Heterogeneity in Knowledge Flows of Regions: Impact on Invention Quality

ABSTRACT

Using patent citations as measures of knowledge flows, we explore how the different types of knowledge flows in a region affect the quality of inventions originating in that region. We leverage a database of worldwide urban centers obtained from remote sensing data to provide a consistent definition for regions worldwide to demonstrate that patent citations provide inconclusive evidence for a clear positive effect of either local knowledge flows or the globalization of R&D on the invention output of regions. Additionally, we highlight the nature of the differences between patent citations classified as 'cited by applicant', 'cited by examiner' and 'cited by other' to contribute to the literature on the use of patent citations as measures of knowledge flows. Finally, our work here suggests that much more research is required to distill any stylized facts on the implications of geography and firm organization on invention outcomes.

Keywords:

Knowledge Flows, Clusters, Invention Quality

Results presented here are still at a preliminary stage. Kindly do not circulate.

Heterogeneity in Knowledge Flows of Regions: Impact on Invention Quality

INTRODUCTION

Empirical studies in the literature on economic geography have used patent citations to demonstrate that knowledge spillovers are geographically localized (Almeida & Kogut, 1999; Branstetter, 2001; Jaffe, Trajtenberg, & Henderson, 1993; Sonn & Storper, 2008). Scholars in this tradition have also used patent citations to show that innovation is more spatially concentrated than production (Feldman, 1994). Patent citations have also been used in the international business literature to demonstrate that multi-national firms (MNCs) gain from the cross-border flow of knowledge and the globalization of R&D (Singh, 2007; Singh & Marx, 2013; Zhao, 2006). This leads to the question of what indeed is the overall effect of geography on innovation outcomes of firms and regions. Are local knowledge flows more important in determining innovation outcomes or are cross-border flows more important? Firms seek to improve innovation outcomes as a way of building a sustainable competitive advantage. Understanding how prior knowledge flows (across firms and geographies) affect innovation outcomes would therefore be valuable to managers in their quest for superior performance.

In this study, we examine how the nature of knowledge spillovers or flows in a region affect the quality of the inventions generated in the region. Specifically, we look at knowledge flows spanning region boundaries and firm boundaries, and ask if and to what extent prior knowledge flows between and within regions and firms affects the quality of patents originating in a region. We use citation data of patents applied for between 2001 and 2012 (and citations received till March 2017) to empirically estimate the relationship between the relative share of knowledge flows of a region and the quality of inventions from that region. We are aware of no prior studies that have examined the effects of the geographical characteristics of knowledge flows on invention outcomes for all major geographic regions in the world. Our work seeks to contribute to the innovation literatures in both economic geography and international business, as well as to the debate on the use of patent citations as measures of knowledge flows.

Knowledge spillovers demonstrate two seemingly conflicting characteristics (Griliches, 1979). While on the one hand, "rent spillovers" exhibit characteristics of a private good (by virtue of knowledge being rival and excludable), the knowledge spillovers associated with R&D activity ("pure spillovers") are non-rival and non-excludable, and hence exhibit characteristics of a public good (Arrow, 1962). This dual nature of knowledge notwithstanding, knowledge flows have been very hard to measure. While Krugman (1991) suggested that knowledge flows may be largely invisible, Jaffe, Trajtenberg, & Henderson (1993) demonstrated that knowledge flows do sometimes leave a paper trail in the form of patent citations. While the knowledge flows literature in both economic geography and international business traditions have built on top of this landmark paper, the use of patent citations as measures of knowledge flows has not been without controversy (Alcácer & Gittelman, 2006; Arora, Belenzon, & Lee, 2017). In the current paper, we continue to build on prior research using patent citations as a proxy for knowledge flows despite recognizing the limitations of using this measure.

The rest of this paper is organized as follows. The following section reviews the literature on knowledge flows in the fields of economic geography and international business. We then define a framework to classify knowledge flows in a region as a process of search across geographical boundaries and firm boundaries. We then motivate our work by demonstrating how a sample of prominent regions fare differently in terms of knowledge flows. In the next section, we position our study as an empirical horse race between the nature of knowledge flows in a region. Our preliminary results are then presented, followed by a discussion of the results. We conclude with limitations, next steps and open questions for further research.

We now review the literature on knowledge flows in two academic traditions - economic geography on the one hand and international business on the other, and suggest that the literature points toward somewhat opposing effects of the nature of knowledge flows on invention quality.

THEORY

Clusters and agglomeration economies have been suggested to play an important role in fostering innovation (Marshall, 1890; Porter, 1990). Agglomeration economies may arise due to labor pooling advantages, economies of specialization of local suppliers, and knowledge spillovers (Krugman, 1991; Porter, 1990). In this study, we focus on the role of knowledge spillovers (or simply, knowledge flows) in making clusters more innovative. Several studies have used patent citations to demonstrate that knowledge flows are localized (Almeida & Kogut, 1999; Jaffe et al., 1993). Audretsch & Feldman (1996b) demonstrated that innovation exhibits a pronounced tendency to cluster spatially even after controlling for the geographic distribution of production. Unsurprisingly, regions have been shown to demonstrate heterogeneity in their inventive output (Agrawal, Cockburn, Galasso, & Oettl, 2014). This argument is further strengthen by Acs, Audretsch, & Feldman (1994) who demonstrated that new product introductions were more geographically concentrated than patents as they derived important inputs from universities and industrial R&D. It seems reasonable therefore, that the nature of knowledge flows in a region may be one possible source of this variation of inventive output across regions.

Some scholars have suggested that location may matter more at the earliest stage of the industry life cycle (Audretsch & Feldman, 1996a) since early stages of industries may be characterized by the importance of tacit knowledge. Innovation, when viewed as recombinant search on prior knowledge, then requires access to a multiple sources of tacit knowledge held by skilled knowledge workers. Since costs of collaboration are lower due to geographic proximity, firms and inventors may exhibit a higher propensity to cluster geographically so as to benefit from knowledge spillovers. Additionally, proximity may promote serendipitous and chance encounters that create an environment conducive for new ideas, newer solutions and to innovation.

Regional institutions such as universities, research labs and national innovation systems (Cooke, 1996; Howells, 1996, 2002; Maskell & Malmberg, 1999) create incentives toward an increased localization of knowledge flows. Political borders, both at the state and country level have been shown to constrain the flows of knowledge (Singh & Marx, 2013). Given the difficulties in transferring knowledge over geographical boundaries on the one hand, and the ease and efficiency of localized knowledge use on the other, the case for the benefits from the localization of knowledge flows appears strong.

Literature in international business has also recognized that regional innovation systems may influence MNCs to take advantage of knowledge spillovers in far off regions (Andersen & Christensen, 2005). Scholars in this tradition have also demonstrated that organizational capabilities (Zhao, 2006), property rights, technological sourcing (Florida, 1997), higher product ownership and interdependence (Pearce, 1999), and access to highly skilled human capital are other alternative factors that drive MNC R&D productivity. In addition, mobility of inventors across political borders has also been shown to help firms improve innovation outcomes (Alnuaimi, Opsahl, & George, 2012). The literature in this tradition seems to unequivocally suggest that there are immense benefits to operating a global R&D pipeline.

A strand of literature that is distinct from the two schools of scholarship discussed above, yet related to the topic at hand suggests that diversity within a region may be a factor that affects the innovative output of regions (Feldman, 1999). Scholars in this tradition argue that related variety (Boschma & Iammarino, 2009; Frenken, Oort, & Verburg, 2007) may help in the generation of new ideas and that diversity between industrial activities may help in the transfer of ideas within regions (Jacobs, 1969). While the end effect argued for by scholars in this tradition is for localization of knowledge flows (similar to other economic geography scholars), the mechanisms they suggest for this effect are similar to those offered by the scholars in the international business literature.

This brief review of the literature in the traditions of economic geography, international business and related variety suggests positive effects of both increased localization and increased globalization of knowledge flows on innovation outcomes.

