

# **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 and demonstrate that patent citations provide 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. 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.

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

## **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 (????), and that innovation is more spatially concentrated than production (?). 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 (???). 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, 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 innovation quality of patents originating in a region. In this study 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 of regions. Our work therefore 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.

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. We then 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.

## THEORY

Knowledge spillovers have been suggested to demonstrate two conflicting characteristics (?). While on the one hand, “rent spillovers” exhibit characteristics of a private good (by virtue of being knowledge that is 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 (?). This dual nature of knowledge spillovers notwithstanding, knowledge flows are also very hard to measure. While ? suggested that knowledge flows may be largely invisible, ? famously 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 (??). In the following, we review the literature in economic geography and international business to conclude that the literature often suggests opposing effects on the nature of knowledge flows on invention quality.

Clusters and agglomeration economies have been suggested to play an important role in fostering innovation (??). Agglomeration economies may arise due to labor pooling advantages, economies of specialization of local suppliers, and knowledge spillovers (??). Several studies have used patent citations to demonstrate that knowledge flows are localized (??). ? demonstrated that innovation exhibits a pronounced tendency to cluster spatially even after controlling for the geographic distribution of production. ? suggested that new product introductions were more geographically concentrated than patents as they derived important inputs from universities and industrial R&D.

Regions, however, demonstrate heterogeneity in their inventive output (?) and the nature of knowledge flows in a region may be only one source of this variation within regions. Location

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. ? 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 may promote serendipitous and chance encounters that create an environment conducive for new ideas, newer solutions and to innovation. Third, diversity within a region has been suggested to benefit innovative output (?). Related variety (??) may help in the generation of new ideas and diversity between industrial activities may help in the transfer of ideas (?). Finally, transfer difficulties arising out of the tacit nature of knowledge, idiosyncratic institutions and regional innovation systems may also contribute to the localization of knowledge flows (????).

Indeed, the literature in the international business tradition has also suggested that regional innovation systems may influence MNC decisions to locate subsidiaries (?). However, scholars have also considered organizational capabilities (?), property rights, technological sourcing (?), higher product ownership and interdependence (?), and access to highly skilled human capital as other alternative factors that drive MNC location choices. Another strand of literature in this tradition has suggested that political borders may constrain the flows of knowledge (?), while mobility of inventors may help firms improve innovation outcomes (?).

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. Agglomerations seem to benefit firms through access to tacit knowledge, while it may also serve to 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 by relative ties with local firms and MNC affiliates. It may seem appropriate therefore to approach this study by pitting these alternative nature of knowledge flows in an empirical horserace. In the following section, we 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 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.

