Heterogeneity in knowledge flows across regions: Investigating patterns and mechanisms

Ashwin Iyenggar, Sai Yayavaram

January 8, 2017

Abstract

We analyze the pattern of knowledge flows by geographical region and by firm for seven prominent regions. We demonstrate that locations display significant heterogeneity with regard to the relative proportions of local and non-local knowledge flows, and within firm and across firm knowledge flows. Specific patterns idiosyncratic distribution of flows are identified by location and suggestions are made for furthering of theory on the causal contributors of knowledge flows.

1 New Results

Table 1: Effect of Geographic Distribution of Citations Made on Citations Received

	(1) Citations Received	(2) Citations Received
Citations Received		
Citations Made to [Same Region, Same Assignee]	0.0000304***	0.000213
	(5.56)	(0.00)
Citations Made to [Same Region, Different Assignee]	0.00000743*	0.0000599
	(2.37)	(0.00)
Citations Made to [Different Region, Same Assignee]	-0.000000348	0.00000249
	(-0.09)	(0.00)
Citations Made to [Different Region, Different Assignee]	-0.00000296***	0.0000135
	(-5.55)	(0.00)
Citations Made to [Other]	-0.00000169	0.0000469
	(-1.54)	(0.01)
Log (Num Patents)	0.0132	-0.462
	(0.60)	(-0.00)
Log (Patent Pool Size)	0.642***	2.382
	(19.36)	(0.01)
Constant	-5.636***	-41.82
	(-23.45)	(.)
ln_r		
Constant	0.390**	
	(3.07)	
ln_s		
Constant	4.448***	
	(25.14)	
Year Dummy	Yes	Yes
Region Fixed Effects	No	Yes
Observations	2624	2624

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2 Introduction

The agglomeration characteristics of economic regions have been highlighted by scholars since long(Marshall, 2009). More recently, scholars have demonstrated through numerous studies that patent citations provide a paper trail of evidence for the existence the knowledge spillovers in economic regions (Almeida and Kogut, 1999; Jaffe et al., 1993), the effects of inventor mobility (e.g., Almeida and Kogut (1999)), of Intellectual Property Rights regime of locations (e.g., Zhao (2006)) and of the role of international geography (e.g., Singh (2007)) on knowledge spillovers. Knowledge spillovers is observed in practice however, to be highly heterogenous across locations, firms and legal regimes. The question of the causal mechanisms leading to knowledge spillovers remains largely unresolved, despite the enormous progress made by prior scholars. In this article, we investigate the patterns of knowledge flow as evidenced by patent citations across geographic regions around the world and use our findings to ascertain plausible causal factors leading to the heterogeneity in knowledge flows.

Literature in the strategy area has highlighted the importance of innovation as a source of competitive advantage in firms. Scholars have highlighted that firms have tended to adopt two distinct strategies in seeking to capture greater advantages in knowledge flows: a) geographic clustering (Porter, 2003), and b) the globalization of R&D (Almeida, 1996). Scholars have in the past conclusively demonstrated that the bay area in California demonstrates strong cluster characteristics in that there are strong flows of knowledge within and across firms from within the same geographical area. While the benefits to such geographical localization of knowledge flows (Porter, 2003) has been celebrated as an important aspect of the superior economic performance of Silicon Valley, there has been little work that has explored the same for emerging innovation regions of Bangalore, Beijing, Israel and Austin. In this article, we analyze the nature of knowledge flows at the level of the region in aggregate rather than focus on specific industries of technologies as have been done in past studies (Lecocq and Van Looy, 2016). In order to understand if these emerging innovation regions are trending toward clustering (Jaffe et al., 1993) or globalization or knowledge flows or

Table 2: Categories of Knowledge Flows

		Geographic Region: Same	Geographic Region: Different
Assignee	Same	Independent Research Centre	Geographic Diversification
	Different	Cluster	Diffusion

both, we categorize all knowledge flows along two dimensions: a) as a relationship between the geographic region of the creator of the knowledge and the geographic region of the user of the knowledge, and b) as a relationship between the firms that create the knowledge (assignee of the cited patent) and those that use the knowledge (assignee of the citing patent). This classification allows us to see knowledge flows in four mutually exclusive but collectively exhaustive categories as illustrated in Table 2

In Table 2, the quadrants on the left column indicate knowledge flows within the region whereas the quadrants in the right column indicate knowledge flows to other regions. We are interested in understanding if the emerging innovation clusters of Bangalore, Beijing, Israel, and Austin show the characteristics of geographical clustering (Jaffe et al., 1993). We use the Boston region and Silicon Valley as leading innovation clusters to inform our reference point in studying the emerging innovation clusters.

