(2018) 48, 193–216

The Review of Regional Studies

The Official Journal of the Southern Regional Science Association

Knowledge Flow Among U.S. Metro Areas: Innovative Activity, Proximity, and the Border Effect*

Nivedita Mukherji and Jonathan Silberman

Department of Economics, Oakland University

Abstract: Knowledge spillovers are critical for innovation and new value creation in an increasingly knowledge-intensive economy. The substantial scholarly attention on knowledge spillovers has shown that there is a rapid distance decay associated with knowledge spillovers and that there is a positive state border effect. We show that the effects of distance, technology proximity, and the state border effect on knowledge flows are dependent on the size of the regions (MSAs) involved in the knowledge flow. Not accounting for innovation size (innovative communities and social relationships) in the flow of knowledge across origin-destination regions results in aggregation bias in the parameter estimates. Knowledge spillovers are more localized for small innovation MSAs than for large ones. Distance is not as much of a resistance factor in knowledge flow for larger innovation metro areas compared with smaller regions. Spatial origin and destination effects due to technology compatibility of neighboring regions do not affect the knowledge flow among large innovation MSAs, but do have an effect when small MSAs interact with large MSAs.

Keywords: knowledge flows, patent citations, innovation, city size, distance, spatial dependence

JEL Codes: O30, O33, R10, R12

1. INTRODUCTION

1.1 Objective

The knowledge flow between citing and cited patents has been used extensively to test for knowledge spillovers since the availability of the NBER Patent Citations Data File (Hall et al., 2001). Papers that study knowledge flows between regions using patent citation data employing a spatial interaction regression model include Maurseth and Verspagen (2002), Singh et al. (2010), Li (2014), Griffith et al. (2007), LeSage et al. (2007), Peri (2005), Hussler (2004), Fischer et al. (2009), and Mukherji and Silberman (2013). We extend this literature by estimating knowledge flows separately for groups of metropolitan areas based on their innovative activity (alternatively called size) as measured by the number of patents issued to the area. This study builds on past works which ignored the potential impact of regions with high (low) concentrations of innovative activity on the absorption and diffusion of knowledge. The flow of knowledge across origin-destination communities is hypothesized to depend on the size of innovative communities. Not accounting for innovation size (innovative communities and social relationships) in the flow of knowledge across origin-destination regions will result in aggregation bias in the parameter estimates. We anticipate that the resistance or separation factors in a spatial interaction model of

*Acknowledgements: The authors gratefully acknowledge the comments of the anonymous referees that substantially improved the paper.

Authors are Associate Professor and Full Professor of Economics at Oakland University, Rochester, MI 48302. *Corresponding Author:* Jonathan Silberman E-mail: silberma@oakland.edu

ISSN 1553-0892, 0048-749X (online)
www.srsa.org/rrs

patent citations will be less (more) significant in knowledge flows between large (small) innovative metropolitan areas. The effect of proximity, measured by physical distance and technology compatibility, is expected to be diminished further after incorporating into the spatial interaction model of knowledge flows spatial effects that capture origin and destination local spillovers.

Our results confirm the hypothesis that innovative region size reduces the importance of physical distance and technology proximity in mediating knowledge flows among U.S. metropolitan areas. The finding is strongest for knowledge flows among large innovative metro areas compared with knowledge flows among small innovative metro areas. Knowledge spillovers are substantially larger and more localized for origin-destination pairs that are small innovative regions compared with origin-destinations pairs that are large innovative regions. Being within 50 miles increases knowledge flow by 209 percent for origin-destination pairs that are small innovative metro areas; while being within 50 miles increases knowledge flow by 73 percent for origin-destination pairs that are large innovative metro areas. The technology compatibility results are similar. Technology compatibility has a much stronger impact on knowledge flow for the origin-destination pairs that are small innovative regions compared with the origin-destination pairs that are large innovative metro areas. State borders do not impact the knowledge flow among large innovative metro areas and local spatial spillovers are also not significant.

Regions with an infrastructure supporting patent production at a high rate have an advantage in absorbing and diffusing knowledge compared with smaller patent production regions. Knowledge spillovers occur most often between origin-destination metro areas that have substantial innovative activity and infrastructure. These results should be considered when formulating public policy to promote innovation and research.

1.2 Literature Review

Knowledge spillovers are critical for innovation and new value creation in an increasingly knowledge-intensive economy and have generated substantial scholarly attention (Paul and Kogut, 1999; Jaffe, 1989; Rosenthal and Strange, 2001; Jaffe et al., 1993; Audretsch and Feldman, 1996). Two approaches are employed to investigate knowledge spillovers using patent citations. The spatial interaction regression (see Maurseth and Verspagen, (2002)), and the matching or choice-sampling method (see Jaffe et al., (1993)).

The spatial interaction model describes origin-destination flows, in this instance the flow of knowledge measured by patent citations. The origin-destination areas are typically political boundaries such as U.S. states, U.S. metropolitan areas, or countries. Kerr and Kominers (2015) study citation flows among postal codes to consider how distances between postal codes influence patent citation rates. The knowledge flow between the two locations depends on the mass at each location measured by the stock of patents and the separation between them. Separation is measured by physical distance and technology compatibility. The origin-destination knowledge flow declines with increasing distance between them, increases with technology compatibility, and is positively associated with the amount of activity (patents) at each location. The use of technology compatibility between the two locations is analogous to controlling for the distribution of technology activities in the choice-sampling method. The focus in the spatial interaction model is on the interaction patterns at the aggregate level rather than the individual level as in the choice-sampling approach. Compared to the matching method, the spatial interaction model avoids selecting a control group and directly estimates distance and border effects.

The key variable in the spatial interaction model to measure localization of knowledge spillovers is the physical distance between the origin-destination pairs. Two measures of distance or mileage between the pairs are typically used: a continuous variable that gives the average effect across all distances and a series of indicator variables for distance bands between origin-destination pairs. The indicator variables show how the effects of knowledge spillovers dissipate with increasing distance. Previous literature has found that proximity, measured by the mileage between origin-destination pairs, is an important determinant of knowledge flows measured by patent citations (Maurseth and Verspagen, 2002; Li, 2014; Peri, 2005; Mukherji and Silberman, 2013).

Our spatial interaction regression model examines knowledge flows among U.S. metropolitan areas using patent citation data. The factors in our model are the separation between the knowledge origin and destination regions measured by physical distance and technological compatibility, the existing stock of knowledge measured by the maximum number of patent citations, and a number of indicator variables that account for factors that have not been measured by the other variables. We include fixed effects for all cited and all citing regions and a variable for the intra-state or border effect. Following LeSage and Fischer (2010), LeSage and Pace (2008), LeSage and Llano (2007), and Scherngell and Barber (2011), we incorporate into the spatial interaction model of knowledge flows spatial effects that captures origin and destination based dependence.

Regions with higher innovative activities are likely to have increased contact rates with other inventors and their ideas. The contact rate for knowledge spillovers or diffusion will be influenced by the number of research communities (i.e. R&D labs, universities, and the number of inventors) that find out about a new discovery (Warnick, 2006). Metropolitan areas with large concentrations of research communities will have higher contact rates within and across regions thereby speeding up the spread of an idea. In the spirit of Carlino et al. (2007), who demonstrated that urban density has a positive impact on innovation productivity, the innovation contact rate represents another view of the economic benefits of agglomeration. A similar explanation is provided by Agrawal et al. (2006) who found that knowledge flows to an inventor's prior location are approximately 50 percent greater than if they had never lived there, suggesting that social relationships, not just physical proximity, are important for determining knowledge flow patterns. It is more likely that inventors relocate to metropolitan areas with similar innovation research infrastructure or other large innovation regions. Consequently, distance and other barriers to knowledge transfer are likely to be less pronounced for such groups.