While agglomerations may seem to benefit firms through access to tacit knowledge, it may also increase the number of competitors who are specialized in the same market segment. MNCs ability to tap into global pipelines (Bathelt, Malmberg, & Maskell, 2004) seem to be affected by local factors, as well as by relative ties with local firms and MNC affiliates. In the absence of a clear theoretical basis to determine the case one way or another, it seems appropriate to approach this study by pitting these alternate types of knowledge flows in an empirical "horse race". In order to provide a theoretical structure to this empirical strategy, we now present a framework that motivates this argument.

Effects of knowledge flows on quality of inventions

We formalize the discussions in the previous sections by categorizing all knowledge flows along two dimensions: first, whether the knowledge flows among inventors are local to a geographical region or not, and second, whether knowledge flows are within the boundary of the firm or not. The motivation behind defining such a structure is to both keep it simple, as well as to focus on aspects that firms and governments may have an ability to influence. Given the nuanced and subtle nature of knowledge flows, and the difficulty in measuring them, firms may be better served with simple and actionable innovation strategies.

Our classification allows us to analyze knowledge flows in four mutually exclusive, but collectively exhaustive categories as illustrated in Figure 1. We next describe each quadrant and discuss how the category of knowledge flow in that quadrant can affect invention quality. Within the context of this framework, we ask what is the net effect of each category of knowledge flow on invention quality, and which category of flow has the largest effect on invention quality.

Insert FIGURE 1 about here.

The top left quadrant, labelled an "Independent Research Center" captures those knowledge flows that are within the region and within the firm. These knowledge flows are not strictly speaking knowledge spillovers, but those that reflect competence building. Since these knowledge flows are both within the region and within the firm, these flows represent local search on two dimensions (within firm and within region). Thus, while the competence that is being built up by the Independent Research Center can be expected to have a positive effect on invention quality, the localness of the search on both dimensions may have a negative effect on invention quality. Figure 2 depicts the knowledge flows for this category (percentage of backward citations from this region that are to the same firm or assignee and same region) across time for five regions: Bangalore, Beijing, Tel Aviv-Yafo, Boston and San Jose (core of "Silicon Valley"). While our empirical analysis covers all the major regions of the world, we chose these five regions as illustrative examples. We note that both San Jose and Boston report a higher proportion of knowledge flows within the same firm in the same region, while Bangalore and Tel Aviv-Yafo have the lowest proportion (fewer than 1%) of their citations from the same firm within the same region.

Insert FIGURE 2 about here.

The quadrant on the bottom left, labelled "Cluster" captures knowledge spillovers within a region. Here firms may be seen as performing local search on one dimension (within regions) but not the other (within firms). Figure 3 depicts the knowledge flows for this category across time for the same five regions. San Jose clearly stands out from the rest, suggesting a higher amount of across firm flows of knowledge in Silicon Valley, a result consistent with several prior studies.

Insert FIGURE 3 about here.

The quadrant on the top right, labelled as "Geographic Diversification" captures local search on the dimension of the firm (across geographies) but not across regions. Innovations that are built on knowledge from several regions can be expected to benefit from the diversity of knowledge across regions. Yet, as in the previous quadrant, there is localness along the dimension of firm and such localness can have a negative effect on invention quality (Rosenkopf & Nerkar, 2001). Figure 4 depicts the knowledge flows for this category across time for the five regions. We note that Bangalore and Beijing have a relatively higher proportion of knowledge flows from same assignees in different locations, thus confirming the role of these regions as R&D outposts of multinational firms.

Insert FIGURE 4 about here.

Finally, the bottom right quadrant labelled "Diffusion" captures high exploration along both dimensions, indicating the development of a global pipeline (Bathelt, Malmberg, & Maskell, 2004). Figure 5 depicts the knowledge flows for this category across time for the five regions. We note that Bangalore, Beijing and Tel Aviv-Yafo have a higher level of knowledge flows from other firms in other regions compared to Boston and San Jose, which is to be expected given that the absolute level of innovative activity in these emerging hotspots is still lower compared to that in Boston and San Jose.

Insert FIGURE 5 about here.

As can be seen from the preceding discussion, prior theory suggests both positive and negative effects for each of these four categories of knowledge flows and it is not clear what the net effect will be on invention quality. Table 1 captures the positive and negative effects suggested by theory. This suggests that prior theory does not provide guidance on which category of knowledge flows will have the highest effect on invention quality. Since theory does not provide us with an answer, we rely on empirical analysis to inform us on the net effect of each category of knowledge flow on invention quality and which category has the highest effect on invention quality.

Insert TABLE 1 about here.

DATA AND METHODS

We use patent citations data from the U.S. Patent Office (USPTO) as provided by patentsview.org. We use U.S. patent data for all firms, technology classes, and all assignees foreign and domestic. Using the U.S. patent data provides us a consistent, reliable and comparable method to discern the knowledge flow patterns for firms and regions across the world. Since the U.S. is one of the largest markets for technology, firms and inventors who patent at all are highly likely to patent in the U.S. Additionally, prior research on knowledge flows in international contexts have also leveraged U.S. patents.

The USPTO provides information on assignee, the location of inventors, and year of patent application. We use this information to identify which patents belong to which region in a given year. Since there can be more than one inventor on a patent, a patent can belong to more than one region. We compare the assignee id and inventor location of each patent with each assignee id and inventor of each backward citation to identify the category of knowledge flow (e.g., same assignee, different location) indicated by the backward citation. Additionally, to map location data of inventors from USPTO to regions, we use urban centers data for worldwide locations from Natural Earth Data that uses remote sensing data to determine urban agglomerations (a process developed in Schneider, Friedl, McIver, & Woodcock (2003)). While it has been common practice to use Metropolitan Statistical Areas (MSA) for analyses related to economic geography in the U.S., an equivalent measure is unavailable for the rest of the world. For comparability and consistency, we choose to use the urban centers definitions from Natural Earth Data for all regions both within U.S. and outside U.S.

Our unit of analysis is the region-year. We do this so as to be able to understand the relationship between the nature of knowledge flows in a region and the invention quality of the region. Controlling for technology at the subcategory level, and for time at the year level, our analysis attempts to answer the question about the inventive performance of regions.

Applicant Citations and Examiner Citations

Much of the prior work analyzing patent citations had pooled all citations made irrespective of the source (i.e., cited by applicant, or cited by examiner, or cited by other) due to lack of systematic classification of the citation source in the USPTO database. However, systematic classification of patent citations is available from 2001 onwards. Scholars have even called for more granular analysis based on citation source. For example, Alcácer & Gittelman (2006) suggest that using pooled citations to make inferences about inventor knowledge may suffer from bias or overinflated significance levels. Scholars have argued that patents cited by the examiner may not represent knowledge flows at all. To be consistent with our objective of measuring knowledge flows without making strong assumptions, we empirically study citations categorized as 'cited by applicant' as well as those categorized as 'cited by examiner'. In addition, in the appendix we also report the results for citations categorized as 'cited by other' as well as all citations irrespective of classification. This decision has the effect of limiting our period of analysis to citing patents applied for from the year 2001 since the data on which citations were added by examiners, applicants, third party, and others is available for patents from only 2001 onwards.

We describe our primary variables using two examples from the data. Figure 6 depicts the measures of knowledge flows across region and assignee boundaries for two regions Bangalore, India and San Jose, California, USA for the year 2010. We construct our four predictor variables along the framework presented in Figure 1. In 2010, Bangalore had fewer same region, different assignee citations made (51) than different region, same assignee citations made (2708), while San Jose had more same region, different assignee citations made (72505) than same region, same assignee citations made (23908). This is consistent with prior studies that have shown that cluster flows are strong in the Silicon Valley region of California, while Bangalore has been an important offshore research center for MNCs. Figure 6 shows how the citations made along the four quadrants add up, and how the citations receive add up across the dependent variables of non-self citations received, and self-citations received.

Insert FIGURE 6 about here.

In Table 2, we present the step-by-step by which we construct our data starting from a patent citation. Starting from a citation as a flow from one patent to another, we expand this record to identify the assignees on each side. Some patents may be assigned to multiple assignees, and in this case we ensure that all entries are retained and all combinations are also retained. In step 3, we expand this record to include all inventors on each patent. As before, no inventors are dropped, and every combination of inventors on the citing and cited patent are maintained. This results in a multiplication of the number of records from the initial patent citation. The following steps identify the geographic region of the inventor on each side and retain only unique region to region flows for a given patent citation. This step may possibly drop duplicate flows so as to capture a flow from one region to another on a patent citation only once.