The investigation of potential mechanisms behind local spillovers is interesting for a number of reasons. Given the wide disparity in the extent of knowledge spillovers across locations, across firms and across IPR regimes it is intriguing to a researcher to find the mechanisms that may lie behind such a phenomenon. A specific flavor of this question is the investigation of the spillover effects of patenting in emerging countries, or those known to have weaker IPR regimes. Specifically, do multinational firms that develop patentable technologies in emerging countries create spillover effects in the host country talent pool, or do the benefits remain localized to within multinational

companies (MNCs)? From a policy perspective, it is valuable to understand the impact of allowing MNCs dominate the patenting process in emerging markets on the quality of the talent pool in the host country. Does a significant group of local inventors develop? Is this affected by the strength of the IPR regime in the host country? Patents data allows us to ask and try and answer this question.

We evaluate the nature of knowledge flows across geographic regions by initially looking at six major regions of the world: San Francisco and greater Bay Area of California, Austin, Texas, the greater Boston area, Tel Aviv, Beijing and Bangalore. The sampling has been made keeping in mind both established and upcoming technological locations.

3 Methods

3.1 Unit of Analysis

Our unit of analysis is the flow of knowledge between locations, between assignees. In order to proceed with empirical work, we make the following decisions. First, we focus our analysis on a region of interest, and flows are observed over time. Second, we map flows onto the two dimensions of region (geography) and assignee. Along each axis we are interested in local and internal flows as against global and external flows of knowledge. Finally, in order that the various regions may be compared on these axes, the flows of knowledge within each quadrant will be normalized to a percent value of the total flows for that region that year.

3.2 Definition of Geographic Regions

We discuss here how we go about defining the geographic regions of interest to this investigation. For locations in the United States, it is standard to use Metropolitan Statistical Areas (MSA) for analyses related to economic geography. The approach is less standard for non-US locations, and this problem is particularly exacerbated by the absence of a similar measure as the MSA. Urban

Table 3: Categories of patent citations

Category	Number of citations
cited by applicant	16,527,942
cited by examiner	17,174,252
cited by other	25,444,463
cited by third party	325
	21,581,784

areas are a reasonable substitute for economic centers, and we therefore determine to use one such definition. Specifically, for MSA of US locations, we obtain data from the US census and for urban areas for world wide locations, we obtain data from Natural Earth Data.

This automatically raises conflicting definitions for locations in the United States. So that the MSA definitions take precedence, we eliminated all data pertaining to US locations from the Natural Earth urban centers data and integrated this with the MSA information. With this we generated a single database of location information for economic centers around the world. The appendix provides visual map-based snapshots of our regions of interest. The regions colored yellow are the ones in focus, while those in purple are neighboring regions outside the region of interest.

We note from the MSA data that the Bay Area of California is actually split between the two MSA regions of San Francisco-Oakland-Hayward, CA and San Jose-Sunnyvale-Santa Clara, CA. we therefore decide to treat the two as two regions for the current analysis. It is possible that we may need analyze the data again clubbing the two in the future. Bangalore is seen as including Hosur, the Boston-Cambridge-Newton MSA includes parts of New Hampshire and Beijing seems to extend a bit to the south. These seem to be reasonable definitions for the respective economic geographies.