In contrast to the spatial interaction models, the matching or choice-sampling method compares the citations to a set of patents and control patents controlling for the preexisting geographic distribution of technological activities. Jaffe et al. (1993) found evidence supporting localized knowledge spillovers at the state and metropolitan statistical area (MSA) levels. Thompson and Fox-Kean (2005) challenged the Jaffe et al. (1993) findings in the selection of the control patents. Jaffe et al. (1993) used three-digit technology classes while Thompson and Fox-Kean (2005) used a finer six-digit technology class. When restricting the control patents to those that have any subclass code in common with the originating patents, Thompson and Fox-Kean (2005) found no evidence supporting localization at the state and MSA levels. Murata et al. (2014) bridged the debate between Jaffe et al. (1993) and Thompson and Fox-Kean (2005) by developing a distance-based test of localized knowledge spillovers that embeds the concept of control patents. They found solid evidence supporting localization even when using finely grained technology controls.

Recent papers using the choice-sampling method confirm the localization of knowledge spillovers and identify a positive state border effect in the U.S. Singh and Marx (2013) simultaneously measured the impact of different geographical borders and proximity on knowledge spillovers using patent citations. They employed a regression framework that uses a choice-based sampling method to estimate the likelihood of citation between random patents. A surprising finding is that U.S. state borders have a robust same-state effect on knowledge spillovers. State borders serve as constraints on knowledge diffusion even after accounting for geographic proximity. Singh and Marx (2013) did not provide an explanation for the significant state border effect and suggested that further insight into the puzzling state border effect will require more research. Belenzon and Schankerman (2013) documented state-level localization effects for knowledge spillovers for patents assigned to only universities. They employed the choice-based sampling method comparing the characteristics of corporate patents that cite university patents to a control group that does not. Controlling for distance, inventors in the same state as the cited university patent are substantially more likely to cite the university's patents than an inventor outside of the state. Possible explanations for the state border effect is the importance of local information in dealing with tacit knowledge and state policy that promotes local commercial development of university innovations.

The rest of the paper is organized as follows. Section 2 describes the data and framework for the analysis. Section 3 presents the knowledge flow regression model and results. Section 4 concludes.

2. DATA AND FRAMEWORK

2.1 Description of Data

Measuring knowledge flows is fraught with difficulties since they often leave no paper trail (Krugman, 1991). Absent a perfect measure, researchers frequently use patent citation statistics (Jaffe et al., 1993). Our study of knowledge flows uses the NBER patent database that includes citations to all patents issued between 1976 and 1999.[1] We use all patents granted to an inventor in the U.S. between 1990 and 1999. The NBER database provides geographical information of the inventors of the issued and cited patents. Information on cited patents is available since 1963. Thus, citing patents are between 1990 and 1999 but the cited patents (backward citations) are between 1963 and 1999. This database does not distinguish between inventor and examiner added citations.[2] The addresses of first inventors are used to aggregate the patent data by metropolitan statistical areas (MSAs or PMSAs).[3] Where ZIP code information is available, the ZipList5 software is used to locate the patent to an MSA or PMSA. In the absence of ZIP code information,

[1] The data description and framework are condensed versions from Mukherji and Silberman (2013).

[2] Thompson and Fox-Kean (2005) used patents issued in January 2003 to examine if patent statistics that control for examiner added citations continue to support the literature's claim of localization of knowledge spillovers. The paper found that in the U.S., inventors are 25 percent more likely to cite patents written in the same geographical area than an examiner. Jaffe et al., (2000) concluded that when examiner added citations are excluded, citations remain an important, albeit noisy, measure of knowledge transfers. Duguet and MacGarvie, (2005) used citations made by French firms and their responses to survey questions regarding technology transfers and found support for the use of patent citations as a measure of knowledge transfers. Absent an alternative unbiased method, citations remain a useful tool. As summarized by Peri (2005), "whereas patents proxy new ideas, citations between patents proxy the diffusion of these ideas through learning."

[3] The use of 'location by the first inventor' is designed by the construction of the NBER Patent and Citation Database. Comprehensive information on the geographic address of multiple inventors involved in the creation of a patent is not available to assign fractional shares.

the name of the MSA and state is used to make the MSA or PMSA assignments. Patents for which addresses are not available are not included in the analysis. The use of metropolitan areas is consistent with the geographic boundary within which knowledge flows, as suggested by Breschi and Lissoni (2001).

We chose for analysis the MSA/PMSAs (referred to as MSA in the remainder of the paper) that were granted more than 1,000 patents in the 1990s, or an average of 100 new patents per year. One hundred and six MSAs/PMSAs satisfied this criterion. Patents granted to these 106 metropolitan areas account for 74 percent of all patents granted during the 1990s in the U.S. by the U.S. Patent and Trademark Office (USPTO), and 87.5 percent of patents granted to all metropolitan areas in the U.S. San Jose, CA had the largest number of patents granted between 1990 and 1999 at 24,986 patents. Nashville, TN ranked 106th with 1,032 patents granted from 1990 to 1999. For each of the 106 MSAs, we calculate the total number of citations made by patents issued in the MSA from 1990 to 1999 to patents granted to each of the other 105 MSAs. This results in 11,130 observations, or citation flows between pairs of metropolitan areas. Only 24 out of the 11,130 observations reported zero citations, which means that 99.8 percent of the pairs of MSAs absorbed knowledge from each other. Since 99.8 percent of the pairs of regions absorbed knowledge from each other, we eliminated the 24 zero citation observations.[4]

To test our hypothesis regarding the impact of a region's innovative activity size on knowledge flows, the 106 metropolitan areas are split into large innovative regions and all other regions. Large innovative metropolitan areas are defined as those whose total number of patents issued from 1990 to 1999 was more than one standard deviation from the average. The mean number of patents granted in the sample is 4,115 and one standard deviation from the mean is 4,446. Thus, the cut off for a large innovative region is 8,561 patents. Thirteen metropolitan areas satisfy this criterion.[5]All other regions are referred to as "small." Four sub-samples are analyzed: large citing and large cited; large citing and small cited; small citing and large cited; and small citing and small cited. Table 1 show the 106 regions, identifies them as either an MSA or a PMSA, displays their number of patents granted in the 1990s, and their patent number rank.[6]

A literature exists that examines inter-metropolitan patenting behavior (Uallachain, 1999; Carlino et al., 2007). Our purpose is to examine the impact of innovative region size on knowledge spillovers measured by patent citation flows across regional pairs. We take as given the inter-metropolitan patenting behavior and control for it in our model by including the maximum number of citations feasible and fixed effects dummy variables for citing and cited regions. We expect that variables that are statistically significant in the inter-metropolitan patenting behavior literature will be statistically different when comparing our large-small innovative regions. Carlino et al. (2007) found that employment density (jobs per square mile) is a positive and statistically significant determinant of inter-metropolitan patenting behavior (patents per capita). In our sample, density is

---

[4] Dropping the 24 city pairs with zero citations enables the use of OLS to estimate the models. Including the zero observations requires use of Poisson or Negative Binomial estimation method and a count specification on the dependent variables that is different than the dependent variable used.

[5] Boston, Chicago, Dallas, Detroit, Houston, Los Angeles, New York, Oakland, Orange County, Philadelphia, Rochester, San Diego, and San Jose.

[6] An MSA (Metropolitan Statistical Area) is a core area with a large population nucleus, plus adjacent communities having a high degree of economic and social integration with that core. Requires a total population of a least 50,000. A PMSA (Primary Metropolitan Statistical Area) meets the MSA requirements, has a population of one million or more, and consists of one or more counties that have a substantial commuting interchange.