Insert TABLE 2 about here.

Citations Received

We restrict our sample to patents applied for between 2001 - 2012, but citations received till 7 March 2017 (this was the most recent available cut of the USPTO data). Our strategy for measuring patent citations is to count citations received all the way till the late date for which we have data. While some scholars have argued that doing so may favor longer standing patents ahead of recent ones since they have had a longer time to accumulate citations, we accommodate this by controlling for the year of citation by using year dummies in all our regression models.

Dependent Variable

Our primary dependent variable is the total count of citations received by patents belonging to a region-year. However, for all models we also report results for total non-self citations received as the dependent variable. This measurement is superior to counting citation received up to a finite time window (say 5 years), because patents from different technology subcategories may exhibit different patterns in citations received over time and setting a fixed time window may end up skewing results. By controlling with year dummies and technology subcategories, the count of citation received till as far as data is available may be less subject to distortion.

Independent Variables

Building on the framework depicted in Figure 1, our independent variables are the percentage of backward citations made to each of the four categories: those to a) same region, same assignee, b) same region, different assignee, c) different region, same assignee, and d) different region, different assignee. Many patents may not make any citations to prior work. In such cases our independent variables are not defined, and those observations automatically get dropped from our analysis. An interesting effect of this shows up when comparing applicant cited patents and examiner cited patents in the following section.

Share of backward citations made to same region and same assignee For each region and for each year, we compute the ratio of the total number of backward citations made to same region and same assignee in that year to the total number of backward citations made in that year.

Share of backward citations made to same region and different assignee For each region and for each year, we compute the ratio of the total number of backward citations made to same region and different assignee in that year to the total number of backward citations made in that year.

Share of backward citations made to different region and same assignee For each region and for each year, we compute the ratio of the total number of backward citations made to different region and same assignee in that year to the total number of backward citations made in that year.

Share of backward citations made to different region and different assignee For each region and for each year, we compute the ratio of the total number of backward citations made to different region and different assignee in that year to the total number of backward citations made in that year.

For any given region-year, we would therefore have the sum of each of the four independent variables to add to 1 (one). We compute the match on region and assignee as follows. For region, we consider the location of each inventor dyad in the citing patent - cited patent record. So if the citing patent had 3 inventors, and the cited patent had 4 inventors, we would consider 12 unique dyads, and how their location as recorded in the patent database matched up. If both inventors' urban center region (based on matched Natural Earth Data definition of urban centers), the knowledge flow would be classified as being within region. If any of the inventors' did not fall within the urban centers definition, then those inventor locations falling within 50 kilometres of each other would be classified as being from the same region, and others classified as being from a different region. We used a full match on the assignee id (as provided by the patentsview.org database) as the basis for determining if the assignees on the citing and cited patents were the same or not.

Control Variables

We control for the total number of citations made in the region-year, the total number of patents in the region-year, the size of the patent pool in the region-year, as well as the percentage of patents in region-year in each technology subcategory as defined by Hall, Jaffe, & Trajtenberg (2001). The patent pool is the total number of patents that belong to a region from 1976 (which is where we have data from) up to the previous year. It is important to control for the size of the patent pool as regions that have a larger patent pool (such as San Jose) will have more patents that can be cited and can therefore have larger within region spillovers as compared to a region that has only a small patent pool. We include region fixed effects and year dummies in all regression models so as to control for region level and year specific effects, if any. Since our dependent variable is a count variable, we use negative binomial regression analysis with fixed effects.

Insert TABLE 3 about here.

Table 3 displays the correlation coefficients and summary statistics of our primary variables in the dataset of applicant citations. Table 4 displays the correlation coefficients and summary statistics of our primary variables in the dataset of examiner citations. The appendix provides similar statistics for the sample of other citations, not including applicant and examiner citations (Table A1), and the sample with all citations (applicant, examiner and other citations, Table A3)

Insert TABLE 4 about here.

RESULTS

We observe from Table 3 that our four predictor variables (variables 4-7) are uncorrelated among themselves and also with the predicted variables (variables 1-3). We note also that the the three measures of invention quality (Total Citations Received, Non-Self Citations Received, and Non-Self Citations Received) are highly correlated. This is known to be the case from prior research.

The results from our analysis for the dataset of applicant cited patents are presented in Table 5. Models 1-3 report results with the dependent variable as the total number of citations received for all regions worldwide, U.S. locations and non-U.S. locations respectively. Models 4-6 report results with the dependent variable as the number of non-self citations received. Since the share of each of the four quadrants of our framework cumulatively add up to one, we report results by using share of citations made to same region, same assignee as the reference category. All models include year dummies, region fixed effects, and controls for technology subcategories based on Hall, Jaffe, & Trajtenberg (2001).

The results in Table 5 suggest that applicant citations do not indicate a significant effect of "cluster" flows on invention quality of regions. This seems to hold for both total citations received as well as for non-self citations received. However the effect of "geographic diversification" is more mixed. Table 5 suggests that the overall positive and significant effect of "geographic diversification" is driven by the effect from U.S. Locations, while Non-U.S. locations do not suggest a significant effect. This maybe because U.S. multinationals may be able to leverage their global network of subsidiaries to drive a higher number of citations for their own patents, something that non-U.S. locations may not be able to emulate. We also note no significant effect for "diffusion" as may have been expected.

Insert TABLE 5 about here.

The results from our analysis for the dataset of examiner cited patents are presented in Table 6. Models 1-3 report results with the dependent variable as the total number of citations received for all regions worldwide, U.S. locations and non-U.S. locations respectively. Models 4-6 report results with the dependent variable as the number of non-self citations received. Examiner citations seem to suggest very different effects for the nature of knowledge flows on invention quality. First, we note that the signs of the coefficients for each of the three categories of flows is reversed as compared to the dataset of applicant citations. Second, we find that the effects from U.S. Locations dominate for both dependent variable measures. This seems to suggest that examiner citations may apply differently than applicant citations.

Insert TABLE 6 about here.

In comparing the results form applicant citations with those from examiner citations, we note a few interesting patterns. First, the number of observations in the applicant citations set is approximately half (8947 observations for the dependent variable as total citations received and sample as all locations) as compared to examiner citations (16464 observations for the dependent variable as total citations received and sample as all locations). This is so because many patents do not make any applicant citations at all, and these observations drop off since the denominator of the independent variables go to zero in this case. An interesting artifact of this effect is the much larger means in citations received for applicant citations (mean of 1763 total citations received in a region-year) as compared with examiner citations (mean of 582 total citations received in a region-year).

Consider the coefficient estimates of Log(Total Citations Made). While both applicant citations and examiner citations report a positive and statistically significant effect, the effect is much more pronounced among the examiner citations. Thus, while a unit increase in Log(Total Citations Made) leads to a 1.019 (coefficient estimate is 0.0198, as is interpreted as the log of the ratio of the dependent variable after and before the unit increase in the independent variable) times increase in total citations received (sample is all locations), the effect is 1.29 times (coefficient estimate is 0.255) for examiner citations. This suggests that while applicants citing more patents improves the patent's invention quality only marginally, that effect is about 29% more pronounced when citations are made by examiners.

We also note across applicant and examiner citations that while Log(Number of Patents) has a large, positive and significant effect on invention quality, Log(Patent Pool Size) has a statistically significant negative effect. This may suggest that the intensity of current patenting activity improves invention quality outcomes, but that the effect withers over time. Regions with a glorious past may therefore only continue to do well so long as their recent patenting activity is high. This may demonstrate the lowering of invention quality of regions that are stagnating in terms of R&D investments.

DISCUSSION

From our preceding analysis of applicant citations, we find that after controlling for geographical region, for time (year of patent application), and for technology, the share of cluster flows are not significant predictors of invention quality as compared to within-firm local flows. On the other hand, we find some effect, though not consistent across all geographies of MNC-Subsidiary knowledge flows positively influencing invention quality of regions. This raises interesting questions about the effect of within-region knowledge flows in influencing invention outcomes. The sample of examiner citations sample suggests a significant but negative effect of knowledge flows on invention outcomes of regions. This seems somewhat consistent with prior research (Alcácer & Gittelman, 2006) in that the effect of examiner added citations seems less clear.