3.3 Mapping geographical co-ordinates to regions

The file named `location.tsv` from patentsview.org contains the latitude and longitude information for all locations referenced in the patentsview.org database. The `location.tsv` associates a `location_id` to each latitude-longitude combination. we use the merged MSA and urban centers information and the geographical information in `location.tsv` to obtain a mapping from each `location_id` used in the patentsview database to the economic geography that it corresponds to. The patentsview.org database defines 128,911 unique location_ids, and our data is able to map 53,424 of those locations to an economic geography region. The rest are assumed to be those locations that fall outside any major urban center in the world from which patents from been filed or been assigned.

3.4 Selecting applicant cited patent citations

The `uspatentcitation.tsv` file from patentsview.org maps every patent-patent level citation that has been made since 1976. This file has 80,728,766 observations. Table 3 provides a break up based on category. The US patent office has been systematically categorizing citations by category since after the year 2000. This explains the many empty citation category entries.

In order that we are consistent with our initial objective of measuring only applicant cited patents as flows of knowledge, we restrict ourselves to those patents categorized as 'cited by applicant'. This decision has the additional effect of limiting our period of analysis to citing patents applied for after the year 2000.

3.5 Expanding the US patent citation

We use the `application.tsv` file to determine the year of application of the citing patent and then use this to add a year field to the uspatentcitation entry. After selecting only those citation entries where the year of application of the citing patent is 2012 or earlier, we are left with 11,822,154

Table 4: Number of citation entries by region of interest (till 2012)

Region	Number of citations
Boston-Cambridge-Newton, MA-NH	4,602,355
San Jose-Sunnyvale-Santa Clara, CA	8,431,536
Bangalore	183,685
Beijing	131,752
Tel Aviv-Yafo	872,578
San Francisco-Oakland-Hayward, CA	9,258,684
Austin-Round Rock, TX	259,503

citation entries. In order to determine internal firm flows or external firm flows of knowledge, we use the assignee_id on each patent to identify similarity or dissimilarity of assignees. There are 5,300,888 unique assignee entries. While a vast majority of patents are assigned to a single assignee, there are a few that are assigned to more than one assignee. In attaching assignee_id to each citing and cited patent on a citation, we create separate entries for each unique assignee on each patent. With this we end up with 12,256,759 citation entries of which 2,869,978 entries have an empty assignee_id for either the citing patent or the cited patent. A future revision could potentially work to reduce the loss due to empty assignee_id. The 12,256,759 citation entries with year and assignee_ids for both citing and cited patents are then expanded to include every inventor on each citing patent and each cited patent. This process is performed using a Python script as the joinby process was turning out to be extremely time consuming on Stata (we had runs of over 40 hours without Stata finishing). At the end of this process, we had 105,369,401 citing-patent-assignee-location to cited-patent-assignee-location citation entries. This formed the master dataset for the further analysis.

4 Analysis of flows

Table 4 captures the number of citing patent-inventor-assignee to cited patent-inventor-assignee flows that form the dataset on which further investigation is conducted. Starting with 23,825,110 entries, we drop duplicate citations between the same patents and the same regions. This leaves us with 5,820,864 entries. In determining if the citing patent assignee and the cited patent assignee