Table 1: Metropolitan Regions Patents Granted in the 1990s

City	MSA/PMSA	State	Patents	Patent Rank
Akron	PMSA	OH	2346	51
Albany	MSA	NY	3777	35
Albuquerque	MSA	NM	1494	74
Allentown	MSA	PA	2126	56
Ann Arbor	PMSA	MI	3049	41
Appleton	MSA	WI	1389	84
Atlanta	MSA	GA	5978	22
Austin	MSA	TX	7652	17
Baltimore	PMSA	MD	4370	30
Baton Rouge	MSA	LA	1560	70
Bergen	PMSA	NJ	3527	38
Binghamton	MSA	NY	1284	88
Boise City	MSA	ID	3607	37
Boston*	PMSA	MA	15867	4
Boulder	PMSA	CO	2649	45
Brazoria	PMSA	TX	1100	101
Bridgeport	PMSA	CT	1782	65
Buffalo	MSA	NY	2002	59
Burlington	MSA	VT	1180	95
Charlotte	MSA	NC	1884	62
Chicago*	PMSA	IL	22627	2
Cincinnati	PMSA	OH	5092	24
Cleveland	PMSA	OH	4654	27
Colorado Springs	MSA	CO	1457	76
Columbus	MSA	OH	2421	47
Dallas*	PMSA	TX	10035	11
Danbury	PMSA	CT	1699	67
Dayton	MSA	OH	2093	57
Denver	PMSA	CO	3773	36
Detroit*	PMSA	MI	13275	5
Dutchess County	PMSA	NY	2215	53
Fort Collins	MSA	CO	1430	79

Table 1 Continued

City	MSA/PMSA	State	Patents	Patent Rank
Fort Lauderdale	PMSA	FL	1731	66
Fort Wayne	MSA	IN	1138	97
Fort Worth	PMSA	TX	2368	49
Grand Rapids	MSA	MI	2649	44
Greensboro	MSA	NC	1595	69
Greenville	MSA	NC	2016	58
Hamilton	PMSA	OH	1501	73
Harrisburg	MSA	PA	1383	85
Hartford	MSA	CT	4023	33
Houston*	PMSA	TX	11153	9
Indianapolis	MSA	IN	4082	32
Johnson City	MSA	TN	1109	100
Kalamazoo	MSA	MI	1070	103
Kansas City	MSA	MO	1665	68
Knoxville	MSA	TN	1320	87
Lancaster	MSA	PA	1033	105
Lawrence	PMSA	KS	1415	81
Los Angeles*	PMSA	CA	17128	3
Louisville	MSA	KY	1141	96
Lowell	PMSA	MA	1249	90
Madison	MSA	WI	1533	71
Melbourne	MSA	FL	1054	104
Memphis	MSA	TN	1094	102
Miami	PMSA	FL	1813	64
Middlesex	PMSA	NJ	7415	21
Milwaukee	PMSA	WI	3804	34
Minneapolis	MSA	MN	8440	14
Monmouth	PMSA	NJ	3330	39
Nashua	PMSA	NH	1236	92
Nashville	MSA	TN	1032	106
Nassau	PMSA	NY	5032	25
New Haven	PMSA	CT	2180	54
New Orleans	MSA	LA	1134	98

Table 1 Continued

City	MSA/PMSA	State	Patents	Patent Rank
New York*	PMSA	NY	11710	8
Newark	PMSA	NJ	7441	20
Norfolk	MSA	VA	1225	93
Oakland*	PMSA	CA	8940	13
Oklahoma City	MSA	OK	1134	99
Orange County*	PMSA	CA	9791	12
Orlando	MSA	FL	1347	86
Peoria	MSA	IL	1209	94
Philadelphia*	PMSA	PA	12030	7
Phoenix	MSA	AZ	7649	18
Pittsburgh	MSA	PA	5599	23
Portland	PMSA	ME	4941	26
Providence	MSA	RI	2270	52
Raleigh	MSA	NC	4499	29
Richmond	MSA	VA	1433	78
Riverside	PMSA	CA	2385	48
Rochester*	MSA	NY	13060	6
Rockford	MSA	IL	1238	91
Sacramento	PMSA	CA	1948	61
Saginaw	MSA	MI	2675	43
Salt Lake City	MSA	UT	3112	40
San Antonio	MSA	TX	1454	77
San Diego*	MSA	CA	10047	10
San Francisco	PMSA	CA	8166	16
San Jose*	PMSA	CA	24986	1
Santa Barbara	MSA	CA	1529	72
Santa Cruz	PMSA	CA	1253	89
Seattle	PMSA	WA	7562	19
St. Louis	MSA	MO	4544	28
Stamford	PMSA	CT	2140	55
Syracuse	MSA	NY	1391	83
Tampa	MSA	FL	2367	50
Toledo	MSA	OH	1474	75

Table 1 Continued

City	MSA/PMSA	State	Patents	Patent Rank
Trenton	PMSA	NJ	1951	60
Tucson	MSA	AZ	1815	63
Tulsa	MSA	OK	1407	82
Ventura	PMSA	CA	2422	46
Washington	PMSA	DC	8201	15
West Palm Beach	MSA	FL	2967	42
Wilmington	PMSA	DE	4201	31
Worcester	PMSA	MA	1426	80

*Large cities used in the analysis

measured as the number of assignees to a patent per square mile, a more precise measure than jobs per square mile. A test of differences in means confirms that the density for large innovative regions (16) is greater than and statistically different from the density for small innovative regions (4) at the 1 percent level or better.

2.2 Framework

Given an available stock of knowledge, K_j, in region j, region i with p_i innovations can make a maximum of p_iK_j references to that knowledge when each innovation uses all available knowledge. The actual use or flow of knowledge, measured as citations made by MSA i (destination) of patents granted to MSA j (origin), that takes place from the stock in region j (knowledge generating) to the innovations in region i (knowledge absorbing) during a period of time depends on p_iK_j and additional factors as follows:

$$c_{ij} = \left(Dist_{ij}^{b}\right)\left(Index_{ij}^{\gamma}\right)\exp\left(\eta D_{state,ij}\right)\exp\left(\delta D_{citing,ij}\zeta D_{cited,j}\right)A_i^{\alpha}B_j^{\beta}\left(p_iP_j\right)^{\Omega} \tag{1}$$

where, c_{ij} denotes citations of patents issued to MSA i cited by MSA j. This equation measures the stock of knowledge K_j through $p_i * P_j$, where p_i indicates the number of patents granted to MSA$_i$ during a certain period of time and P_j indicates the number of patents available for citations from MSA$_{j.}$ Therefore, $p_i * P_j$ gives the maximum number of citations that can be made by MSA$_i$ of MSA$_j$'s past patents. The model in Equation (1) is a classical spatial interaction model that describes origin-destination flows. $Dist_{ij}$ is the geographical distance between the two regions, $Index_{ij}$ is a measure of the technological compatibility between the two regions, $D_{State,ij}$ is a regional border indicator variable, and $D_{Citing,i}$ and $D_{Cited,j}$ reflect other regional characteristics that may influence an MSA's ability to absorb knowledge from other regions ($D_{Citing,i}$) and diffuse knowledge to other regions ($D_{Cited,j}$). Examples of these characteristics are the industrial structure, the number of inventors, the diversity of the economy, and the entrepreneurial culture. Since the indicator variables take the values of 0 or 1, they are modeled as exponential functions. Furthermore, the literature on spatial interactions with origin-destination effects shows that the interaction between a dyad not only depends on the properties of the individual units comprising the dyad, but the regions neighboring them as well. These spatial effects can have a significant impact on the interactions between the two units. These additional spatial origin and destination effects emanating from regions surrounding the knowledge generating and absorbing regions are captured by the variables B_j and A_i, respectively. α, β, Ω are elasticities of knowledge flow to spatial destination effects, spatial origin effects, and maximum possible references to knowledge generating region's knowledge, respectively.

Measurement of the technology compatibility index, $Index_{ij}$, follows the one developed by Maurseth and Verspagen (2002). In calculating the compatibility index, we use the two-digit patent classification system used in the NBER patent citation data.[7] If two MSAs are technologically close, the compatibility index between them will be high. In regression analysis, the coefficient of this index reflects the number of citations that are due to the similarities between the MSAs in technological space. To investigate distance in the context of knowledge flow between MSAs, we use a continuous variable and distance dummy variables. Distance, measured as the mileage between pairs of MSAs, gives the average effect across all distances. Distance dummy variables find the impact of different distances on knowledge flow, for example, whether the impact of being within 50 miles of each other is different than being between 2,500 miles.