Knowledge flows are hard to measure, and knowledge is characterized by conflicting qualities of being tacit and coded. Additionally, knowledge flows demonstrate characteristics of a public good some of the time and that of a private good at other times. The central challenge therefore seems to be to find a way to disentangle the opposing qualities of knowledge. This may perhaps explain the different results we observe between the two dependent variable measures of total citations received and non-self citations received. In fact, the 2005 debate in the American Economic Review highlights how minor changes in measures can dramatically alter results when measuring knowledge and its effects (Henderson, Jaffe, & Trajtenberg, 2005; Thompson & Fox-Kean, 2005). The question then is if there is an alternative means to measure knowledge flows more reliably, or if the problem lies with the research methodology used herein. Maurseth & Verspagen (2002) have suggested a regional compatibility index for capturing technological linkages. An alternate approach of understanding knowledge flows relies on the movement of people with the assumption that knowledge is embedded within an individual. While patent citations arguably mapped codified knowledge, mobility captures a measure of tacit knowledge (Polanyi, 2015). However even in this tradition, scholars have suggested opposite effects of knowledge flows. While Almeida & Kogut (1997) show that mobility patterns of star inventors directly affect the geographic patterns of knowledge spillovers, Song et al. (2003) illustrate that mobile engineers who join a firm with stronger path dependence are unlikely to build on the knowledge of their previous firms. While geography may be seen as providing a platform to organize these interactions amongst inventors, it seems that we may need to consider other dimensions of geography so as to fully understand the effect on invention outcomes (Bunnell & Coe, 2001). Universities, research consortia, and other institutions may play a significant role in affecting invention outcomes.

It would seem appropriate to raise two measurement related issues here. First, despite well known issues in the use of patents (Griliches, 1990; Scherer, 1984) and in the use of patent citations to demonstrate localization of knowledge spillovers (Arora, Belenzon, & Lee, 2017; Thompson & Fox-Kean, 2005), much work in understanding knowledge spillovers has continued to rely on patent citations. While our analysis in this paper is singularly dependent on data of United States patents, we are careful to not make strong assumptions about patent citations reflecting underlying knowledge flows. Wherever possible, we consider competing hypotheses built on opposing assumptions about what patent citations may capture. Second, innovations are clearly not the same as inventions. The measure of an innovation lies in its acceptance in the marketplace. An invention may therefore only represent an early event in the innovation process. While we recognize that managers and policy makers are interested in innovation outcomes, we continue to be limited by data availability that allows us to only make claims about invention quality. Despite those caveats, pursuing the questions posed above can provide not just interesting trade-offs to explore, but indeed create opportunities for a larger research agenda on actionable strategies for firms seeking to improve their innovation outcomes.

LIMITATIONS

In addition to the customary disclaimers about using patent citations as measures of knowledge flows, we would like to highlight a couple of limitations that may stem from the methodology we adopted. While the use of patent citations as a measure of knowledge flows has been popular in the literature, this may nevertheless be subject to error (Arora, Belenzon, & Lee, 2017). Our definition of regions is dependent on the latitude/longitude assignment in the patentsview.org data and on the urban centers definition in the Natural Earth Data. Any systematic biases in the definition of regions can create biases in measures of within region and across region knowledge flows.

Methodologically, there may be mechanisms operating at multiple levels that we are not capturing. This issue may be beyond a simple ommitted variable problem, since emergent effects from lower level phenomena may not be assumed to have similar characteristics as the underlying lower-level phenomena. Knowledge flows may therefore need to be understood as emergent from underlying effects so this may be modeled correctly. Finally, since the our unit of analysis is the Region-Year, the implications that may be drawn for firms and inventors making location choices may be somewhat limited. This may be overcome by investigating the impact on invention quality at the Inventor-Region-Year or the Firm-Region-Year level.

FURTHER WORK

While still at a preliminary stage, our analysis seem to suggest that knowledge flows are extremely sensitive to the nature of measurement deployed. This casts a doubt on the widely accepted idea that local knowledge spillovers are an important source of agglomeration economies. On the other, it also brings us to bear on the important issue of measuring knowledge flows. One way to try and untangle the cumulative non-effect could be to explore contingent variables such as technology relatedness that may explain different effects on invention outcomes. Another potential extension of the study could be to conduct empirical analysis at the level of firm-year rather than at the region-year. This may help explain the variation in invention quality through potential firm level characteristics. Yet another direction we could consider would be to explore technology as the third dimension (within and outside technological domain, Rosenkopf & Nerkar (2001)) in addition to those of within/outside region and within/outside firm. This may provide us with a more nuanced understanding of the factors affecting invention quality. Future studies could potentially examine other measures of invention outcomes such as breakthrough inventions. Finally, while our work suggests multiple effects of knowledge flows on invention quality, it is not quite as clear why this may be the case. We have attempted to provide some direction by building from the theory on both economic geography and international business, but there is clearly much more to be done. We hope that our current work spurs further research in this direction.

REFERENCES

Acs, Z. J., Audretsch, D. B., & Feldman, M. P. 1994. R&D spillovers and recipient firm size. ***The Review of Economics and Statistics***, 76(2): 336–340.

Agrawal, A., Cockburn, I., Galasso, A., & Oettl, A. 2014. Why are some regions more innovative than others? The role of small firms in the presence of large labs. ***Journal of Urban Economics***, 81: 149 – 165.

Alcácer, J., & Gittelman, M. 2006. Patent citations as a measure of knowledge flows: The influence of examiner citations. ***Review of Economics and Statistics***, 88(4): 774–779.

Almeida, P., & Kogut, B. 1997. The exploration of technological diversity and geographic localization in innovation: Start-up firms in the semiconductor industry. ***Small Business Economics***, 9(1): 21–31.

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

Alnuaimi, T., Opsahl, T., & George, G. 2012. Innovating in the periphery: The impact of local and foreign inventor mobility on the value of indian patents. ***Research Policy***, 41(9): 1534 – 1543.

Andersen, P. H., & Christensen, P. R. 2005. ***From localized to corporate excellence: How do MNCs extract, combine and disseminate sticky knowledge from regional innovation systems***. DRUID Working Paper 16, Danish Research Unit on Industrial Dynamics.

Arora, A., Belenzon, S., & Lee, H. 2017. ***Reversed citations and the localization of knowledge spillovers***. Working Paper 23036, National Bureau of Economic Research.

Arrow, K. J. 1962. The economic implications of learning by doing. ***The Review of Economic Studies***, 29(3): 155–173.

Audretsch, D. B., & Feldman, M. P. 1996a. Innovative clusters and the industry life cycle. ***Review of Industrial Organization***, 11(2): 253–273.

Audretsch, D. B., & Feldman, M. P. 1996b. R&D spillovers and the geography of innovation and production. ***The American Economic Review***, 86(3): 630–640.

Bathelt, H., Malmberg, A., & Maskell, P. 2004. Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. ***Progress in Human Geography***, 28(1): 31–56.

Boschma, R., & Iammarino, S. 2009. Related variety, trade linkages, and regional growth in italy. ***Economic Geography***, 85(3): 289–311.

Branstetter, L. G. 2001. Are knowledge spillovers international or intranational in scope?: Microeconometric evidence from the U.S. and Japan. ***Journal of International Economics***, 53(1): 53 – 79.

Bunnell, T. G., & Coe, N. M. 2001. Spaces and scales of innovation. ***Progress in Human Geography***, 25(4): 569–589.

Cooke, P. 1996. Regional innovation systems: An evolutionary approach. In H. Baraczyk, P. Cooke, & R. Heidenreich (Eds.), ***Regional Innovation Systems***. London: UCL Press.

Feldman, M. P. 1994. ***The geography of innovation***. Boston: Kluwer Academic Publishers.

Feldman, M. P. 1999. The new economics of innovation, spillovers and agglomeration: Areview of empirical studies. ***Economics of Innovation and New Technology***, 8(1-2): 5–25.

