

Heterogeneity in Knowledge Flows of Regions: Impact on Invention Quality

ABSTRACT

Using patent citations as measures of knowledge flows, we explore if 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. As some other recent research in the area (?), our preliminary results find inconclusive evidence for a clear positive effect of local knowledge flows on invention output of regions, or the globalization of R&D on the invention output of regions. 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.

Heterogeneity in Knowledge Flows of Regions: Impact on Invention Quality

Empirical studies in the literature on economic geography have used patent citations to suggest that knowledge spillovers are geographically localized (Almeida & Kogut, 1999; Bottazzi & Peri, 2003; Branstetter, 2001; Jaffe, Trajtenberg, & Henderson, 1993; Maurseth & Verspagen, 2002; Sonn & Storper, 2008), and that innovation is more spatially concentrated than production (Feldman, 1994). On the other hand, 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 net effect of geography on innovation outcomes of firms and regions. 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, we¹ 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 innovation quality of patents originating in a region. In this study we use citation data of patents applied for between 2001 and 2012 to empirically estimate the relationship between knowledge flows of a region and 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 of regions. Our work therefore seeks to contribute to the innovation literatures in both economic geography and international business.

It would seem appropriate to raise two measurement related issues here soon after we propose our research question. 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

¹This paper is the outcome of the joint work by my co-author and me. While the changes since the version submitted to a conference in February 2017 (sms2017.pdf, provided as supplementary material) are without consultation with my co-author, I have used the plural “we” throughout this paper to preserve continuity with the previous version and compatibility with future versions that are going to be jointly authored.

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.

Correction The next section reviews the theory from the geography of innovation and innovation in international business. We then define a framework to classify knowledge flows in a region and motivate our work by demonstrating how different regions fare differently in terms of knowledge flows. We build on the theory and this framework to propose our hypotheses. We then propose our research design, and follow it up with a description of our data and methods. 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.

THEORY AND HYPOTHESES

Griliches (1979) drew our attention toward two types of knowledge spillovers - “rent spillovers” that are rival and excludable knowledge flows that occur from the exchange of goods, and “pure or idea-creating spillovers” that are non-rival and non-excludable knowledge that is typically associated with R&D activity. While rent spillovers may be characterized as a private good, pure spillovers exhibit characteristics of a public good (Arrow, 1962). This dual nature of knowledge is additionally complicated by the fact that knowledge flows may be largely invisible (Krugman,

1991), and only sometimes do knowledge flows leave a paper trail in the form of patent citations (Jaffe et al., 1993). In the following sections, we review the literature in economic geography and international business **Correction** cognizant of these fundamental limitations relating to the nature and measurement of knowledge.

Economic geography of innovation

Correction - What is the purpose of this section Scholars of economics and strategy have for long recognized that clusters and agglomeration economies 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). Several studies have used patent citations to demonstrate that knowledge flows are localized (Almeida & Kogut, 1999; Jaffe, Trajtenberg, & Henderson, 1993). After controlling for the geographic distribution of production, Audretsch & Feldman (1996b) demonstrated that innovation exhibits a pronounced tendency to cluster spatially. Acs, Audretsch, & Feldman (1994) found that new product introductions were more geographically concentrated than patents and that they derived important inputs from universities and industrial R&D.

Regions, however vary in their inventive output (Agrawal, Cockburn, Galasso, & Oettl, 2014) and the nature of knowledge flows in a region may be only one source of this variation within regions. Location however has been suggested to matter more at the earliest stage of the industry life cycle, with life-cycle production costs featuring more prominently in mature goods and industries. Audretsch & Feldman (1996a) suggest that early stages of industry life cycles are characterized by the importance of tacit knowledge, and may hence exhibit higher a propensity to cluster to benefit from knowledge spillovers.

Multiple reasons have been attributed for the localized nature of knowledge spillovers. First, costs of collaboration are lower due to geographic proximity. Second, proximity has been argued to promote serendipitous and chance encounters that create the environment for new solutions. Third, diversity within a region has also been suggested to benefit innovative output (Feldman,

1999). Scholars have suggested that related variety (Boschma & Iammarino, 2009; Frenken, Oort, & Verburg, 2007) may benefit in the generation of new ideas and that diversity between industrial activities may benefit the transfer of ideas (Jacobs, 1969). Additionally, scholars have suggested that the tacit nature of knowledge, institutions and regional innovation systems may all contribute to the localization of knowledge flows (Cooke, 1996; Howells, 1996, 2002; Maskell & Malmberg, 1999).

Correction - Summarize, the reader is left hanging

Globalization of innovation

Regional innovation systems have also been suggested as influencing MNC decisions to locate subsidiaries (Andersen & Christensen, 2005). However, scholars have also considered organizational capabilities (Zhao, 2006), property rights, technological sourcing (Florida, 1997), higher product ownership and interdependence (Pearce, 1999), and access to highly skilled human capital. On the other hand, political borders have been found to constrain the flows of knowledge (Singh & Marx, 2013), but mobility of inventors helps firms improve innovation outcomes (Alnuaimi, Opsahl, & George, 2012).