Types of spatial cross-section models that account for spillovers from neighboring regions include the spatial autoregressive model ($y = \lambda Wy + X\beta + u$), the spatial lag of X model ($y = X\beta + WX\theta + u$), the spatial error model ($y = X\beta + \rho Wu + v$), and the general spatial model ($y = \lambda W_1 y + X\beta + \rho W_2 u + v$), where y represents the vector of dependent variables, X represents the matrix of explanatory variables, W is a spatial weight matrix, u and v are error terms. In this paper, we use the spatial lag of X model to capture the notion that citations of patents made by a region depend not only on the economic conditions, such as technological compatibility of the dyad, but also the characteristics of the neighbors of the citing and cited MSAs. That is, if an MSA is surrounded by other MSAs that are similar, the neighbors are likely to have similar types of innovators and economic relationships, and this can lead to a higher amount of knowledge flows between the dyad. This is a local spillover where endogenous interaction and feedback effects are not present. Endogenous interaction and feedback refers to when an impact on a neighbor region also impacts the neighbors of the neighbor, LeSage (2014).

To introduce the spatial factors in the estimation of the flow of knowledge between the MSAs we follow the origin-destination filtering approach used by LeSage and Pace (2008). To select the MSAs that neighbor the origin and destination MSAs, a spatial weight matrix, W, is used. This matrix uses the traditional inverse distance approach to model that interactions and spatial effects are stronger at closer distances. Using k to represent an MSA$_j$ that neighbors the origin MSA$_j$, d_{kj} to represent the inverse distance between MSAs k and j, the origin effects for origin $_j$ is given by $B_j = \sum_k \frac{1}{d_{kj}} B_{kj}$. The destination effects can be addressed similarly. Using the same spatial matrix W, the neighbors of the destination MSA are identified. The destination effect gives the weighted values of the compatibility indexes of the neighbors of the destination MSA and is given by $A_i = \sum_z \frac{1}{d_{zi}} A_{iz}$. The origin and destination effects are obtained from the multiplication of the Kronecker products $W \otimes I$ and $I \otimes W$, respectively, and the vector of the compatibility index. For the full-sample, W is a 106x106 inverse distance spatial weight matrix and I is a 106x106 identity matrix. For smaller subsets such as large MSAs only, the matrices are

[7] The NBER patent citation database has six classes of patents: chemicals, computers, drugs and medical, electrical, mechanical, and other. Within each patent class there are six sub-categories resulting in 36 patent classes used in calculating the technology compatibility index.

13x13. Weight matrices typically assign a value of 0 to regions beyond a certain geographical space surrounding the destination and origin regions. To determine a reasonable estimate of the region within which such spatial effects may be operational, the full model is first estimated without the spatial effects. The results from the regression with the distance dummy variables will be used to inform the distance beyond which the weight matrix assigns a 0 value.[8]

To reduce heteroskedasticity in the patent product term, $p_i * P_j$, the dependent variable is transformed to the logarithm of the actual citations as a share of the entire pool of potential citations, that is $\ln \frac{c_{ij}}{p_i P_j}$. The ratio of actual to all potential citations is also $p_i * P_j$ used by Duguet and MacGarvie (2005), among others. Although our citation data is count data, the transformation of the data converts the citation counts into a continuous variable. Consequently, the use of OLS is justifiable in our case.

To obtain an estimate of the direct effect of the distance dummy variables directly when multiple dummy variables are included, we use the variable transformation suggested by Hirschberg and Lye (2001). The transformed equation is described below.

$$(2)\quad \ln\left(\frac{c_{ij}}{p_i P_i}\right) = \bar{y} + \sum_d \beta_d \left(D_{d,i} - \frac{n_d}{n_{g2500}} D_{g2500,i}\right) + \gamma\left(\ln\left(Index_{ij}\right) - \overline{\ln Index}\right) + \sum_k \delta_k\left(D_{citing,k} - D_{citing,M}\right) + \zeta_l\left(D_{cited,l} - D_{cited,M}\right) + \eta\left(D_{state,i} - \frac{n_S}{n_{NS}} D_{diffstate,i}\right) + \alpha\left(\ln \sum_z \frac{1}{d_{iz}} A_{iz} - \overline{\ln A}\right) + \beta\left(\ln \sum_k \frac{1}{d_{kj}} B_{kj} - \overline{\ln B}\right) + \varepsilon_{ij}$$

In this transformed equation, D_d denotes the distance dummy variables D_{50}, D_{100}, D_{200}, D_{300}, D_{400}, D_{500}, D_{600}, D_{700}, D_{800}, D_{900}, D_{1000}, D_{1500}, D_{2000}, D_{2500}, and D_{g2500}. D_{50} takes the value 1 when the cited MSA is within 50 miles of the citing MSA, D_{100} takes the value 1 when it is within 51 to 100 miles from the citing MSA, and so on for the other distance dummies. The variable D_{g2500} takes the value 1 for distances greater than 2500 miles from the citing MSA. This variable is used for the variable transformations. The term $\frac{n_d}{n_{g2500}}$ gives the proportion of cited MSAs that are d miles away from the citing MSA to the cited MSAs that are greater than 2500 miles away. Similarly, M, denotes the MSA (Akron or Boston) used to transform the MSA dummy variables. In the above equation, $\overline{lnIndex}$ is the mean of the logarithm of the compatibility index and $\overline{lnA}$ and $\overline{lnB}$ denote the means of lnA and lnB, respectively. $\frac{n_S}{n_{NS}}$ gives the ratio of MSA pairs that belong to the same state to those that do not belong to the same state (DiffState). The constant of this regression equals the mean of the dependent variable. $\bar{y}$ in Equation (2) denotes the mean of the dependent variable. Furthermore, the coefficient of the transformed dummy variable exactly equals the coefficient of the variable without the transformation thus allowing us to generate direct estimates of each of the dummy variables. This transformation also allows for an easier interpretation of the coefficients of the dummy variables since such interpretations are difficult when multiple categories of dummy variables are included in a regression. Given that the constant of the regression equals the mean of the dependent variable, the values of the coefficients are interpreted as the amounts they add to or subtract from the mean of the dependent variable. Definitions and descriptive statistics for the regression variables are reported in Tables 2 and 3.

---

[8] The distance indicator values reported in Table 4 column 2 show that citations are above average for MSA pairs within 700 miles and becomes negative at 800 miles. It remains positive and significant up to 800 miles for pairs of large MSAs. We used these results to assign a 0 value in the weight matrix for MSAs that are more than 800 miles apart.