Florida, R. 1997. The globalization of R&D: Results of a survey of foreign-affiliated R&D laboratories in the usa. ***Research Policy***, 26(1): 85 – 103.

Frenken, K., Oort, F. V., & Verburg, T. 2007. Related variety, unrelated variety and regional economic growth. ***Regional Studies***, 41(5): 685–697.

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

Griliches, Z. 1990. Patent statistics as economic indicators: A survey. ***Journal of Economic Literature***, 28(4): 1661–1707.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. 2001. ***The NBER patent citation data file: Lessons, insights and methodological tools***. Working Paper 8498, National Bureau of Economic Research.

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

Howells, J. 1996. Tacit knowledge. ***Technology Analysis & Strategic Management***, 8(2): 91–106.

Howells, J. R. L. 2002. Tacit knowledge, innovation and economic geography. ***Urban Studies***, 39(5-6): 871–884.

Jacobs, J. 1969. ***The economy of cities***. New York: Random House.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. ***The Quarterly Journal of Economics***, 108(3): 577–598.

Krugman, P. R. 1991. ***Geography and trade***. The MIT press.

Marshall, A. 1890. ***Principles of economics***. London: Macmillan, first Ed.

Maskell, P., & Malmberg, A. 1999. Localised learning and industrial competitiveness. ***Cambridge Journal of Economics***, 23(2): 167–185.

Maurseth, P. B., & Verspagen, B. 2002. Knowledge spillovers in Europe: A patent citations analysis. ***Scandinavian Journal of Economics***, 104(4): 531–545.

Pearce, R. D. 1999. Decentralised R&D and strategic competitiveness: Globalised approaches to generation and use of technology in multinational enterprises (mnes). ***Research Policy***, 28(2–3): 157 – 178.

Polanyi, M. 2015. ***Personal knowledge: Towards a post-critical philosophy***. University of Chicago Press.

Porter, M. E. 1990. ***The competitve advantage of nations***. New York: Free Press.

Rosenkopf, L., & Nerkar, A. 2001. Beyond local search: Boundary-spanning, exploration, and impact in the optical disk industry. ***Strategic Management Journal***, 22(4): 287–306.

Scherer, F. 1984. Using linked patent and R&D data to measure interindustry technology flows. In Z. Griliches (Ed.), ***R&D, patents, and productivity***, 417–464. University of Chicago Press.

Schneider, A., Friedl, M. A., McIver, D. K., & Woodcock, C. E. 2003. Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data. ***Photogrammetric Engineering & Remote Sensing***, 69(12): 1377–1386.

Singh, J. 2007. Asymmetry of knowledge spillovers between mncs and host country firms. ***Journal of International Business Studies***, 38(5): 764–786.

Singh, J., & Marx, M. 2013. Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity. ***Management Science***, 59(9): 2056–2078.

Song, J., Almeida, P., & Wu, G. 2003. Learning–by–hiring: When is mobility more likely to facilitate interfirm knowledge transfer? ***Management Science***, 49(4): 351–365.

Sonn, J. W., & Storper, M. 2008. The increasing importance of geographical proximity in knowledge production: An analysis of US patent citations, 1975–1997. ***Environment and Planning A***, 40(5): 1020–1039.

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

Zhao, M. 2006. Conducting R&D in countries with weak intellectual property rights protection. ***Management Science***, 52(8): 1185–1199.

APPENDIX: ADDITIONAL RESULTS

We also explored two additional samples of the data - the sample of non-applicant, non-examiner citations (we refer to this as other citations), and the sample of all citations (including applicant made, examiner made and other citations). They are presented below for completeness, but have not been exhaustively analyzed.

The correlation and summary statistics of the primary variables for the non-applicant, non-examiner citations sample is presented in Table A1. The regression results from this sample is presented in Table A2. Models 1-3 report results with the dependent variable as the total number of citations received for all regions worldwide, U.S. locations and non-U.S. locations respectively. Models 4-6 report results with the dependent variable as the number of non-self citations received. The divergence between results for the two outcome measures of total citations received and non-self citations received are similar to that observed in Table A4 on the dataset of all citations.

Insert TABLE A1 about here.

Insert TABLE A2 about here.

Table A3 and Table A4 present the correlations and regression results respectively, for the sample of all citations, including applicant citations, examiner citations and other citations.

Insert TABLE A3 about here.

Insert TABLE A4 about here.

FIGURE 1

Categories of Knowledge Flows

	Same Region	Different Region
Same Assignee	Independent Research Center	Geographic Diversification
Different Assignee	Cluster	Diffusion

