Center for Climate, Regional, Environmental and Trade Economics
www.create.illinois.edu

The role of interregional and inter-sectoral knowledge spillovers on regional knowledge creation across US metropolitan counties

Orsa Kekezi[a], Sandy Dall'erba[b], Dongwoo Kang[c]

[a] Swedish Institute for Social Research (SOFI), Stockholm University, 106 91 Stockholm, Sweden and Centre for Entrepreneurship and Spatial Economics (CEnSE), Jönköping International Business School, 551 11 Jönköping, Sweden, orsa.kekezi@sofi.su.se
[b] Department of Agricultural and Consumer Economics and Center for Climate, Regional, Envionmental and Trade Economics, University of Illinois at Urbana-Champaign, IL 61801, USA, dallerba@illinois.edu
[c] Korea Labor Institute, Sejong National Research Complex, 370, Sicheong-daero, Sejong-si, 30147, Korea, dwkang1982@kli.re.kr

CREATE Discussion Paper 2-21
July, 2021

Abstract: Knowledge accumulation and its spillovers are important determinants of the regional economic growth process in the U.S. As such, this paper relies on a regional knowledge production function to examine the heterogeneous determinants of knowledge creation across 5 U.S. manufacturing sectors and 853 metropolitan counties. Using a Tobit model with State fixed effects**,** our results indicate that local intra-sectoral and inter-sectoral R&D investments by the private sector as well as university R&D play a key role in knowledge creation across all sectors under study. We also find that the role of short-distance vs. long-distance interregional spillovers on knowledge creation varies greatly across sectors. These key features improve the design of future local and national innovation policies.

Climate, Regional, Environmental and Trade Economics
65-67 Mumford Hall
1301 West Gregory Drive
Urbana, IL, 61801

1. Introduction

As knowledge accumulation and its spillovers are recognized as important determinants of economic growth (Romer 1986, Grossman and Helpman 1994, Jaffe 1989), the knowledge production function literature has paid an increasing amount of attention to the role and the geography of knowledge spillovers (Audretsch and Feldman 2004, Anselin et al. 1997, Acs et al. 2002, Bode 2004, Autant-Bernard 2012). While geographical proximity facilitates the flow of knowledge (Audretsch and Feldman 2004, Jaffe et al. 1993), other mechanisms such as non-market interactions (Glaeser and Scheinkman 2000), technological proximity (Maggioni et al. 2011), labor mobility (Almeida and Kogut 1999), social networks (Breschi and Lissoni 2009, Crescenzi et al. 2016), and patent citations (Peri 2005, Kang and Dall'erba 2016a, b) promote the diffusion of knowledge across space also. As such, a large amount of the more recent research in this area has challenged the traditional view that local knowledge flows are the main determinants of local innovation (Peri 2005, Ponds et al. 2010).

At the same time, a growing number of studies has followed Glaeser et al. (1992) in debating the relative importance of intra-sectoral and inter-sectoral knowledge spillovers on the creation of innovation, productivity, employment growth and, ultimately, urban agglomeration. Despite the extensive literature in this area, several contributions show that the conclusions on what type matters more are rather heterogeneous and depend on the sector studied as well as level of spatial aggregation (Groot et al. 2016). This debate is relevant to the regional knowledge production function literature because, until recently, most empirical studies used data aggregated across sectors (Anselin et al. 1997, Fischer and Varga 2003, Bode 2004, Parent and LeSage 2008). When (partially) controlled for, sectoral heterogeneity is modeled through sectoral dummies (Ponds et al. 2010) or through a variable reporting the share of value added produced by the manufacturing sector in the region (Bottazzi and Peri 2003). Clear evidence of the sectoral heterogeneity present in the knowledge spillovers is, to our knowledge, very limited. For instance, Jaffe (1989) and Anselin et al. (2000) differentiate the localized knowledge spillovers by sector but only capture intra-sectoral spillovers. Autant-Bernard and LeSage (2011) demonstrate the significant impact of inter-sectoral spillovers of private R&D among French metropolitan areas. However, their panel model is averaged across all sectors so that the marginal effect of intersectoral spillovers is not reported by sector. More recently, Acemoglu et al. (2016) highlight the importance of inter-sectoral spillovers of knowledge in the U.S. as captured through a matrix of patent creation- patent citations. Their findings indicate that the most important spillovers come from within the industry. Yet, their work is performed at the national level, hence provides no guidance on how the geographical distance between the origin and destination locations of these spillovers may affect knowledge creation differently across sectors.

This paper contributes to this literature by identifying the singular role of intra- and intersectoral knowledge spillovers on knowledge creation (patent counts) by sector. In addition, to account for geographical proximity, we classify these spillovers into three categories: i) local spillovers (within the county), ii) short-distanced inter-regional spillover (from neighboring counties located on a 50-mile radius), and iii) spillovers from the rest of the US (beyond 50 miles). More specifically, we focus on the five most innovative manufacturing sectors in the U.S.: 1) Chemical, 2) Drugs & Medical, 3) Mechanical, 4) Computer & Communication, 5) Electrical & Electronic. They represent about 82% of our patent data drawn from the US Patent and Trade Office (USPTO 2010). We study how the existing stock of research inputs impacts new knowledge creation by relying on an interregional innovation network of patent creation-patent citation. This approach and our reliance on the actual patent network allows us to recognize the complex and spatially dependent nature of innovation and to improve our understanding of industrial innovation dynamics.

Our sample covers 853 metropolitan counties. It allows us to get more detailed results on the regional and sectoral knowledge production process compared to the existing literature where the data are aggregated at the U.S. state level (e.g. Peri, 2005), across Metropolitan Statistical Areas (Anselin et al. 2000) or even at the national level (Acemoglu et al. 2016). Furthermore, we use a panel Tobit model with time and State fixed effects to control for the case where no knowledge output is recorded and for cross-sectional unobservable heterogeneity (Wooldridge 2010). Last but not least, knowledge spillovers are all based on the data collected by Lai et al. (2013). This dataset has been previously used in the literature in different contexts ((Autor et al. 2016, Moretti and Wilson 2017, Verhoeven et al. 2016, Galasso and Schankerman 2018). This dataset tracks the actual flows of knowledge from the place where they are created to the place(s) where they are cited. Compared to knowledge spillovers based on geographical proximity (Anselin et al. 1997, Bode 2004) or collaborative work (Ponds et al. 2010, Crescenzi et al. 2016), the major advantage of capturing the directionality of the flows of knowledge is to allow us to explicitly identify the role of externalities on knowledge output.

Relying on a Tobit estimation with fixed effects at the State level, our county-level results support the importance of geographical proximity for knowledge creation and indicate that both local intra-sectoral as well as local inter-sectoral spillovers are important determinants of knowledge production. This result is valid across all the sectors. Regarding the importance of inter-regional private and university spillovers, we observe large sectoral heterogeneity as the various sectors under study benefit differently from the knowledge created in other regions whether they are close-by or remote. These results strongly suggest that the cumulative process of scientific discovery is heterogenous and complex.

The remainder of the paper is organized as follows: Section 2 reviews the literature focusing on local and distant knowledge spillovers and their role on knowledge creation and innovation. Section 3 describes our knowledge production function, the strategy of modeling intra- and inter-sectoral knowledge spillovers across counties and the relevant data. The estimation results and their interpretation are reported in Section 4, and the robustness tests are presented in Section 5. The last section closes with some concluding remarks.

2. Literature Review

2.1. Local and distant knowledge spillovers

Most studies on knowledge creation and innovation focus on local knowledge spillovers (Jaffe 1986, Jaffe et al. 1993, Feldman 1994, Anselin et al. 1997). Their local extent is usually explained by two types of externalities. The first one is Marshall-Arrow-Romer (MAR) externalities that emphasize industrial specialization within the same or similar sectors. It allows to lower the cost of communication and transaction, thereby facilitating knowledge spillovers among firms (Audretsch and Feldman 2004). A well-known example is the Silicon Valley cluster where knowledge flows across high-technology and internet firms are galvanized through non-market interactions and inter-firm mobility of skilled workers (Saxenian 1994)[1]. The second type of local externalities, based on Jacobs (1969), derives from industrial diversity. She argues that a diverse knowledge coming from external sectors can complement a specific sector's knowledge and thus

[1] **Note, however, that this paper does not include the "Internet publishing and broadcasting and web search portals" industry (NAICS code 51913) that most of the firms in Silicon Valley belong to. The list of manufacturing sectors we focus on appears in table 1.**

facilitate innovation. As geographical proximity contributes to the exchange of ideas (Feldman and Kogler 2010) and activities have become more clustered over time (Glaeser et al. 1992), both intra- and inter-sectoral spillovers have played an increasing role in the creation of knowledge (Beaudry and Schiffauerova 2009, Groot et al. 2016).

