f_{ii}\right)$, one draws the final conclusions on the stochastic properties of the equilibrium through the network:

$$\mathbb{C}\text{ov}\left(\log S_i^*, \log S_j^*\right) = \begin{cases} (d_i\sigma)^2 & \text{if } i = j \\ \psi d_i d_j g_{ij} & \text{if } i \neq j \text{ and } g_{ij} \neq 0 \\ 0 & \text{if } i \neq j \text{ and } g_{ij} = 0 \end{cases} \tag{A.13}$$

(the latter two cases could be collapsed, but are distinguished for illustrative purposes). In words, the strategic choices of two firms that are connected in a network correlate *only* thanks to the common shocks. The lack of complete information – the fact that firms are not omniscient – impedes that correlation arises between far-away firms thanks to indirect strategic dependence. An extension of this result is that also the other inputs' choices are uncorrelated with the R&D of indirect neighbors. Recall that in the CNE, $\bar{X}_{iq} = h\left(\bar{S}_i, \omega_i\right)$ which implies that in the Bayes-Nash equilibrium, $X_{iq}^* = h\left(S_i^*, \omega_i\right)$.

The properties expressed in (A.13) correspond to Lemmas 1-2 in the text, for the simpler model case. They are the rationale of the IV strategy employed in this paper, motivating conditions (1.22) and (1.23) for indirect connections of second degree. The strategy relies on two assumptions being true: that (A.3) is a good description of the underlying variance of the shocks in the network, and that the information set on which firms base their strategic consideration is limited to their own observation of ω_i. The analogous of Lemma 3 from the text is the consideration that

$$\lambda\mathbb{C}\text{ov}\left(\log S_j^*, \omega_i\right) \propto \lambda\left(1 - \sum_{k=1}^{K}\beta_k + \lambda\right) \tag{A.14}$$

which is negative if and only if $\sum_{k=1}^{K}\beta_k - 1 < \lambda < 0$; that is for small negative values of λ.

A.3 Higher-order correlation

Suppose that the actual description of the data generation process of ω_i is

$$\boldsymbol{\omega} \sim \mathcal{N}\left(0, \varsigma^2\mathbf{I} + \psi_1\mathbf{G} + \psi_2\mathbf{G}^2\right) \tag{A.15}$$

so that also the common shocks of indirect friends are correlated. In this case the Bayes-Nash equilibrium choice of R&D would read as

$$\log S_i^* = \sum_{j=1}^{N} f_{ij} + \left(\tilde{\rho} + \tilde{\lambda}\right) \sum_{j \neq i}^{N} f_{ij} \left[g_{ij} \frac{\psi_1}{\varsigma^2} + \left(g^2\right)_{ij} \frac{\psi_2}{\varsigma^2} \right] \omega_i + \left(\tilde{\rho} + \tilde{\lambda}\right) f_{ii} \omega_i \tag{A.16}$$

where $\left(g^2\right)_{ij}$ is the ij-th element of matrix $\mathbf{G}^2$. In this case, the choices of R&D of indirect friends are correlated to each other – because their underlying shocks are.

This would make the 2SLS estimates of the production function's parameters inconsistent, although the total asymptotic bias of $\hat{\delta}$ can be shown to be smaller than the OLS one. Since indirect friends of third degree would have, in equilibrium, uncorrelated choices of R&D, they can form the basis for an alternative instrument. Their R&D is still correlated to the one of the very connected neighbors thanks to (A.15). Analogous results apply if the underlying correlation extends to even higher orders d, in which case the appropriate instrument is based on indirect connections of degree $d+1$. If ψ_d is small, however, it can be worthwile to trade off some small bias in exchange for a stronger power of a lower order instrument.

A.4 Extended information set

Suppose instead that the true data generation process for ω_i is still (A.3). However, firms are also able observe the realization of ω_j of their direct connections, but not of other firms. Presumably, connected inventors would give access to different types of information. In this case, the Bayes-Nash equilibrium would be given by

$$\log S_i^* = \sum_{j=1}^{N} f_{ij} + \left(\tilde{\rho} + \tilde{\lambda}\right) \sum_{j \neq i}^{N} f_{ij} \mathbb{E} \left[\omega_j \middle| \, \omega_i; \left\{\omega_{k \neq i,j}\right\}_{k: g_{ik} \neq 0} \right] + \left(\tilde{\rho} + \tilde{\lambda}\right) f_{ii} \omega_i \tag{A.17}$$

where the expression inside the expected value operator is a complex function of the network and of the information set. The intuition is, however, simple: in equilibrium, firms still attempt to internalize the shocks received by their indirect friends. They do so by formulating predictions on the basis of the observed set of $\{\omega_j\}$. However large this effect is on the equilibrium, it also makes – in line of principle – the 2SLS estimates based on 2nd degree indirect friends inconsistent. An instrument based on the 3rd degree is instead adequate: firms cannot predict the shocks received by 3rd-degree distant nodes in the network, while the equilibrium strategies of the latter are still correlated to the ones of direct connections.

A.5 General Case, Proofs of Propositions 1-2

I now outline the properties of the model under general assumptions on the cross-correlation of the shocks and the informational structure of the game: this allows to prove Propositions 1 and 2 from the text. In order to formally characterize the result, it is convenient to introduce two "General Assumptions"; that is, two constant properties of the model that are flexible in some specific details (such as the spatial extent of exogenous cross-correlation). Like in text, I use the notation d_{ij} to indicate the degree of distance between firms i and j.

General Assumption 1. *The cross-correlation in the shocks extends up to C degrees of distance, that is*

$$\boldsymbol{\omega} \sim \mathcal{N}\left(0, \varsigma^2 \boldsymbol{I} + \psi_1 \boldsymbol{G} + \cdots + \psi_C \boldsymbol{G}^C\right) \tag{A.18}$$

and thus $\mathbb{C}\mathrm{ov}\left(\omega_i, \omega_j\right) = 0$ if $d_{ij} > C$.

General Assumption 2. *In stage 2 of the game, the set $\boldsymbol{\Omega}_i$ of shocks ω_j observed by firm i does not include firms that are located at distances higher than L. Formally, the following property applies to the set $\boldsymbol{\Omega}_i = \{\omega_j : \omega_j$ is observed by firm $i\}$.*

$$d_{ij} > L \Rightarrow \omega_j \notin \boldsymbol{\Omega}_i \tag{A.19}$$

The equilibrium profile of R&D investment choices is, for every firm i, as follows.

$$\log S_i^* = \sum_{j=1}^{N} f_{ij} + \left(\tilde{\rho} + \tilde{\lambda}\right) \sum_{j=1}^{N} f_{ij} \mathbb{E}\left[\omega_j \middle| \boldsymbol{\Omega}_i\right] \tag{A.20}$$

Thus, one can write $\log S_i^* = g\left(\boldsymbol{\Omega}_i\right)$. Propositions 1 and 2 with their relative proofs follow.

Proposition 1. $\mathbb{C}\mathrm{ov}\left(\omega_i, \log S_j^*\right) = 0$ *if* $d_{ij} > C + L$; $\mathbb{C}\mathrm{ov}\left(\log S_i^*, \log S_j^*\right) = 0$ *if* $d_{ij} > C + 2L$.

Proof. For any two firms i and j such that $d_{ij} = D > C + L$, take any of their shortest paths of length D. Order the intermediate connections along the path: $\ell = 0, \ldots, D$ where (without loss of generality) $i = 0$ and $j = D$. By General Assumption 2 and the definition of path in a network, $\omega_\ell \notin \boldsymbol{\Omega}_j$ if $\ell < L$. Thus, the shortest path connecting ω_i with any element $Q \in \boldsymbol{\Omega}_j$ has length $D - L$. Since $D - L > C$, their correlation is zero by General Assumption 1, implying $\mathbb{C}\mathrm{ov}\left(\omega_i, \log S_j^*\right) = 0$ because of equation (A.20). If this is true for the shortest path connecting i and j, so it is for any other path. By analogous reasoning, suppose that $d_{ij} = D > C + 2L$, and take the shortest path between i and j as defined earlier. In addition to the considerations above, $\omega_\ell \notin \boldsymbol{\Omega}_i$ if $\ell > L$, hence the shortest path connecting any element $P \in \boldsymbol{\Omega}_i$ with another element $Q \in \boldsymbol{\Omega}_j$ has length $D - 2L > C$. Hence, firms i and j are in equilibrium functions of mutually independent sets of random variables, which completes the proof.

Proposition 2. $\mathbb{C}\mathrm{ov}\left(\log X_{ik}^*, \log S_j^*\right) = 0$ *if* $d_{ij} \geq C + 2L$ *for* $k = 1, \ldots, K$.

Proof. Recall from the First Order Conditions that in equilibrium, $\bar{X}_{ki}^* = \frac{\beta_k \kappa_i}{\gamma \xi_k} S_i^* e^{-\omega_i}$. It follows that $\log \bar{X}_{ki}^* = \log \frac{\beta_k \kappa_i}{\gamma \xi_k} + \log S_i^* - \omega_i = \log \frac{\beta_k \kappa_i}{\gamma \xi_k} + g\left(\boldsymbol{\Omega}_i\right) - \omega_i$. Since ω_i is listed in $\boldsymbol{\Omega}_i$, by an analysis similar to the one conducted to demonstrate Proposition 1 one can conclude that $\log X_{ik}^*$ and $\log S_j^*$ are independent, which proves the statement.

A.6 Dynamic Model

Suppose that interest lies in a model with two time periods $t = 0, 1$:

$$Y_{it} = \tilde{A}_i \left(\prod_{k=1}^{K} X_{k_{it}}^{\beta_k} \right) S_{it}^{\gamma} \left(\prod_{j=1}^{N} S_{jt}^{g_{(ij)t}} \right)^{\delta} e^{\lambda \omega_{it} + \varepsilon_{it}} \tag{A.21}$$

$$S_{it} = \sum_{s=0}^{\infty} \theta^s R_{i(t-s)} \tag{A.22}$$

where S_{it} is the *stock* of R&D, R_{it} is the *flow*, while θ is the stock's depreciation parameter. The main results from the static version of the model transfer to the stock variable S_{it} of the dynamic model as well. In $t = 0$, the analysis is identical to the one in the one-shot game, with $R_{i0} = S_{i0}$. In $t = 1$ instead, the firms' objective function reads as follows.

$$Y_{i1} = \tilde{A}_i \left(\prod_{k=1}^{K} X_{k_{i1}}^{\beta_k} \right) (\theta R_{i0} + R_{i1})^{\gamma} \left(\prod_{j=1}^{N} (\theta R_{j0} + R_{j1})^{g_{(ij)1}} \right)^{\delta} e^{\lambda \omega_{i1} + \varepsilon_{i1}} \tag{A.23}$$

In $t = 1$ firms choose R_{i1} and not S_{i1}; the analog of the F.O.C. (A.6) here is

$$\frac{\partial \pi_{i1}}{\partial R_{i1}} = \gamma \tilde{A}_{i1} \left(\prod_{k=1}^{K} \bar{X}_{k_{i1}}^{\beta_k} \right) \left(\theta R_{i0} + \bar{R}_{i1} \right)^{\gamma - 1} \left(\prod_{j=1}^{N} \left(R_{i0} + \bar{R}_{i1} \right)^{g_{(ij)1}} \right)^{\delta} e^{\lambda \omega_{i1} + \frac{\sigma^2}{2}} - \kappa_i e^{-\omega_{i1}} = 0 \tag{A.24}$$

which is also solved as in the standard case, finding the equilibrium value of log $(\theta R_{i0} + R_{i1}^*)$ first and then determining R_{i1}^* residually (notice that the marginal product of R&D does not differ across the two models). The equilibrium cross-covariance of R&D flows in the network retains the same properties as in the static case, and for R&D stocks it reads as:

$$\mathbb{C}\text{ov}\left[S_{i1}, S_{j1}\right] = \mathbb{C}\text{ov}\left[R_{i1}, R_{j1}\right] + \theta \mathbb{C}\text{ov}\left[R_{i1}, R_{j0}\right] + \theta \mathbb{C}\text{ov}\left[R_{i0}, R_{j1}\right] + \theta^2 \mathbb{C}\text{ov}\left[R_{i0}, R_{j0}\right]$$

there is no reason why the cross-correlation should extend to higher distances in the network.

This intuition can be generalized to longer time horizons and does not require any assumption on the autocorrelation of the shock ω_{it}. It does require, however, that connections $g_{(ij)t}$ are not severed (but connections are otherwise allowed to appear and vary in intensity over time). To see why, suppose that firms i, j and k are all connected at $t = 0$. As a result, all their R&D variables are correlated with each other. If one link is lost, say between i and k, their R&D stocks are still correlated (because of past choices) and annot serve as reciprocal instruments for j. This is a minor concern in the case of networks, like the one analyzed in this work, which tend to tighten and become denser over time.