Table 2: Variable Descriptions

Variable	Description
c_{ij}	Total citations of MSA j's patents by 1990-99 patents of MSA i
p_i	Total patents of MSA i for 1990-99
P_j	Total patents granted to MSA j from 1963-1999
$ln\ [c_{ij} / (p_i * P_j)]$	ln of actual to maximum potential citations made by MSA i of MSA j's patents
Index	Technological compatibility index for MSAs i and j.
Border	Equals 1 when the MSAs are from the same state; 0 otherwise
Distance	Distance in miles between two MSAs
D_{50}	Equals 1 when MSAs are 0 to 50 miles apart; 0 otherwise
D_{100}	Equals 1 when MSAs are 51 to 100 miles apart; 0 otherwise
D_{200}	Equals 1 when MSAs are 101 to 200 miles apart; 0 otherwise
D_{300}	Equals 1 when MSAs are 201 to 300 miles apart; 0 otherwise
D_{400}	Equals 1 when MSAs are 301 to 400 miles apart; 0 otherwise
D_{500}	Equals 1 when MSAs are 401 to 500 miles apart; 0 otherwise
D_{600}	Equals 1 when MSAs are 501 to 600 miles apart; 0 otherwise
D_{700}	Equals 1 when MSAs are 601 to 700 miles apart; 0 otherwise
D_{800}	Equals 1 when MSAs are 701 to 800 miles apart; 0 otherwise
D_{900}	Equals 1 when MSAs are 801 to 900 miles apart; 0 otherwise
D_{1000}	Equals 1 when MSAs are 901 to 1000 miles apart; 0 otherwise
D_{1500}	Equals 1 when MSAs are 1001 to 1500 miles apart; 0 otherwise
D_{2000}	Equals 1 when MSAs are 1501 to 2000 miles apart; 0 otherwise
D_{2500}	Equals 1 when MSAs are 2001 to 2500 miles apart; 0 otherwise
D_{g2500}	Equals 1 when MSAs are greater than 2500 miles apart; 0 otherwise

3. KNOWLEDGE FLOW REGRESSION RESULTS AND IMPACT OF MSA SIZE

The estimated regression coefficients from Equation (2) are reported in Tables 4 through 8. Results for the full sample are in Table 4. Table 5 shows patent citation flow between pairs of large metropolitan areas. Table 6 displays patent citation flow for pairs of large MSAs citing patents from small MSAs. The patent citation flow for pairs of small MSAs citing patents from large MSAs is shown in Table 7. Table 8 displays the patent citation flow among pairs of small MSAs. Columns 1 and 2 provide regression results without local spatial origin and destination spillover impacts and Columns 3 and 4 show the results with local spatial origin and destination effects. Columns 1 and 4 include the distance between the areas measured as a continuous variable; Columns 2 and 3 have distance between the areas measured as a series of dummy variables. The main results from the regression estimation relate to the border effect, the impact of proximity and technology compatibility, and the effect of local spatial origin and destination spillovers. The results are discussed below.

Table 3: Descriptive Statistics

Variable	Obs	Mean	Std. Dev	Min	Max
Large Citing and Large Cited					
Border	156	0.15	0.36	0.00	1.00
Index	156	0.47	0.15	0.18	0.82
Distances	156	1,574	988	31.00	3,128
Citations	156	1,851	1,299	307	8,307
Patents of Citing MSA	156	13,896	4,832	8,940	24,986
Patents of Cited MSA	156	13,896	4,832	8,940	24,986
$ln \frac{c_{ij}}{p_i P_j}$	156	-12.34	0.49	-13.56	-10.88
Small Citing and Large Cited					
Border	1,209	0.05	0.22	0.00	1.00
Index	1,209	0.48	0.12	0.24	0.91
Distances	1,209	1,452	927	12.00	3,159
Citations	1,208	390	495	16.00	7,596
Patents of Citing MSA	1,209	2,748	1,958	1,032	8,440
Patents of Cited MSA	1,209	13,896	4,818	8,940	24,986
$ln \frac{c_{ij}}{p_i P_j}$	1,208	-12.36	0.54	-14.55	-10.16
Large Citing and Small Cited					
Border	1,209	0.05	0.22	0.00	1.00
Index	1,209	0.50	0.12	0.23	0.93
Distances	1,209	1,452	927	11.00	3,157
Citations	1,209	348	425	19.00	6,072
Patents of Citing MSA	1,209	13,896	4,818	8,940	24,986
Patents of Cited MSA	1,209	2,748	1,958	1,032	8,440
$ln \frac{c_{ij}}{p_i P_j}$	1,209	-12.46	0.56	-14.46	-10.46
Small Citing and Small Cited					
Border	8,556	0.03	0.17	0.00	1.00
Index	8,556	0.51	0.12	0.23	0.97
Distances	8,556	1,174	791	11.00	3,334
Citations	8,534	74	119	0.00	2,755
Patents of Citing MSA	8,556	2,748	1,958	1,032	8,440
Patents of Cited MSA	8,556	2,748	1,958	1,032	8,440
$ln \frac{c_{ij}}{p_i P_j}$	8,533	12.48	0.70	-15.65	-8.65

Table 4: Full Sample

	(1) Non-Spatial Distance	(2) Non-Spatial Distance Dummies	(3) Spatial Distance Dummies	(4) Spatial Distance
Border	0.217∗∗∗	0.138∗∗∗	0.079∗∗∗	0.161∗∗∗
	(6.54)	(4.54)	(2.63)	(4.85)
Ln Index	1.344∗∗∗	1.328∗∗∗	1.239∗∗∗	1.262∗∗∗
	(54.90)	(54.54)	(44.09)	(44.51)
Ln Distance	-0.110∗∗∗			-0.104∗∗∗
	(-14.58)			(-12.94)
D_{50}		0.886∗∗∗	0.886∗∗∗	
		(12.94)	(12.56)	
D_{100}		0.317∗∗∗	0.285∗∗∗	
		(8.87)	(8.07)	
D_{200}		0.218∗∗∗	0.197∗∗∗	
		(9.70)	(8.56)	
D_{300}		0.130∗∗∗	0.114∗∗∗	
		(5.49)	(4.73)	
D_{400}		0.148∗∗∗	0.123∗∗∗	
		(6.40)	(5.20)	
D_{500}		0.087∗∗∗	0.064∗∗∗	
		(3.87)	(2.82)	
D_{600}		0.052∗∗	0.034	
		(2.45)	(1.61)	
D_{700}		0.032∗	0.028	
		(1.75)	(1.52)	
D_{800}		-0.026	-0.027	
		(-1.30)	(-1.41)	
D_{900}		-0.119∗∗∗	-0.118∗∗∗	
		(-5.27)	(-5.27)	
D_{1000}		-0.123∗∗∗	0.106∗∗∗	
		(-5.05)	(-4.44)	
D_{1500}		-0.123∗∗∗	-0.119∗∗∗	
		(-11.31)	(-11.34)	
D_{2000}		-0.062∗∗∗	-0.033∗∗	
		(-4.14)	(-2.17)	
D_{2500}		-0.042∗∗∗	-0.011	
		(-2.78)	(-0.72)	
Spatial Origin			0.374∗∗∗	0.337∗∗∗
			(8.71)	(7.92)
Spatial Destination			0.135∗∗∗	0.100∗∗
			(3.04)	(2.27)
N	11,129	11,129	11,103	11,103
R^2	0.690	0.697	0.513	0.501
adj. R^2	0.684	0.691	0.503	0.491

T-statistics are in parentheses ∗ $p < 0.1$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$

3.1 Border Effect

The results for the full sample show a positive and statistically significant border effect, a finding consistent with prior research. After the full sample is split by MSA innovation size, we

Table 5: Large Citing and Large Cited

	(1) Non-Spatial Distance	(2) Non-Spatial Distance Dummies	(3) Spatial Distance Dummies	(4) Spatial Distance
Border	0.121	0.121	-0.012	0.103
	(1.09)	(-0.37)	(-0.14)	(1.34)
Ln Index	0.566***	0.449***	0.435***	0.483***
	(9.24)	(7.64)	(7.53)	(9.74)
Ln Distance	-0.077***			-0.057**
	(-2.76)			(-2.22)
D_{50}		0.628***	0.551***	
		(4.14)	(3.73)	
D_{100}		0.279***	0.199**	
		(3.62)	(2.06)	
D_{200}		0.418***	0.354***	
		(3.80)	(2.88)	
D_{300}		0.237***	0.164**	
		(4.05)	(2.05)	
D_{400}		0.251***	0.214**	
		(2.98)	(2.40)	
D_{500}		0.162	0.138	
		(1.58)	(1.37)	
D_{600}		-	-	
D_{700}		0.202**	0.171*	
		(2.26)	(1.96)	
D_{800}		0.122**	0.130**	
		(2.18)	(2.23)	
D_{900}		-	-	
D_{1000}		-0.170	-0.141	
		(-1.18)	(-1.00)	
D_{1500}		-0.402***	0.339***	
		(-5.49)	(-3.95)	
D_{2000}		-0.283***	-0.225**	
		(-3.67)	(-2.48)	
D_{2500}		-0.059	-0.049	
		(-0.98)	(-0.90)	
Spatial Origin			0.160	0.265***
			(1.51)	(2.77)
Spatial Destination			0.043	0.141*
			(0.48)	(1.70)
N	156	156	156	156
R^2	0.858	0.893	0.897	0.878
adj. R^2	0.829	0.858	0.861	0.850