FIGURE 2

Knowledge Flows Within Regions and Within Assignees

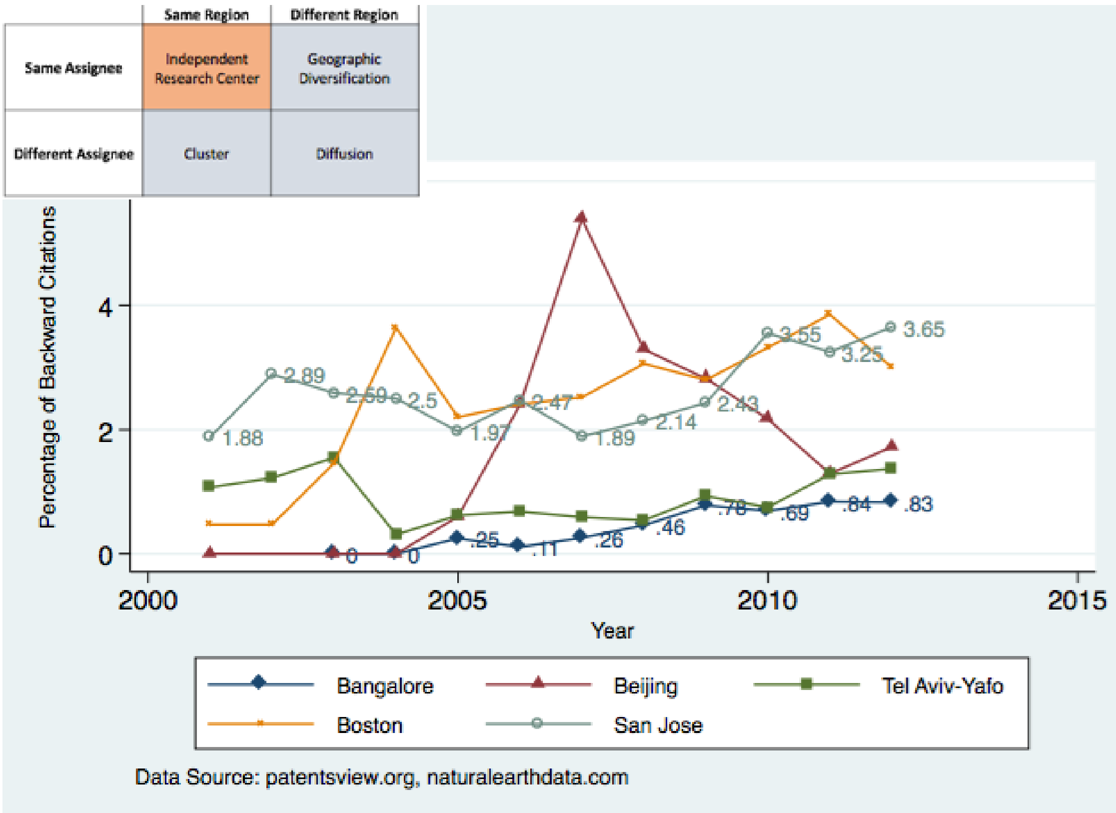

FIGURE 3

Knowledge Flows Within Regions and Across Assignees

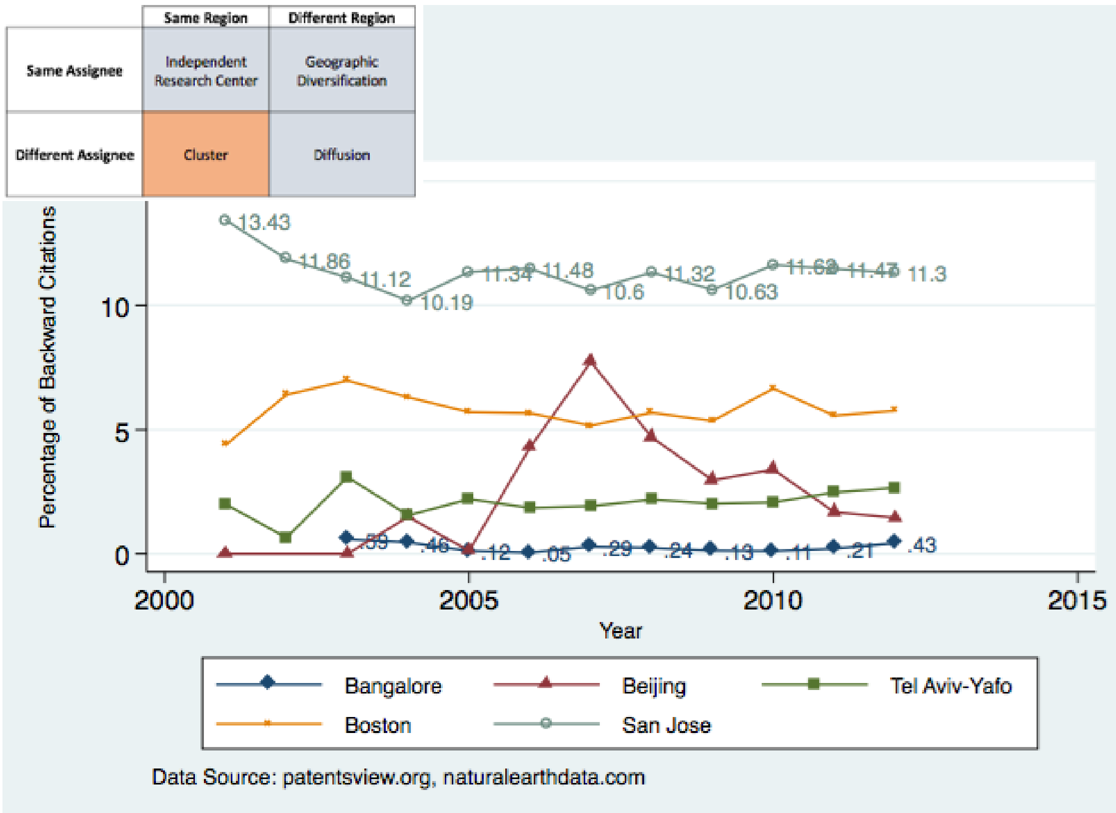

FIGURE 4

Knowledge Flows Across Regions and Within Assignees

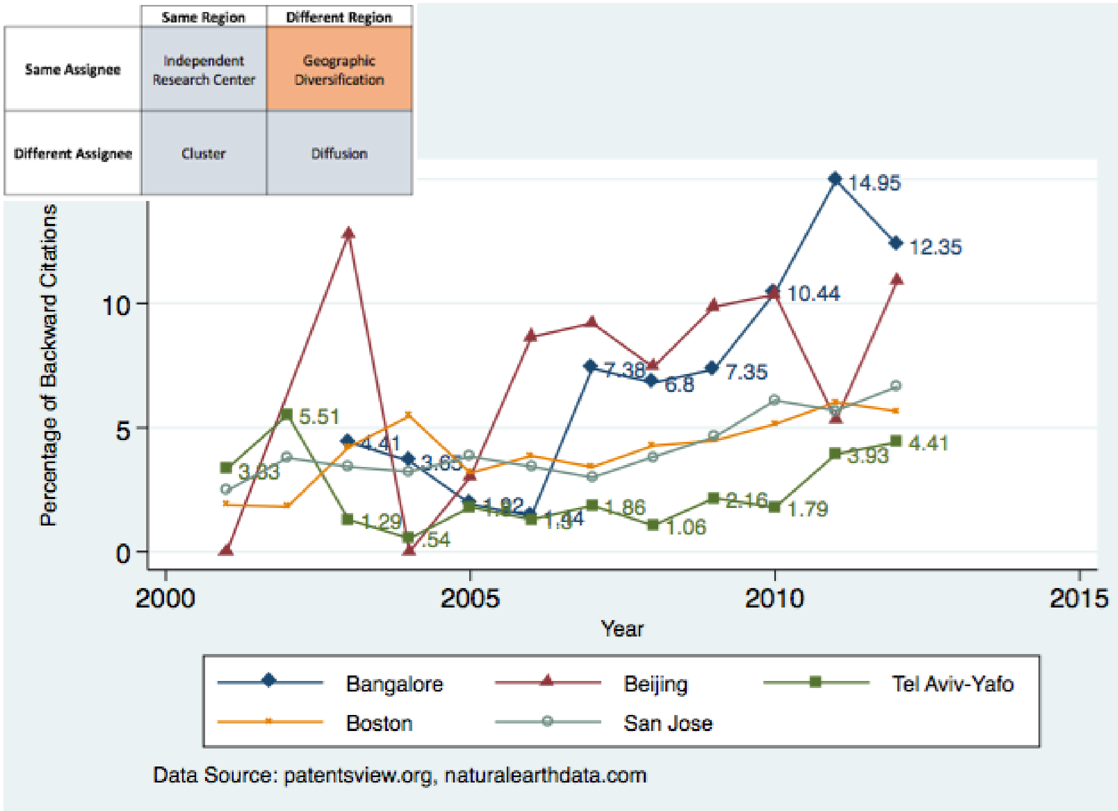

FIGURE 5

Knowledge Flows Across Regions and Across Assignees

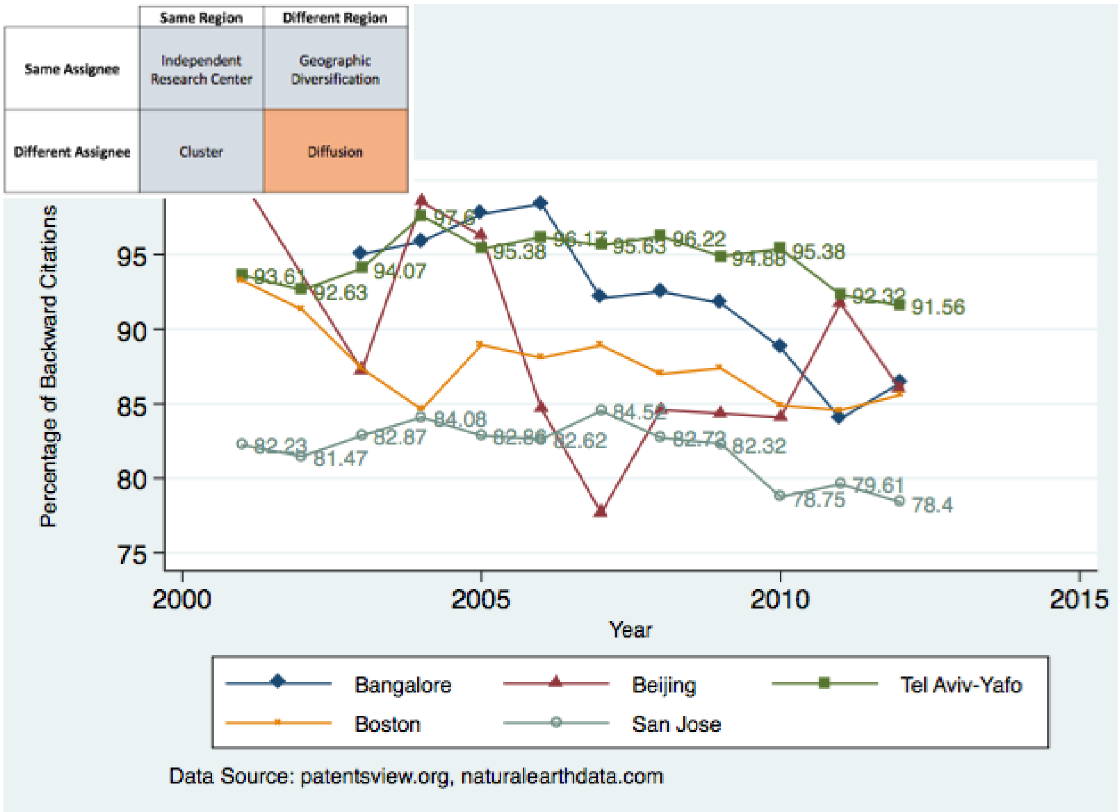

FIGURE 6

Two Examples of Variable Construction From the Data

Citations Made			Unit of Analysis: Region-Year Example: Bangalore-2010	Citations Received
	Same Region	Different Region	**Total Number of Patents:** 1114 **Total Number of Citations Made** = 25750 (=191 + 51 + 2708 + 22800) **Total Number of Citations Received** = 4415	
Same Assignee	Number of Citations Made = 191	Number of Citations Made = 2708		Number of Self Citations Received = 1407
Different Assignee	Number of Citations Made = 51	Number of Citations Made = 22800		Number of Non-Self Citations Received = 3008