Appendix B

Properties of the GMM Model

B.1 Proof of Proposition 4 from Chapter 2

Proposition 4. *Let s_i^* be given by equation 2.17 and Hypothesis 3 hold. Then, the moments given in equations 2.18 and 2.19 identify $(\vartheta, \mu(\cdot))$.*

Proof. Consider two agents, 1 and 2, whose degree of connection is 3 or higher, *i.e.* $g_{12} = 0$ and $\sum_{h=1}^{N} g_{1h} g_{h2} = 0$. For notational simplicity, assume $dim(\mathbf{x}_i) = 1$. Following Roehrig (1988), define f and f^* as

$$f = \begin{bmatrix} s_1^* - \vartheta \sum_{h=1}^{N} g_{1h} - \mu(\mathbf{x}_1) - \eta_1 \\ \left(s_1^* - \vartheta \sum_{h=1}^{N} g_{1h} - \mu(\mathbf{x}_1)\right)\left(s_2^* - \vartheta \sum_{h=1}^{N} g_{2h} - \mu(\mathbf{x}_2)\right) \end{bmatrix}$$

$$f^* = \begin{bmatrix} s_1^* - \vartheta^* \sum_{h=1}^{N} g_{1h} - \mu^*(\mathbf{x}_1) - \eta_1^* \\ \left(s_1^* - \vartheta^* \sum_{h=1}^{N} g_{1h} - \mu^*(\mathbf{x}_1)\right)\left(s_2^* - \vartheta^* \sum_{h=1}^{N} g_{2h} - \mu^*(\mathbf{x}_2)\right) - \varepsilon_{12}^* \end{bmatrix}$$

where ε_{12} is the residual from the conditional covariance equation, which fundamentally depends on the correlation of the vector η. Here, matrices N_1 and N_2 from Roehrig have a correspondence in the following two matrices:

$$N_1 = \begin{bmatrix} -\mu^{*\prime}(\mathbf{x}_1) & 0 & 1 & 0 & -\vartheta^* g_{13} & \dots & -\vartheta^* g_{1N} \\ -\mu'(\mathbf{x}_1) & 0 & 1 & 0 & -\vartheta g_{13} & \dots & -\vartheta g_{1N} \\ -\mu'(\mathbf{x}_1)\eta_3 & -\mu'(\mathbf{x}_3) & \eta_3 & \eta_1 & -\vartheta(g_{13}\eta_2 + g_{23}\eta_1) & \dots & -\vartheta(g_{1N}\eta_2 + g_{2N}\eta_1) \end{bmatrix}$$

$$N_2 = \begin{bmatrix} -\mu^{*\prime}(\mathbf{x}_1)\eta_3^* & -\mu^{*\prime}(\mathbf{x}_3) & \eta_3^* & \eta_1^* & -\vartheta^*(g_{13}\eta_2^* + g_{23}\eta_1^*) & \dots & -\vartheta^*(g_{1N}\eta_2^* + g_{2N}\eta_1^*) \\ -\mu'(\mathbf{x}_1) & 0 & 1 & 0 & -\vartheta g_{13} & \dots & -\vartheta g_{1N} \\ -\mu'(\mathbf{x}_1)\eta_3 & -\mu'(\mathbf{x}_3) & \eta_3 & \eta_1 & -\vartheta(g_{13}\eta_2 + g_{23}\eta_1) & \dots & -\vartheta(g_{1N}\eta_2 + g_{2N}\eta_1) \end{bmatrix}$$

given that $rank(N_i) < G + 1 = 3$ for $i = 1, 2$, with some algebra one can show that $\mu^{*\prime}(\cdot) = \mu'(\cdot)$ and $\vartheta^* = \vartheta$. Therefore, Condition 3.2 holds, and by Theorem 3.1 in Roehrig, $(\vartheta, \mu(\cdot))$ are identified.

B.2 Asymptotic Normality of $\hat{\boldsymbol{\theta}}$

I show here that $\sqrt{T_N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\boldsymbol{\theta}}\right)$, where

$$V_{\boldsymbol{\theta}} = \left(M_0' W_0 M_0\right)^{-1} M_0' W_0 V_0 W_0 M_0 \left(M_0' W_0 M_0\right)^{-1} \tag{B.1}$$

$$W_0 \equiv \lim_{N\to\infty} W_{T_N} \tag{B.2}$$

$$M_0 \equiv \lim_{N\to\infty} \frac{1}{T_N} \sum_{i=1}^{T_N} \frac{\partial m_i\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\theta}'} \tag{B.3}$$

$$V_0 \equiv \lim_{N\to\infty} \frac{1}{T_N} \sum_{i=1}^{T_N} \sum_{i=1}^{T_N} \mathbb{E}\left[m_i\left(\boldsymbol{\theta}_0\right) m_j\left(\boldsymbol{\theta}_0\right)\right] \tag{B.4}$$

Before proceeding, it is useful to define the following two indicators

$$q_{ij} = \mathbb{I}\left[g_{ij} > 0\right]$$
$$\tilde{q}_{ij} = q_{ij} + \mathbb{I}\left[i = j\right]$$

as well as the following, intuitive graph-theoretic result.

Lemma B.1. *Let triads T_1 and T_2 be given by nodes A_i, B_i, and C_i, $i = 1, 2$, where the central node is B_i. Define the distance between triads T_1 and T_2, $d\left(T_1, T_2\right)$, as the minimum distance existing between any two nodes from each of the triads. Then, it holds the implication that $d\left(T_1, T_2\right) \leq 2 \Rightarrow d\left(B_1, B_2\right) \leq 4$.*

Proof. Suppose $d\left(T_1, T_2\right) \geq 2$. Then, $\max\left\{d\left(A_1, A_2\right), d\left(A_1, C_2\right), d\left(C_1, A_2\right), d\left(C_1, C_2\right)\right\} \leq 2$ and $\max\left\{d\left(B_1, A_2\right), d\left(B_1, C_2\right), d\left(A_1, B_2\right), d\left(C_1, B_2\right)\right\} \leq 3$, and therefore $d\left(B_1, B_2\right) \leq 4$.

In addition, let $m\left(1\left(i\right), 2\left(i\right), 3\left(i\right), \boldsymbol{\theta}\right) \equiv m_i\left(\boldsymbol{\theta}\right)$ and make the following assumption.

Assumption 4. *The covariance of $m_i\left(\boldsymbol{\theta}\right)$ is bounded.*

$$\mathbb{E}\left[m\left(k, h, l, \boldsymbol{\theta}\right) m\left(k', h', l', \boldsymbol{\theta}\right)\right] < \infty$$

To show that V_0 is well defined, it is convenient to rewrite the variance of $\overline{m}_{T_N}\left(\boldsymbol{\theta}\right)$ in terms of the number of nodes.

$$Var\left(\overline{m}_{T_N}\left(\boldsymbol{\theta}\right)\right) = Var\left(\sum_{h=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} q_{hk} q_{hl}\left(1-\tilde{q}_{kl}\right) m\left(k, h, l, \boldsymbol{\theta}\right)\right)$$
$$= \sum_{h=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N}\sum_{h'=1}^{N}\sum_{k'=1}^{N}\sum_{l'=1}^{N} q_{hk} q_{hl}\left(1-\tilde{q}_{kl}\right) q_{h'k'} q_{h'l'}\left(1-\tilde{q}_{k'l'}\right) \mathbb{E}\left[m\left(k, h, l, \boldsymbol{\theta}\right) m\left(k', h', l', \boldsymbol{\theta}\right)\right]$$

It is necessary to show that this sum over N^6 terms grows at most at a rate N. Because of Assumption 4 it suffices to show that the number of nonzero terms in the preceding sum does not grow faster than N. Consider the case of node-specific exogenous dependence and two open triad respectively composed of nodes (A,B,C) and (D,E,F). In order to have a nonzero term it is required that either: *i*) a node of the first triad is also a node in the second triad; *ii*) a node of the first triad is connected to a node in the second triad; *iii*) a node of the first triad is connected to another node, which is itself connected to a node in the second triad. If the minimum distance between any two nodes from the two triads exceeds 2, then $\mathbb{E}\left[m\left(k,h,l,\boldsymbol{\theta}\right)m\left(k',h',l',\boldsymbol{\theta}\right)\right]=0$. Therefore, the number of pairs of open triads with nonzero covariance terms, defined here as $\tilde{N}$, equals

$$\begin{aligned}\tilde{N}=&\sum_{h=1}^{N}\sum_{k=1}^{N}\sum_{l=k+1}^{N}\sum_{h'=1}^{N}\sum_{k'=1}^{N}\sum_{l'=k'+1}^{N}q_{hk}q_{hl}\left(1-\tilde{q}_{kl}\right)q_{h'k'}q_{h'l'}\left(1-\tilde{q}_{k'l'}\right)\\&\cdot\mathbb{I}\left\{\mathbb{I}\left(h=h'\right)+\mathbb{I}\left(h=k'\right)+\mathbb{I}\left(h=l'\right)+\mathbb{I}\left(k=h'\right)+\mathbb{I}\left(k=k'\right)+\mathbb{I}\left(k=l'\right)\right.\\&+\mathbb{I}\left(l=h'\right)+\mathbb{I}\left(l=k'\right)+\mathbb{I}\left(l=l'\right)\\&+\sum_{m=1}^{N}\left(q_{h'm}+q_{k'm}+q_{l'm}\right)+\sum_{m'=1}^{N}\left(q_{hm'}+q_{km'}+q_{lm'}\right)\\&\left.+\sum_{m=1}^{N}\sum_{n=1}^{N}\left(q_{h'm}q_{mn}+q_{k'm}q_{mn}+q_{l'm}q_{mn}\right)\sum_{m'=1}^{N}\sum_{n'=1}^{N}\left(q_{hm'}q_{m'n'}+q_{km'}q_{m'n'}+q_{lm'}q_{m'n'}\right)\right\}\end{aligned}$$

notice that there is no double counting (i.e. (A,B,C) and (C,B,A) are counted only once).

Term $\tilde{N}$ is bounded by a multiple of N. In order to see this notice that, for each node, the maximum number of open triads for which it can potentially be the central one is, by Assumption 2, $\binom{G^*}{2}$. By Lemma B.1, if the distance between two triads is not larger than 2, then the distance between their two central nodes is not larger than 4. This fact, combined with Assumption 2, ensures that the maximum number of nodes at distance not larger than 4 and belonging to sufficiently close open triads equals $1+G^*+G^{*2}+G^{*3}+G^{*4}$. Hence, for each node that is central to any open triad, the maximum number of open triads that can potentially be within distance 2 is never larger than $\binom{G^*}{2}\left(1+G^*+G^{*2}+G^{*3}+G^{*4}\right)$. By aggregating over all nodes, it follows that

$$\tilde{N}\leq\binom{G^*}{2}\left(1+G^*+G^{*2}+G^{*3}+G^{*4}\right)N=O\left(N\right)$$

Consequently, it is immediate to show that $Var\left(\overline{m}_{T_N}\left(\boldsymbol{\theta}\right)\right)=O\left(T_N^{-1}\right)$, and therefore $V_0<\infty$. Showing $M_0<\infty$ straightforward under Assumption 2. Importantly, this result is obtained under the hypothesis of node-specific common-shocks. If there is no exogenous dependence or if it is edge-specific instead, then the number of nonzero covariance terms is even smaller; hence the result of finite asymptotic variance trivially holds in both cases.

B.3 Computation of the Standard Errors

The analytically appropriate estimation of the Variance-Covariance matrix of the estimated parameters $\hat{\boldsymbol{\theta}}$ would depend on the hypothesis made on the data generation process driving $\boldsymbol{\omega}$. Consider for example the case of node-specific shocks: let T_N denote the number of open triads and W_{T_N} the weighting matrix used in the estimation of $\hat{\boldsymbol{\theta}}$. The asymptotic variance would be estimated as

$$\hat{V}_{\boldsymbol{\theta}} \equiv \left(\hat{M}'_{T_N} W_{T_N} \hat{M}_{T_N}\right)^{-1} \hat{M}'_{T_N} W_{T_N} \hat{V}_{T_N} W_{T_N} \hat{M}_{T_N} \left(\hat{M}'_{T_N} W_{T_N} \hat{M}_{T_N}\right)^{-1} \tag{B.5}$$

where

$$\hat{M}_{T_N} \equiv \frac{1}{T_N} \sum_{i=1}^{T_N} \frac{\partial m_i \left(\hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} \tag{B.6}$$

$$\hat{V}_{T_N} \equiv \frac{1}{T_N} \sum_{i=1}^{T_N} \sum_{i=1}^{T_N} m_i \left(\hat{\boldsymbol{\theta}}\right) m_j \left(\hat{\boldsymbol{\theta}}\right)' \mathbb{I}\left[d\left(i,j\right) \leq 2\right] \tag{B.7}$$

and $d(i,j)$ is the minimum distance between any two nodes in the open triads i and j. In words, if the two triads have a node in common, two directly connected nodes, or two nodes indirectly connected at path length 2, the product of their moments computed at $\hat{\boldsymbol{\theta}}$ would be added to the sum in B.7. Analogously, with edge-specific shocks only the moments relative to triads located at distance 1 or lower are multiplied and then added to $\hat{V}_{T_N}$, while in the case of no common shocks the required maximum distance is reduced to 0 (which is the case when triads have at least one node in common).