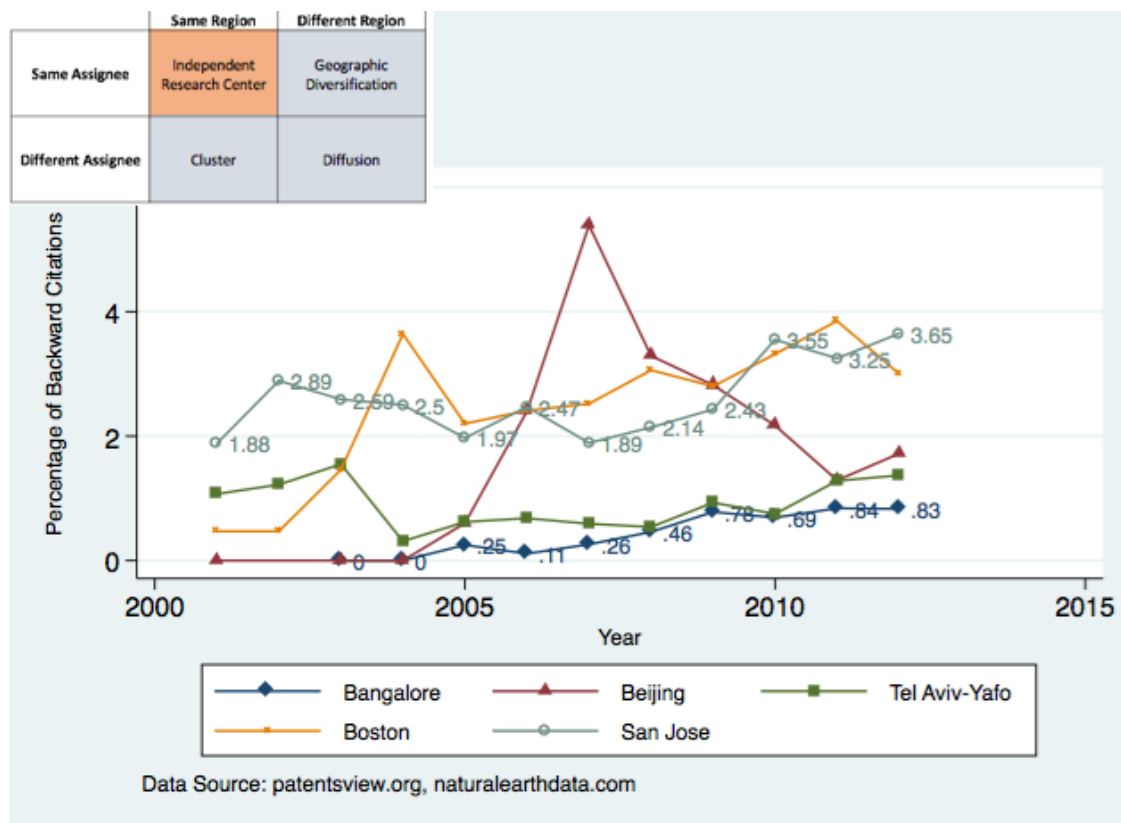
**FIGURE 1:** Categories of knowledge flows

	<b>Same Region</b>	<b>Different Region</b>
<b>Same Assignee</b>	Independent Research Center	Geographic Diversification
<b>Different Assignee</b>	Cluster	Diffusion

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.

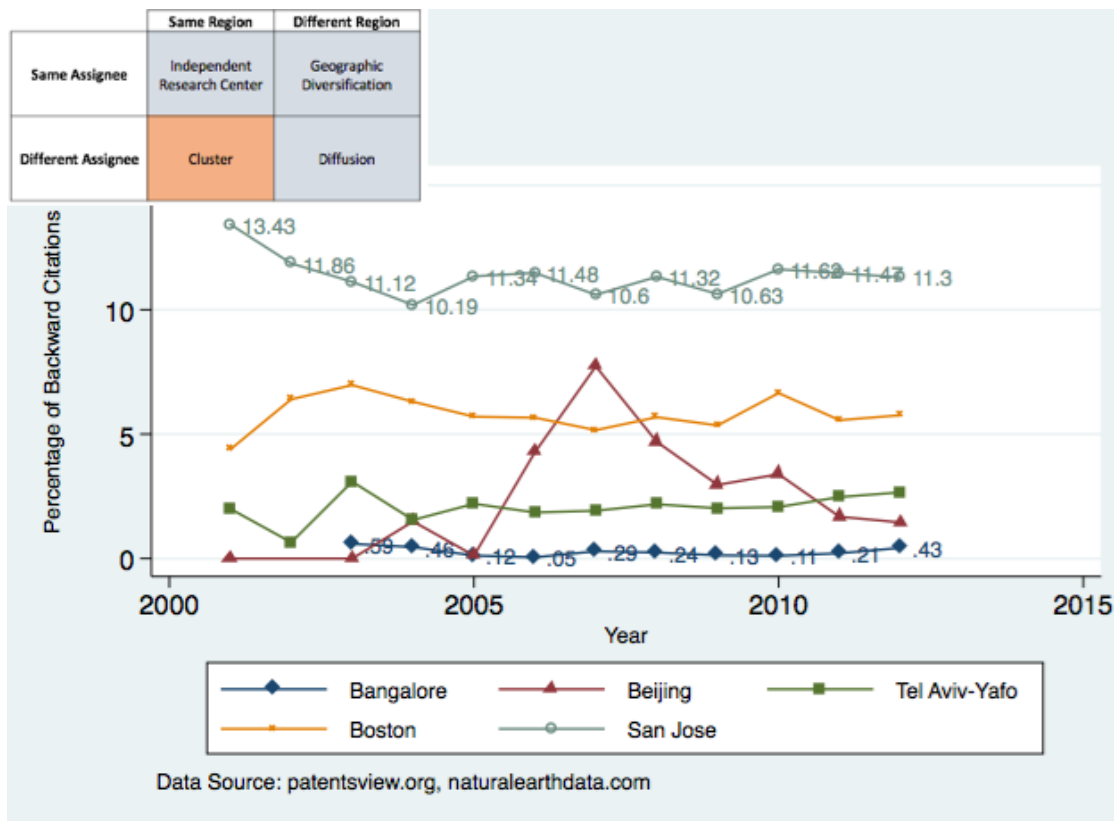
**FIGURE 2:** Knowledge flows within regions and within assignees