Other studies have also paid attention to knowledge emanating from geographically distant sources (Owen-Smith and Powell 2004, Trippl et al. 2009). The literature has demonstrated that firms with limited access to distant knowledge pools tend to be less innovative and generate less output than their peers (Feldman and Kogler 2010, Moreno and Miguélez 2012). For Maskell *et al.* (2006), it is the complementarity between the local knowledge pool and distant sources of knowledge that will promote regional innovation growth. Because each region has its own industry-mix and exploits local and distant knowledge pools differently (Feldman and Kogler 2010), one should expect the relative role of distant intra- and inter-sectoral knowledge spillovers on local innovation to vary across sectors. To our knowledge, no previous study investigates this issue, hence this paper fills this gap.

2.2. Accounting for sectoral heterogeneity in regional and sectoral knowledge spillovers

Only a handful of studies focus on the differences in regional knowledge production across sectors. Using U.S. state level data, the seminal work of Jaffe (1989) investigates the influence of university research on corporate patents across four different sectors. His study finds that the Drugs, Chemical and Mechanical sectors benefit from intra-sectoral university research that is taking place locally. Based on more detailed MSA (Metropolitan Statistical Areas) data, Anselin et al. (2000) also investigate how local (within MSA) university research spills over to four industrial sectors and, unlike the previous study, highlight that interregional (beyond the MSA boundaries) spillovers of university research play a key role for some sectors. The authors find that the local effects from university research for the Drugs and Chemical sector are not significant, but the interregional externalities have a strong impact in innovation. They justify their result by noticing that most chemistry departments focus on basic research for which spatial proximity is not of high importance. Their findings indicate also that while the Electronic and Instruments sectors enjoy significant local and inter-regional spillovers, the Machinery sector is the only one that benefits from long-distance research spillovers.

One important element that is missing from the aforementioned studies is the presence of spillovers of private R&D, whether they take place within or across sectors, as they have been found to promote innovation as well (Wallsten 2001, Orlando 2004). Autant-Bernard and LeSage (2011) account for the capacity of both private and public intra- and interregional R&D spillovers to promote knowledge across 11 sectors of 94 French regions. They conclude that Jacobian externalities dominate MAR externalities when they emanate from private R&D efforts. This result holds true whether they look at intra- or interregional spillovers. Local Jacobs and MAR externalities have roughly a similar role on innovation when they come from public R&D efforts while at the interregional level only the MAR externalities matter. However, they do not account for any form of heterogeneity across sectors. In addition, the spatial extent of the knowledge spillovers is modeled on geographical contiguity only so that neither the geographical extent nor the directionality of the knowledge flows are captured in their work.

A recent study focusing on the innovation network within and across sectors is Acemoglu et al. (2016) who focus on the 1975-1984 network in the U.S. The authors find that most of the patents are cited within the sector they belong to. Intersectoral spillovers take place mostly within the parent sector sub-sectors belong to (e.g. computer peripherals citing from computer communication) and in a few cases across parent sectors (e.g. chemical sector citing from the drug & medical sector). However, their contribution does not provide us with a sense of the geographical extent of these spillovers as the results are for the nation as a whole and they do not rely on econometric techniques. Finally, based on a matrix of patent creation-patent citation, Cai and Li (2018) highlight a "technology network" and measure the applicability of the knowledge created in one sector to patenting in other sectors. Their results provide guidance on the amount of R&D to provide by sector.

As such, this paper remedies to the gaps listed above by tracking the various types (intra- vs interregional, intra-vs intersectoral, private vs. public R&D) of knowledge flows that exist. The relevant data and specific modeling strategy are described in the next section.

3. Empirical Model and Data

3.1. Regional Knowledge Production Function and Tobit model

Our empirical model relies on a regional approach of Griliches (1979) knowledge production function using US county-level panel data. The knowledge production function is assumed to follow a Cobb-Douglas functional form as depicted in Equation (1) where y_{iht} is the knowledge output of sector h in county i at time t, $x_{k,iht}$ is the kth knowledge input, β_k is the elasticity of the output with respect to the corresponding input and μ_{ih} and v_{iht} represent an individual specific effect and an error term respectively.

$$y_{iht} = \prod_k x_{k,iht}^{\beta_k} \cdot e^{\mu_{ih}} \cdot e^{v_{iht}} \qquad (1)$$

The logarithm transformation of Equation (1) leads to a log-linear model that is widely used in empirical studies of the knowledge production function (Anselin et al. 1997, Acs et al. 2002, Fischer and Varga 2003, Bode 2004).

As usual in the literature, we use patent data as a proxy for knowledge output (Parent and LeSage 2008, Autant-Bernard and LeSage 2011) and work with patent applications (Cincera 1997, Ramani et al. 2008) instead of granted patents because the former are closer in time to knowledge creation. Patent application data are retrieved from the database constructed by Lai et al. (2013).[2] In order to allocate the patent data across counties, we use the fractional counting method suggested by Jaffe et al. (1993). When a patent is created by N inventors, 1/N fraction of the patent is attributed to each inventor. Each 1/N fractional patent is geocoded to its associated county based on the address of the inventor. As a result, the patent data is a rational number[3]. Besides location, we also focus on the knowledge created across five manufacturing sectors: (i) chemical, (ii) drugs & medical, (iii) mechanical, (iv) computer & communication, and (v) electrical & electronic. Thus, we classify the patent applications into these five sectors based on the North American Industry Classification System (NAICS) defined in 2002 (Table 1).[4] According to Kang and Dall'erba (2016a, 2016b), the metropolitan regions have a greater propensity to innovate (an average of 150.9 patents in the MSA counties vs. 3.4 in non-MSA counties over 2003-2005) and their knowledge production mechanism is different from that of the non-metropolitan regions. Therefore, we focus on the metropolitan counties only. There were 853 of them across the 3,109 continental US counties in 2000.

[Insert Table 1 here]

---

[2] The files can be downloaded here https://github.com/funginstitute/downloads

[3] We rely on the USPCS-NAICS 2002 concordance file developed by the US Patent and Trademark Office (USPTO) to transfer all the data in NAICS format.

[4] The geographical allocation of the patent data could also be based on the address of the assignee(s). However, large companies use the address of their headquarter to file patents, which is not always the place where research took place. Using the inventor's address to geocode the creation and citation of patent data can lead to a similar problem (Autant-Bernard and LeSage 2011), but the size of the error is smaller as we assume that inventors live close to their workplace.

Since the minimum value of observed patent data is zero[5], we follow Cameron and Trivedi (2009) and rely on a Tobit model for our empirical estimation where $\boldsymbol{Patent}^{*}_{\boldsymbol{iht}}$ is the unobservable latent value of patent application and $\boldsymbol{Patent}_{\boldsymbol{iht}}$ is the observed patent application so that:

$$ln\boldsymbol{Patent}_{\boldsymbol{iht}} = \begin{cases} \ln\left(\boldsymbol{Patent}^{*}_{\boldsymbol{iht}}\right) & if\ \ \boldsymbol{Patent}^{*}_{\boldsymbol{iht}} > 0 \\ 0 & if\ \ \boldsymbol{Patent}^{*}_{\boldsymbol{iht}} \leq 0 \end{cases}$$

The second reason for the choice of a Tobit model comes from Autant-Bernard and LeSage (2011) who argue that patenting is an uncertain process. Indeed, even if R&D investments and innovation take place, patenting is a strategic decision that may or may not happen as it depends on other factors such as the cost and benefits of filing. Because we expect unobserved spatial heterogeneity to be present in our panel dataset that covers 2001-2008, we specify our panel data Tobit model with individual specific effects as follows:

$$ln\boldsymbol{Patent}_{\boldsymbol{iht}} = \boldsymbol{x}'_{\boldsymbol{iht}}\boldsymbol{\beta} + \boldsymbol{\mu}_{\boldsymbol{ih}} + \boldsymbol{v}_{\boldsymbol{iht}}$$

Under the assumption of υ following a normal distribution (0, σ_v^2), the log likelihood function of the model above is:

$$\log L_h = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[I_{iht}\log\Phi\left(\frac{y_{iht} - x'_{iht}\beta - \mu_{ih}}{\sigma}\right) + (1 - I_{iht})\left(\log\phi\left(\frac{-x'_{iht}\beta - \mu_{ih}}{\sigma}\right) - \log\sigma_v\right)\right]$$

Where $\Phi(.)$ and $\phi(.)$ denote the standard normal c.d.f. and p.d.f. respectively, and $I_{iht} = \begin{cases} I.\,y_{iht} \text{ if } y_{iht} > 0 \\ 0 \text{ if } y_{iht} \leq 0 \end{cases}$.

MLE maximizes this log-likelihood with respect to β, σ_v^2 and $\mu_1, \ldots, \mu_N$, hence the fixed-effects panel Tobit model suffers from the incidental parameter problem (Neyman and Scott 1948, Lancaster 2000) rendering the estimated coefficients inconsistent unless T approaches infinity

[5] If the variables have a minimum value of zero, we add one before log transformation.