In practice, in the empirical analysis I adopt a general procedure for the calculation of the standard errors: one which does not depend on the stochastic hypothesis made about $\boldsymbol{\omega}$. This consists in an approximation based on a clustering scheme. It has the advantages, among the others, of computing standard errors conservatively and of maintaining an agnostic stance about the data generation process, thus keeping consistency across models estimated under different hypotheses. This clustering scheme is based on an assignment rule that associates open triads to communities of nodes identified by the Louvain algorithm (see Chapter 1). The assignment rule is summarized as follows. Consider a set of communities $C = \{c_1, \ldots, c_C\}$ and an assignment rule $p : F \to C$ that associates every firm $f = 1, \ldots, N$ from a set F $(f \in F)$ to one specific community. Consider then, from the set of open triads I, any element $i \in I$, composed of three nodes $(1(i), 2(i), 3(i))$ of which $2(i)$ is the central node. The assignment rule $q : I \to C$ associating open triads to communities is based on the most frequent community to which the nodes of that triad are assigned. Ties – those unlikely cases in which all three nodes belong to a different community – are resolved in favor of the central node of the triad. Consequently, the assignment rule takes the following formal representation.

$$q(i) = \begin{cases} p(1(i)) = p(3(i)) & \text{if } p(1(i)) = p(3(i)) \neq p(2(i)) \\ p(2(i)) & \text{otherwise} \end{cases}$$

Armed with a rule that assigns triads to communities, the alternative (approximated) central component of the variance covariance matrix can be calculated as

$$\tilde{V}_{T_N} \equiv \frac{1}{T_N} \sum_{i=1}^{T_N} \sum_{i=1}^{T_N} m_i \left(\hat{\boldsymbol{\theta}}\right) m_j \left(\hat{\boldsymbol{\theta}}\right)' \mathbb{I}\left[q\left(i\right) = q\left(j\right)\right] \tag{B.8}$$

that is, the product of the moments associated to any two triads i and j is added to $\tilde{V}_{T_N}$ if i and j are associated to the same community. This rule is very simple to implement as a clustering method on standard econometric packages. With this procedure, the cross-correlation of triads that are particularly "close" in the network space is typically taken into account for the calculation of the standard errors – so to approximate (B.7) – as long as the original communities are appropriately defined as clusters of tightly connected nodes. If two triads assigned to the same community have orthogonal generalized residuals, the resulting increase in the calculated standard errors is, at worst, negligible: in this respect, the proposed method is conservative. The only downside, which is typical of all clustering schemes (including the one adopted in the reduced form estimates from Chapter 1) is that the cross-correlation between two units of observation – here, open triads – that are assigned to different clusters is not taken into account for the calculation of the standard errors. In order to minimize this problem I employ few, very large communities when implementing the clustering scheme. They are the six communities identified by the application of the Louvain algorithm with Resolution 1; these are displayed in Figure E.6.

Appendix C

Data and Connection Measures

In this appendix I briefly expand on the construction of the dataset and the main variables employed in the analysis, with emphasis on the construction of the connection measures.

C.1 Data

The main panel of firms has been reconstructed by Bloom et al. (2013) (BSV) by selecting firms from COMPUSTAT with at least one entry in the "Segment" complementary dataset. The latter breaks down sales by line of business for specific firms. The main variables employed in the estimation of the production function are constructed according to standard methodologies. In particular, monetary values are deflated using appropriate price indices (for output/sales, 4-digits level price deflators are employed for each line of sales in "Segment"). The stock of R&D is derived from flows using the perpetual inventory method with a 15% depreciation parameter. I refer to the online appendix from BSV for the details.

Firm-level identifiers are matched to patents as per the NBER patent dataset developed until 2006; see Hall et al. (2001) for the details. All the observed patents for each firm i in the entire time interval under analysis are broken down into 426 patent classes defined by the USPTO. Following Jaffe, BSV calculate the $TECH$ weights as the uncentered correlation of two firms' technological allocation of patents:

$$TECH_{ij} = \frac{\left(T_i T_j'\right)}{\left(T_i T_i'\right)^{\frac{1}{2}} \left(T_j T_j'\right)^{\frac{1}{2}}}$$

where $T_i = (T_1, \ldots, T_{426})$ is the vector that collects the shares of patents of each firm across the 426 patent classes. Notice that these weights are constant over time. The Jaffe measure of technological proximity is constructed as the average of all other firms' R&D stock weighted by the $TECH$ measures, $Spilltech_{it} = \sum_j TECH_{ij} S_{jt}$. It enters logarithmically in the estimation of the Cobb-Douglas production function. To facilitate comparisons, I employ the same variables in my estimates, with no variations.

C.2 Connection Measures

To calculate the connection measures, I need information on *i*) the disambiguated identity of all the actual inventors who signed all the patents attributed to the firms, *ii*) their patent coauthorship relationships; *iii*) the time interval in which each inventor is associated to a firm. I obtain information on *i*) and *ii*) thanks to the match of the patent identifiers from the USPTO across the NBER and the HPND datasets. I rely on the work performed by the authors of the HPND dataset for the quality of their disambiguation algorithm, see Li et al. (2014) for details. However, I have no direct information about *iii*). In order to associate individuals to firms, I use indirect information extrapolated from the patent data.

In particular, I can establish to which firm are assigned patents, that are signed by specific individual inventors. By defining the time interval in which every individual is observed to collaborate on patents for a specific firm, I can provide an approximate time interval that defines their mutual association. Define $\underline{p}_{im}$ as the first year when inventor i is observed patenting (application year) for firm m. Similarly, $\overline{p}_{im}$ is the last year. The assignment rule between the inventor and the firm in year t is

$$f_{(mi)t} = \begin{cases} 1 & \text{if } t \in \left[\underline{p}_{im} - 1, \overline{p}_{im} + 1\right] \\ 0 & \text{otherwise} \end{cases}$$

which is extended one year in the past relative to $\underline{p}_{im}$ and one year in the future relative to $\overline{p}_{im}$. This choice is based on the presumption that every collaboration does not begin immediately the year the first patent is being applied for, and does not terminate immediately after the last patent. Clearly, this may miss years in which inventors, while not producing patents, are still part of an organization. This would be relevant (and generate problems of measurement error) especially if these idle inventors were connected to individuals in other firms. However, it is also arguable that idle inventors are not really active in the process of knowledge exchange and creation. Such a restricted time window really captures the size of the R&D-performing team of a firm, whether it is made of regular employees or, say, academic collaborators. It is reassuring that the results are very robust to perturbations in this assignment rule (these results are available upon request).

One can collect all the binary indicators $f_{(mi)t}$ in a matrix $\mathbf{F}_t$ which has N rows (number of firms in the data) and M_t columns (the number of inventors at time t). To calculate the connection measures, one should first obtain the binary and symmetric adjacency matrix $\mathbf{P}_t$ of coauthors at time t. It is a matrix of dimension $M_t \times M_t$ where $p_{(ij)t} = p_{(ji)t} = 1$ if the two inventors i and j have at least one joint patent at $t+1$. Define $\mathcal{B}(\cdot)$ as a boolean operator that applied to matrices, returns other matrices whose entries are equal to 1 for positive corresponding entries in the argument and 0 otherwise. One can easily calculate the asymmetric $N \times N$ matrix that counts the reciprocal connections between inventors across firms at time t:

$$\begin{aligned} \mathbf{K}_t &= \mathbf{F}_t \cdot \mathcal{B}\left(\mathbf{P}_t \mathbf{F}_t'\right) \\ &= \mathcal{B}\left(\mathbf{F}_t \mathbf{P}_t\right) \cdot \mathbf{F}_t' \end{aligned}$$

and obtain the numerator of the expression within parentheses in (1.14) for every pair of firms as $k_{(ij)t} + k_{(ji)t}$. Notice that the diagonal elements of $\mathbf{K}_t$ denote the total number of inventors assigned to one firm in year t. Hence, the denominator of the aforementioned argument of (1.14) can be obtained as $k_{(ii)t} + k_{(jj)t}$. Therefore

$$c^f_{(ij)t} = c^f_{(ji)t} = f\left(\frac{k_{(ij)t} + k_{(ji)t}}{k_{(ii)t} + k_{(jj)t}}\right)$$

and $g_{(ij)t} = \sqrt{\frac{k_{(ij)t}+k_{(ji)t}}{k_{(ii)t}+k_{(jj)t}}}$, per (1.15).

Finally, it is worth mentioning how I compute the matrix $\mathbf{H}$ that collects the weights $h_{(ik)t}$ for the second-degree instrument (the case of the third-degree instrument is analogous):

$$\mathbf{H}_t = \mathbf{G}_t^2 \circ (\mathbf{1} - \mathbb{I}[\mathbf{I} + \mathbf{G}_t])$$

where the symbol $\circ$ denotes the Hadamard (pointwise) multiplication and the indicator function $\mathbb{I}[\cdot]$ is also taken pointwise on its matrix argument. To better see why, consider that for $i \neq k$, $h_{(ik)t} = \sum_{j\neq i} \left(g_{(ij)t} g_{(jk)t}\right) \mathbb{I}\left[g_{(ik)t} = 0\right]$. Therefore the expression for the instrument, that is the i-th entry of the column vector $\mathbf{H}_t \mathbf{s}_t$ (where $\mathbf{s}_t$ is the vector of log R&D) reads as

$$\begin{aligned} \sum_{k\neq i} h_{(ik)t} \log S_{kt} &= \sum_{k\neq i}\left(\sum_{j\neq i}\left(g_{(ij)t} g_{(jk)t}\right)\mathbb{I}\left[g_{(ik)t} = 0\right]\right)\log S_{kt} \\ &= \sum_{j\neq i} g_{(ij)t} \sum_{k\neq i,j} \underbrace{g_{(jk)t}\mathbb{I}\left[g_{(ik)t} = 0\right]}_{=\tilde{h}^i_{(jk)t}} \log S_{kt} \end{aligned}$$

which corresponds to the definition given in (1.24-1.25), as expected.

C.3 Geographic Control and Proximity Measures

An empirical concern of the analysis is that patent coauthorship relationships may simply capture the fact that inventors live close to one another. Therefore, measures of connection might reflect the fact that firms have their R&D labs in the most innovative areas – something that might have a direct impact on innovation and productivity. To clear this concern I calculate a measure of R&D spillovers that is weighted against the relative spatial proximity of two R&D teams. These weights are called *proximity* measures and they are conceptually similar to the connection measures. In lieu of a patent coauthorship relationship, however, two inventors are identified as being "linked" if they are "neighbors" in spatial terms, that is they are observed to patent from the same Core Based Statistical Area (CBSA) in a given year. I obtain this information from patent data, that report the ZIP code of the address of residence of each signing inventor. Proximity measures read as

$$b_{(ij)t} = \frac{(\#\text{inventors of firms } i \text{ and } j \text{ overlapping on the same CBSAs at } t)}{(\#\text{ inv.s of firm } i \text{ at } t) + (\#\text{ inv.s of firm } j \text{ at } t)}$$

and they are calculated with a procedure that is analogous to the one of connection measures. The actual control employed in the regressions is also analogous to the variable of connection-induced spillovers, and it is defined as $Geospills_{it} = \sum_j b_{(ij)t} \log S_{jt}$.

An interesting finding that can be drawn from the calculation of the proximity measures is that there are virtually no patent coauthors that live in the same CBSAs *and* are coauthors in different firms. This is a hint to the fact that much of cross-firms connections are caused by inventor mobility. This fact can well explain why the geography-based control has little statistical and economic significance in most regressions. It is also consistent with the finding from Agrawal et al. (2010) according to whom, in selected R&D intensive cities, the vast majority of the patent citations are observed within the firm and not across firms. It must be stressed, however, that the nature of the data do not allow to make strong conclusion on the importance of local vs. distant networks of inventors: small and private firm, however R&D intensive, are excluded from the sample. Phenomena like the one on which much of the narrative about the Silicon Valley has been built – the story of talented employees leaving a larger firm and launching their own high-tech start-up – cannot be captured by these data.

Appendix D

Alternative Connections Measures

In this appendix I discuss alternative connection measures $g^{alt}_{(ij)t}$ and the estimates obtained from their application to model (1.18). I focus in particular on four alternatives.

1. **Linear Connection**. I use a pure linear connection measure $c_{(ij)t}$ (that is, in (1.14) $f(\cdot)$ is an identity function). This measure does not give disproportionate importance to few connected inventors that are part of two large R&D teams.

2. **Second Degree Connections**. I define "connected" inventors as not simply those individuals who are patent-coauthors of someone in the other firm, but also coauthors of coauthors of someone in the other firm (second-degree coauthors). I take the square root of the corresponding measure, which I call $g^{2dg}_{(ij)t}$. Relative to the baseline, this alternative measure downplays those connected scientists who do not develop many bonds in the firm they are assigned to (occasional inventors).

3. **Asymmetric "Receiving" Connections**. I abandon the framework of directed networks and I consider the possibility that spillover relationships are asymmetric between firms. In particular, I suppose that the degree of a firm's access to the knowledge of another depends only by its own share of connected inventors: $c^{asr}_{(ij)t} = k_{(ij)t}/k_{(ii)t}$. In the estimation, however, I use its square root $g^{asr}_{(ij)t} = \sqrt{c^{asr}_{(ij)t}}$. This measure gives more importance to smaller, well connected firms in the process of ideas exchange.

4. **Asymmetric "Spilling" Connections**. An alternative economic assumption is that spillovers do not depend on active acquisition of knowledge by well-connected firms, but rather by their passive access to naturally leaked information. In this case it would be more advantageous to have access to as many inventors as possible in the "spilling" firm. The connection measure is in this case defined as $g^{ass}_{(ij)t} = \sqrt{c^{ass}_{(ij)t}}$ with $c^{ass}_{(ij)t} = k_{(ji)t}/k_{(jj)t}$. This measure gives more relevance to firms that are well connected to larger ones.

For any of these measures $g^{alt}_{(ij)t}$ the spillover variable is still constructed as $\sum_j g^{alt}_{(ij)t} \log S_{jt}$.