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

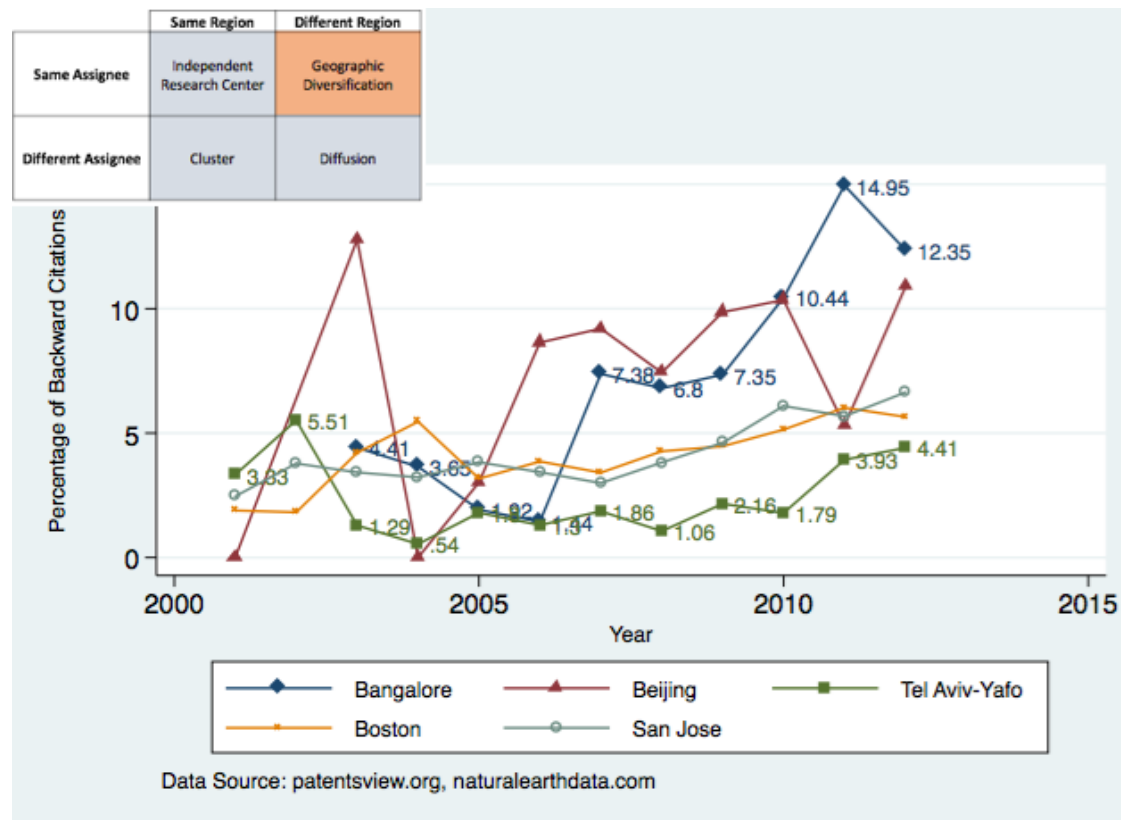
across firm flows of knowledge in Silicon Valley, a result consistent with several prior studies.