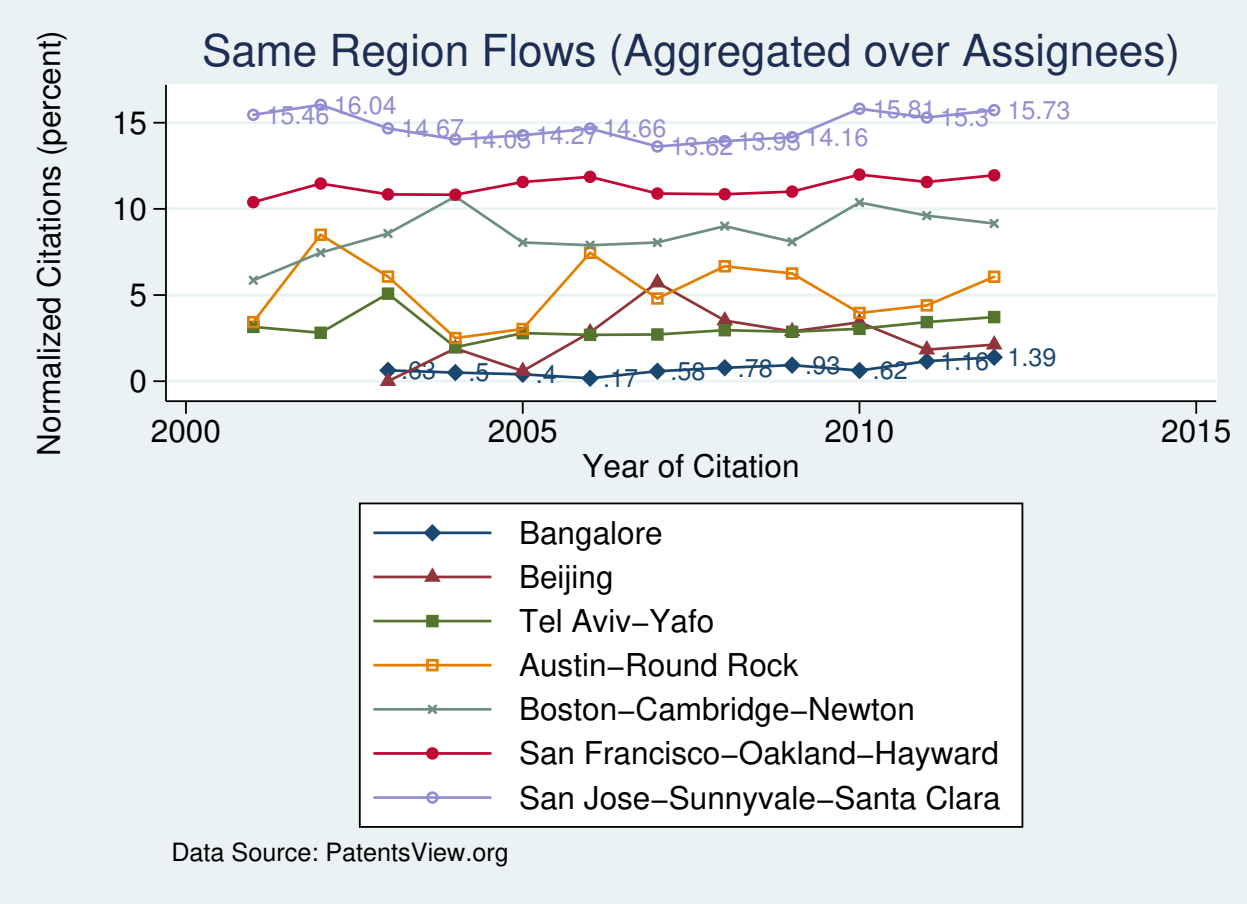

Figure 1: Local Knowledge Flows by Region

match, we drop all those entries where assignee information is unavailable for either side. That leaves us with 5,058,782 entries. We similarly drop those entries where either location region is unknown. We do so because it would seem incorrect to conclude that two locations differ when one location is undefined. After this step, we have 4,661,422 entries in our data set where conclusions can be clearly made about whether the assignees match and if the locations match between citing patent inventors and cited patent inventors. With this dataset, we calculate the normalized scores in our 2x2, and two aggregate measures - the first for same location flows across assignees, and the second for same assignee flows across locations. All six scores are expressed in percentages rounded to the nearest integer. The results are plotted in linear scale and are evident in the following graphs.