Citations Made			Unit of Analysis: Region-Year Example: San Jose-2010	Citations Received
	Same Region	Different Region	**Total Number of Patents:** 14261 **Total Number of Citations Made** = 625664 (=23908 + 72505 + 38735 + 490516) **Total Number of Citations Received** = 114143 (=48267 + 65876)	
Same Assignee	Number of Citations Made = 23908	Number of Citations Made = 38735		Number of Self Citations Received = 48267
Different Assignee	Number of Citations Made = 72505	Number of Citations Made = 490516		Number of Non-Self Citations Received = 65876

TABLE 1

Effects of Knowledge Flows on Quality of Inventions

	Positive Effects	**Negative Effects**
Same Region, Same Assignee	Competence building	Local search on both dimensions (within firm and within region)
Same Region, Different Assignee	Knowledge spillovers	Local search on one dimension (within region)
Different Region, Same Assignee	Benefits of geographic diversification	Local search on one dimension (within firm)
Different Region, Different Assignee	High exploration, global pipeline	Excessive exploration? (Outside firm and outside region)

TABLE 2

Steps Followed in Constructing Region-Year Level Dataset

Step	Properties of Citing Patent						Flow	Properties of Cited Patent					
1						Patent-ID	←	Patent-ID					
2					Assignee	Patent-ID	←	Patent-ID	Assignee				
	In the above, when more than one assignee is defined for any patent, all assignees are retained for both citing patent and cited patent												
3				Inventor	Assignee	Patent-ID	←	Patent-ID	Assignee	Inventor			
	In the above step, all inventors are retained for both citing patent and cited patent												
4			Inventor Region	Inventor	Assignee	Patent-ID	←	Patent-ID	Assignee	Inventor	Inventor Region		
5	**Drop Citing Inventor and Cited Inventor**												
6			Inventor Region		Assignee	Patent-ID	←	Patent-ID	Assignee		Inventor Region		
7	**Drop duplicate observations of <Citing Inventor Region, Cited Inventor Region> for the same Patent Citation**												
8			Inventor Region		Assignee	Patent-ID	←	Patent-ID	Assignee		Inventor Region		
9		Application Year	Inventor Region		Assignee	Patent-ID	←	Patent-ID	Assignee		Inventor Region		
10	**Drop Citing Patent-ID and Cited Patent-ID**												
11		Application Year	Inventor Region		Assignee		←		Assignee		Inventor Region		
12	**Agglomerate at <Citing Application Year, Citing Inventor Region> to compute Independent Variables**												
13	**Agglomerate at <Citing Application Year, Cited Inventor Region> to compute Dependent Variables**												

TABLE 3

Correlations and Summary Statistics for the Sample of Applicant Made Citations

Variables	**Mean**	**Std. Dev.**	1	2	3	4	5	6	7	8	9	10
1. Total Citations Received	1762.906	8670.540	1.00									
2. Non-Self Citations Received	1423.084	7358.189	0.99	1.00								
3. Self Citations Received	339.822	1754.63	0.79	0.70	1.00							
4. Share Citations Made[Same Region, Same Assignee]	0.016	0.042	0.05	0.04	0.08	1.00						
5. Share Citations Made[Same Region, Different Assignee]	0.016	0.045	0.14	0.13	0.13	0.08	1.00					
6. Share Citations Made[Different Region, Same Assignee]	0.065	0.11	-0.02	-0.03	0.01	0.19	-0.04	1.00				
7. Share Citations Made[Different Region, Different Assignee]	0.903	0.132	-0.04	-0.04	-0.07	-0.51	-0.33	-0.88	1.00			
8. Log (Total Citations Made)	4.891	2.397	0.26	0.23	0.33	0.10	0.11	0.07	-0.13	1.00		
9. Log (Number of Patents)	3.807	1.928	0.40	0.38	0.38	0.13	0.15	0.02	-0.11	0.70	1.00	
10. Log (Patent Pool Size)	6.586	2.02	0.36	0.34	0.35	0.14	0.19	0.01	-0.12	0.69	0.94	1.00
N					8947							

Note: The above statistics were computed on the sample that was included in the regression with dependent variable as total citations received

TABLE 4

Correlations and Summary Statistics for the Sample of Examiner Made Citations

Variables	**Mean**	**Std. Dev.**	1	2	3	4	5	6	7	8	9	10
1. Total Citations Received	581.998	3264.788	1.00									
2. Non-Self Citations Received	490.539	2829.21	1.00	1.00								
3. Self Citations Received	91.459	478.033	0.92	0.90	1.00							
4. Share Citations Made[Same Region, Same Assignee]	0.023	0.051	0.06	0.06	0.08	1.00						
5. Share Citations Made[Same Region, Different Assignee]	0.013	0.038	0.16	0.15	0.16	0.06	1.00					
6. Share Citations Made[Different Region, Same Assignee]	0.069	0.112	-0.00	-0.00	0.01	0.25	-0.02	1.00				
7. Share Citations Made[Different Region, Different Assignee]	0.895	0.14	-0.06	-0.06	-0.08	-0.58	-0.28	-0.89	1.00			
8. Log (Total Citations Made)	4.352	2.162	0.39	0.38	0.43	0.11	0.13	0.03	-0.11	1.00		
9. Log (Number of Patents)	2.855	2.018	0.40	0.39	0.43	0.14	0.14	0.04	-0.12	0.96	1.00	
10. Log (Patent Pool Size)	5.495	2.249	0.35	0.34	0.38	0.16	0.17	0.04	-0.14	0.88	0.92	1.00
N					16464							

Note: The above statistics were computed on the sample that was included in the regression with dependent variable as total citations received

TABLE 5

Negative Binomial Regression Analysis of Invention Quality for the Sample of Applicant Made Citations

Dependent Variable	Total Citations Received			Non-Self Citations Received		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All Locations	U.S. Locations	Non-U.S. Locations	All Locations	U.S. Locations	Non-U.S. Locations
Share Citations Made[Same Region, Different Assignee]	0.150	0.113	0.148	0.261	0.238	0.274
	(0.405)	(0.713)	(0.506)	(0.114)	(0.413)	(0.206)
Share Citations Made[Different Region, Same Assignee]	0.365	0.568	0.284	0.175	0.312	0.143
	(0.020)	(0.018)	(0.168)	(0.248)	(0.174)	(0.493)
Share Citations Made[Different Region, Different Assignee]	0.178	0.316	0.0835	0.0835	0.183	0.0659
	(0.170)	(0.102)	(0.631)	(0.497)	(0.317)	(0.704)
Log (Total Citations Made)	0.0198	0.0132	0.0252	0.0144	0.00658	0.0184
	(0.000)	(0.025)	(0.000)	(0.000)	(0.233)	(0.002)
Log (Number of Patents)	0.783	0.848	0.830	0.780	0.824	0.843
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (Patent Pool Size)	-0.117	-0.285	-0.113	-0.0903	-0.213	-0.116
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-1.192	0.102	-1.553	-1.071	0.0237	-1.522
	(0.000)	(0.688)	(0.000)	(0.000)	(0.924)	(0.000)
Observations	8947	3798	5149	8860	3782	5078
Groups	1309	518	791	1281	512	769

Reference category is Share Citations Made[Same Region, Same Assignee]

p-values in parentheses

All models include region fixed effects, year dummies and technology subcategory controls