Table D.1: Alternative Connection Measures, Production Function, 1981-2001

	(Linear)		(2nd Degree)	
	OLS	2SLS	OLS	2SLS
Private R&D (γ)	0.0408	0.0411	0.0418	0.0385
	(0.0126)	(0.0121)	(0.0121)	(0.0130)
Spillovers (δ)	0.0825	0.0765	0.0042	0.0079
	(0.0155)	(0.0397)	(0.0012)	(0.0028)
Capital	0.1553	0.1557	0.1542	0.1486
	(0.0211)	(0.0210)	(0.0213)	(0.0213)
Labor	0.6419	0.6415	0.6454	0.6530
	(0.0282)	(0.0276)	(0.0285)	(0.0284)
Jaffe Tech. Proximity	0.0283	0.0311	0.0316	0.0020
	(0.1046)	(0.1092)	(0.1083)	(0.1165)
Fixed Effects	YES	YES	YES	YES
Only Network	YES	YES	YES	YES
Instrument	NO	3rd deg.	NO	3rd deg.
No. of Communities ($\times$ Year)	20	20	20	20
No. of Observations	6914	6914	6914	6914

	(As. Receiving)		(As. Spilling)	
	OLS	2SLS	OLS	2SLS
Private R&D (γ)	0.0429	0.0442	0.0431	0.0403
	(0.0126)	(0.0137)	(0.0124)	(0.0125)
Spillovers (δ)	0.0088	0.0046	0.0057	0.0119
	(0.0020)	(0.0063)	(0.0022)	(0.0030)
Capital	0.1567	0.1586	0.1568	0.1526
	(0.0208)	(0.0214)	(0.0210)	(0.0209)
Labor	0.6398	0.6381	0.6421	0.6483
	(0.0275)	(0.0272)	(0.0277)	(0.0277)
Jaffe Tech. Proximity	0.0345	0.0498	0.0371	0.0059
	(0.1088)	(0.1203)	(0.1094)	(0.1157)
Fixed Effects	YES	YES	YES	YES
Only Network	YES	YES	YES	YES
Instrument	NO	3rd deg.	NO	3rd deg.
No. of Communities ($\times$ Year)	20	20	20	20
No. of Observations	6914	6914	6914	6914

Table D.1 shows the results from the estimation of model (1.18) using these alternative connection measures. For each measure, both the OLS and the 2SLS estimates are shown. All estimates are restricted to the network only, include the full set of fixed effects, and the 2SLS estimates are based on the third degree instrument. Point estimates vary in magnitude because of the rescaling implied by different measures. Noticeably, for the "Second Degree" and the "Asymmetric Spilling" measures, 2SLS estimates of δ are larger than OLS and significant at the 1% level. For the "Linear" and the "Asymmetric Receiving" measures instead they are smaller, and of the other two cases only for the Linear one is $\hat{\delta}_{2SLS}$ statistically significant (at the 10% level).

Table D.2: MPR and MSR: Comparative Prospect

	Baseline Measure	Alternative Measures		
		Linear	2nd Degree	As. Spilling
Mean of the Row Sum $\bar{g}_{it}$	0.440	0.065	1.049	0.675
(Standard Deviation)	(0.178)	(0.084)	(1.422)	(1.127)
Marginal Private Return	44%	47%	44%	46%
Marginal Social Return	10.4%	11.6%	8.4%	2.7%

Table D.2 offers a comparative prospectus of summary statistics about the row sum of connections, as well as the calculated MPRs and MSRs, across the baseline connection measure as well as the three alternative measures whose 2SLS estimates are estimated statistically significant (Linear, Second Degree, Asymmetric Spilling). The MPRs and the MSRs, in particular, are calculated using the 2SLS estimates for $\hat{\gamma}$ and $\hat{\delta}$. As expected, the calculated MPRs are similar across the four connection measures. The MSRs, instead, are similar only for the baseline, the Linear and the Second Degree measures. The MSR calculated from the Linear measure is actually larger than in the baseline case, but is obtained from a much less precise estimate of δ. By contrast, the MSR calculated for the Asymmetric Spilling measure is much smaller in magnitude, essentially because larger values of $g^{ass}_{(ij)t}$ are by construction associated with lower values of S_j (see the decomposition in (1.31)).

Appendix E

Graphical Description of the Network

This appendix collects, in the next few pages, some visual representations of the network in the form of graphs. For ease of comparison, all nodes (firms) are placed in the same position and have the same size across all figures. Node size is a positive function of the total strength of links that are summed over the years. In Figures from E.1 to E.5 nodes are also distinguished by various shades of the same orange color; in particular, darker colors indicate larger measures of network centrality associated with a node. The names of selected firms are apposed to some of the largest, most central nodes. A brief introduction or commentary for each of the following graphs is given in the list below.

- Figure E.1 displays the network in 1985.

- Figure E.2 displays the network in 1990.

- Figure E.3 displays the network in 1995.

- Figure E.4 displays the network in 2000.

- Figure E.5 displays the "pooled" network, which results from aggregating all connections over time for all nodes.

- Figure E.6 displays the communities obtained by applying the Louvain algorithm on the "pooled" network with maximum (1) resolution. The top hierarchy of communities is composed of six groups. The semiconductor/ICT, mechanical, biotech/pharmaceutical and chemical industries are clearly identifiable as separate communities; in addition there are two smaller, mixed groups whose nodes are dispersed across the graph.

- Figure E.7 displays the communities obtained by applying the Louvain algorithm with the lower 0.6 resolution. The resulting partition is the one that is used to cluster standard errors across all empirical estimates featured in this work.