T-statistics are in parentheses * p < 0.1, ** p < 0.05, *** p < 0.01

Table 6: Large Citing and Small Cited

	(1) Non-Spatial Distance	(2) Non-Spatial Distance Dummies	(3) Spatial Distance Dummies	(4) Spatial Distance
Border	0.134∗∗	0.111∗	0.118∗∗	0.145∗∗
	(2.16)	(1.94)	(2.11)	(2.38)
Ln Index	1.163∗∗∗	1.121∗∗∗	0.953∗∗∗	0.980∗∗∗
	(22.34)	(22.42)	(17.16)	(17.14)
Ln Distance	-0.103∗∗∗			-0.058∗∗∗
	(-7.38)			(-3.82)
D_{50}		0.577∗∗∗	0.473∗∗∗	
		(7.10)	(5.65)	
D_{100}		0.387∗∗∗	0.273∗∗∗	
		(5.18)	(3.57)	
D_{200}		0.205∗∗∗	0.127∗∗∗	
		(4.38)	(2.62)	
D_{300}		0.112∗∗	0.026	
		(2.40)	(0.55)	
D_{400}		0.126∗∗∗	0.061	
		(2.92)	(1.36)	
D_{500}		0.174∗∗∗	0.120∗	
		(2.74)	(1.83)	
D_{600}		0.110∗∗	0.036	
		(2.35)	(0.73)	
D_{700}		0.104∗	0.046	
		(1.76)	(0.81)	
D_{800}		0.015	-0.012	
		(0.37)	(-0.29)	
D_{900}		-0.085∗	-0.111∗∗∗	
		(-1.95)	(-2.69)	
D_{1000}		-0.158∗∗∗	-0.150∗∗∗	
		(-2.97)	(-2.89)	
D_{1500}		-0.136∗∗∗	-0.125∗∗∗	
		(-5.12)	(-4.72)	
D_{2000}		-0.042	-0.009	
		(-1.26)	(-0.27)	
D_{2500}		-0.115∗∗∗	-0.023	
		(-3.89)	(-0.72)	
Spatial Origin			0.598∗∗∗	0.569∗∗∗
			(6.08)	(5.96)
Spatial Destination			0.230∗∗∗	0.269∗∗∗
			(2.58)	(3.02)
N	1,209	1,209	1,209	1,209
R^2	0.706	0.722	0.732	0.719
adj. R^2	0.678	0.691	0.702	0.691

T-statistics are in parentheses ∗ $p < 0.1$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$

find in the basic model with no spatial spillover and a continuous distance variable, Column 1, that the border effect dummy variable is positive and statistically significant in all regressions except for large MSAs citing large MSAs' patents. Once the distance dummy variables and spatial origin and destination spillover are included in the regression model, Column 3, the border effect is not statistically different from zero in two of the sub-samples: large citing, large cited and small citing,

Table 7: Small Citing and Large Cited

	(1) Non-Spatial Distance	(2) Non-Spatial Distance Dummies	(3) Spatial Distance Dummies	(4) Spatial Distance
Border	0.163∗∗ (2.38)	0.124∗∗ (2.03)	0.133∗∗ (2.22)	0.174∗∗∗ (2.58)
Ln Index	1.023∗∗∗ (21.15)	0.992∗∗∗ (20.83)	0.831∗∗∗ (14.74)	0.867∗∗∗ (15.11)
Ln Distance	-0.117∗∗∗ (-7.95)			-0.077∗∗∗ (-4.89)
D_{50}		0.718∗∗∗ (9.44)	0.624∗∗∗ (8.29)	
D_{100}		0.443∗∗∗ (5.45)	0.339∗∗∗ (4.22)	
D_{200}		0.259∗∗∗ (5.26)	0.177∗∗∗ (3.54)	
D_{300}		0.122∗∗ (2.32)	0.041 (0.80)	
D_{400}		0.173∗∗∗ (3.94)	0.112∗∗ (2.45)	
D_{500}		0.145∗∗ (2.03)	0.090 (1.21)	
D_{600}		0.050 (1.05)	-0.016 (-0.34)	
D_{700}		-0.011 (-0.24)	-0.061 (-1.33)	
D_{800}		-0.022 (-0.53)	-0.052 (-1.28)	
D_{900}		-0.041 (-1.00)	-0.069∗ (-1.74)	
D_{1000}		-0.167∗∗ (-2.58)	-0.153∗∗ (-2.38)	
D_{1500}		-0.146∗∗∗ (-5.38)	-0.138∗∗∗ (-5.13)	
D_{2000}		-0.070∗∗ (-2.08)	-0.041 (-1.24)	
D_{2500}		-0.110∗∗∗ (-4.05)	-0.026 (-0.87)	
Spatial Origin			0.321∗∗∗ (3.84)	0.342∗∗∗ (4.05)
Spatial Destination			0.450∗∗∗ (4.18)	0.370∗∗∗ (3.56)
N	1,208	1,208	1,208	1,208
R^2	0.696	0.716	0.725	0.706
adj. R^2	0.666	0.684	0.694	0.677

T-statistics are in parentheses ∗ $p < 0.1$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$

Table 8: Small Citing and Small Cited

	(1) Non-Spatial Distance	(2) Non-Spatial Distance Dummies	(3) Spatial Distance Dummies	(4) Spatial Distance
Border	0.186***	0.065	0.053	0.186***
	(4.25)	(1.59)	(1.30)	(4.22)
Ln Index	1.438***	1.432***	1.372***	1.392***
	(47.92)	(47.99)	(39.78)	(40.02)
Ln Distance	-0.146***			-0.131***
	(-15.38)			(-12.57)
D_{50}		1.162***	1.130***	
		(12.13)	(11.74)	
D_{100}		0.355***	0.315***	
		(7.99)	(7.11)	
D_{200}		0.249***	0.220***	
		(9.43)	(8.05)	
D_{300}		0.172***	0.148***	
		(6.11)	(5.15)	
D_{400}		0.170***	0.144***	
		(5.96)	(4.90)	
D_{500}		0.091***	0.066***	
		(3.63)	(2.61)	
D_{600}		0.072***	0.051**	
		(3.00)	(2.10)	
D_{700}		0.052**	0.039*	
		(2.48)	(1.83)	
D_{800}		-0.013	-0.023	
		(-0.57)	(-1.00)	
D_{900}		-0.112***	-0.116***	
		(-4.31)	(-4.46)	
D_{1000}		-0.102***	-0.101***	
		(-3.70)	(-3.69)	
D_{1500}		-0.131***	-0.120***	
		(-10.69)	(-9.78)	
D_{2000}		-0.059***	-0.040**	
		(-3.26)	(-2.19)	
D_{2500}		-0.054***	-0.020	
		(-2.81)	(-0.95)	
Spatial Origin			0.256***	0.214***
			(5.03)	(4.21)
Spatial Destination			0.071	0.024
			(1.36)	(0.46)
N	8,533	8,533	8,533	8,533
R^2	0.473	0.484	0.486	0.474
adj. R^2	0.461	0.472	0.473	0.462

T-statistics are in parentheses $*$ $p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$

small cited. These results show that being in the same state has no impact on knowledge flow or that crossing a state border does not reduce knowledge flow when the city pairs are large-large or small-small. This result is in contrast with the positive and significant border effect found by Singh and Marx (2013), Belenzon and Schankerman (2013), and Li (2014). Our results control for MSA innovation size, distance measured as dummy variables, and local spatial spillovers which reduces the border impact when comparing the full sample to the split samples.