Correction - Literature review is not sufficient. What is the point that you are making?

Inter-firm knowledge flows and geography

The review of the literature in the economic geography and international business traditions suggests both positive and negative effects of localization of knowledge spillovers on innovation outcomes. On the one hand agglomerations may seem to benefit firms through access to tacit knowledge, while on the other hand it may also increase the number of competitors who are specialized in the same market segment (?). MNC subsidiary location choices seem to be affected by local factors, as well as relative ties with local firms and MNC affiliates(?). **Correction - Write more.** In the following section, we present a simple framework with which to address this phenomenon.

Effects of knowledge flows on quality of inventions

We formalize the discussions in the previous sections by categorizing all knowledge flows incorporated into an invention 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 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. This also 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.

Invention quality at the independent research center Firms operating in silos, i.e., where citations are made to other prior work within the same firm in the same region may benefit from specialization. On the other hand, the literature from economic geography suggests that related variety is associated with superior innovations (Boschma & Iammarino, 2009; Frenken et al., 2007). It would seem, from prior literature that the cross-pollinating effect of related variety would have the larger effect on innovation outcomes(?). **Correction - More arguments need to be provided** This leads to our first hypothesis.

Hypothesis 1: An increase in the proportion of same firm, same region knowledge flows will be associated with a lower level of invention quality.

Invention quality in a cluster The literature on clusters suggests benefits through two different mechanisms. Marshallian externalities are concerned with intraindustry economies of localization, while Jacobs' externalities refer to interindustry exchanges between different technologies. While some studies have suggested that Marshall-Arrow-Romer (MAR) externalities are considered more prevalent (Baptista & Swann, 1998), others have suggested that Jacobs' externalities dictate local knowledge spillover processes (Feldman, 1999). Still other studies have found both types of externalities to be significant (Capello, 2002). Ibrahim, Fallah, & Reilly (2009) suggest that within metropolitan regions, the two may not be mutually exclusive. The theory therefore suggests a higher potential to generate radical, disruptive innovative output when diverse sectoral knowledge bases are combined. On the other hand, theory would suggest incremental innovation in building innovations on top of specialized knowledge (Schumpeter, 1942). Since a cluster itself does not lend itself to a clear definition of one type or the other, we suggest two contradictory hypotheses.

Hypothesis 2a: A higher proportion of different firm, same region knowledge flows to a region will be associated with a higher level of invention quality of that region

Hypothesis 2b: A higher proportion of different firm, same region knowledge flows to a region will not be associated with a change in invention quality of that region

Invention quality under geographical diversification The international business literature has suggested several different effects in MNC's structuring of R&D across units. While agglomeration benefits have been suggested to be beneficial (Cantwell & Iammarino, 2000; Nachum, 2000; Pantzalis, 2001), scholars have also suggested potential threats from highly specialized firms in similar market niches. While the net effect on innovation under geographical diversification is not clear, innovation performance may itself only benefit. This leads to our third hypothesis.

Hypothesis 3: A higher proportion of same firm, different region knowledge flows to a region will be associated with a higher level of invention quality of that region

Invention quality under diffusion Breschi & Lissoni (2005, 2009) argue that many localized knowledge spillover studies use spatial proximity as a proxy for social proximity, and that higher innovation outcomes are determined by the access to knowledge provided by social proximity. The effect of spatial proximity itself would be greatly reduced when social proximity is accounted for. The absence of social proximity under conditions of diffusion would therefore diminish the opportunities for positive knowledge externalities. **Correction - How do you determine between no effect and negative effect.** This leads us to our final hypothesis.

Hypothesis 4: A higher proportion of different firm, different region knowledge flows to a region will not be associated with a change in invention quality of that region

DATA AND METHODS

We use patent citations data from the U.S. Patent Office (USPTO) as provided by patentsview.org. 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 and inventor location of each patent with each assignee 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 Schnei-

der, 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.

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 inferences about inventor knowledge using pooled citations may suffer from bias or overinflated significance levels. On the other hand 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 strong assumptions, we do not restrict ourselves to those citations categorized as 'cited by applicant' or those categorized as 'cited by examiner'. We use three data sets - a) the pooled data set of all applicant and examiner citations (this is our primary data set), b) the data set of only examiner citations, and c) the data set of only applicant citations. This decision has the additional effect of limiting our period of analysis to citing patents applied for after the year 2000 since the data on which citations were added by examiners is available for patents from only 2001 onwards.

Citations Received

Correction - Revert this to counting all citations to date, and justify it We restrict our sample to patents applied for between 2001 - 2012, but citations received till 2015. Our default strategy for measuring patent citations is to count citations received until and including the fifth annual

anniversary of the patent application. This has the effect of biasing against patent applications from 2011 and 2012. This problem may be resolved by updating the dataset to the version updated to March 2017. For paucity of time, the dataset was not updated in time for the submission of this paper.