(Greene 2008). Four options are proposed to circumvent this problem. The first assumes that the μ_i effects are independent of the regressors x_{it} and the parameters can be consistently estimated by a random effect model (Wooldridge 2010). However, we believe this assumption is too strong in our case. The second option consists of estimating the fixed effects model which requires T≥8. Heckman and Macurdy (1980) apply the fixed effect MLE to study female labor supply when T = 8 and conclude that the estimators are inconsistent. Another option is the semiparametric estimators initiated in Honoré (1992) for truncated regression models and further developed in (Honoré et al. 2000) for censored regression models.

Here, we select yet another option, discussed in Wooldridge (2010, p. 709), which consists in a conditional Tobit model with time fixed effects and State fixed effects. We justify the use of the time fixed effects through the surge in innovative activities that three of the five sectors have experienced over the study period. Indeed, for the Drugs & Medical industry, the Computer & Communication and the Electrical & Electronic industries, the average number of patents has increased between 17.4% and 25.1% while the stock of private knowledge has increased between 65.8% and 346% (see table 1). When it comes to the State fixed effects, their presence allows us to control for differences in State-level policies and innovative milieus. One well-known example is Entreprise Zones (Ham et al. 2011) but many states also have R&D tax credit incentives (Wilson 2009), state-specific corporate tax rates and non-compete laws (Greenstone and Looney 2011).

Among the independent variables, the stock of knowledge is a major factor of the knowledge production function (Griliches 1979). Here, the county-level knowledge stock is modeled through lagged expenditures in Research and Development (R&D) using the perpetual inventory method (Equation 2) as in Mancusi (2008). In the equation, S_{iht} and RD_{iht} represent the stock of knowledge and the R&D expenditure in county i, sector h at time t. All R&D expenditures are converted in constant 2008 U.S. dollars using each sector's Producer Price Index from the U.S. Bureau of Labor Statistics.[6] We assume a 15% depreciation rate (δ) following Okubo et al. (2006) and Mancusi (2008). In order to calculate the knowledge stock of the initial year, we approximate the industry specific growth rate of R&D expenditures (g) by the average of the annual growth rate over 1990-1999 across the U.S. continental counties. This approach is used for each individual sector as in Mancusi (2008).

$$S_{iht} = (1-\delta) \cdot S_{iht-1} + RD_{iht-1} \ \ and \ \ S_{ih1990} = \left(\frac{RD_{ih1990}}{\delta+g}\right) \quad (2)$$

We model two types of regional ***knowledge stocks: (i) private and (ii) academic R&D***. The private knowledge stock in sector h ($\boldsymbol{Private_{iht}}$) is approximated by the R&D expenditure of private companies collected from Standard and Poor's Compustat (Standard & Poor's 2011). The dataset from Lai et al. (2013) links the raw assignee from patent records with the name and address of the assignee . Here, we use the address of these companies and their NAICS codes to allocate the R&D expenditures across counties and sectors. $\boldsymbol{Private_{iqt}}$ captures R&D expenditures in the four sectors q which are not the sector of interest h. Thus, if significant, the coefficients associated to $Private_{iht}$ and $Private_{iqt}$ measure the importance of intra- and inter-industry externalities on knowledge creation respectively[7].

The regional academic knowledge stock ($\boldsymbol{Univ_{it}}$) is measured by the total amount of R&D spent across universities and colleges according to the National Science Foundation's Survey of R&D expenditures (National Center for Science and Engineering Statistics 2013). In order to match this type of expenditure to a specific county, we use the address of the institutions. Since one academic

[6] The average of the Producer Price Index (PPI) of the NAICS sectors reported in Table 1 is used to calculate the annual PPI of each of our five manufacturing sectors.

field can contribute to several of the five economic sectors under study (e.g. electrical engineering is relevant to mechanical, computer & communication, electrical & electronic), it is difficult to distinguish between intra- and inter-sectoral externalities emanating from the academia and therefore we sum all academic R&D expenditures.

Besides R&D, it is well known that ***human capital*** plays an important role in knowledge creation (Audretsch and Feldman 2004). For instance, Sorensen (1999) unveils that productivity gains from R&D investment become profitable only once human capital reaches a threshold level. If human capital levels are too low, R&D is unprofitable. In order to measure the level of human capital available by county and industrial sector, we use the total number of Graduate (Master's and Doctoral) or professional degree holders who are 25 years and over ($\boldsymbol{Graduate_{iht-1}}$). Shambaugh et al. (2017) have shown that around 70% of patent holders have a least a Master's degree. The use of a one year lag is common in the knowledge production literature to alleviate any possible endogeneity problem (Ponds et al. 2010, Nesta and Saviotti 2005). The data comes from the 2000-2007 Integrated Public Use Microdata Series (IPUMS) developed by Ruggles et al. (2010). Since IPUMS classifies the occupation of the workers according to NAICS, we can easily allocate the number of degree holders by sector. IPUMS is surveyed based on the Public Use Microdata Area (PUMA), thus we match the location of the PUMA with that of the counties based on their 2000 U.S. Census boundaries.[8]

In addition, we control for several county-specific conditions. Regional differences in the economic ***size of each sector*** are captured by the total number of employees in each sector

[8] If a PUMA area consists of more than one county, we allocate the number of degree holders of each sector proportionally to the counties' total number of degree holders (for all industrial sectors) of which data come from the 2000 U.S. Census. This approach is used for the years 2000, 2005, 2006, 2007. IPUMS provides personal information at the state level only for the years 2001-2004. As a result, we first calculate the total sum of degree holders by state and by sector and then distribute it across counties proportionally to their average number of degree holders over the years 2000 and 2005.

($\boldsymbol{Emp_{iht-1}}$). This variable is constructed based on the same method and data as the human capital variable. We also control for the share of large firms in a county's economy ($\boldsymbol{Large_{it-1}}$). Since small firms capitalize better than large firms on the knowledge created in university laboratories according to (Acs et al. 1994), more small firms in a county can be conducive to regional knowledge creation. On the other hand, as large firms contribute to the higher level of agglomeration in a local economy (Acs and Armington 2004), their presence could be more beneficial to regional knowledge creation. We examine the relative role of small or large firms on regional knowledge creation by including the share of establishments with at least 500 employees in our model. This cut-off is used by the 2000 County Business Patterns to define small businesses and it has been used for similar purposes by Acs and Audretsch (1988) and (Anselin et al. 1997).

The degree of ***industrial diversity*** is also included to control for the general economic structure of each locality ($\boldsymbol{Diversity_{it-1}}$). It is measured through the index developed by Duranton and Puga (2000). The presence of this variable is necessary to capture the net effect of MAR vs. Jacobs externalities on innovation as more diverse places appear, by definition, to benefit more from the latter type. The calculation of this variable is reported in Equation (3) where s_{iht} is the share of industry h in county *i's* employment at time *t* and s_{ht} the share of industry h in employment at the national level. The number of employees is measured across 13 industries.[9] This variable changes over time and space but not by sector.

[9] The 13 industries are based on the 2000 US Census classification: 1) Agriculture, forestry, fishing and hunting, and mining, 2) Construction, 3) Manufacturing, 4) Wholesale trade, 5) Retail trade, 6) Transportation and warehousing, and utilities, 7) Information, 8) Finance, insurance, real estate, and rental and leasing, 9) Professional, scientific, management, administrative, and waste management services, 10) Educational, health and social services, 11) Arts, entertainment, recreation, accommodation and food services, 12) Other services (except public administration, 13) Public administration or Industries not classified. We use the 2000 U.S. Census for the diversity index of 2000 and the County Business Patterns for the index over 2003-2007. For the year 2002, the Census Bureau does not provide the number of employees for several sectors. We

$$Diversity_{it} = 1/\sum_h |s_{iht} - s_{ht}| \quad (3)$$

In addition, we account for 1) ***intra-regional and inter-sectoral spillovers***, 2) ***interregional and intra-sectoral spillovers***, 3) ***interregional and inter-sectoral spillovers***. Intra-regional and intra-sectoral spillovers are already accounted for in $\boldsymbol{Private_{iht}}$ since they capture expenses within the same county and sector as the dependent variable. We decide to have the intersectoral knowledge spillovers emanate from private R&D expenses only because, as noted above, university R&D spending in one academic field can contribute to innovation across several sectors.

The three types of spillovers above are modeled based on Lai et al. (2013). Since this data allow us to track the patent creation-citation flows between all 3,109 US continental counties as well as the industrial sector of both the cited and citing patents, we first construct 25 (5×5 sectors) technological network matrices across the 3,109×3,109 counties and then use the 853×3,109 sub-matrices to capture the knowledge flowing to the 853 metropolitan counties only. The fractional counting method is used here too so that we capture all 1/(O×D) knowledge flows between the number of inventors at the place origin O and their peers at the destination D for any pair of origin-destination sectors. This patent creation-patent citation matrix is noted C_{ij}^{hq}. It represents the citation flows from sector h county i to sector q in the MSA county j. This matrix is the basis for the intraregional spillovers as well as the interregional spillovers (below 50 miles and above 50 miles respectively) that will be described further below Here, we use the sum of the patent citation flows over 1997-2000 to model the spillovers of knowledge stocks every single year over 2001-

fill these data with the corresponding values from the 2003 County Business Patterns data and use an average of 2000 and 2002 as a proxy for 2001.