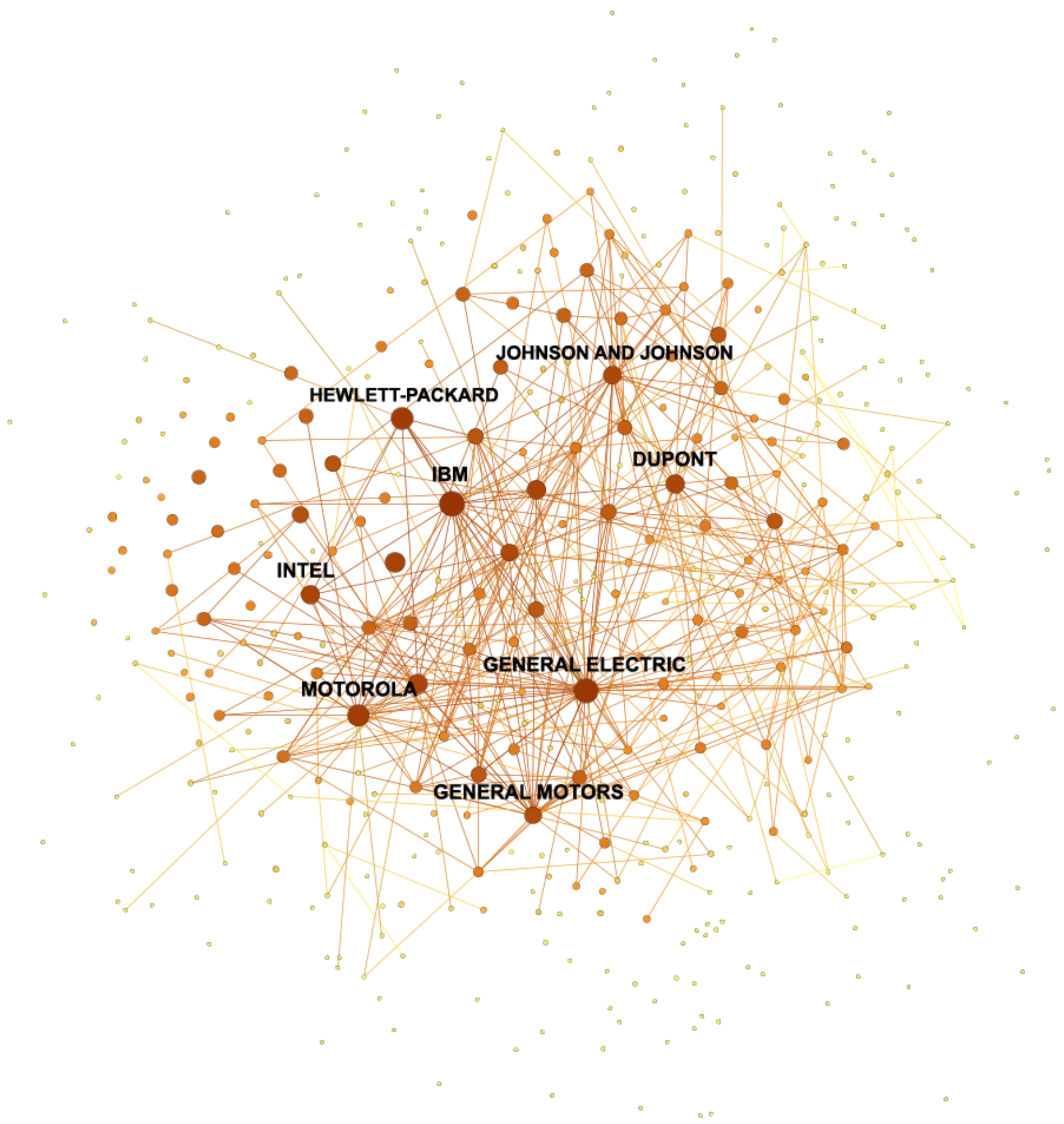


Figure E.1: The Network in 1985

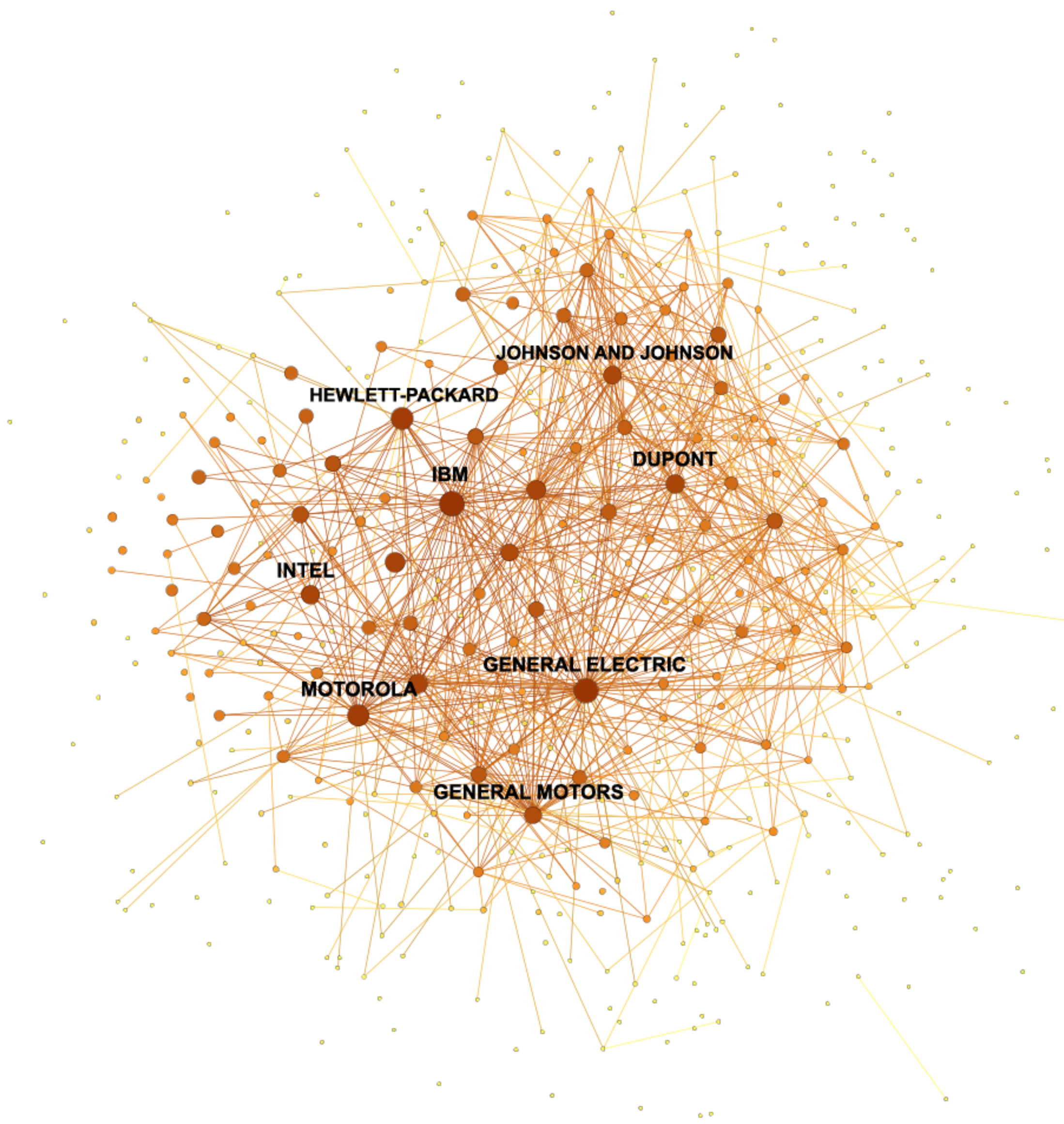


Figure E.2: The Network in 1990

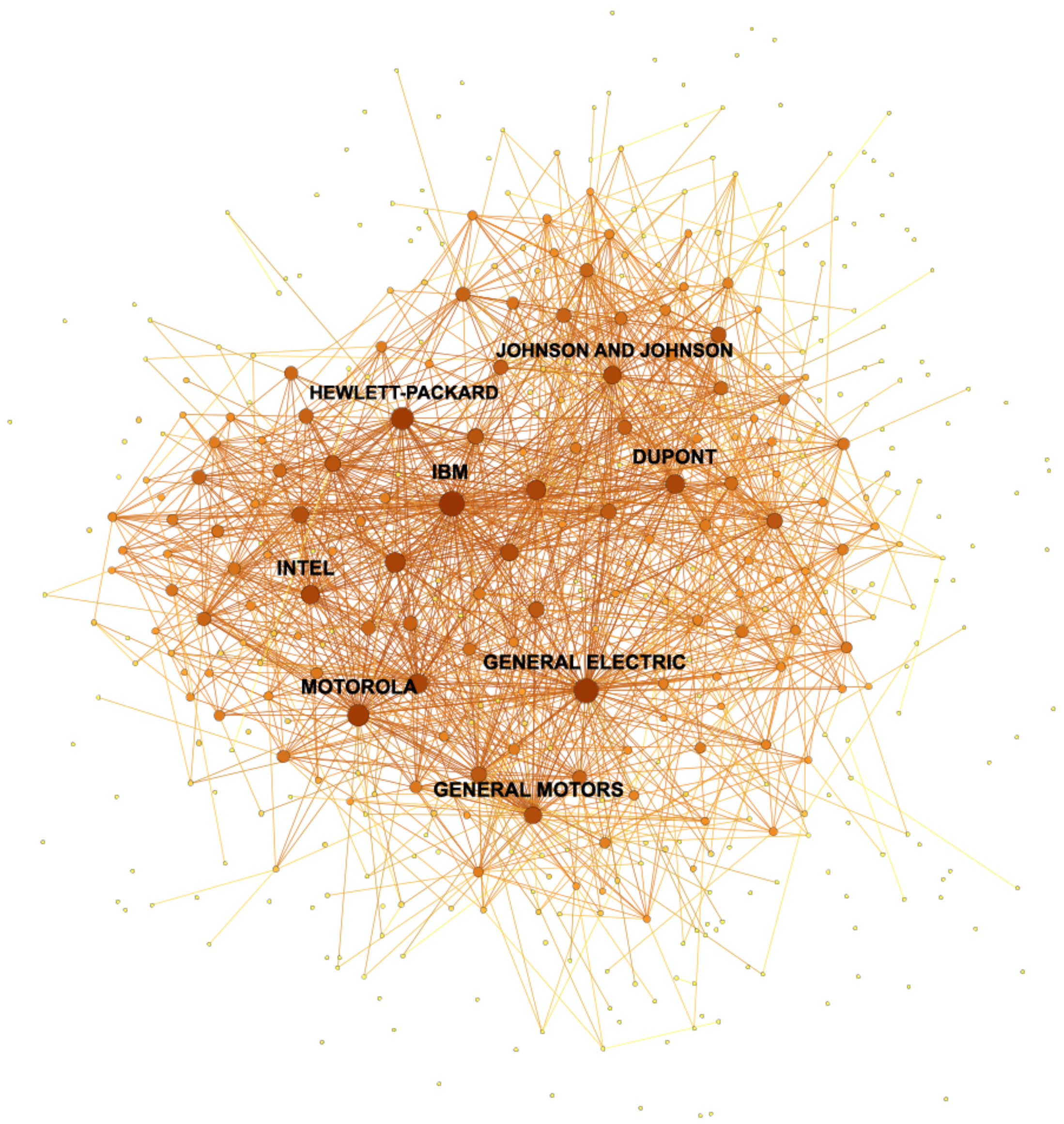


Figure E.3: The Network in 1995

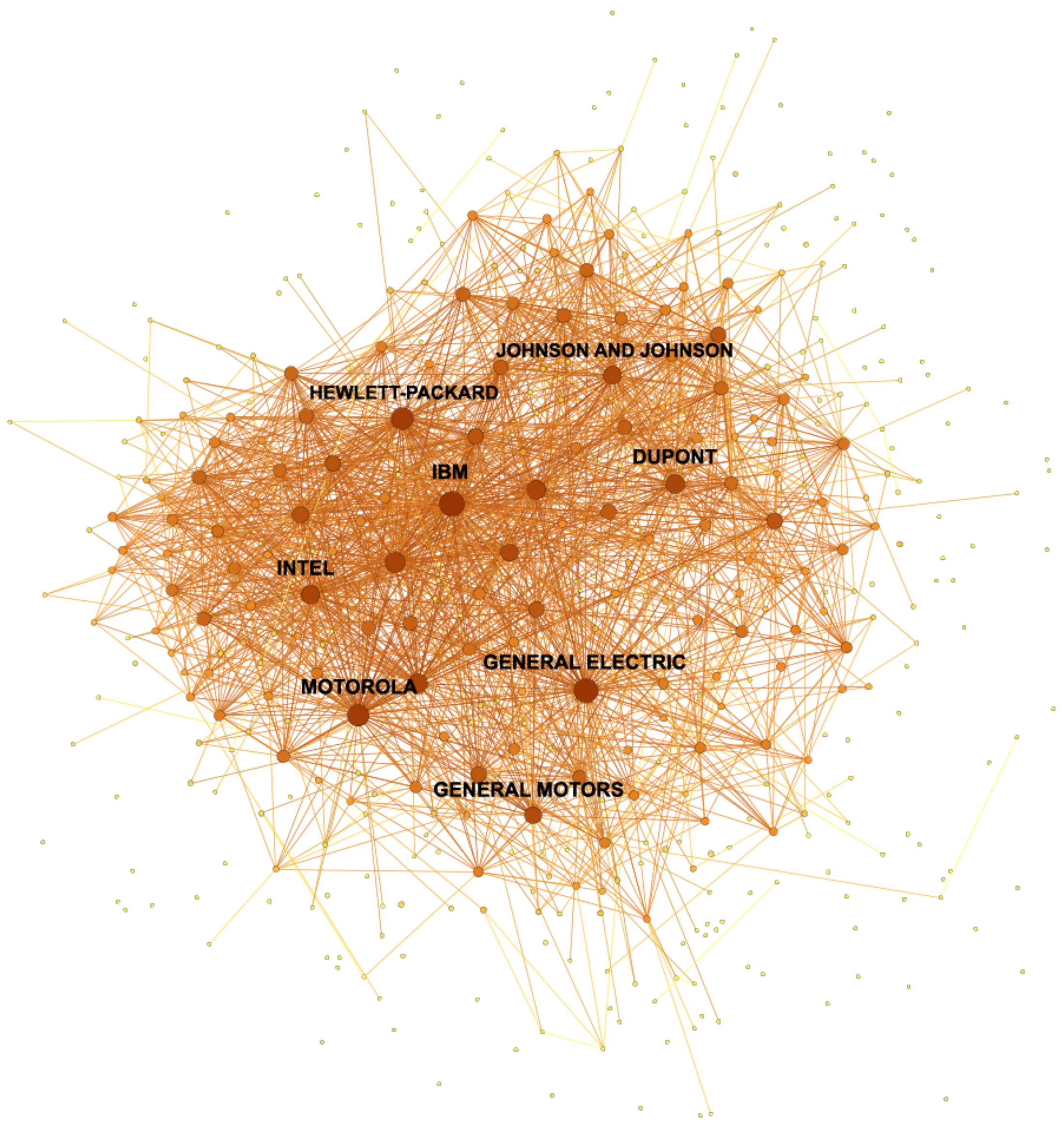


Figure E.4: The Network in 2000

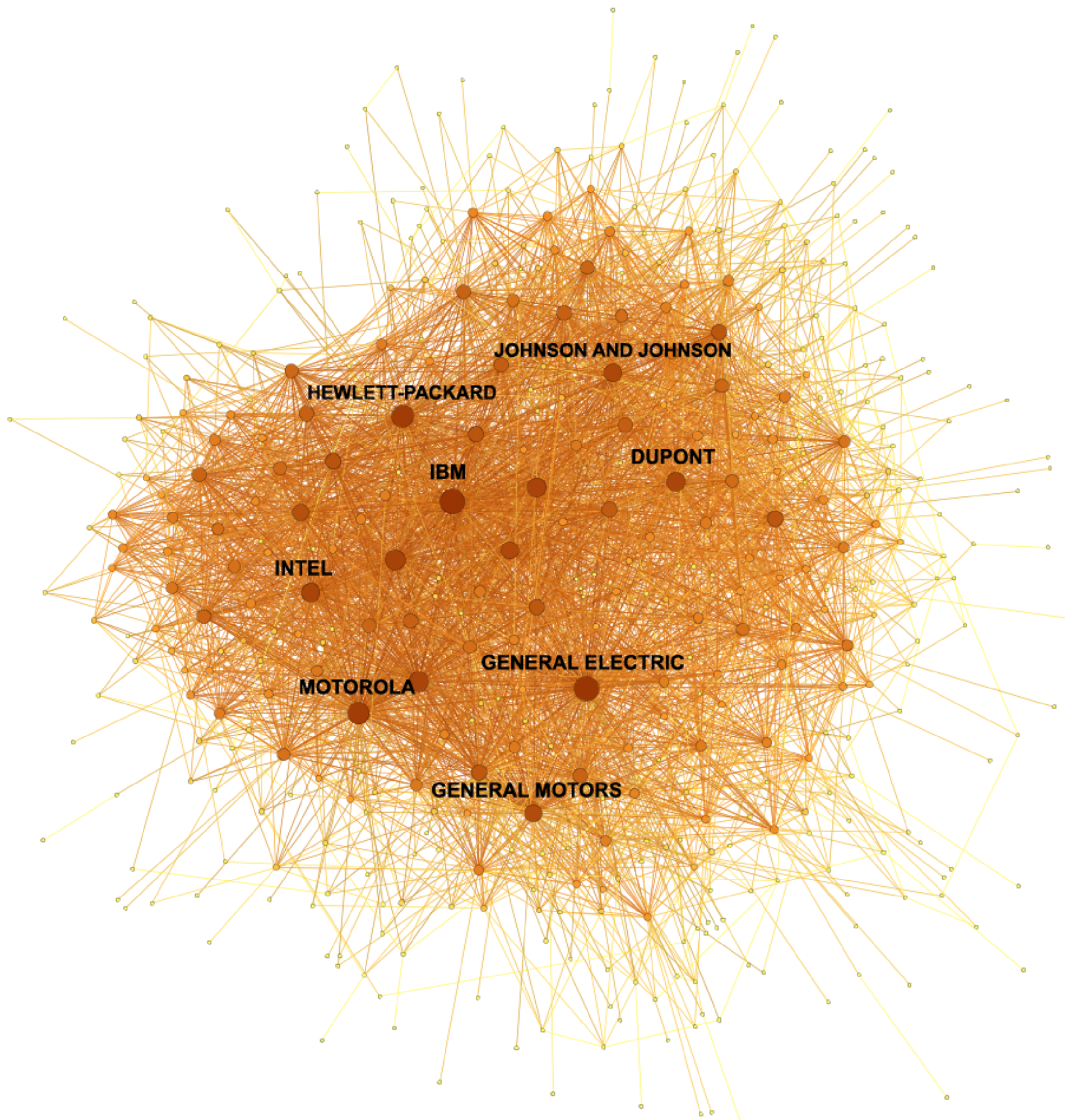


Figure E.5: The “Pooled” Network

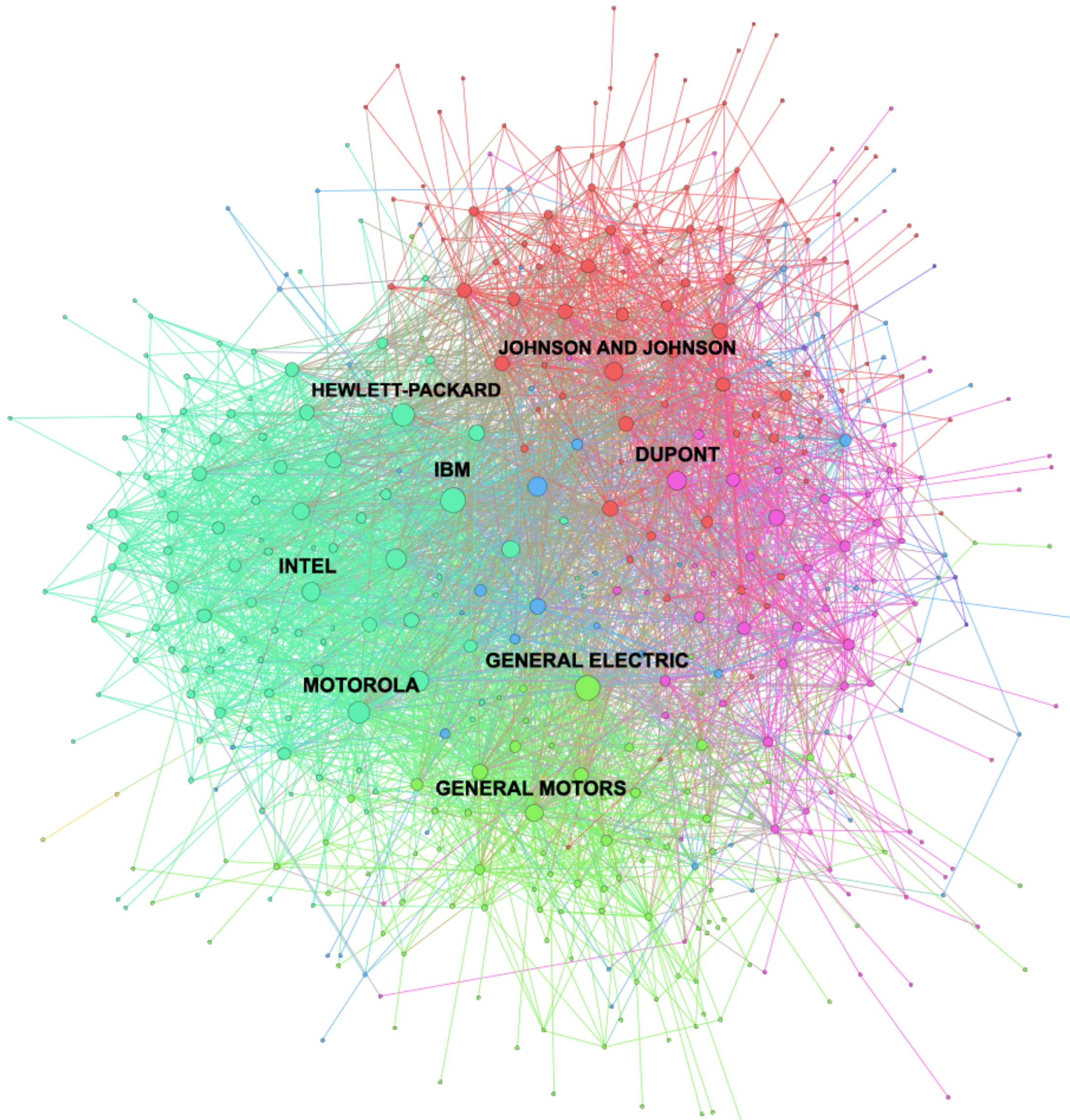


Figure E.6: Network Communities, Resolution 1

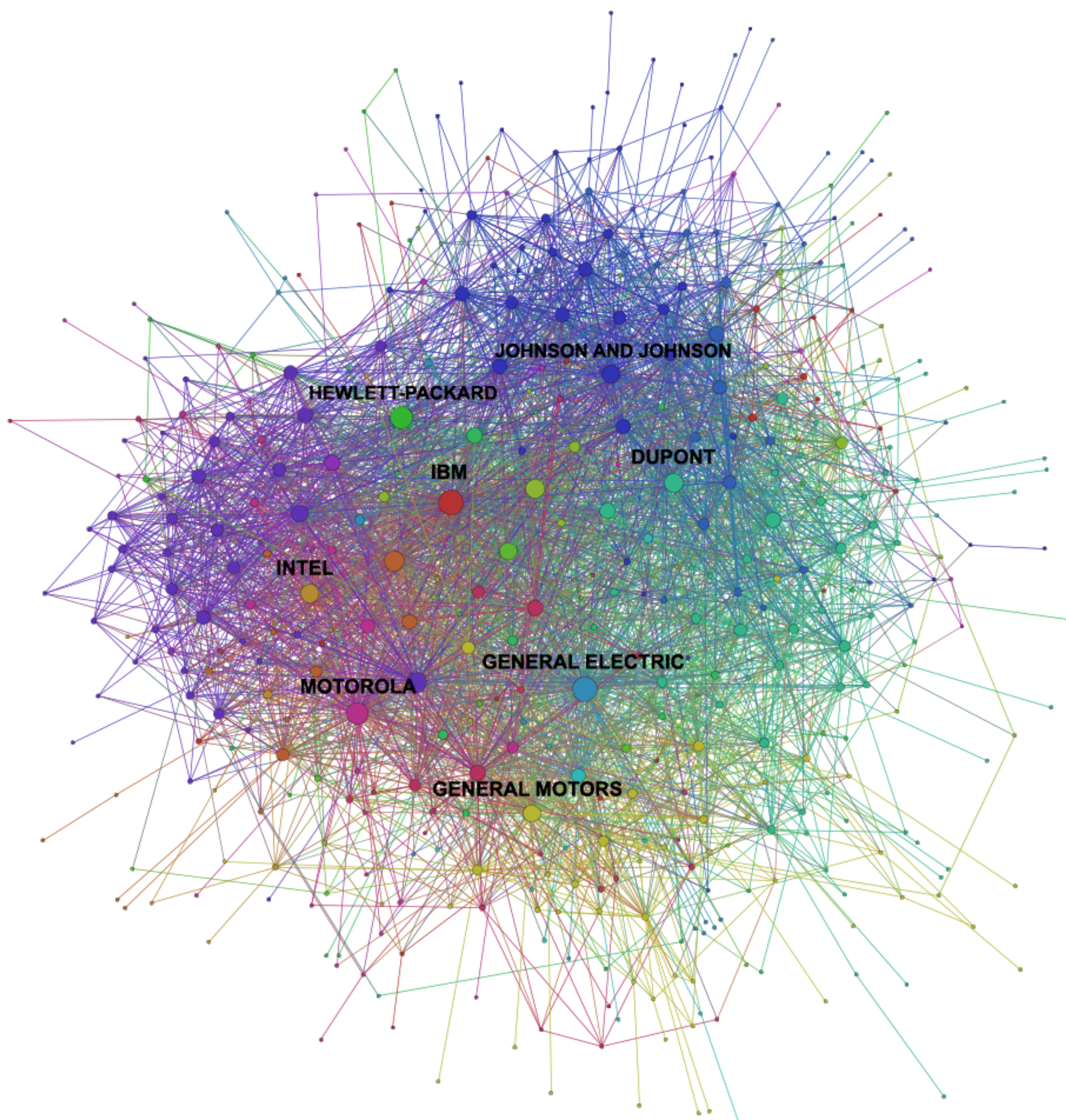


Figure E.7: Network Communities, Resolution 0.6

Appendix F

Extra Event Analyses of Superstar Migrations

In this appendix I report additional figures displaying results from the event study analysis described in Chapter 3, that are relative to outcome measures whose analysis has not been explicitly presented in the text, or that are restricted in more detail at the level of specific superstar groups (so to display the standard errors associated to event dummy estimates). These results are not presented in the main text for the sake of synthesis, as – among the other things – their interpretation does not alter the main qualitative conclusions discussed in Chapter 3. The list of Figures associated with these results is given below.