Dependent Variable

Correction - Revert this to counting all citations to date, get rid of fifth annual anniversary

Our primary dependent variable is the total count of citations 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. Citations are counted as those received up to and including the fifth annual anniversary of the patent filing or until 2015, whichever is earlier. In the section on additional results, we also provide results with an alternative measure of citations received with the dataset on applicant only citations. Here we count all citations received till 2015 rather than truncate after the fifth annual anniversary of patent application filing date.

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, same assignee.

Percentage of backward citations made to same region and same assignee We compute this variable as follows. For each region, 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.

Percentage of backward citations made to same region and different assignee We compute this variable as follows. For each region, 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.

Percentage of backward citations made to different region and same assignee We compute this variable as follows. For each region, 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.

Percentage of backward citations made to different region and different assignee We compute this variable as follows. For each region, 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). **Correction - Percentage and fraction, both terminologies are used.**

Stick to one

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 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.

Table ?? displays the correlation of our primary variables in the dataset of applicant and examiner citations. Summary statistics for the variables in the dataset of applicant and examiner citations is provided in Table ?. Similar statistics for the applicant only citations dataset and the examiner only citations dataset have been provided in the section on additional results.

RESULTS

The preliminary results from our analysis for the dataset of applicant and examiner cited patents are presented in Table ?. 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 initial observation is that results seem to vary depending on the chosen dependent variable, though the signs of the coefficients are the same. Across the models we find that knowledge flows under cluster and geographical diversification have a significant impact on the quality of inventions as measured by the number of citations received. However, the differences in the results between the two dependent variables suggests that there is an underlying effect that the current model is not capturing.

Let us set aside the inconsistency in results between the two measures of the dependent variable, and just focus on the results for total citations received. H1 suggested a negative effect for within firm, within region flows on invention quality. The results however suggest a positive effect and significant effect in models 1 and 3. H1 is therefore not supported. H2a and H2b suggested that potentially opposite claims for higher cluster knowledge flows. The results are mixed across the samples, suggesting perhaps that different effects are at play under different conditions. H3 predicted a positive effect for the globalized diversification driven knowledge flows. The results here are consistent with our hypotheses, across all models and across datasets. Finally, we had predicted a non-effect for the diffusion knowledge flow on invention performance. Again the results for the total citations received dependent variable hold.

Additional Results

Table ?? displays the correlation of our primary variables in the dataset of examiner only citations. Summary statistics for the variables in the dataset of examiner only citations is provided in Table ??.

Insert TABLE 1 about here.

Insert TABLE 2 about here.

Insert TABLE 3 about here.

Insert TABLE 4 about here.

The additional results from our analysis for the dataset of examiner only 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. 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 ?? on the dataset with applicant and examiner citations.

Insert TABLE 5 about here.

Table 6 displays the correlation of our primary variables in the dataset of applicant only citations. Summary statistics for the variables in the dataset of applicant only citations is provided in Table 7.

Insert TABLE 6 about here.

Insert TABLE 7 about here.

Insert TABLE 8 about here.

Insert TABLE 9 about here.

The additional results from our analysis for the dataset of applicant only cited patents are presented in Table 10. 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 ?? on the dataset with applicant and examiner citations.

Insert TABLE 10 about here.

The additional results from our analysis for the dataset of applicant only cited patents with citations received counted till 2015 are presented in Table 11. 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. Interestingly here, we note that results between the two dependent variable measures are not particularly different. However, these set of measures seem

to suggest different effects to that observed in Table 10 on the dataset with applicant only citations but where citations were counted until the fifth annual anniversary of the patent filing date. This suggests that we may need to investigate the time effect on citations more closely.

Insert TABLE 11 about here.

DISCUSSION

Knowledge flows are hard to measure, and knowledge is characterized by by opposing qualities of being tacit and coded, and of behaving like a public or a private good. The central challenge therefore seems to be to find a way to disentangle the opposing natures of knowledge. It is possibly this that is behind 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 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.

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.

CONCLUSION

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. A potential extension of the study could be to conduct empirical analysis at the level of firm-year rather than at the region-year. We could additionally look at the additional dimension of technology (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 streams economic geography, and international business, but there is clearly a lot 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.
- Baptista, R., & Swann, P. 1998. Do firms in clusters innovate more? *Research Policy*, 27(5): 525 – 540.
- 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.
- Bottazzi, L., & Peri, G. 2003. Innovation and spillovers in regions: Evidence from European patent data. *European Economic Review*, 47(4): 687 – 710.

- Branstetter, L. G. 2001. Are knowledge spillovers international or intranational in scope?: Microeconomic evidence from the U.S. and Japan. *Journal of International Economics*, 53(1): 53 – 79.
- Breschi, S., & Lissoni, F. 2005. “Cross-Firm” inventors and social networks: Localized knowledge spillovers revisited. *Annales d'Économie et de Statistique*, 2005(79/80): 189–209.
- Breschi, S., & Lissoni, F. 2009. Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, 9(4): 439.
- Bunnell, T. G., & Coe, N. M. 2001. Spaces and scales of innovation. *Progress in Human Geography*, 25(4): 569–589.
- Cantwell, J., & Iammarino, S. 2000. Multinational corporations and the location of technological innovation in the uk regions. *Regional Studies*, 34(4): 317–332.
- Capello, R. 2002. Entrepreneurship and spatial externalities: Theory and measurement. *The Annals of Regional Science*, 36(3): 387–402.
- 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: A review 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.