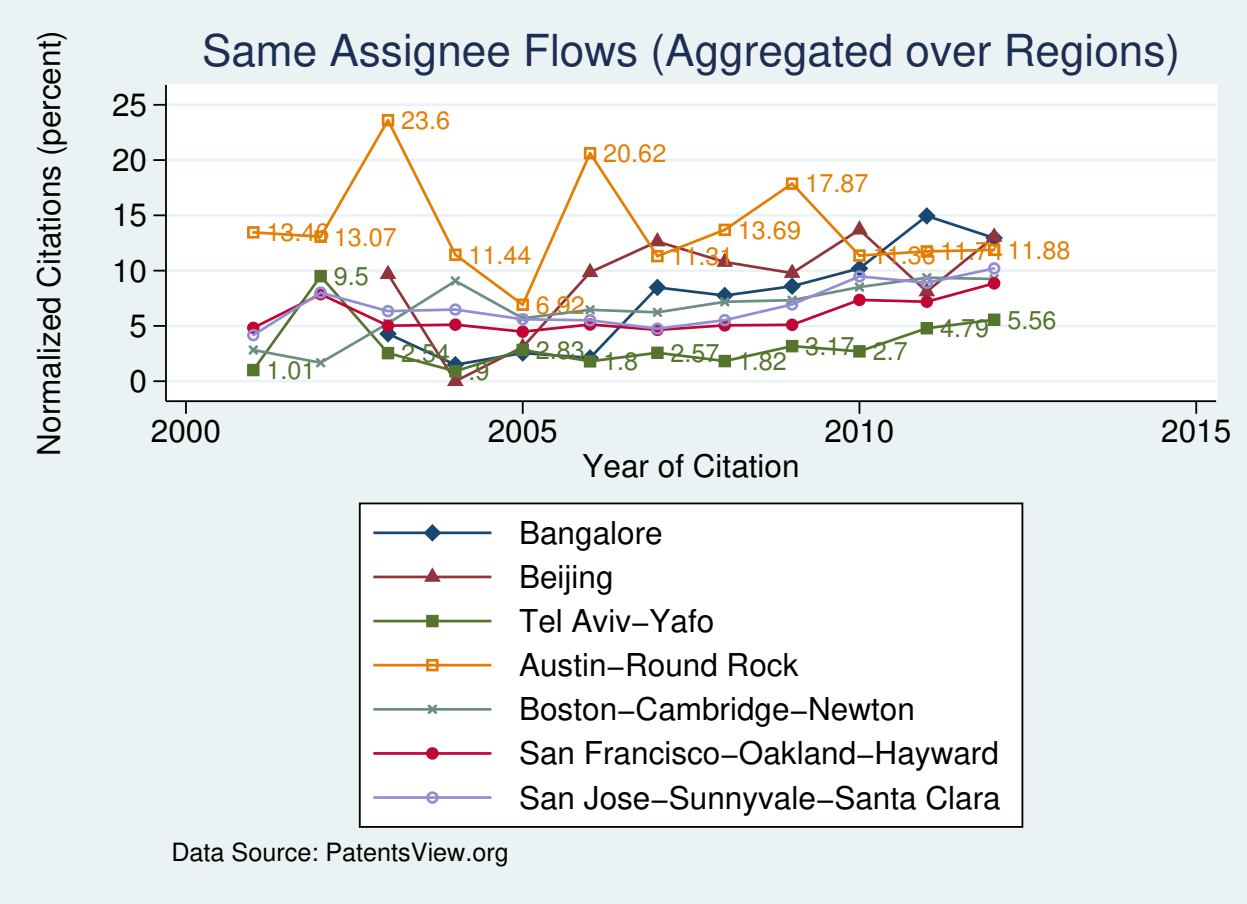

Figure 2: Internal Knowledge Flows by Region

5 Preliminary Results

Figure 1 clearly indicates that the phenomenon of knowledge spillovers is strongly evident in the two California locations in our sample, and to a smaller extent in the extended Boston region. But clearly, the localization of knowledge spillovers is not observed in the three emerging country locations of Bangalore, Tel Aviv and Beijing. Figure fig:SameRegionDiffAssigneeFlows reiterates the gulf between the two California locations and the rest. On the other hand from Figure 2, we note that the Bangalore region has seen an increasing share of flows to non-local but internal flows. This maybe proof of an increasing amount of research being done in Bangalore by global multinationals for their corporate headquarters. It is also interesting to note that while Beijing and Bangalore are at a similar level (14 percent) in 2012 on this count, Bangalore has seen a steady