- Figure F.1: Network-level Estimates, Average Patent Shares $\bar{\tilde{P}}_{ict}$.
- Figure F.2: Network-level Estimates, Cit. Weighted Average Patent Counts $\bar{B}_{ict}$.
- Figure F.3: Network-level Estimates, Cit. Weighted Patent Shares $\tilde{B}_{ict}$.
- Figure F.4: Network-level Estimates, Cit. Weighted Average Patent Shares $\bar{\tilde{B}}_{ict}$.
- Figure F.5: City-level Estimates, Average Patent Counts $\bar{C}_{ict}$.
- Figure F.6: City-level Estimates, Patent Shares $\tilde{C}_{ict}$.
- Figure F.7: City-level Estimates, Cit. Weighted Average Patent Counts $\bar{G}_{ict}$.
- Figure F.8: City-level Estimates, Cit. Weighted Patent Shares $\tilde{G}_{ict}$.
- Figure F.9: Network-level Estimates, Top 0.1% Group, Patent Counts P^H_{ict}.
- Figure F.10: Network-level Estimates, Top 0.1% Group, Average Patent Counts $\bar{P}^H_{ict}$.
- Figure F.11: City-level Estimates, Top 0.1% Group, Patent Counts C^H_{ict}.
- Figure F.12: City-level Estimates, Top 0.1% Group, Average Patent Counts $\bar{C}^H_{ict}$.