2003 and the sum of the patent citation flows over 2001-2004 for the spillovers measured each year over 2004-2008. It is important that the two windows regarding the patent citation flows are neither too short or too long. The reason is that a short window leads to a low number of citations because there is not enough time to allow it to be cited. On the other hand, if the period is too long, we cannot ensure that the patents from too far back are still relevant in today's knowledge creation. This split also ensures that there is no overlap between the two creation-citation matrices and their time lag with the measurement of the dependent variable avoids any bias coming from reverse causality. We do not rely on past data for the measurement of the stock of knowledge created outside of *j* ($\boldsymbol{Private_j}$) because the same 15% depreciation rate used in Eq. (2) applies to R&D expenditure in *j* and obsolete R&D is not expected to have any impact on knowledge creation.

We make use of matrix C_{ij}^{hq} above to model first the *intra-regional and inter-sectoral spillovers* from private knowledge stock (intra-regional and inter-sectoral externalities are noted $ln\sum_{q\neq h}^{n}\boldsymbol{p_{iiqh}}\,\boldsymbol{Private_{iqt}}$ or ln Private_q in our tables of results for simplicity) as follows:

$$ln\sum_{q\neq h}^{n}\boldsymbol{p_{iiqh}}\,\boldsymbol{Private_{iqt}} = \begin{cases} \ln\sum_{q\neq h} C_{ii}^{hq,1997-2000} \cdot Private_{iqt} & \text{for } t = 2001,\cdots,2003 \\ \ln\sum_{q\neq h} C_{ii}^{hq,2001-2004} \cdot Private_{iqt} & \text{for } t = 2004,\cdots,2008 \end{cases} \tag{4}$$

We follow Kang and Dall'erba (2016a) and normalize the column sums of this matrix to represent the frequency of the citation flows from sector *q* to *h* within the MSA county *i*.

The interregional and intra-sectoral spillovers are modeled as in equations (5) and (6). We distinguish the singular role of short- vs. long-distance interregional knowledge spillovers. The former (equation 5) have a spatial extent limited to 50 miles as in Anselin et al. (1997) and Mukherji and Silberman (2013) since it corresponds to the average daily U.S. commuting distance (Rapino and Fields 2013, Smallen 2004). Distant interregional spillovers (equation 6) correspond to externalities taking place from 50 miles to any farther counties. Distances are based on the great circle distance between the centroids of counties *i* and *j*.

$$ln \sum_{i \neq j}^{N} \boldsymbol{p_{ijhh}}\, \boldsymbol{Private_{jht}} = \text{ ln Private} \leq 50_h =$$

$$\begin{cases} \ln \sum_{j \neq i} C_{ij}^{hh,1997-2000} \cdot Private_{jht} \cdot 1(d(i,j) \leq 50\, miles) & \text{for } t = 2001, \cdots, 2003 \\ \ln \sum_{j \neq i} C_{ij}^{hh,2001-2004} \cdot Private_{jht} \cdot 1(d(i,j) \leq 50\, miles) & \text{for } t = 2004, \cdots, 2008 \end{cases} \tag{5}$$

$$ln \sum_{i \neq j}^{N} \boldsymbol{P_{ijhh}}\, \boldsymbol{Private_{jht}} = \text{ln Private} > 50_h =$$

$$\begin{cases} \ln \sum_{j \neq i} C_{ij}^{hh,1997-2000} \cdot \ln Private_{jht} \cdot 1(d(i,j) > 50\, miles) & \text{for } t = 2001, \cdots, 2003 \\ \ln \sum_{j \neq i} C_{ij}^{hh,2001-2004} \cdot \ln Private_{jht} \cdot 1(d(i,j) > 50\, miles) & \text{for } t = 2004, \cdots, 2008 \end{cases} \tag{6}$$

Where C_{ij}^{hh} is also column standardized so it captures the share of patents created in *i* that cite and rely on patents previously created in *j*.

Interregional Jacobian externalities below that threshold ($ln \sum_{q \neq h}^{n} \sum_{i \neq j}^{N} \boldsymbol{p_{ijqh}}\, \boldsymbol{Private_{jqt}}$ or ln Private $\leq 50_q$) and above it ($ln \sum_{q \neq h}^{n} \sum_{i \neq j}^{N} \boldsymbol{P_{ijqh}}\, \boldsymbol{Private_{jqt}}$ or ln Private $> 50_q$) are built on the same model as equations (5) and (6) but they are captured through the normalized patent creation-citation flows C^{hq} from the other 4 sectors to sector *h*. The same holds true for the definition of the local spillovers (noted ln $\boldsymbol{Univ_{iht}}$), short-distance ($ln \sum_{i \neq j}^{N} \boldsymbol{p_{ijhh}}\, \boldsymbol{Univ_{jht}}$ or ln Univ $\leq$ 50) and long-distance ($ln \sum_{i \neq j}^{N} \boldsymbol{P_{ijhh}}\, \boldsymbol{Univ_{jht}}$ or ln Univ $>$ 50) spillovers of academic knowledge although, as mentioned above, they are not disaggregated by sector.

Tables 2 shows the descriptive statistics for the first and last years of all the variables. It is obvious that the Chemical and Drugs & Medical (hereafter Drugs) sectors generate relatively less patents than the other sectors while the high-tech sectors of Computer & Communication (hereafter Computer) and Electrical & Electronic (hereafter Electrical) display the largest mean values.

[Insert Table 2 here]

In summary, the full model we will estimate can be written as:

$$ln\boldsymbol{Patent_{iht}} = \beta_0 + \beta_1 ln\boldsymbol{Private_{iht}} + \beta_2 ln \textstyle\sum_{q\neq h}^{n} \boldsymbol{p_{iiqh}}\, \boldsymbol{Private_{iqt}} + \beta_3 \ln \boldsymbol{Univ_{it}} +$$

$$\beta_4 \ln \boldsymbol{Graduate_{iht-1}} + \beta_5 ln\boldsymbol{Emp_{iht-1}} + \beta_6 ln\boldsymbol{Large_{it-1}} + \beta_7 ln\boldsymbol{Diversity_{it-1}} +$$

$$\beta_8\, ln\, \boldsymbol{Intra_{iht}} + \beta_9\, ln \textstyle\sum_{i\neq j}^{N} \boldsymbol{p_{ijhh}}\, \boldsymbol{Private_{jht}} + \beta_{10}\, ln \textstyle\sum_{q\neq h}^{n} \textstyle\sum_{i\neq j}^{N} \boldsymbol{p_{ijqh}}\, \boldsymbol{Private_{jqt}} +$$

$$\beta_{11}\, ln \textstyle\sum_{i\neq j}^{N} \boldsymbol{p_{ijhh}}\, \boldsymbol{Univ_{jt}} + \beta_{12}\, ln \textstyle\sum_{i\neq j}^{N} \boldsymbol{P_{ijhh}}\, \boldsymbol{Private_{jht}} + \beta_{13}\, ln \textstyle\sum_{q\neq h}^{n} \textstyle\sum_{i\neq j}^{N} \boldsymbol{P_{ijqh}}\, \boldsymbol{Private_{jqt}} +$$

$$\beta_{14}\, ln \textstyle\sum_{i\neq j}^{N} \boldsymbol{P_{ijhh}}\, \boldsymbol{Univ_{jt}} + \mu_{Sh} + \theta_{ht} + v_{iht} \qquad (7)$$

where $v_{iht} \sim N(0, \sigma_v^2)$; $i = 1, \cdots, 853$; $h = 1, \ldots, 5; t = 2001, \ldots, 2008$; and μ_{Sh} and θ_{ht} capture the State and time fixed effects respectively.

4. Estimation Results

Table 3 shows the Maximum Likelihood estimation results of the fixed-effects Tobit models with a 50-mile distance cut-off. The Hausman test is significant across all specifications, indicating that the State and time fixed effect model is preferred over the random effect model. It confirms our expectations that the covariates are not uncorrelated with the fixed effects. We also report σ, the estimate of the standard deviation of ln(Patent).

All specifications indicate that the local stock of private R&D leads to significant and positive intra-regional intra-sectoral (ln Private_h) and inter-sectoral (ln Private_q) externalities on regional knowledge creation. Furthermore, the latter displays a greater elasticity than the former at the 5% significance level (one-tailed test) for sectors 1, 3, 4, and 5 at the 5% level but not for sector 2. These results confirm the importance of geographical proximity and associated face-to-face interactions to facilitate knowledge creation and knowledge spillovers (Glaeser and Scheinkman

2000). In addition, our results confirm our expectations by reporting that spending in education promotes knowledge creation (Jaffe 1989, Anselin et al. 2000, Kang and Dall'erba 2016a).