- Ibrahim, S. E., Fallah, M. H., & Reilly, R. R. 2009. Localized sources of knowledge and the effect of knowledge spillovers: An empirical study of inventors in the telecommunications industry. *Journal of Economic Geography*, 9(3): 405.
- 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.
- Nachum, L. 2000. Economic geography and the location of TNCs: Financial and professional service FDI to the USA. *Journal of International Business Studies*, 31(3): 367–385.
- Pantzalis, C. 2001. Does location matter? an empirical analysis of geographic scope and MNC market valuation. *Journal of International Business Studies*, 32(1): 133–155.
- 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 competitive 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.
- Schumpeter, J. A. 1942. *Capitalism, socialism and democracy*. New York, NY: Harper.
- 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.

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

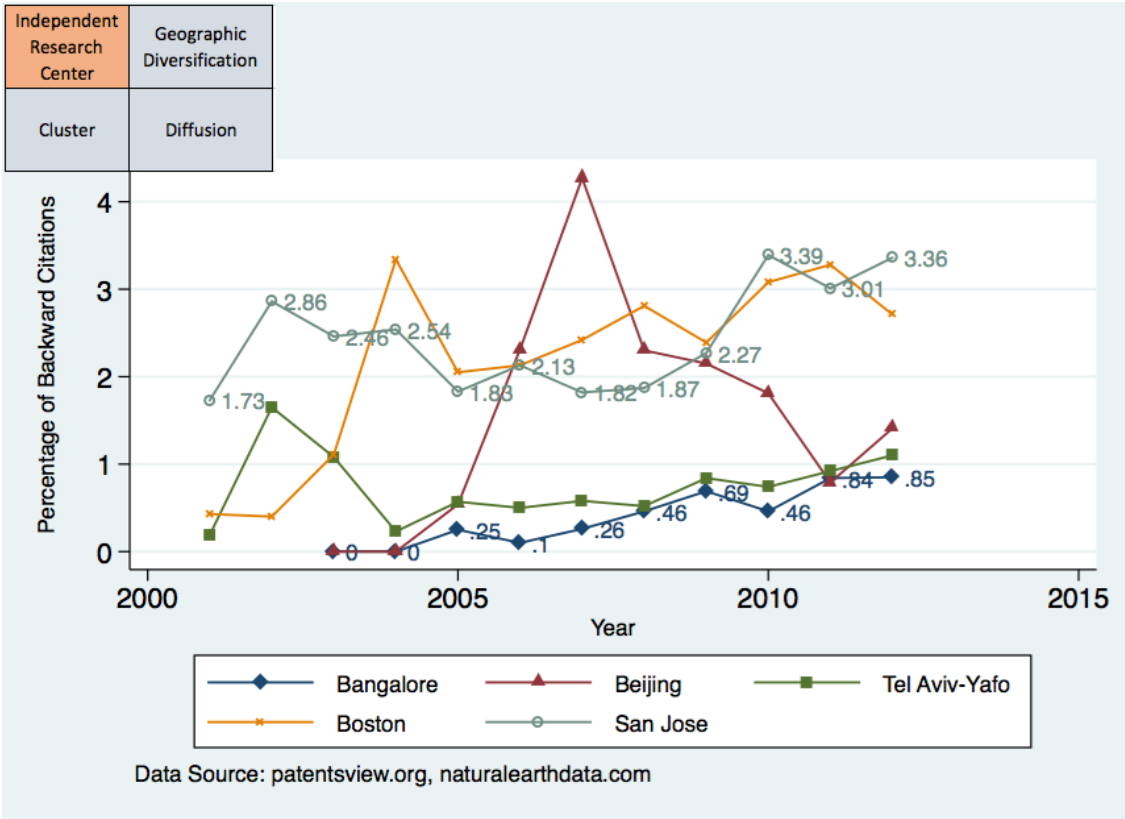


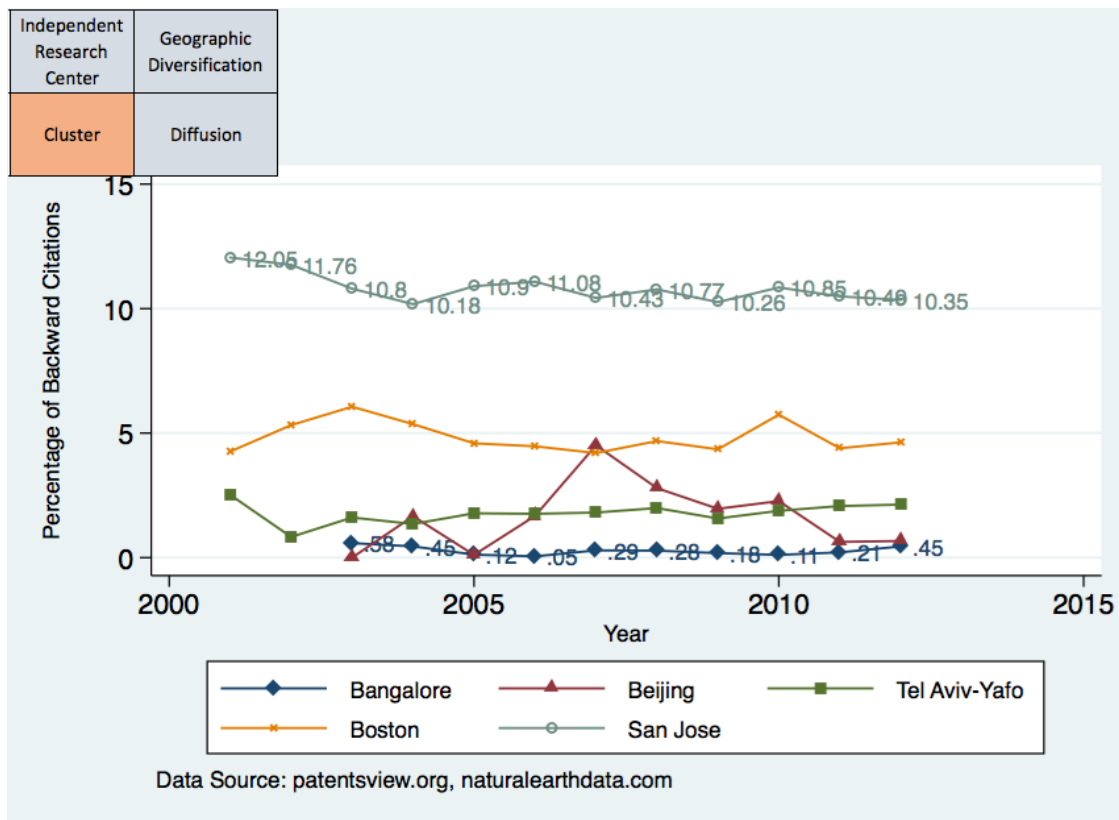
FIGURE 3: Knowledge flows within regions and across assignees

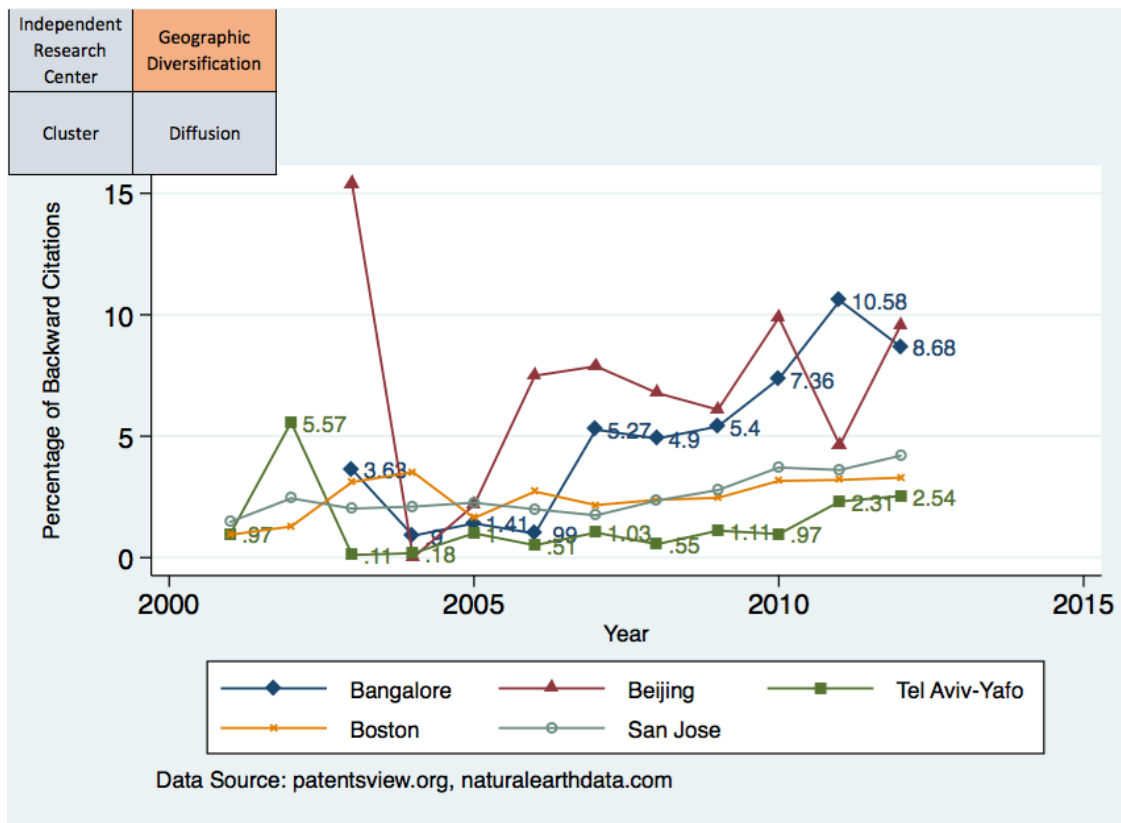
FIGURE 4: Knowledge flows across regions and within assignees