**FIGURE 3:** Knowledge flows within regions and across assignees



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 (?). 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.

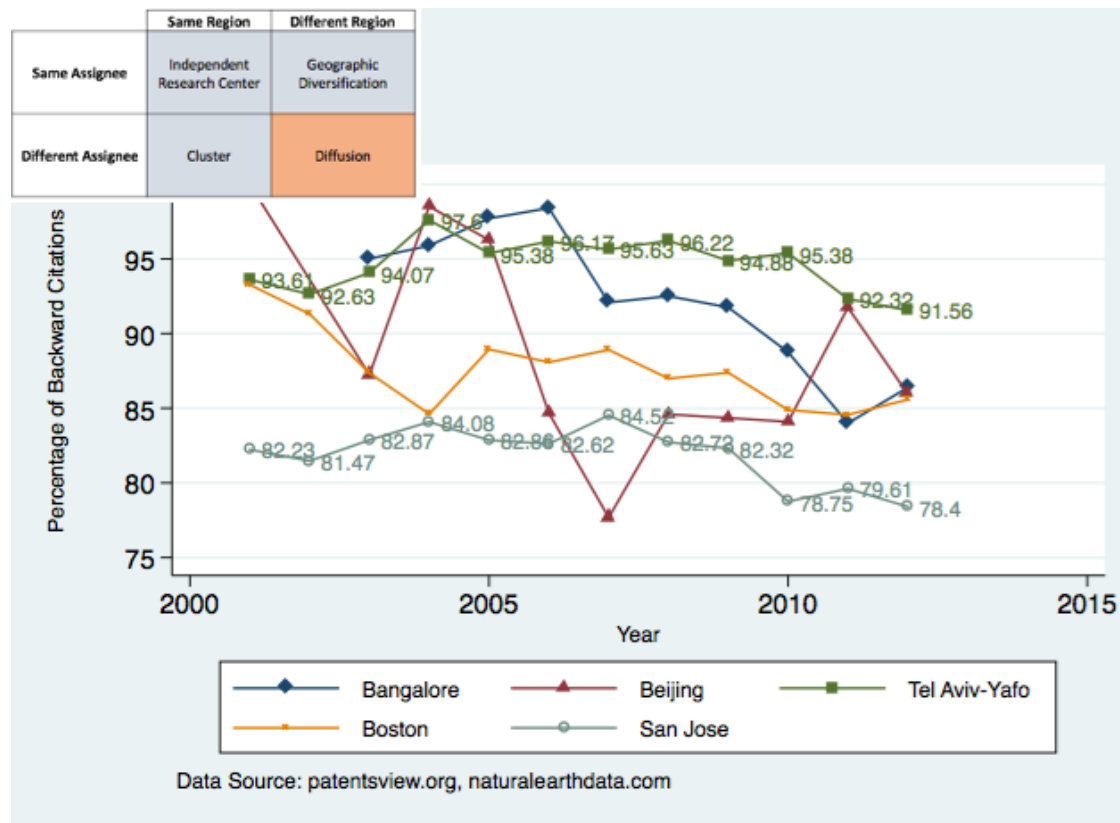
Finally, the bottom right quadrant labelled “Diffusion” captures high exploration along both dimensions, indicating the development of a global pipeline (?). Figure 5 depicts the knowledge flows for this category across time for the five regions. We note that Bangalore, Beijing and Tel

**FIGURE 4:** Knowledge flows across regions and within assignees

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.

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 9 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.



**FIGURE 5:** Knowledge flows across regions and across assignees

## 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 ?). 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

**TABLE 1:** Effects of knowledge flows on quality of inventions

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

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, ? 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.

## 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 or more 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. Citations are counted as those received up to the last date to which we have data.

## 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.

***Share 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.

***Share 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.

***Share 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.

***Share 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). 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 ?. 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.