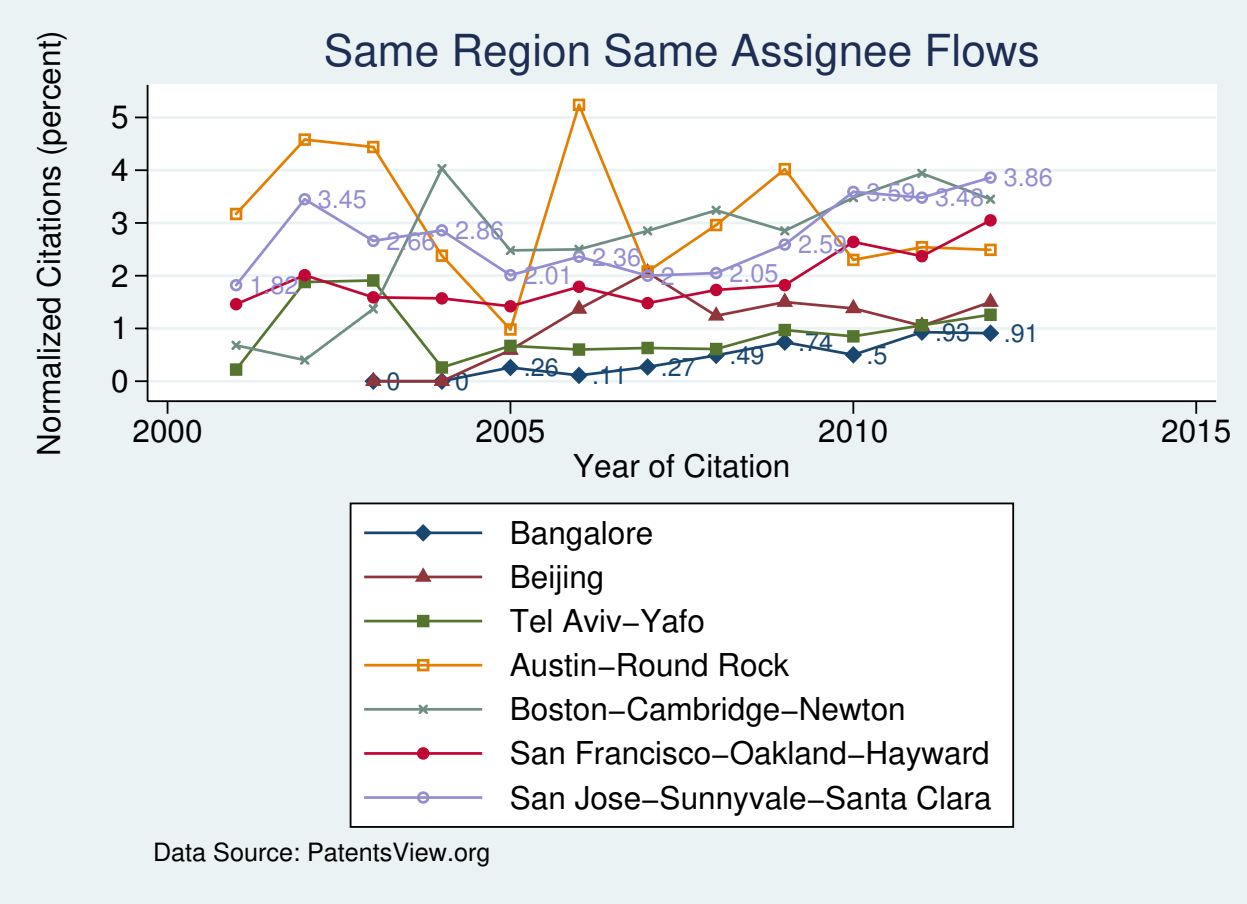

Figure 3: Local and Internal Flows by Region

increase while Beijing has somewhat stalled since 2005.

Additionally, Figure 6 demonstrates the stark difference between Tel Aviv-Yafo and San Jose-Sunnyvale-Santa Clara, CA on the extent of non-local external flows. Tel Aviv-Yafo seems to be strongly integrated into the external knowledge network but not integrated much locally. Figure 5 reiterates the Bangalore position as the outsourcing destination for global R&D.