We also find a significant positive role of the number of graduate degree holders and employees in a sector on the patenting activity of the same sector. These results confirm the economies of scale that can be achieved with spatial agglomeration in conjunction with specialization and MAR externalities. While the elasticity of human capital ranges from across all sectors, the role of employment is particularly acute in the Mechanical sector compared to the other sectors. Finally, we find that the greater is the presence of large establishments the more knowledge is created across all sectors but particularly Computer & Communications. Acs and Armington (2004) indicate that large firms lead to a greater local labor pool which contributes to agglomeration economies and, as seen earlier, innovation. The industrial diversity within the county also shows a positive effect on the patenting activity for all sectors.

[Insert Table 3 here]

The results related to all types of interregional spillovers appear in the middle part of table 3. Unlike the marginal effect of the local variables, spillovers display very different magnitudes and significance level across sectors. For instance, interregional spillovers do not display a significant role in the Chemical industry (column 1). This result reveals the dominance of localized face-to-face contacts in this industry as highlighted by Mariani (2000) in the European context. On the other hand, both short-distance intersectoral and long-distance intrasectoral spillovers have a significant impact on patenting in the Drug & Medical industry. This result indicates that geographical proximity is not a requirement to transfer basic research knowledge in this sector.

This result is in tune with the findings of Gittelman (2007) that indicate that the collaboration network of the U.S. biotechnology industry is spread geographically. The results indicate also that local (ln Private_q) and short distance (ln $\text{Private} \leq 50_q$) intersectoral spillovers do not display a statistically different marginal return, indicating that this type of externalities go beyond the boundaries of the county of interest. This result is in sharp contrast with the intrasectoral spillovers (ln Private_h) of which magnitude is statistically above the one of the long-distance spillovers (ln $\text{Private} > 50_h$). Yet, the significance level of the latter matters as it indicates that the nation-wide knowledge pipelines this sector relies on are just as important as those built on interactions with close neighbors (Audretsch and Feldman 1996, Sonn and Storper 2008).

This later result holds true for the Mechanical sector too. Its estimates are reported in column 3. This sector and Computer & Communications are the only ones for which the short-distance spillovers emanating from the university (ln Univ ≤50) are significant. In the case of Mechanical, their order of magnitude is the same as the one of the local spillovers (ln Univ). On the other hand, long-distance university spillovers (ln Univ > 50) do not display a significant impact. These results corroborate with Mansfield (1995) who finds that geographical proximity of academic research plays a greater role in fields that require applied R&D. Autant-Bernard and LeSage (2011) conclude also that the spillovers arising from the university R&D are localized. However, our results contradict the ones of Jaffe (1989) and Anselin et al. (2000) who find no evidence of localized spillovers of university research in the Machinery sector. The difference in the results may come from the spatial units under study as they use state and MSA level data respectively, which leads to more aggregate results than our county-level approach. We also find the presence of a competitive effect across nearby counties as intra-sectoral spillovers of private R&D investments (ln $\text{Private} \leq 50_h$) display a negative elasticity.

When it comes to Computer & Communications (column 4), all the interregional spillovers that display a significant marginal effect are limited to short-distance. While the intrasectoral spillovers display a negative impact, the sign does not hold in the robustness tests we display further below. On the other hand, we find that intersectoral spillovers positively support local patenting. Their marginal effect is statistically greater than the one of the local intersectoral and intrasectoral spillovers, which reflects that companies in this sector rely on a network that extends to the counties nearby. These results are in line with those of Kang and Dall'erba (2016b) and Anselin et al (2000) who also conclude that the Computer & Communications industry, experiences significant short-distance interregional knowledge spillovers. However, their estimates do not differentiate intra- vs intersectoral externalities.

Finally, the results for the Electrical and Electronic sector (column 5) show that intersectoral short-distance spillovers display a significant impact on patenting. Their magnitude is not as large as the magnitude of their respective local effects though, which suggests that frequent face-to-face contacts are still the main mechanism to facilitate the diffusion of ideas in this sector.

To summarize the findings regarding spillovers, the Tobit estimations suggest that that none of the sectors observe positive intra-sectoral spillovers within 50 miles ($\ln \text{Private} \leq 50_h$). However, innovation in the Drugs and Medical as well as the Mechanical sectors benefits from long-distance intra-sectoral spillovers ($\ln \text{Private} > 50_h$). Regarding inter-sectoral spillovers, results show that they do not operate long-distant and only benefit three of the sectors within the 50 miles radius. Similar results are found for university R&D which also only show positive results for two sectors within the 50-mile radius. These findings stress the sectoral heterogeneity of knowledge spillovers on knowledge creation.

5. Robustness checks

Our first robustness check focuses on the fixed effects Tobit model where the interregional spillovers are pooled (no split above/below 50 miles). The results are displayed in table 4. The significant role of intrasectoral spillovers in the Drugs & Medical as well as the Mechanical sectors confirm the results found earlier. Their magnitude is very similar to the one of $\ln \text{Private} > 50_h$ in table 3, which suggests it is mostly long-distance spillovers that drive the current results. We also find that intersectoral spillovers have a significant role in patenting in Drugs and Medical as well as Computer and Communications. Based on the results of table 3, it is very likely that it is short-distance spillovers that drive this finding for the latter industry.

[Insert Table 4 here]

The second robustness check consists in averaging the time periods noted in Eq. (7) into two time periods, 2001-2003 and 2004-2008, and use the matrices of patent creation-patent citation described in (Eqs. 4-6) only once over the corresponding time period. The estimates are reported in table 5. Since the local marginal effects meet our expectations, we focus on the results for the interregional spillovers. While they are relatively consistent with those of Table 3, a few differences com to light. First, the negative impact of $\ln \text{Private} < 50_h$ in the Mechanical and Computers & Communications sectors is not significant anymore. Second, the short-distance intra-sectoral spillovers are not significant in the Electrical & Electronic sector anymore. The short-distance university spillovers for Computers and Communication become non-significant too. On the other hand, we observe statistically significant results for long-distance intra-sectoral spillovers in Computers and Communication as well as long-distance inter-sectoral spillovers for the Drugs

& Medical industry. Note that the variance of any coefficient has increased compared to Table 3 since the sample is four times smaller.

Finally, the last robustness check we perform consists in changing the cut-off of short- vs long-distance spillovers to 75 miles as the U.S. Department of Transportation reports that as many as 3.3 million Americans are ‘‘stretch commuters’’ traveling more than 50 miles one-way to work. Stretch commuters living in rural areas drive up to 99 miles daily according to Smallen (2004). When we run our estimates with this new cut-off, we find that all the results are very consistent with those displayed here[10].

[Insert Table 5 here]

6. Conclusion

The regional knowledge production literature has given an increasing amount of attention to the role of spatial spillovers on knowledge creation. However, the bulk of empirical studies relies on an aggregated approach that masks the differences in the marginal effect of intrasectoral and intersectoral R&D investments on knowledge creation across sectors. The few exceptions (e.g. Jaffe (1989), Anselin et al. (2000) have highlighted the presence of sectoral heterogeneity in the size of the localized knowledge spillovers emanating from university research. However, they have not investigated how intersectoral and distant interregional knowledge spillovers matter. Autant-Bernard and LeSage (2011) have remedied to this problem but at the cost of providing estimates averaged across all sectors. Furthermore, a large amount of the literature models the flows of regional spillovers of knowledge based on variables such as geographical proximity (Audretsch

[10] Complete results available from the authors upon request.

and Feldman 2004, Jaffe et al. 1993), technological proximity (Maggioni et al. 2011), or social/professional networks (Breschi and Lissoni 2009, Crescenzi et al. 2016). These approaches do not explicitly account for the directionality of the flows, which can be challenging when trying to establish causality in an econometric model.

This paper contributes to this literature by examining the heterogeneous role of intra- and interregional as well as intra- and inter-sectoral knowledge spillovers across 5 U.S. manufacturing sectors that cover 82% of the patents recorded in USPTO. In addition, interregional spillovers are measured through a matrix of patent creation – patent citation as in Peri (2005) and Kang and Dall'erba (2016a) that allows us to explicitly account for the directionality of the flows of knowledge.

Measured over 853 US metropolitan counties and in the frame of a Tobit model with State fixed effects, our results show that both intra-sectoral and inter-sectoral spillovers taking place within a county are important determinants of knowledge production. It implies that frequent face-to-face contacts are still an important factor for the creation of new knowledge. When it comes to the interregional spillovers of private and university knowledge, the relative role of each type depends on the industry under study suggesting that there is strong heterogeneity across sectors on the mechanisms of how new knowledge is created. These differences are not only visible in terms of different sensitivity to geographical proximity, but also depending on whether spillovers arise from the same industry, from others and/or from academic research.