Table 1 displays the correlation coefficients and summary statistics of our primary variables in the dataset of applicant citations. Table 2 displays the correlation coefficients and summary statistics of our primary variables in the dataset of examiner citations. The appendix provides similar statistics for the dataset of other citations (Table 5), and the entire citation dataset (Table 7)

## RESULTS

The results from our analysis for the dataset of applicant cited patents are presented in Table 3. 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 ?.

The results in Table 3 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 3 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.

The results from our analysis for the dataset of examiner cited patents are presented in Table 4. 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.

## DISCUSSION

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 be able to 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 (??). 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. ? 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 (?). However even in this tradition,



**TABLE 4:** Negative binomial regression analysis of invention quality for applicant citations

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

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

scholars have suggested opposite effects of knowledge flows. While ? show that mobility patterns of star inventors directly affect the geographic patterns of knowledge spillovers, ? 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 (?). 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 (??) and in the use of patent citations to demonstrate localization of knowledge spillovers (??), 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 (?). 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, ?) 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.

## APPENDIX: ADDITIONAL RESULTS

The additional results from our analysis for the dataset of applicant only 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. 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 8 on the dataset of all citations.

Table 7 displays the correlation of our primary variables in the dataset of all citations. Summary statistics for the variables in the dataset of all citations is provided in Table 7. Similar statistics for the applicant only citations dataset and the examiner only citations dataset have been provided in the section on additional results.



**TABLE 7:** Negative binomial regression analysis of invention quality for other citations

Dependent Variable Model	Total Citations Received			Non-Self Citations Received		
	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All	U.S.	Non-U.S.	All	U.S.	Non-U.S.
	Locations	Locations	Locations	Locations	Locations	Locations
Share Citations Made[Same Region, Different Assignee]	-1.176 (0.001)	-0.944 (0.050)	-1.351 (0.013)	-0.678 (0.077)	-0.674 (0.197)	-0.588 (0.301)
Share Citations Made[Different Region, Same Assignee]	-0.238 (0.290)	-0.433 (0.180)	-0.106 (0.741)	0.0426 (0.868)	-0.0363 (0.921)	0.131 (0.716)
Share Citations Made[Different Region, Different Assignee]	-0.607 (0.002)	-0.415 (0.114)	-0.720 (0.012)	0.0246 (0.911)	0.250 (0.411)	-0.0737 (0.817)
Log (Total Citations Made)	0.228 (0.000)	0.186 (0.000)	0.212 (0.000)	0.164 (0.000)	0.113 (0.000)	0.156 (0.000)
Log (Number of Patents)	0.554 (0.000)	0.586 (0.000)	0.646 (0.000)	0.610 (0.000)	0.633 (0.000)	0.726 (0.000)
Log (Patent Pool Size)	-0.0522 (0.002)	-0.101 (0.001)	-0.102 (0.000)	-0.00475 (0.792)	-0.0298 (0.339)	-0.0801 (0.001)
Constant	0.254 (0.218)	0.452 (0.119)	0.342 (0.259)	-0.569 (0.015)	-0.396 (0.232)	-0.567 (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 8:** Correlations and summary statistics for all citations with dependent variable as total citations received

Variables	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11	12
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. Share Citations Made[Same Region]	0.03	0.053	0.16	0.15	0.18	0.77	0.71	0.10	-0.54	1.00				
9. Share Citations Made[Same Assignee]	0.086	0.114	0.02	0.02	0.04	0.49	-0.00	0.95	-0.96	0.35	1.00			
10. Log (Total Citations Made)	5.331	2.463	0.33	0.31	0.37	0.15	0.17	0.07	-0.16	0.22	0.11	1.00		
11. Log (Number of Patents)	2.753	2.04	0.35	0.34	0.38	0.18	0.19	0.05	-0.16	0.25	0.11	0.94	1.00	
12. Log (Patent Pool Size)	5.381	2.286	0.31	0.30	0.34	0.20	0.23	0.04	-0.16	0.28	0.10	0.88	0.92	1.00
<b>N</b>							17166							



**TABLE 9:** Negative binomial regression analysis of invention quality for all citations

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