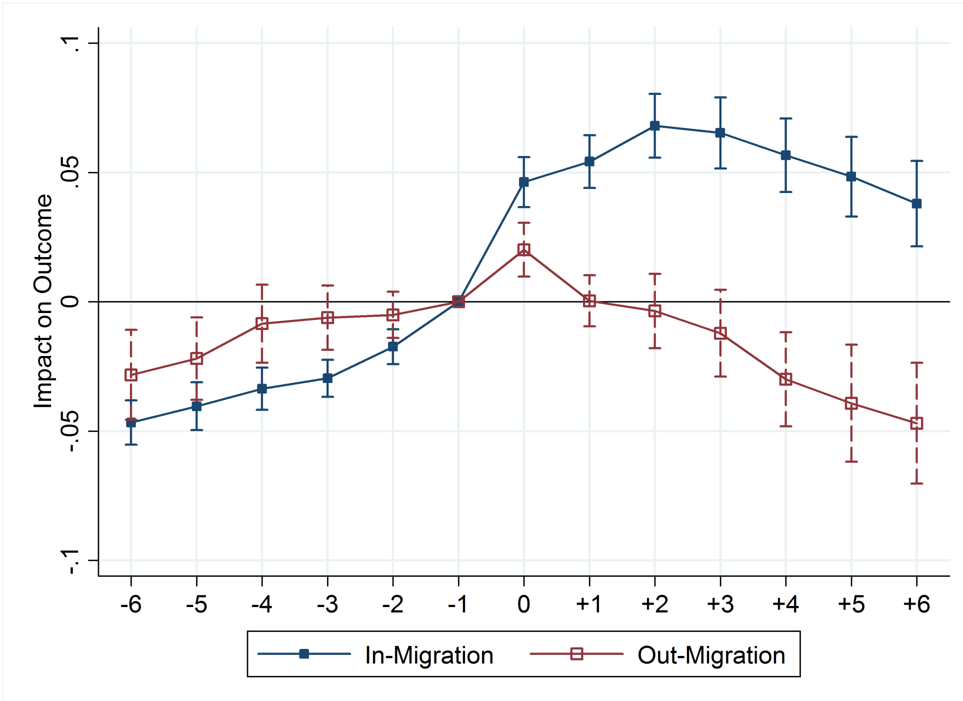


Figure F.1: Network-level Estimates, Average Patent Shares

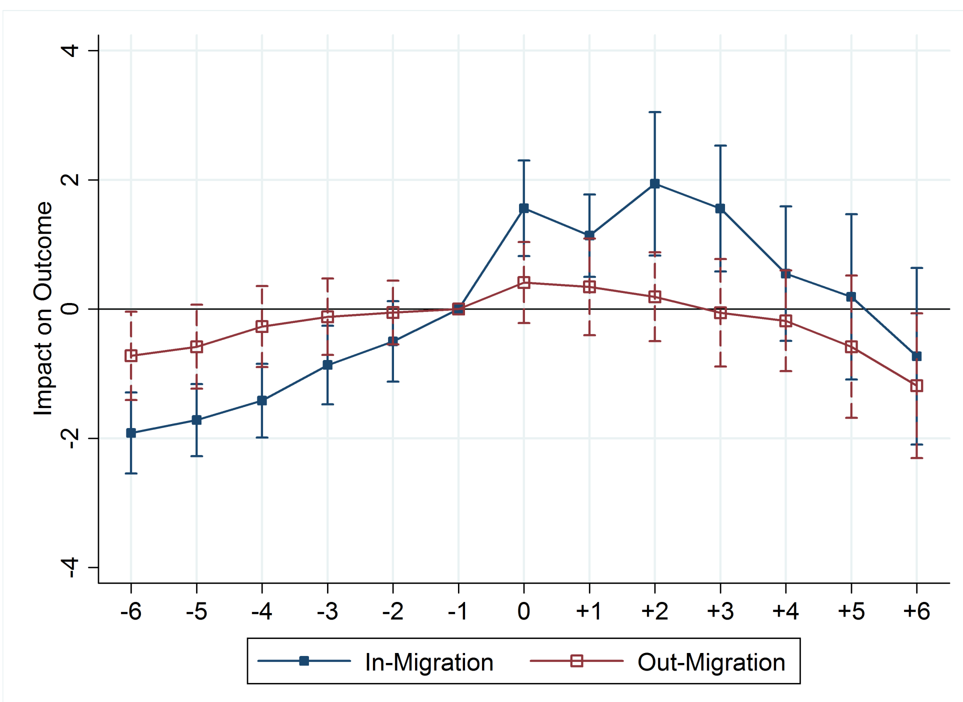


Figure F.2: Network-level Estimates, Cit. Weighted Average Patent Counts

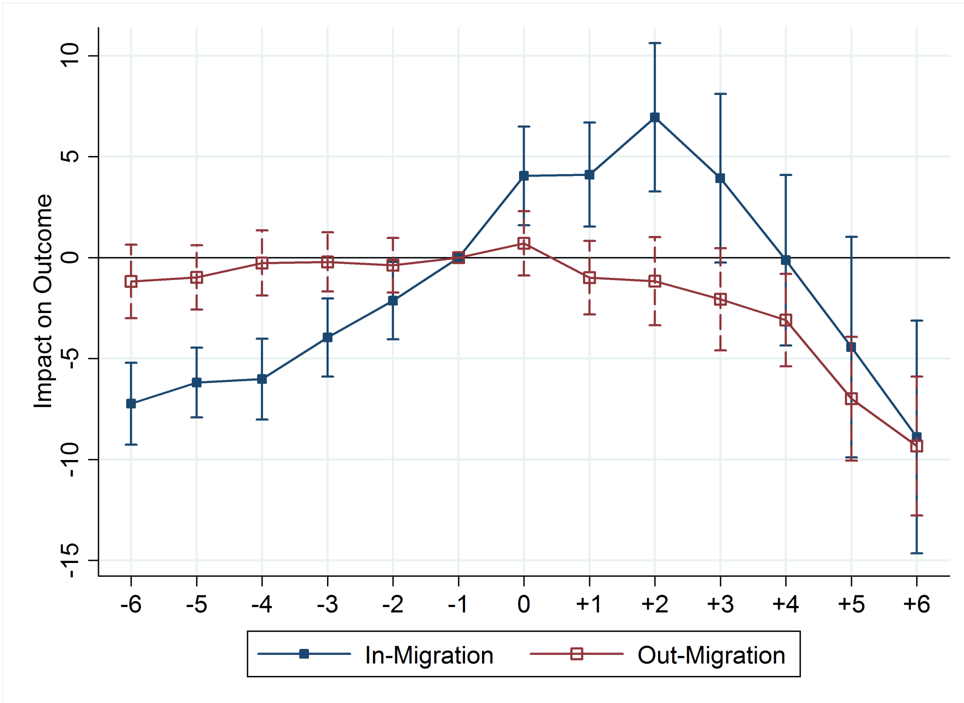


Figure F.3: Network-level Estimates, Cit. Weighted Patent Shares

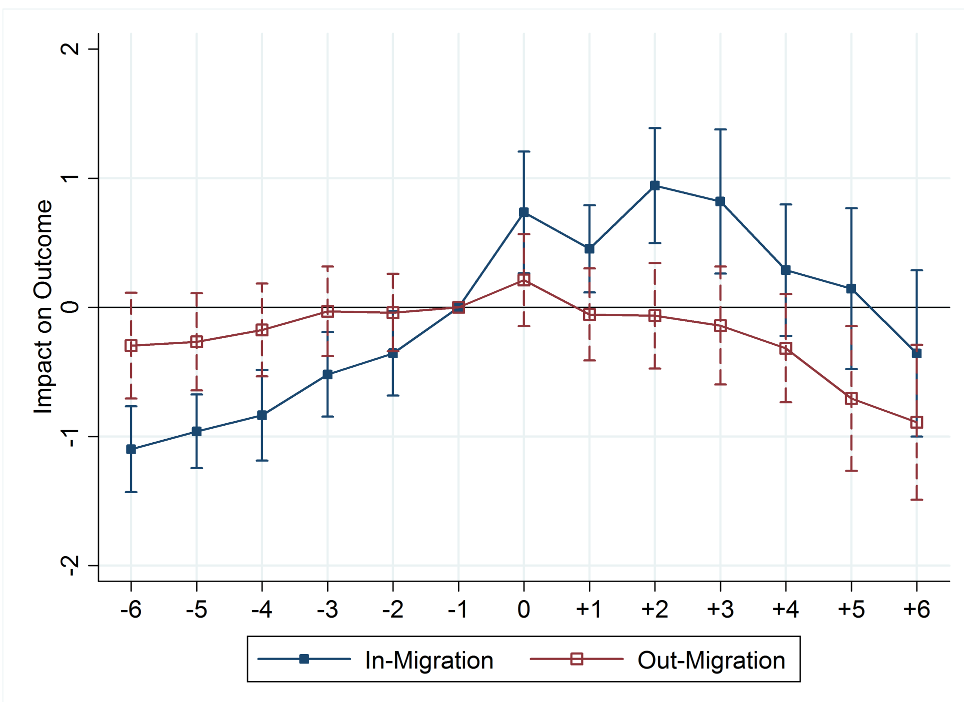


Figure F.4: Network-level Estimates, Cit. Weighted Average Patent Shares

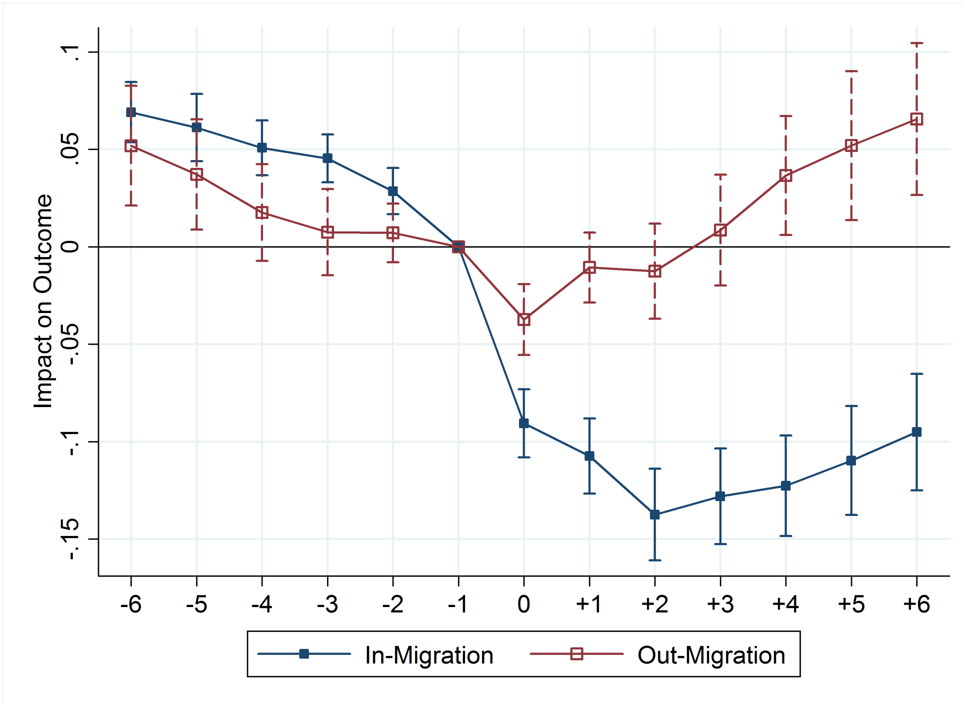


Figure F.5: City-level Estimates, Average Patent Counts

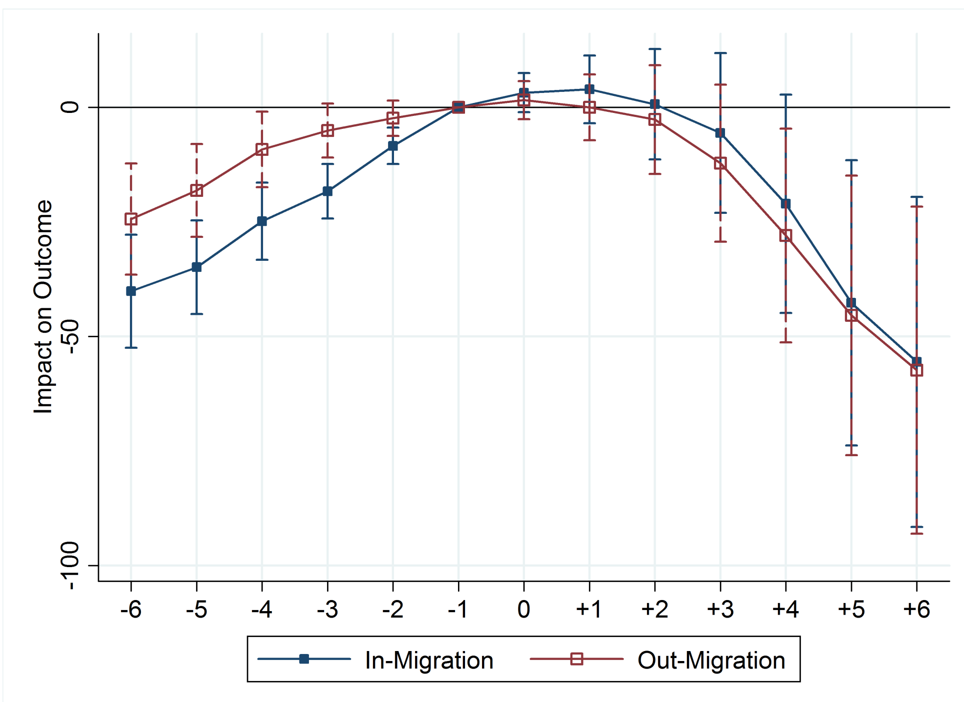


Figure F.6: City-level Estimates, Patent Shares

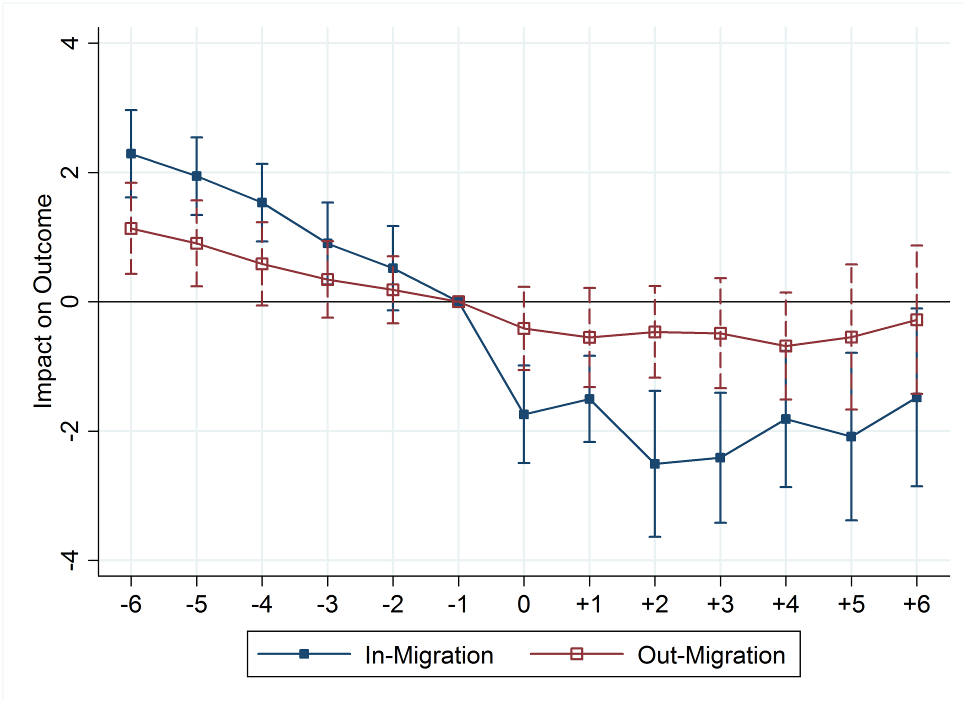


Figure F.7: City-level Estimates, Cit. Weighted Average Patent Counts

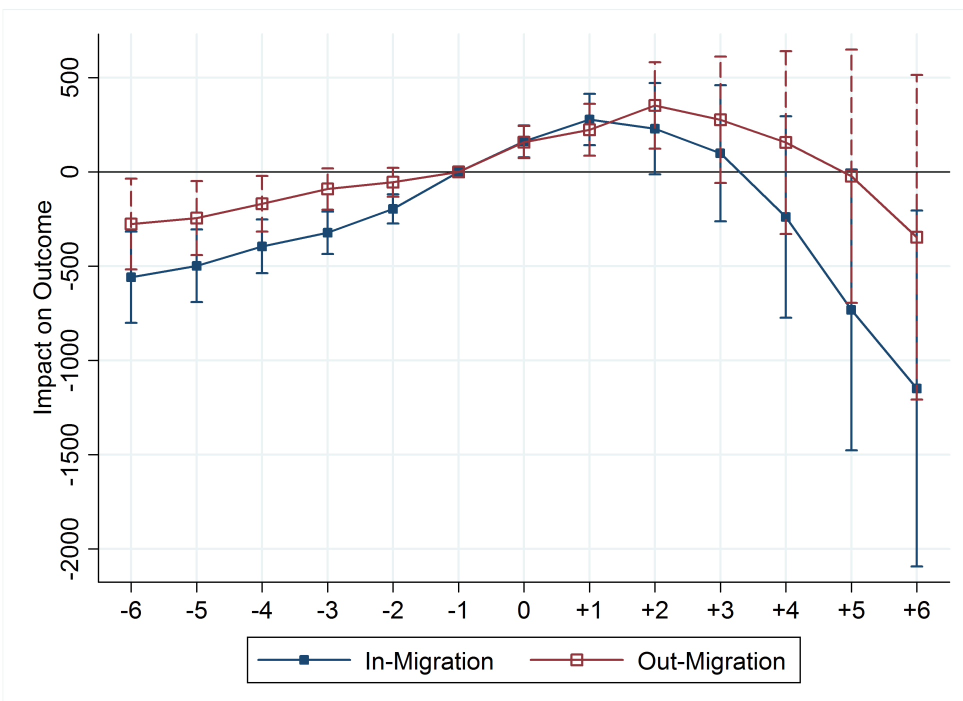


Figure F.8: City-level Estimates, Cit. Weighted Patent Shares

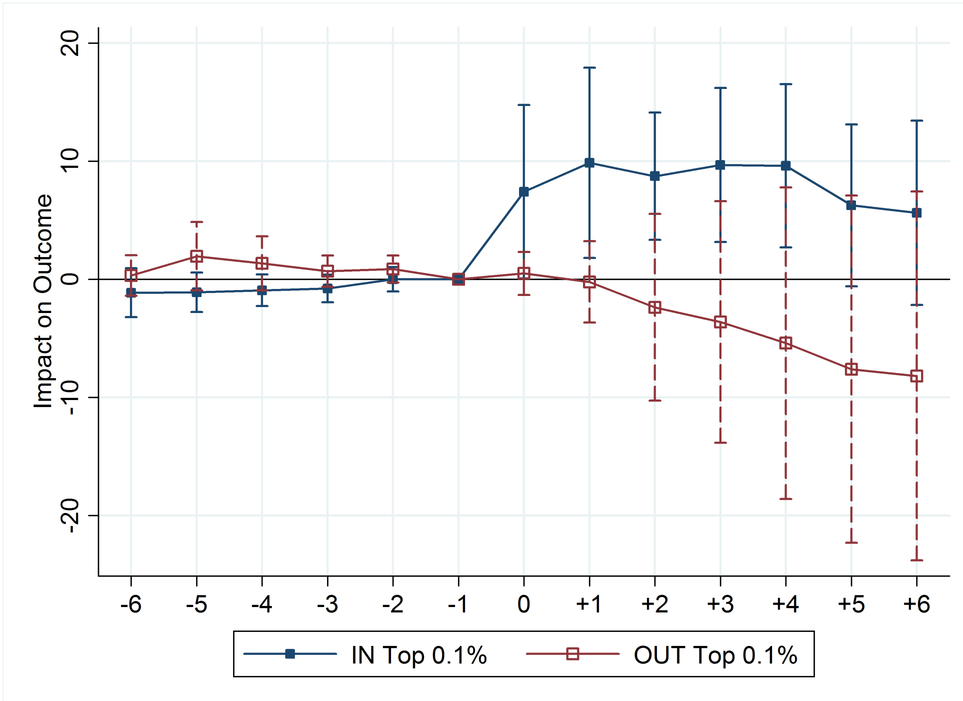


Figure F.9: Network-level Estimates, Patent Counts, Top 0.1%

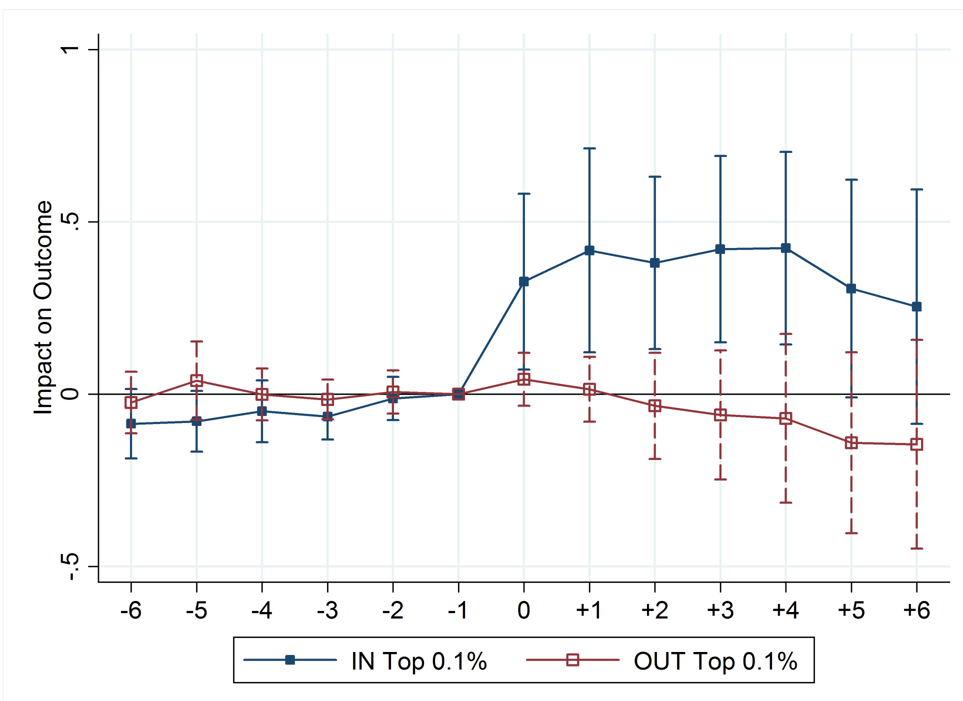


Figure F.10: Network-level Estimates, Average Patent Counts, Top 0.1%

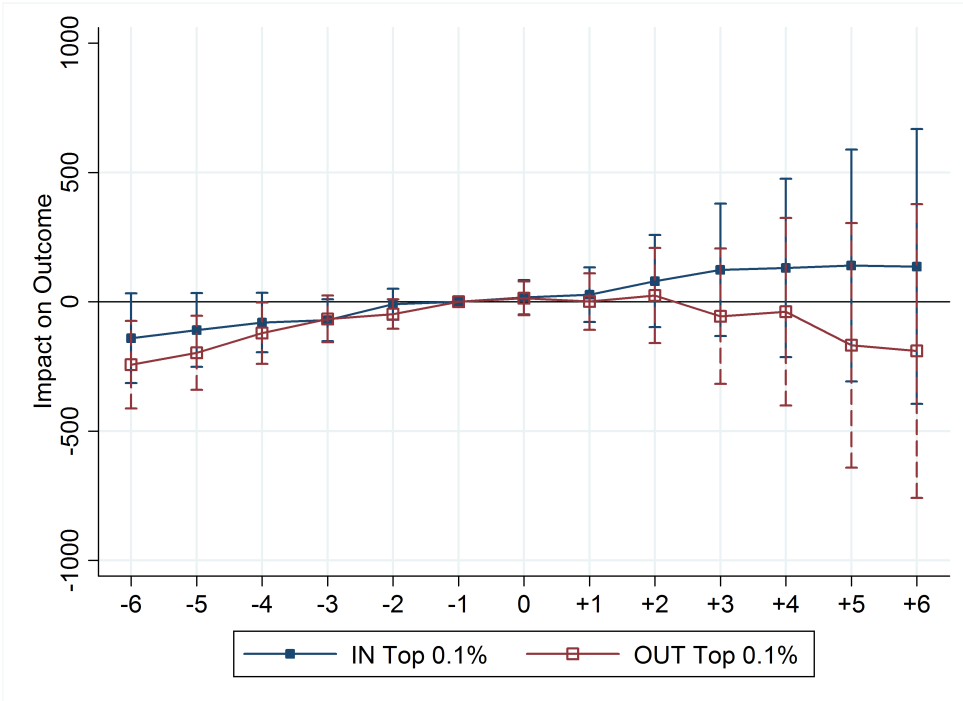


Figure F.11: City-level Estimates, Patent Counts, Top 0.1%

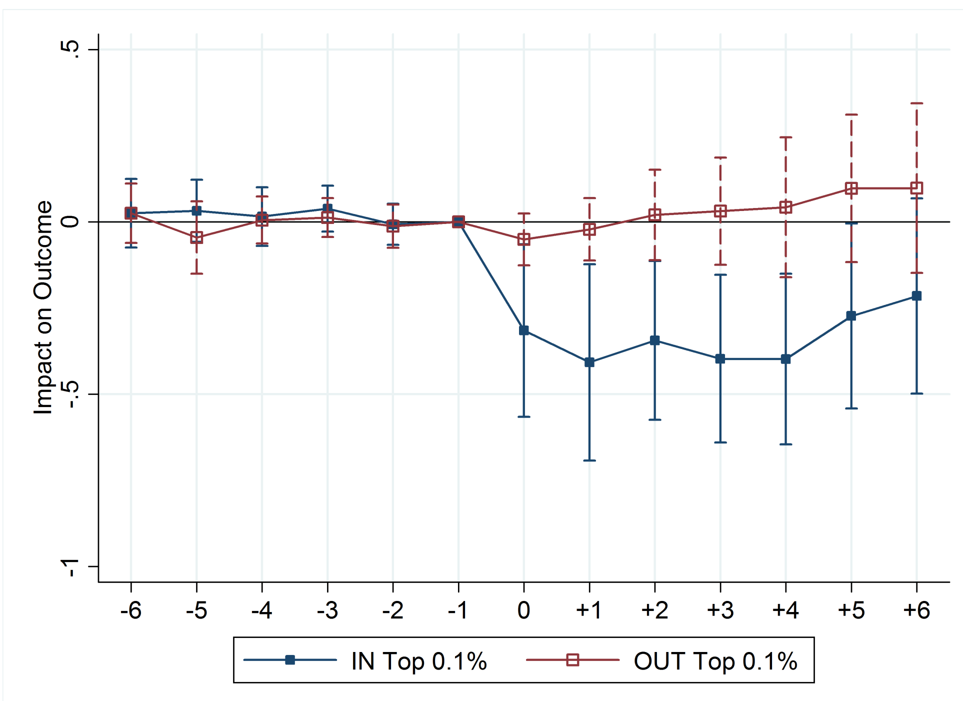


Figure F.12: City-level Estimates, Average Patent Counts, Top 0.1%

Bibliography

Agrawal, Ajay, Iain Cockburn, and Carlos Rosell, “Not invented here? Innovation in company towns,” *Journal of Urban Economics*, 2010, *67* (1), 78–89.

Akcigit, Ufuk, Salomé Baslandze, and Stefanie Stantcheva, “Taxation and the International Migration of Inventors,” 2015. Mimeo.

Angrist, Joshua D., “The perils of peer effects,” *Labour Economics*, 2014, *30*, 98–108.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang, “Superstar Extinction,” *The Quarterly Journal of Economics*, 2010, *125* (2), 549–589.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, 2008, *2008* (10), P10008.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, “Identifying technology spillovers and product market rivalry,” *Econometrica*, 2013, *81* (4), 1347–1393.

Blundell, Richard, Rachel Griffith, and John Van Reenen, “Dynamic count data models of technological innovation,” *The Economic Journal*, 1995, *105* (429), 333–344.

Bonhomme, Stéphane and Jean-Marc Robin, “Consistent noisy independent component analysis,” *Journal of Econometrics*, 2009, *149* (1), 12–25.

_ **and** _ , “Generalized non-parametric deconvolution with an application to earnings dynamics,” *The Review of Economic Studies*, 2010, *77* (2), 491–533.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, “Identification of peer effects through social networks,” *Journal of econometrics*, 2009, *150* (1), 41–55.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, “Peer effects and social networks in education,” *The Review of Economic Studies*, 2009, *76* (4), 1239–1267.

Chandrasekhar, Arun and Randall Lewis, “Econometrics of sampled networks,” 2011. Mimeo.

Christakis, Nicholas A. and James H. Fowler, "Social contagion theory: examining dynamic social networks and human behavior," *Statistics in medicine*, 2013, *32* (4), 556–577.

Conley, Timothy G. and Christopher R. Udry, "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, 2010, *100* (1), 35–69.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli, "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2010, *2* (2), 241–275.

Gort, Michael and Steven Klepper, "Time Paths in the Diffusion of Product Innovations," *The Economic Journal*, 1982, *92* (367), 630–653.

Graham, Bryan, "An empirical model of network formation: detecting homophily when agents are heterogeneous," 2014. Mimeo.

Graham, Bryan S., "Identifying social interactions through conditional variance restrictions," *Econometrica*, 2008, *76* (3), 643–660.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 2010, *118* (3), 536–598.

Griliches, Zvi, "Research expenditures, education, and the aggregate agricultural production function," *The American Economic Review*, 1964, *54* (6), 961–974.