TABLE 6

Negative Binomial Regression Analysis of Invention Quality for the Sample of Examiner Made Citations

Dependent Variable	Total Citations Received			Non-Self Citations Received		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All Locations	U.S. Locations	Non-U.S. Locations	All Locations	U.S. Locations	Non-U.S. Locations
Share Citations Made[Same Region, Different Assignee]	-0.555	-0.987	-0.114	-0.364	-0.570	-0.0850
	(0.010)	(0.001)	(0.716)	(0.114)	(0.060)	(0.803)
Share Citations Made[Different Region, Same Assignee]	-0.466	-0.582	-0.290	-0.497	-0.389	-0.453
	(0.002)	(0.003)	(0.174)	(0.002)	(0.071)	(0.054)
Share Citations Made[Different Region, Different Assignee]	-0.944	-0.705	-0.953	-0.677	-0.274	-0.796
	(0.000)	(0.000)	(0.000)	(0.000)	(0.110)	(0.000)
Log (Total Citations Made)	0.255	0.200	0.299	0.248	0.179	0.300
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (Number of Patents)	0.586	0.680	0.589	0.572	0.685	0.567
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (Patent Pool Size)	-0.112	-0.274	-0.120	-0.0863	-0.252	-0.0974
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	1.200	2.443	0.821	0.913	2.067	0.607
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
Observations	16464	6246	10218	16410	6244	10166
Groups	1839	596	1243	1820	595	1225

Reference category is Share Citations Made[Same Region, Same Assignee]

p-values in parentheses

All models include region fixed effects, year dummies and technology subcategory controls

TABLE A1

Correlations and Summary Statistics for the Sample of non-Applicant, non-Examiner Citations Made (Other Citations)

Variables	**Mean**	**Std. Dev.**	1	2	3	4	5	6	7	8	9	10
1. Total Citations Received	571.832	4349.983	1.00									
2. Non-Self Citations Received	448.739	3606.807	1.00	1.00								
3. Self Citations Received	123.093	828.482	0.92	0.87	1.00							
4. Share Citations Made[Same Region, Same Assignee]	0.02	0.05	0.05	0.04	0.08	1.00						
5. Share Citations Made[Same Region, Different Assignee]	0.016	0.043	0.10	0.10	0.12	0.07	1.00					
6. Share Citations Made[Different Region, Same Assignee]	0.078	0.126	-0.01	-0.01	-0.00	0.16	-0.05	1.00				
7. Share Citations Made[Different Region, Different Assignee]	0.887	0.148	-0.04	-0.03	-0.06	-0.49	-0.27	-0.89	1.00			
8. Log (Total Citations Made)	5.027	2.392	0.27	0.26	0.32	0.10	0.13	0.02	-0.09	1.00		
9. Log (Number of Patents)	3.297	1.964	0.27	0.26	0.31	0.17	0.14	0.04	-0.13	0.84	1.00	
10. Log (Patent Pool Size)	5.983	2.121	0.24	0.22	0.27	0.18	0.18	0.03	-0.14	0.79	0.93	1.00
N					13047							

Note: The above statistics were computed on the sample that was included in the regression with dependent variable as total citations received

TABLE A2

Negative Binomial Regression Analysis of Invention Quality for the Sample of non-Applicant, non-Examiner Citations Made (Other Citations)

Dependent Variable	Total Citations Received			Non-Self Citations Received		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All Locations	U.S. Locations	Non-U.S. Locations	All Locations	U.S. Locations	Non-U.S. Locations
Share Citations Made[Same Region, Different Assignee]	-1.176	-0.944	-1.351	-0.678	-0.674	-0.588
	(0.001)	(0.050)	(0.013)	(0.077)	(0.197)	(0.301)
Share Citations Made[Different Region, Same Assignee]	-0.238	-0.433	-0.106	0.0426	-0.0363	0.131
	(0.290)	(0.180)	(0.741)	(0.868)	(0.921)	(0.716)
Share Citations Made[Different Region, Different Assignee]	-0.607	-0.415	-0.720	0.0246	0.250	-0.0737
	(0.002)	(0.114)	(0.012)	(0.911)	(0.411)	(0.817)
Log (Total Citations Made)	0.228	0.186	0.212	0.164	0.113	0.156
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (Number of Patents)	0.554	0.586	0.646	0.610	0.633	0.726
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (Patent Pool Size)	-0.0522	-0.101	-0.102	-0.00475	-0.0298	-0.0801
	(0.002)	(0.001)	(0.000)	(0.792)	(0.339)	(0.001)
Constant	0.254	0.452	0.342	-0.569	-0.396	-0.567
	(0.218)	(0.119)	(0.259)	(0.015)	(0.232)	(0.092)
Observations	13047	5618	7429	12924	5586	7338
Groups	1516	573	943	1484	566	918

Reference category is Share Citations Made[Same Region, Same Assignee]

p-values in parentheses

All models include region fixed effects, year dummies and technology subcategory controls

TABLE A3

Correlations and Summary Statistics for the Sample of All Citations Made (including Applicant, Examiner, and Others)

Variables	**Mean**	**Std. Dev.**	1	2	3	4	5	6	7	8	9	10
1. Total Citations Received	1993.201	12786.711	1.00									
2. Non-Self Citations Received	1626.983	10883.951	1.00	1.00								
3. Self Citations Received	366.218	2214.878	0.88	0.83	1.00							
4. Share Citations Made[Same Region, Same Assignee]	0.017	0.038	0.08	0.07	0.10	1.00						
5. Share Citations Made[Same Region, Different Assignee]	0.013	0.034	0.16	0.15	0.17	0.11	1.00					
6. Share Citations Made[Different Region, Same Assignee]	0.068	0.101	-0.00	-0.01	0.01	0.18	-0.04	1.00				
7. Share Citations Made[Different Region, Different Assignee]	0.901	0.119	-0.07	-0.06	-0.09	-0.50	-0.28	-0.89	1.00			
8. Log (Total Citations Made)	5.331	2.463	0.33	0.31	0.37	0.15	0.17	0.07	-0.16	1.00		
9. Log (Number of Patents)	2.753	2.04	0.35	0.34	0.38	0.18	0.19	0.05	-0.16	0.94	1.00	
10. Log (Patent Pool Size)	5.381	2.286	0.31	0.30	0.34	0.20	0.23	0.04	-0.16	0.88	0.92	1.00
N					17166							

Note: The above statistics were computed on the sample that was included in the regression with dependent variable as total citations received

TABLE A4

Negative Binomial Regression Analysis of Invention Quality for the Sample of All Citations Made (including Applicant, Examiner, and Others)

Dependent Variable	Total Citations Received			Non-Self Citations Received		
Model	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All Locations	U.S. Locations	Non-U.S. Locations	All Locations	U.S. Locations	Non-U.S. Locations
Share Citations Made[Same Region, Different Assignee]	-0.516	0.0657	-1.308	-0.604	0.189	-1.419
	(0.054)	(0.859)	(0.001)	(0.028)	(0.617)	(0.000)
Share Citations Made[Different Region, Same Assignee]	-0.220	-0.0596	-0.563	-0.559	-0.166	-0.920
	(0.273)	(0.837)	(0.036)	(0.008)	(0.584)	(0.001)
Share Citations Made[Different Region, Different Assignee]	-0.588	-0.149	-1.065	-0.579	0.0762	-1.049
	(0.001)	(0.545)	(0.000)	(0.001)	(0.766)	(0.000)
Log (Total Citations Made)	0.168	0.162	0.174	0.120	0.105	0.138
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (Number of Patents)	0.615	0.673	0.668	0.650	0.714	0.696
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (Patent Pool Size)	-0.104	-0.287	-0.113	-0.0703	-0.219	-0.101
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.470	1.314	0.568	0.501	1.002	0.588
	(0.009)	(0.000)	(0.019)	(0.007)	(0.000)	(0.023)
Observations	17166	6373	10793	17120	6371	10749
Groups	1929	601	1328	1914	600	1314

Reference category is Share Citations Made[Same Region, Same Assignee]

p-values in parentheses

All models include region fixed effects, year dummies and technology subcategory controls