FIGURE 5: Knowledge flows across regions and across assignees

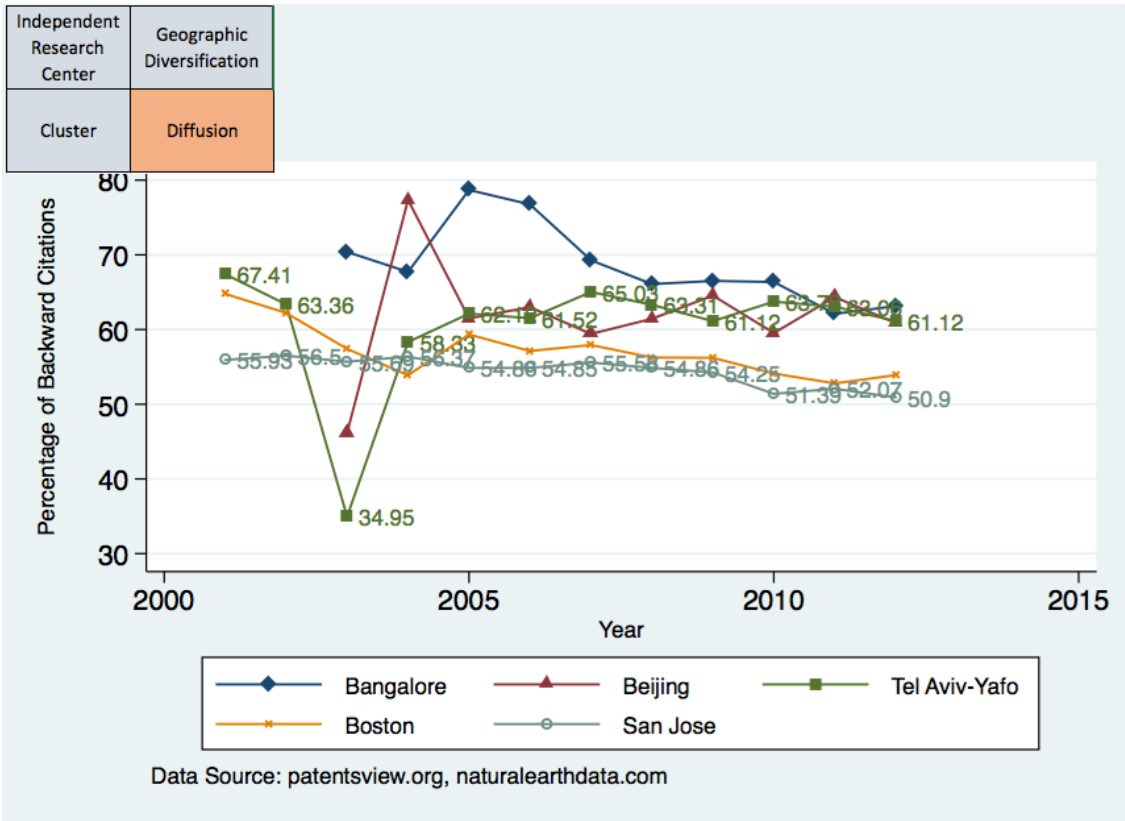


TABLE 1: Correlation table for examiner only data set with DV as Total Citations Received (distance calculated)

Variables	Citations Received	Non-Self Citations Received
Citations Received	1.00	
Non-Self Citations Received	1.00	1.00
Self Citations Received	0.92	0.90
Share Citations Made[Same Region, Same Assignee]	0.06	0.06
Share Citations Made[Same Region, Different Assignee]	0.16	0.16
Share Citations Made[Different Region, Same Assignee]	-0.00	-0.01
Share Citations Made[Different Region, Different Assignee]	-0.07	-0.06
Share Citations Made[Other]	.	.
Share Citations Made[Same Region]	0.14	0.14
Share Citations Made[Same Assignee]	0.02	0.02
Log (Total Citations Made)	0.40	0.39
Log (Num Patents)	0.40	0.39
Log (Patent Pool Size)	0.35	0.34

TABLE 2: Summary statistics for examiner only data set with DV as Total Citations Received (distance calculated)