Our estimation results suggest three important implications for policy-makers interested in more efficient innovations strategies. First, both intra-sectoral and inter-sectoral externalities matter for innovation. It is surprising to see how much the local stock of intersectoral knowledge affects patenting in some sectors. As a result, various innovation strategies can be compared with each other and ranked accordingly only when the cumulative process of scientific discovery across industries has been accounted for. Second, because close geographic proximity matters a great deal for the innovative capacity of each industry, policy makers need to facilitate both academic and private sector R&D through, for instance, Entreprise Zones and R&D tax credit incentives (Ham et al, 2011), and help build local networks of university-industry collaboration (Ponds et al. 2010). Third, innovation policies need to be more attentive to the heterogeneity in the source and spatial extent of the interregional spillovers that affect patenting in each industry. The recommendations based on the traditional aggregated approach mask this heterogeneity and can lead to inefficient policies whereby, for instance, long-distance network would be supported while an industry, like Computer & Communications industry, depends on local and short-distance spillovers only.

List of references

Acemoglu, Daron, Ufuk Akcigit, and William R. Kerr. 2016. "Innovation network." *Proceedings of the National Academy of Sciences* 113 (41):11483-11488.

Acs, Zoltan, Luc Anselin, and Attila Varga. 2002. "Patents and Innovation Counts as Measures of Regional Production of New Knowledge." *Research Policy* 31 (7):1069-1085.

Acs, Zoltan, and Catherine Armington. 2004. "Employment Growth and Entrepreneurial Activity in Cities." *Regional Studies* 38 (8):911-927.

Acs, Zoltan J., and David B. Audretsch. 1988. "Innovation in Large and Small Firms: An Empirical Analysis." *The American Economic Review* 78 (4):678-690.

Acs, Zoltan J., David B. Audretsch, and Maryann P. Feldman. 1994. "R & D Spillovers and Recipient Firm Size." *The Review of Economics and Statistics* 76 (2):336-340.

Almeida, Paul , and Bruce Kogut. 1999. "Localization of Knowledge and the Mobility of Engineers in Regional Networks." *Management Science* 45 (7):905-917.

Anselin, Luc, Attila Varga, and Zoltan Acs. 1997. "Local Geographic Spillovers between University Research and High Technology Innovations." *Journal of Urban Economics* 42 (3):422-448.

Anselin, Luc, Attila Varga, and Zoltan Acs. 2000. "Geographical Spillovers and University Research: A Spatial Econometric Perspective." *Growth and Change* 31 (4):501-515.

Audretsch, David, and Maryann Feldman. 2004. "Knowledge Spillovers and the Geography of Innovation." In *Handbook of Regional and Urban Economics*, edited by J Vernon Henderson and Jacques-Francois Thisse, 2713 - 2739. Amsterdam: Elsevier.

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

Autant-Bernard, Corinne. 2001. "Science and knowledge flows: evidence from the French case." *Research Policy* 30 (7):1069-1078.

Autant-Bernard, Corinne. 2012. "Spatial Econometrics of Innovation: Recent Contributions and Research Perspectives." *Spatial Economic Analysis* 7 (4):403-419.

Autant-Bernard, Corinne, and James LeSage. 2011. "Quantifying knowledge spillovers using spatial econometric models." *Journal of Regional Science* 51 (3):471-496.

Autor, David, David Dorn, Gordon H Hanson, Gary Pisano, and Pian Shu. 2016. Foreign competition and domestic innovation: Evidence from US patents. National Bureau of Economic Research.

Beaudry, Catherine, and Andrea Schiffauerova. 2009. "Who's right, Marshall or Jacobs? The localization versus urbanization debate." *Research Policy* 38 (2):318-337.

Bode, Eckhardt. 2004. "The spatial pattern of localized R&D spillovers: an empirical investigation for Germany." *Journal of Economic Geography* 4 (1):43-64.

Bottazzi, Laura, and Giovanni Peri. 2003. "Innovation and spillovers in regions: Evidence from European patent data." *European Economic Review* 47 (4):687-710.

Breschi, Stefano, and Francesco Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of Economic Geography* 9 (4):439-468.

Cai, Jie, and Nan Li. 2018. "Growth Through Inter-sectoral Knowledge Linkages." *The Review of Economic Studies* 86 (5):1827-1866.

Cameron, Adrian Colin, and Pravin K Trivedi. 2009. *Microeconometrics Using Stata*. Vol. 5. Texas: Stata press College Station.

Cincera, Michele. 1997. "Patents, R&D, and technological spillovers at the firm level: some evidence from econometric count models for panel data." *Journal of Applied Econometrics*:265-280.
Crescenzi, Riccardo, Max Nathan, and Andrés Rodríguez-Pose. 2016. "Do inventors talk to strangers? On proximity and collaborative knowledge creation." *Research Policy* 45 (1):177-194.
Duranton, Gilles, and Diego Puga. 2000. "Diversity and Specialisation in Cities: Why, Where and When Does it Matter?" *Urban Studies* 37 (3):533-555.
Feldman, Maryann, and Dieter Kogler. 2010. "Chapter 8 - Stylized Facts in the Geography of Innovation." In *Handbook of the Economics of Innovation*, edited by Bronwyn H. Hall and Nathan Rosenberg, 381-410. North-Holland.
Feldman, Maryann P. 1994. *The Geography of Innovation*. Dordrecht: Kluwer Academic.
Fischer, Manfred, and Attila Varga. 2003. "Spatial knowledge spillovers and university research: Evidence from Austria." *The Annals of Regional Science* 37 (2):303-322.
Galasso, Alberto, and Mark Schankerman. 2018. "Patent rights, innovation, and firm exit." *The RAND Journal of Economics* 49 (1):64-86.
Gittelman, Michelle 2007. "Does Geography Matter for Science-Based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology." *Organization Science* 18 (4):724-741.
Glaeser, Edward L, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. 1992. "Growth in Cities." *Journal of Political Economy*:1126–1152.
Glaeser, Edward L, and Jose Scheinkman. 2000. Non-Market Interactions. National Bureau of Economic Research.
Greene, William H. 2008. *Econometric analysis*: Granite Hill Publishers.
Greenstone, Michael, and Adam Looney. 2011. "A dozen economic facts about innovation." *The Hamilton Project*:3.
Griliches, Zvi. 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." *The Bell Journal of Economics* 10 (1):92-116.
Groot, Henri L.F., Jacques Poot, and Martijn J. Smit. 2016. "Which agglomeration externalities matter most and why?" *Journal of Economic Surveys* 30 (4):756-782.
Grossman, Gene M., and Elhanan Helpman. 1994. "Endogenous Innovation in the Theory of Growth." *Journal of Economic Perspectives* 8 (1):23-44.
Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. In *NBER Woring Paper Series 8498*. Cambridge, MA: National Bureau of Economic Research.
Ham, John C., Charles Swenson, Ayşe İmrohoroğlu, and Heonjae Song. 2011. "Government programs can improve local labor markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Community." *Journal of Public Economics* 95 (7):779-797.
Heckman, James J., and Thomas E. Macurdy. 1980. "A Life Cycle Model of Female Labour Supply." *The Review of Economic Studies* 47 (1):47-74.
Honoré, Bo E. 1992. "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects." *Econometrica* 60 (3):533-565.
Honoré, Bo E., Ekaterini Kyriazidou, and J. L. Powell. 2000. "Estimation of tobit-type models with individual specific effects." *Econometric Reviews* 19 (3):341-366.
Jacobs, Jane. 1969. *The Economy of Cities* New York: Random House, Inc.

Jaffe, Adam B. 1986. "Technological Opportunity and Spillovers of R&D: Evidence From Firms' Patents, Profits and Market Value." *American Economic Review* 76:984-1001.
Jaffe, Adam B. 1989. "Real Effects of Academic Research." *The American Economic Review* 79 (5):957-970.
Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108 (3):577-598.
Kang, Dongwoo, and Sandy Dall'erba. 2016a. "An Examination of the Role of Local and Distant Knowledge Spillovers on the US Regional Knowledge Creation." *International Regional Science Review* 39 (4):355-385.
Kang, Dongwoo, and Sandy Dall'erba. 2016b. "Exploring the spatially varying innovation capacity of the US counties in the framework of Griliches' knowledge production function: a mixed GWR approach." *Journal of Geographical Systems* 18 (2):125-157.
Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Felming. 2013. Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010). Harvard Dataverse.
Lancaster, Tony. 2000. "The incidental parameter problem since 1948." *Journal of Econometrics* 95 (2):391-413.
Maggioni, Mario A., Teodora Erika Uberti, and Stefano Usai. 2011. "Treating Patents as Relational Data: Knowledge Transfers and Spillovers across Italian Provinces." *Industry and Innovation* 18 (1):39-67.
Mancusi, Maria Luisa. 2008. "International spillovers and absorptive capacity: A cross-country cross-sector analysis based on patents and citations." *Journal of International Economics* 76 (2):155-165.
Mansfield, Edwin. 1995. "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing." *The Review of Economics and Statistics* 77 (1):55-65.
Mariani, Myriam. 2000. "Title." WP No. 2/20-019, Economic Research Institute on Innovation and Technology (MERIT).
Meyer-Krahmer, Frieder, and Ulrich Schmoch. 1998. "Science-based technologies: university–industry interactions in four fields." *Research Policy* 27 (8):835-851.
Moreno, Rosina, and Ernest Miguélez. 2012. "A relational approach to the geography of innovation: a typology of regions." *Journal of Economic Surveys* 26 (3):492-516.
Moretti, Enrico, and Daniel J Wilson. 2017. "The effect of state taxes on the geographical location of top earners: evidence from star scientists." *American Economic Review* 107 (7):1858-1903.
Mukherji, Nivedita, and Jonathan Silberman. 2013. "Absorptive Capacity, Knowledge Flows, and Innovation in U.S. Metropolitan Areas." *Journal of Regional Science* 53 (3):392-417.
National Center for Science and Engineering Statistics. 2013. Guide for Public Use Data Files National Science Foundation's Higher Education Research and Development Survey: Fiscal Year 2011. In *Yearly*. Arlington, Virginia NSF National Center for Science and Engineering Statistics.
Nesta, Lionel, and Pier Paolo Saviotti. 2005. "Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical industry." *The Journal of Industrial Economics* 53 (1):123-142.