_ , "Issues in assessing the contribution of research and development to productivity growth," *The Bell Journal of Economics*, 1979, *10* (1), 92–116.

_ , "The search for R&D spillovers," *The Scandinavian Journal of Economics*, 1992, *94*, 29–47. Supplement.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, "The NBER patent citation data file: Lessons, insights and methodological tools," 2001. NBER Working Paper.

_ **, Jacques Mairesse, and Pierre Mohnen**, "Measuring the Returns to R&D," in B. H. Hall and N. Rosenberg, eds., *Handbook of the Economics of Innovation*, Elsevier, 2010.

Hausman, Jerry A., Bronwyn H. Hall, and Zvi Griliches, "Econometric models for count data with an application to the patents-R&D relationship," *Econometrica*, 1984, *52* (4), 909–938.

Jaffe, Adam B., "Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value," *American Economic Review*, 1986, *76* (5), 984–1001.

_ , "Characterizing the technological position of firms, with application to quantifying technological opportunity and research spillovers," *Research Policy*, 1989, *18* (2), 87–97.

_ **, Manuel Trajtenberg, and Rebecca Henderson**, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

Kelejian, Harry H. and Ingmar R. Prucha, "On the asymptotic distribution of the Moran I test statistic with applications," *Journal of Econometrics*, 2001, *104* (2), 219–257.

Klepper, Steven, "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *The American Economic Review*, 1996, *86* (3), 562–583.

Kranton, Rachel, Martin D'Amours, and Yann Bramoullé, "Strategic Interaction and Networks," *American Economic Review*, 2014, *104* (3), 898–930.

Krugman, Paul, *Geography and Trade. Gaston Eyskens Lecture Series*, Cambridge, MA: MIT Press, 1991.

Li, Guan-Cheng, Ronald Lai, Alexander D'Amour, David M. Doolin, Ye Sun, Vetle I. Torvik, Amy Z. Yu, and Lee Fleming, "Disambiguation and co-authorship networks of the US patent inventor database (1975–2010)," *Research Policy*, 2014, *43* (6), 941–955.

Lychagin, Sergey, Joris Pinkse, Margaret E. Slade, and John Van Reenen, "Spillovers in space: does geography matter?," 2010. Forthcoming in *The Journal of Industrial Economics*.

Manresa, Elena, "Estimating the Structure of Social Interactions Using Panel Data," 2014. Mimeo.

Manski, Charles F., "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 1993, *60* (3), 531–542.

Marshall, Alfred, *Principles of Economics*, London: Macmillan., 1890. (First Edition ed.).

Mas, Alexandre and Enrico Moretti, "Peers at work," *American Economic Review*, 2009, *99* (1), 112–145.

Moretti, Enrico, "Workers' education, spillovers, and productivity: evidence from plant-level production functions," *American Economic Review*, 2004, *94* (1), 656–690.

_ , "Local labor markets," *Handbook of labor economics*, 2011, *4*, 1237–1313.

_ **and Daniel J. Wilson**, "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists," 2014. Mimeo.

Pereda-Fernández, Santiago, "Social spillovers in the classroom: Identification, estimation and policy analysis," 2015. Mimeo.

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

_ , "Wages, rents and amenities: differences among workers and regions," *Economic Inquiry*, 1988, *26* (1), 23–41.

Roehrig, Charles S., "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 1988, *56* (2), 433–447.

Rosen, Sherwin, "Wage based indexes of urban quality of life," in P. N. Miezkowski and M. R. Straszheim, eds., *Current Issues in Urban Economics*, Johns Hopkins University Press, Baltimore, MD, 1979.

Serafinelli, Michel, "Good Firms, Worker Flows and Productivity," 2013. Mimeo.

Syverson, Chad, "What determines productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–365.

Thompson, Peter and Melanie Fox-Kean, "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *The American Economic Review*, 2005, *95* (1), 450–460.

Williamson, Oliver, "The Theory of the Firm as Governance Structure: From Choice to Contract," *The American Journal of Sociology*, 1981, *87* (3), 548–577.