Variable	Mean	Std. Dev.	N
Citations Received	612.014	3348.657	15624
Non-Self Citations Received	515.836	2902.025	15624
Self Citations Received	96.177	490.241	15624
Share Citations Made[Same Region, Same Assignee]	0.024	0.051	15624
Share Citations Made[Same Region, Different Assignee]	0.013	0.038	15624
Share Citations Made[Different Region, Same Assignee]	0.069	0.11	15624
Share Citations Made[Different Region, Different Assignee]	0.893	0.138	15624
Share Citations Made[Other]	0	0	15624
Share Citations Made[Same Region]	0.037	0.065	15624
Share Citations Made[Same Assignee]	0.093	0.132	15624
Log (Total Citations Made)	4.448	2.158	15624
Log (Num Patents)	2.949	2.006	15624
Log (Patent Pool Size)	5.643	2.216	15624

TABLE 3: Correlation table for examiner only data set with DV as Non-Self Citations Received (distance calculated)

Variables	Citations Received	Non-Self Citations Received
Citations Received	1.00	
Non-Self Citations Received	1.00	1.00
Self Citations Received	0.92	0.90
Share Citations Made[Same Region, Same Assignee]	0.06	0.06
Share Citations Made[Same Region, Different Assignee]	0.16	0.16
Share Citations Made[Different Region, Same Assignee]	-0.00	-0.01
Share Citations Made[Different Region, Different Assignee]	-0.07	-0.06
Share Citations Made[Other]	.	.
Share Citations Made[Same Region]	0.15	0.14
Share Citations Made[Same Assignee]	0.02	0.02
Log (Total Citations Made)	0.40	0.39
Log (Num Patents)	0.40	0.39
Log (Patent Pool Size)	0.35	0.34

TABLE 4: Summary statistics for examiner only data set with DV as Non-Self Citations Received (distance calculated)

Variable	Mean	Std. Dev.	N
Citations Received	613.621	3352.913	15583
Non-Self Citations Received	517.194	2905.72	15583
Self Citations Received	96.427	490.861	15583
Share Citations Made[Same Region, Same Assignee]	0.024	0.05	15583
Share Citations Made[Same Region, Different Assignee]	0.013	0.037	15583
Share Citations Made[Different Region, Same Assignee]	0.069	0.109	15583
Share Citations Made[Different Region, Different Assignee]	0.894	0.137	15583
Share Citations Made[Other]	0	0	15583
Share Citations Made[Same Region]	0.037	0.064	15583
Share Citations Made[Same Assignee]	0.093	0.131	15583
Log (Total Citations Made)	4.455	2.157	15583
Log (Num Patents)	2.956	2.004	15583
Log (Patent Pool Size)	5.653	2.21	15583

TABLE 5: NB Regression Analysis of Invention Quality for Examiner Citations Only (Distance Calculated)

	(1) Citations Received	(2) Citations Received	Citati
main			
Share Citations Made[Same Region, Same Assignee]	0.906 (0.000)	0.551 (0.000)	
Share Citations Made[Same Region, Different Assignee]	0.454 (0.012)	-0.238 (0.311)	
Share Citations Made[Different Region, Same Assignee]	0.329 (0.000)	0.0909 (0.261)	
Log (Total Citations Made)	0.437 (0.000)	0.227 (0.000)	
Log (Num Patents)	0.348 (0.000)	0.621 (0.000)	
Log (Patent Pool Size)	-0.0537 (0.000)	-0.209 (0.000)	
Constant	0.355 (0.000)	1.806 (0.000)	
Observations	15624	6204	
Groups	1727	594	
Sample	All Locations (UC)	US Locations (UC)	Non-US

p-values in parentheses

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

TABLE 6: Correlation table for applicant only data set with DV as Total Citations Received (distance calculated)

Variables	Citations Received	Non-Self Citations Received
Citations Received	1.00	
Non-Self Citations Received	0.99	1.00
Self Citations Received	0.79	0.70
Share Citations Made[Same Region, Same Assignee]	0.05	0.04
Share Citations Made[Same Region, Different Assignee]	0.14	0.13
Share Citations Made[Different Region, Same Assignee]	-0.02	-0.03
Share Citations Made[Different Region, Different Assignee]	-0.04	-0.04
Share Citations Made[Other]	.	.
Share Citations Made[Same Region]	0.13	0.12
Share Citations Made[Same Assignee]	-0.00	-0.01
Log (Total Citations Made)	0.26	0.23
Log (Num Patents)	0.40	0.38
Log (Patent Pool Size)	0.36	0.34

TABLE 7: Summary statistics for applicant only data set with DV as Total Citations Received (distance calculated)

Variable	Mean	Std. Dev.	N
Citations Received	1762.906	8670.540	8947
Non-Self Citations Received	1423.084	7358.189	8947
Self Citations Received	339.822	1754.63	8947
Share Citations Made[Same Region, Same Assignee]	0.016	0.043	8947
Share Citations Made[Same Region, Different Assignee]	0.016	0.045	8947
Share Citations Made[Different Region, Same Assignee]	0.065	0.11	8947
Share Citations Made[Different Region, Different Assignee]	0.903	0.133	8947
Share Citations Made[Other]	0	0	8947
Share Citations Made[Same Region]	0.032	0.065	8947
Share Citations Made[Same Assignee]	0.081	0.125	8947
Log (Total Citations Made)	4.891	2.397	8947
Log (Num Patents)	3.807	1.928	8947
Log (Patent Pool Size)	6.586	2.02	8947