The size of the border regression coefficient is smaller in the split samples and not statistically different from zero in two instances. The lesser impact of the border effect is consistent with the hypothesis that crossing a U.S. state should not matter for knowledge flow because, unlike European regions, no significant differences in language, culture, and political institutions exist across U.S. states. The small impact of the border dummy suggests that physical distance has a stronger impact on knowledge flow than the presence of a regional border for a country like the U.S. State borders do constrain knowledge diffusion for the large citing, small cited and small citing, large cited city pairs. This finding is consistent with previous literature and shows that when city pairs are different in their innovation size, being in the same state increases knowledge flow. The positive, statistically significant border effect means that unobserved forces are at work that represent spillovers within a state for these MSA pairs.

3.2 Distance

The literature supports the localization of knowledge spillovers and suggests that there is an extremely rapid distance decay associated with local knowledge spillovers (Audretsch and Feldman, 1996; Jaffe et al., 1993). Singh and Marx (2013) found that knowledge flows are greatest when the source and recipient are co-located within the same MSA and that the distance effect gradually falls with distance, but is positive at all distance categories even for distances between 4,000 and 6,000 miles. Belenzon and Schankerman (2013) reported that the estimated distance effects are robust and the impact of distance dies out after about 100 to 150 miles, and Li (2014) reported a statistically significant negative distance when it is measured as a continuous variable.

A summary of the distance effects across the four sub-samples and the full sample is displayed in Table 9. The full sample results show a negative and significant distance continuous variable and positive indicator distance variables up to 500 miles with spatial spillovers. The results for the sub-samples using distance dummy variables, without local spatial spillovers (Column 2), indicate an above average number of citations made to patents granted within 500 to 800 miles. These results show that MSAs within a broad geographic area surrounding an MSA positively impact the number of patents granted to it, a finding consistent with prior research. Once spatial origin and destination spillovers are included, the impact of distance on knowledge flows is diminished. Importantly, the impact of distance on knowledge flow is found to be dependent on the relative sizes of the MSAs involved. When two large MSAs are involved, the positive effect of proximity exists up to 800 miles, not accounting for spatial spillovers. Once spatial spillovers are accounted for, the effect of distance remains positive up to 800 miles, but the impact of distance is smaller. When small MSAs are involved, either as citing or as cited MSAs, distance plays a more significant role. Without spatial dependency effects taken into consideration, citations are positive when small MSAs are within 700 miles. When all MSAs involved in knowledge transfer are small and spatial spillovers are included, the positive effect of distance extends to 700 miles but the impact is smaller. These results demonstrate that when considering the effect of physical distance on knowledge transfers, MSA innovation size must be taken into consideration. Knowledge spillovers are more localized for small MSAs than for large ones. Distance is not as much of a resistance factor in knowledge flow for larger metropolitan areas compared to smaller regions as shown by the log of distance continuous variable being smaller and less significant for larger regions compared with smaller regions (-0.057 vs -0.131 from Column 4).

Table 9: Impact of Distance on Knowledge Flow: Spatial Spillover with Distance Indicator Variables

	Small Citing Small Cited	Small Citing Large Cited	Large Citing Large Cited	Large Citing Small Cited	Full Sample
Distance Continuous Variable	-0.131	-0.077	-0.057	-0.058	-0.104
Distance Indicator Variable Positive Up To	700 Miles	400 Miles	800 Miles	200 Miles	500 Miles
Impact of Distance Up To 50 Miles	209%	87%	73%	60%	142%
Impact of Distance 51-100 Miles	37%	40%	22%	31%	33%

Due to the semi-log nature of the regression, a coefficient of β for a dummy variable implies a $(e^{\beta} - 1) * 100$ percent change for the dependent variable. Being within 50 miles has a substantial positive impact on knowledge flow for small MSAs citing patents from small MSAs, increasing the flow by 209 percent. The localization of knowledge flows for these small MSAs decays rapidly, increasing knowledge flow by 37 percent for distances between 51 and 100 miles, and by 24 percent for distances between 101 and 200 miles. For large MSAs, citing patents from other large MSAs within 50 miles increases knowledge flow by 73 percent and within 51 to 100 miles increases knowledge flow by 22 percent.

3.3 Technology Compatibility

The coefficient for technological compatibility, as measured by the variable *Index*, is lowest (0.435) for a pair of large metropolitan areas and greatest (1.372) for a pair of smaller ones. The full sample estimate is 1.239. A possible explanation is that larger MSAs are more technologically diverse and hence capable of absorbing more diverse knowledge than smaller ones. To test this hypothesis we compare the mean technology compatibility index in both samples. Our expectation is that the technology compatibility index will be less in the large to large sample compared with the small to small sample. If two MSAs are technologically close, the compatibility index between them is high. A means difference test confirms that the large to large mean technology compatibility index, 0.467, is statistically smaller that the small to small mean index, 0.512. As expected, the standard deviation is substantially greater in the large to large sample compared with the small to small sample (0.146 vs 0.123).

The appropriate level of classification of patents is an issue in the selection of control patents using the choice-sampling method of measuring knowledge spillovers. The use of three-digit or finer six-digit technology class impacts the findings on localization of spillovers (Thompson and Fox-Kean, 2005; Murata et al., 2014). We examine the impact of the level of classification in the spatial interaction regression method by comparing the results of one-digit and two-digit technology classification in calculating the technology compatibility index and the resulting impact. The results using the two-digit classification (reported in the paper) compared with the one-digit classification are: the regression coefficient on the index variable is substantially larger; the regression coefficients on the distance indicator variables are slightly smaller; the regression coefficients on the spatial origin and destination effects are slightly smaller; and the border effect variables are slightly larger.[9] The overall results of the analysis remain unchanged. While technology classification is an important issue in the choice-sampling method, it does not materially impact the main findings using the spatial interaction regression method on the localization of knowledge spillovers.

3.4 Local Spatial Spillover Impacts

To address the potential for spatial spillover effects emanating from economic and technological conditions of neighboring regions, we use the spatial lag of *X* model. In this framework, spatial weights are applied to the model's explanatory variables. In the knowledge flow model, the key explanatory variables include the technological compatibility index, distance, a state border indicator, and citing MSA and cited MSA indicator variables. The results reported in Tables 4-8 include only results for the spatially weighted technological compatibility index, shown in columns 3 and 4. Including spatially weighted MSA indicator variables along with the unweighted indicator variables led to extreme multicollinearity problems. The coefficients of the other explanatory variables do not change with the inclusion of this spatial effect. Additionally, since including the own MSA indicator variable is more important than those of the neighbors, the results reported do not include any spatially weighted citing or cited MSA variables.

The origin and destination local spillover variables, based on the technology compatibility index, are positive but not statistically significant for the large-to-large MSA pairs when distance dummy variables are included. There are no local spillovers in knowledge flow between large MSA pairs. This result is consistent with the finding that technology compatibility exerts a small impact on knowledge flow between pairs of MSAs that are large. For the large and small dyads, the local spillover is stronger for regions surrounding the smaller MSAs. The spatial origin spillover is highest when large MSAs cite smaller ones, indicating that small MSAs surrounded by MSAs with high technological compatibility are more successful in disseminating their knowledge to larger MSAs. Researchers in the large MSA are likely more aware of knowledge in small MSAs if there are other surrounding small MSAs with similar technology compatibility. Reinforcing this impact is the spatial destination spillover being largest for small MSAs citing larger MSA knowledge indicating that small MSAs surrounded by others with high technology compatibility are more successful in absorbing knowledge from large MSAs.