Neyman, Jerzy, and Elizabeth L Scott. 1948. "Consistent estimates based on partially consistent observations." *Econometrica* 16 (1):1-32.
Okubo, Sumiye, Carol A. Robbins, Carol E. Moylan, Brian K. Sliker, Laura I. Schultz, and Lisa S. Mataloni. 2006. R&D Satellite Account: Preliminary Estimates. Washington, DC: Bureau of Economic Analysis, US Department of Commerce.
Orlando, Michael J. 2004. "Measuring Spillovers from Industrial R&D: On the Importance of Geographic and Technological Proximity." *The RAND Journal of Economics* 35 (4):777-786.
Owen-Smith, Jason, and Walter W Powell. 2004. "Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community." *Organization Science* 15 (1):5-21.
Parent, Olivier , and James LeSage. 2008. "Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers." *Journal of Applied Econometrics* 23 (2):235-256.
Peri, Giovanni. 2005. "Determinants of Knowledge Flows and Their Effect on Innovation." *The Review of Economics and Statistics* 87 (2):308-322.
Ponds, Roderik, Frank van Oort, and Koen Frenken. 2010. "Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach." *Journal of Economic Geography* 10 (2):231-255.
Ramani, Shyama V., Mhamed-Ali El-Aroui, and Myriam Carrère. 2008. "On estimating a knowledge production function at the firm and sector level using patent statistics." *Research Policy* 37 (9):1568-1578.
Rapino, Melanie A., and Alison K. Fields. 2013. "Title." Working Paper 2013-03.
Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5):1002-1037.
Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.
Saxenian, AnnaLee. 1994. *Regional Advantage*. Cambridge, MA: Harvard University Press.
Shambaugh, Jay, Ryan Nunn, and Becca Portman. 2017. "Eleven facts about innovation and patents." *The Hamilton Proj ect (Brookings Institution)*.
Smallen, Dave. 2004. "3.3 Million Americans are Stretch Commuters Traveling at Least 50 Miles One-Way to Work." United States Department of Transportation, accessed November 16. https://www.rita.dot.gov/bts/sites/default/files/rita_archives/bts_press_releases/2004/bts010_04/html/bts010_04.html.
Sonn, Jung Won, and Michael Storper. 2008. "The Increasing Importance of Geographical Proximity in Knowledge Production: An Analysis of US Patent Citations, 1975–1997." *Environment and Planning A: Economy and Space* 40 (5):1020-1039.
Sorensen, Anders. 1999. "R&D, Learning, and Phases of Economic Growth." *Journal of Economic Growth* 4 (4):429-445.
Standard & Poor's. 2011. "Standard & Poor's Compustat® Xpressfeed Understanding the Data." In: The McGraw-Hill Companies, Inc.
Trippl, Michaela, Franz Tödtling, and Lukas Lengauer. 2009. "Knowledge Sourcing Beyond Buzz and Pipelines: Evidence from the Vienna Software Sector." *Economic Geography* 85 (4):443-462.

USPTO. 2010. Patents BIB: Selected Bibliographic Information From US Patents Issued 1969 to Present. edited by the US Patent & Trademark Office. Alexandria, VA.
Verhoeven, Dennis, Jurriën Bakker, and Reinhilde Veugelers. 2016. "Measuring technological novelty with patent-based indicators." *Research Policy* 45 (3):707-723.
Wallsten, Scott J. 2001. "An empirical test of geographic knowledge spillovers using geographic information systems and firm-level data." *Regional Science and Urban Economics* 31 (5):571-599.
Wilson, Daniel J. 2009. "Beggar Thy Neighbor? The In-State, Out-of-State, and Aggregate Effects of R&D Tax Credits." *The Review of Economics and Statistics* 91 (2):431-436.
Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. Cambridge: MIT press.

Tables

Table 1 Classification of Industrial sectors

Sector	NAICS 2002	Description
Chemical	325	Chemicals
	3251	Basic Chemicals
	3252	Resin, Synthetic Rubber, and Artificial and Synthetic Fibers and Filaments
	3253, 3255, 3256, 3259	Other Chemical Product and Preparation
Drugs & Medical	3254	Pharmaceutical and Medicines
	3391	Medical Equipment and Supplies
Mechanical	333	Machinery
	336	Transportation Equipment
	3361, 3162, 3363	Motor Vehicles, Trailers and Parts
	3364	Aerospace Product and Parts
	3365, 3366, 3369	Other Transportation Equipment
Computer & Communication	334	Computer and Electronic Products
	3341	Computer and Peripheral Equipment
	3342	Communications Equipment
Electrical & Electronic	3344	Semiconductors and Other Electronic Components
	3345	Navigational, Measuring, Electromedical, and Control Instruments
	3343, 3346	Other Computer and Electronic Products
	335	Electrical Equipment, Appliances, and Components

Note: Classification is based on four-digit NAICS 2002 codes. When four-digit codes are not available for patents and private R&D expenditures, we use their three-digit codes. This classification is defined to make the patent data compatible with the patent citation data.

Table 2 Descriptive Statistics (continued): Five manufacturing sectors

	Chemical		Drugs & Medical		Mechanical		Computer & Communication		Electrical & Electronic	
Variables	Mean 2001	Mean 2008	Mean 2001	Mean 2008	Mean 2001	Mean 2008	Mean 2001	Mean 2008	Mean 2001	Mean 2008
Patents_h	11.278	10.967	10.111	11.880	15.721	17.346	30.540	38.216	37.001	40.707
Private_h	9.69×10^7	5.76×10^7	1.93×10^8	3.20×10^8	1.90×10^8	1.74×10^8	2.23×10^7	9.96×10^7	6.30×10^7	1.76×10^8
Private_q	4.68×10^8	7.69×10^8	3.72×10^8	5.07×10^8	3.75×10^8	6.53×10^8	5.43×10^8	7.27×10^8	5.02×10^8	6.51×10^8
Univ	9.66×10^7	2.29×10^8	9.66×10^7	2.29×10^8	9.66×10^7	2.29×10^8	9.66×10^7	2.29×10^8	9.66×10^7	2.29×10^8
Grad_h	92.248	88.387	148.765	211.466	281.553	328.224	144.259	116.766	210.967	228.002
Emp_h	725.931	765.650	854.023	970.648	3,444.377	3,256.495	761.133	491.021	1,769.287	1,427.791
Large	13.690	14.979	13.690	14.979	13.690	14.979	13.690	14.979	13.690	14.979
Diversity	5.541	3.292	4.541	3.292	4.541	3.292	4.541	3.292	4.541	3.292
$\text{Private} \leq 50_h$	65,097.92	113,014.3	161,861.3	2,303,204	225,103.1	153,028.2	32,201.6	26,867.27	396,780.9	269,864.4
$\text{Private} \leq 50_q$	101,899.9	93,855.84	125,742.5	608,469.1	1,222,438	1,078,569	5,405.042	123,132.3	454,583.9	751,299.8
$\text{Univ} \leq 50$	1,472,010	3,123,482	1,001,198	4,975,501	456,260.1	5,049,239	964,874.6	3,634,151	785,929.5	3,537,675
$\text{Private} > 50_h$	742,849.2	400,242.3	4,091,602	5.62×10^7	7,450,628	1.04×10^7	685,553.7	2,821,082	2,495,875	4,672,901
$\text{Private} > 50_q$	2,702,108	933,815	4.34×10^7	3.22×10^7	3,353,677	7,530,756	1,944,051	5,519,791	2,697,406	9,936,554

Univ > 50	8,659,084	1.54×10^7	1.74×10^7	2.37×10^7	9,715,928	1.38×10^7	1.89×10^7	2.56×10^7	1.95×10^7	3.00×10^7
Private spillovers$_h$	807,947.1	513,256.6	4,253,463	5.85×10^7	7,675,731	1.06×10^7	717,755.3	2,847,949	2,892,656	4,942,765
Private spillovers$_q$	2,804,008	1,027,671	4.36×10^7	3.28×10^7	4,576,115	8,609,325	1,949,457	5,642,924	3,151,990	1.07×10^7
University spillovers	1.01×10^6	1.85×10^7	1.84×10^7	2.87×10^7	1.02×10^7	1.89×10^7	1.98×10^7	2.93×10^7	2.03×10^7	3.35×10^7

Note: h stands for the sector recorded in the column, q stands for the other four sectors. ≤50 means a spillover taking place between the county of interest and any county located within 50 miles. >50 means a spillover taking place between the county of interest and any county located farther away than 50 miles.