TABLE 8: Correlation table for applicant only data set with DV as Non-Self Citations Received (distance calculated)

Variables	Citations Received	Non-Self Citations Received
Citations Received	1.00	
Non-Self Citations Received	0.99	1.00
Self Citations Received	0.79	0.70
Share Citations Made[Same Region, Same Assignee]	0.05	0.04
Share Citations Made[Same Region, Different Assignee]	0.14	0.13
Share Citations Made[Different Region, Same Assignee]	-0.02	-0.03
Share Citations Made[Different Region, Different Assignee]	-0.05	-0.04
Share Citations Made[Other]	.	.
Share Citations Made[Same Region]	0.13	0.12
Share Citations Made[Same Assignee]	-0.00	-0.01
Log (Total Citations Made)	0.26	0.23
Log (Num Patents)	0.40	0.38
Log (Patent Pool Size)	0.36	0.35

TABLE 9: Summary statistics for applicant only data set with DV as Non-Self Citations Received (distance calculated)

Variable	Mean	Std. Dev.	N
Citations Received	1780.181	8711.249	8860
Non-Self Citations Received	1437.058	7392.873	8860
Self Citations Received	343.123	1762.906	8860
Share Citations Made[Same Region, Same Assignee]	0.016	0.042	8860
Share Citations Made[Same Region, Different Assignee]	0.017	0.044	8860
Share Citations Made[Different Region, Same Assignee]	0.065	0.108	8860
Share Citations Made[Different Region, Different Assignee]	0.903	0.131	8860
Share Citations Made[Other]	0	0	8860
Share Citations Made[Same Region]	0.032	0.064	8860
Share Citations Made[Same Assignee]	0.08	0.123	8860
Log (Total Citations Made)	4.914	2.394	8860
Log (Num Patents)	3.836	1.913	8860
Log (Patent Pool Size)	6.622	1.993	8860

TABLE 10: NB Regression Analysis of Invention Quality for Applicant Citations Only (Distance Calculated)

	(1) Citations Received	(2) Citations Received	Citati
main			
Share Citations Made[Same Region, Same Assignee]	-0.167 (0.194)	-0.311 (0.107)	
Share Citations Made[Same Region, Different Assignee]	-0.0321 (0.774)	-0.213 (0.306)	
Share Citations Made[Different Region, Same Assignee]	0.185 (0.002)	0.257 (0.005)	
Log (Total Citations Made)	0.0198 (0.000)	0.0138 (0.019)	
Log (Num Patents)	0.783 (0.000)	0.847 (0.000)	
Log (Patent Pool Size)	-0.117 (0.000)	-0.284 (0.000)	
Constant	-1.014 (0.000)	0.412 (0.015)	
Observations	8947	3794	
Groups	1309	517	
Sample	All Locations (UC)	US Locations (UC)	Non-US

p-values in parentheses

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

TABLE 11: Regression Analysis of Invention Quality* for Applicant Only Citations

	(1) Total Citations Received	(2) Total Citations Received	(3) Total Citations Received	(4) Non-Self Citations Received	(5) Non-Self Citations Received	(6) Non-Self Citations Received
Share Citations Made[Same Region, Same Assignee]	-0.169 (0.218)	-0.104 (0.623)	-0.134 (0.454)	-0.232 (0.105)	-0.183 (0.392)	-0.316 (0.123)
Share Citations Made[Same Region, Different Assignee]	-0.149 (0.215)	-0.191 (0.458)	-0.0919 (0.468)	0.0391 (0.733)	0.0945 (0.721)	0.0505 (0.699)
Share Citations Made[Different Region, Same Assignee]	0.217 (0.012)	0.266 (0.037)	0.348 (0.004)	0.195 (0.035)	0.263 (0.041)	0.174 (0.212)
Share Citations Made[Different Region, Different Assignee]	0.00828 (0.789)	0.0315 (0.499)	0.0179 (0.661)	0.00989 (0.762)	0.0203 (0.675)	0.00983 (0.826)
Log (Total Citations Made)	0.0180 (0.000)	0.0198 (0.001)	0.0134 (0.014)	0.0150 (0.000)	0.0139 (0.023)	0.0107 (0.075)
Log (Num Patents)	0.785 (0.000)	0.811 (0.000)	0.830 (0.000)	0.772 (0.000)	0.795 (0.000)	0.828 (0.000)
Log (Patent Pool Size)	-0.111 (0.075)	-0.227 (0.000)	-0.0944 (0.950)	-0.0453 (0.039)	-0.108 (0.006)	-0.0685 (0.017)
Observations	9241	3885	5356	8879	3732	5147
Groups	1314	529	785	1199	478	721
Sample	All Locations	U.S. Locations	Non-U.S. Locations	All Locations	U.S. Locations	Non-U.S. Locations

p-values in parentheses

All regressions use negative binomial estimation on the region-year panel

* Citations received are counted all the way to end 2016 for which data is available

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