4. CONCLUSION

The results for the sub-samples split by number of patents (innovation size) confirm our hypothesis that the geographic and technology resistance factors are less for knowledge flows between pairs of metropolitan areas that are both large and greater for pairs of MSAs that are both small. Innovative MSA size reduces the importance of spatial and technology proximity in mediating knowledge flows among regions. Moreover, technology compatibility local spillovers do not impact the knowledge flow among large MSAs, while they do impact knowledge flow when

[9] The results of using the one digit classification are not shown in the interest of brevity, and are available upon request.

small MSAs are involved in knowledge flow with a large MSA. Knowledge spillovers are highly localized among smaller MSAs with a very rapid decay. In the small to small sample, distances less than 50 miles increase knowledge flows by 209 percent, while distances between 51 and 100 miles increase knowledge flows by only 37 percent. Not considering innovation size in knowledge flow among city origin-destination pairs results in aggregation bias that masks significant differences in the determinants of knowledge flow.

For the small-to-small and large-to-large MSA pairs, the state border effect of being within the state is not statistically different from zero when fine-grained distance indicator variables are included. While we do find a positive and significant border effect for the other two sub-samples, they represent only 22 percent of the MSA pairs and there may be a California effect since five of the large innovation MSAs are in that state. Existing studies of the state border effect in the U.S. have not had an explanation for the significant effect. Singh et al. (2010) concluded that “future research should investigate why state borders seem to play a surprisingly robust role in shaping knowledge diffusion patterns.” Our results show that state borders do not play a substantial role in shaping knowledge diffusion patterns once fine-grained distance effects, spatial local spillover, and innovation size of metropolitan areas are accounted for.

Regions with an existing infrastructure that supports patent production at a high rate have an advantage in absorbing knowledge flow compared with smaller patent production regions. Our result is consistent with Packalen and Bhattacharya (2015) who found that larger city size provides considerable advantage in inventive activities. That advantage manifests itself in lowering the impact of proximity in mediating knowledge flows. The substantial difference in the localization of knowledge spillovers depending on innovative region size has implications for state innovation policy. State investment in innovation and research and development is based, in part, on the idea that knowledge spillovers are localized within a state. This idea is accurate only for investments in small innovative regions, but not for large innovative regions. Large innovative regions may, however, generate substantial knowledge spillovers within the region, an area for future research. The insignificant state border effect also suggests that state investments will not have as much knowledge spillovers as anticipated by policy makers.

REFERENCES

Agrawal, Ajay, Iain Cockburn, and John McHale. (2006) “Gone but Not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships,” *Journal of Economic Geography*, 6, 571–591.

Audretsch, David and Maryann P. Feldman. (1996) “R&D Spillovers and the Geography of Innovation and Production,” *American Economic Review*, 86, 630–640.

Belenzon, Sharon and Mark Schankerman. (2013) “Spreading the Word: Geography, Policy, and Knowledge Spillovers,” *Review of Economics and Statistics*, 95, 884–903.

Breschi, Stefano and Francesco Lissoni. (2001) “Knowledge Spillovers and Local Innovation Systems: A Critical Survey,” *Industrial and Corporate Change*, 10, 975–1005.

Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt. (2007) “Urban Density and the Rate of Invention,” *Journal of Urban Economics*, 61, 389–419.

Duguet, Emmanuel and Megan MacGarvie. (2005) “How Well Do Patent Citations Measure Knowledge Spillovers? Evidence from French Innovation Surveys,” *Economics of Innovation and New Technology*, 14, 375–393.

Fischer, Manfred M., Thomas Scherngell, and Eva Jansenberger. (2009) "Geographic Localisation of Knowledge Spillovers: Evidence from High-Tech Patent Citations in Europe," *Annals of Regional Science*, 43, 839–858.

Griffith, Rachel, Sokbae Lee, and John Reenen. (2007) " Is Distance Dying at Last? Falling Home Bias in Fixed Effects Models of Patent Citations," *National Bureau of Economic Research Working Paper 13338*: Cambridge, Massachusetts.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. (2001) "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," *National Bureau of Economic Research Working Paper 8498*: Cambridge, Massachusetts.

Hirschberg, Joe and Jenny Lye. (2001) "The Interpretation of Multiple Dummy Variable Coefficients: An Application to Industry Effects in Wage Equations," *Applied Economic Letters*, 8, 701–707.

O'hUallachain, Breandan (1999) "Patent Places: Size Matters," *Journal of Regional Science*, 39, 613–636.

Hussler, Caroline. (2004) "Culture and Knowledge Spillovers in Europe: New Perspectives for Innovation and Convergence Policies?," *Economics of Innovation and New Technology*, 13, 523–541.

Jaffe, Adam. (1989) "Real Effects of Academic Research," *American Economic Review*, 79, 957–970.

Jaffe, Adam, Manuel Trajtenberg, and Michael Fogarty. (2000) "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *American Economic Review*, 90, 215–218.

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

Kerr, William R. and Scott Duke Kominers. (2015) "Agglomerative Forces and Cluster Shapes," *Review of Economics and Statistics*, 97, 877–899.

Krugman, Paul. (1991) *Geography and Trade*. MIT Press: Cambridge, Massachusetts.

LeSage, James P. (2014) "What Regional Scientists Need to Know about Spatial Econometrics," *Review of Regional Studies*, 44, 13–32.

LeSage, James P. and Carlos Llano. (2007) "A Spatial Interaction Model with Spatially Structured Origin and Destination Effects," *Journal of Geographical Systems*, 15, 265–289.

LeSage, James P. and Manfred M. Fischer. (2010) *Spatial Econometric Method for Modeling Origin-Destination Flows Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, Berlin: Springer.

LeSage, James P. Manfred M. Fischer, and Thomas Scherngell. (2007) "Knowledge Spillovers across Europe: Evidence from a Poisson Spatial Interaction Model with Spatial Effects," *Papers in Regional Science*, 86, 393–421.

LeSage, James P. and R. Kelley Pace. (2008) "Spatial Econometric Modeling of Origin-Destination Flows," *Journal of Regional Science*, 48, 941–967.

Li, Yao. (2014) Borders and Distance in Knowledge Spillovers: Dying over Time or Dying with Age? - Evidence from Patent Citations," *European Economic Review*, 71, 152–172.

Maurseth, Per Botolf and Bart Verspagen. (2002) "Knowledge Spillovers in Europe: A Patent Citation Analysis," *Scandinavian Journal of Economics*, 104, 531–545.

Mukherji, Nivedita and Jonathan Silberman. (2013) "Absorptive Capacity, Knowledge Flows, and Innovation in U.S. Metropolitan Areas," *Journal of Regional Science*, 53, 392–417.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamota, and Ryuichi Tamura. (2014) "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach," *Review of Economics and Statistics*, 96, 967–985.

Packalen, Mikko and Jay Bhattacharya. (2015) "Cities and Ideas," *National Bureau of Economic Research Working Paper 20921*: Cambridge, Massachusetts.

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

Peri, Giovanni. (2005) "Determinants of Knowledge Flows and Their Effect on Innovation," *Review of Economics and Statistics*, 87, 308–322.

Rosenthal, Stuart and William Strange. (2001) "The Determinants of Agglomeration," *Journal of Urban Economics*, 50, 191–229.

Scherngell, Thomas and Michael J. Barber. (2011) "Distinct Spatial Characteristics of Industrial and Public Research Collaborations: Evidence from the Fifth EU Framework Programme," *Annals of Regional Science*, 46, 246–266.

Singh, Jasjit and Matt Marx. (2013) "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity," *Management Science*, 59, 2056–2078.

Singh, Jasjit Matt Marx, and Lee Fleming. (2010) "Patent Citations and the Geography of Knowledge Spillovers: Disentangling the Role of State Borders, Metropolitan Boundaries and Distance," *INSEAD Working Paper Series*: Fountainebleau, France.

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

Warnick, Walter. (2006) "Global Discovery: Increasing the Pace of Knowledge Diffusion to Increase the Pace of Science," in "American Association for the Advancement of Science Annual Meeting," pp. 1–10.