Table 3 Fixed Effects Tobit Model with 50-mile distance cut-off spillovers

Dep.: ln Patent	Chemical	Drugs & Medical	Mechanical	Comp. & Comm.	Electrical &
$\ln \text{Private}_h$	0.022***	0.026***	0.018***	0.030***	0.027***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
$\ln \text{Private}_q$	0.051***	0.029***	0.040***	0.042***	0.035***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
ln Univ	0.051***	0.038***	0.021***	0.035***	0.031***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\ln \text{Grad}_h$	0.206***	0.132***	0.128***	0.180***	0.169***
	(0.011)	(0.012)	(0.011)	(0.011)	(0.010)
$\ln \text{Emp}_h$	0.162***	0.275***	0.423***	0.242***	0.341***
	(0.017)	(0.018)	(0.021)	(0.012)	(0.017)
ln Large	0.325***	0.458***	0.183***	0.546***	0.343***
	(0.076)	(0.074)	(0.067)	(0.072)	(0.063)
ln Diversity	0.723***	0.626***	0.448***	0.495***	0.463***
	(0.074)	(0.073)	(0.058)	(0.069)	(0.061)
$\ln \text{Private} \leq 50_h$	0.003	-0.008	**-0.017***	**-0.029****	0.013
	(0.014)	(0.013)	**(0.009)**	**(0.013)**	(0.008)
$\ln \text{Private} \leq 50_q$	-0.014	**0.030****	-0.010	**0.083*****	**0.029*****
	(0.019)	**(0.015)**	(0.011)	**(0.015)**	**(0.011)**
$\ln \text{Univ} \leq 50$	-0.001	0.004	**0.021*****	**0.018*****	0.005
	(0.008)	(0.007)	**(0.006)**	**(0.006)**	(0.006)
$\ln \text{Private} > 50_h$	0.001	**0.016****	**0.016*****	0.008	0.004
	(0.007)	**(0.007)**	**(0.006)**	(0.005)	(0.004)
$\ln \text{Private} > 50_q$	-0.006	0.007	-0.004	0.006	-0.007
	(0.009)	(0.005)	(0.006)	(0.006)	(0.005)
ln Univ > 50	0.007	-0.005	-0.005	-0.003	0.002
	(0.005)	(0.006)	(0.005)	(0.004)	(0.004)
Year & State FE	Yes	Yes	Yes	Yes	Yes
Constant	-7.598***	-8.839***	-5.451***	-9.394***	-5.562***
	(0.355)	(0.372)	(0.229)	(0.308)	(0.235)
σ	4.032	3.992	1.538	3.040	1.821
	(0.088)	(0.091)	(0.037)	(0.069)	(0.042)
Counties	853	853	853	853	853
Time periods	8	8	8	8	8
Observations	6,824	6,824	6,824	6,824	6,824
Hausman p-value	0.000	0.000	0.000	0.000	0.000

Note: Robust standard errors in parenthesis. *** p<0.01. ** p<0.05. * p<0.1

h stands for the sector recorded in the column, *q* stands for the other four sectors. ≤50 means a spillover taking place between the county of interest and any county located within 50 miles. >50 means a spillover taking place between the county of interest and any county located farther away than 50 miles.

Table 4 Fixed Effects Tobit Model with aggregated inter-regional spillovers

Dep.: ln Patent	Chemical	Drugs & Medical	Mechanical	Comp. & Comm.	Electrical & Electronic
ln Private_h	0.022***	0.026***	0.018***	0.030***	0.027***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
ln Private_q	0.051***	0.029***	0.040***	0.041***	0.035***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
ln Univ	0.050***	0.038***	0.020***	0.035***	0.031***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ln Grad_h	0.206***	0.132***	0.129***	0.180***	0.170***
	(0.011)	(0.012)	(0.011)	(0.011)	(0.010)
ln Emp_h	0.162***	0.275***	0.424***	0.242***	0.338***
	(0.017)	(0.018)	(0.021)	(0.012)	(0.017)
ln Large	0.327***	0.458***	0.178***	0.543***	0.332***
	(0.076)	(0.074)	(0.067)	(0.072)	(0.063)
ln Diversity	0.720***	0.627***	0.450***	0.495***	0.448***
	(0.073)	(0.073)	(0.058)	(0.069)	(0.061)
ln Private spillovers_h	0.002	**0.014****	**0.011****	0.002	0.007
	(0.007)	**(0.006)**	**(0.005)**	(0.005)	(0.004)
ln Private spillovers_q	-0.009	**0.010***	-0.008	**0.015*****	0.005
	(0.008)	**(0.005)**	(0.005)	**(0.006)**	(0.004)
ln University spillovers	0.007	-0.004	0.001	0.002	-0.000
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
Year & State FE	Yes	Yes	Yes	Yes	Yes
Constant	-7.596***	-8.839***	-5.441***	-6.414***	-5.488***
	(0.355)	(0.372)	(0.229)	(0.308)	(0.234)
σ	4.032	3.994	1.540	3.054	1.828
	(0.088)	(0.091)	(0.229)	(0.069)	(0.042)
Counties	853	853	853	853	853
Time periods	8	8	8	8	8
Observations	6,824	6,824	6,824	6,824	6,824
Hausman p-value	0.000	0.000	0.000	0.000	0.000

Note: Robust standard errors in parenthesis. *** p<0.01. ** p<0.05. * p<0.1

h stands for the sector recorded in the column, *q* stands for the other four sectors.

Table 5 Fixed Effects Tobit Model with 50-mile distance cut-off spillovers (T=2)

Dep.: ln Patent	Chemical	Drugs & Medical	Mechanical	Comp. & Comm.	Electrical & Electronic
ln Private_h	0.022***	0.022***	0.017***	0.028***	0.022***
	(0.006)	(0.005)	(0.003)	(0.005)	(0.005)
ln Private_q	0.051***	0.028***	0.037***	0.036***	0.038***
	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)
ln Univ	0.051***	0.036***	0.019***	0.034***	0.027***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)
ln Grad_h	0.311***	0.118***	0.206***	0.213***	0.225***
	(0.030)	(0.028)	(0.027)	(0.028)	(0.026)
ln Emp_h	0.222***	0.527***	0.386***	0.449***	0.466***
	(0.046)	(0.046)	(0.042)	(0.036)	(0.042)
ln Large	0.384**	0.396***	0.101	0.439***	0.229*
	(0.160)	(0.136)	(0.124)	(0.158)	(0.127)
ln Diversity	0.727***	0.327**	0.383***	0.287**	0.362***
	(0.163)	(0.150)	(0.105)	(0.143)	(0.123)
ln $\text{Private} \leq 50_h$	-0.019	-0.011	-0.011	-0.008	0.014
	(0.023)	(0.020)	(0.014)	(0.021)	(0.012)
ln $\text{Private} \leq 50_q$	0.004	**0.065*** **	-0.012	**0.058** **	0.025
	(0.036)	**(0.016)**	(0.017)	**(0.025)**	(0.016)
ln Univ ≤ 50	0.001	-0.010	**0.016***	0.015	0.012
	(0.013)	(0.012)	**(0.010)**	(0.012)	(0.010)
ln $\text{Private} > 50_h$	0.006	0.015	0.011	**0.021** **	0.010
	(0.012)	(0.010)	(0.007)	**(0.009)**	(0.007)
ln $\text{Private} > 50_q$	0.010	**0.015***	-0.001	-0.009	-0.012
	(0.013)	**(0.009)**	(0.008)	(0.012)	(0.008)
ln Univ > 50	0.002	-0.007	-0.007	-0.003	-0.002
	(0.009)	(0.007)	(0.005)	(0.007)	(0.006)
Year & State FE	Yes	Yes	Yes	Yes	Yes
Constant	-6.811***	-7.010***	-4.184***	-5.141***	-4.818***
	(0.586)	(0.521)	(0.375)	(0.489)	(0.375)
σ	2.585	2.220	0.918	1.995	1.299
	(0.112)	(0.099)	(0.049)	(0.096)	(0.062)
Counties	853	853	853	853	853
Time periods	2	2	2	2	2
Observations	1,706	1,706	1,706	1,706	1,706
Hausman p-value	0.000	0.000	0.000	0.000	0.000

Note: Robust Sstandard errors in parenthesis. *** p<0.01. ** p<0.05. * p<0.1

h stands for the sector recorded in the column, q stands for the other four sectors. ≤50 means a spillover taking place between the county of interest and any county located within 50 miles. >50 means a spillover taking place between the county of interest and any county located farther away than 50 miles.