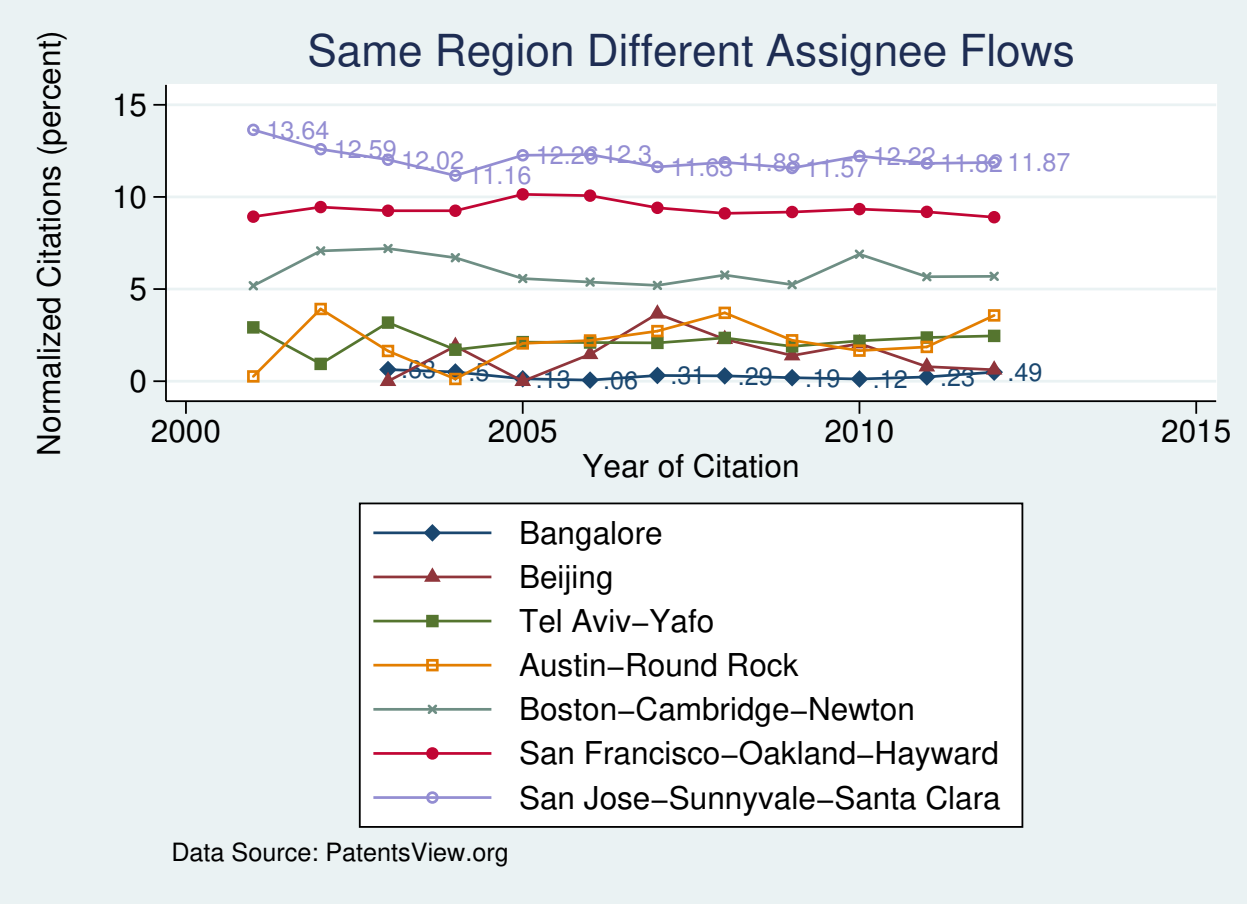

Figure 4: Local and External Flows by Region

6 Main Results

7 Conclusion

We started this article attempting to understand some of the factors that may explain the heterogeneity in knowledge flows across regions. While we have yet to get there, we have in the process demonstrated strong support for both the existence of the heterogeneity in knowledge flows across regions, as well as identifying some patterns relative to certain locations. We found that the San Francisco and San Jose MSA regions have the highest proportion of local knowledge flows, while Tel Aviv has the least. We also found that most of the flows for Tel Aviv patents are from different firms in different locations. Bangalore flows were higher to same firms in other locations, and

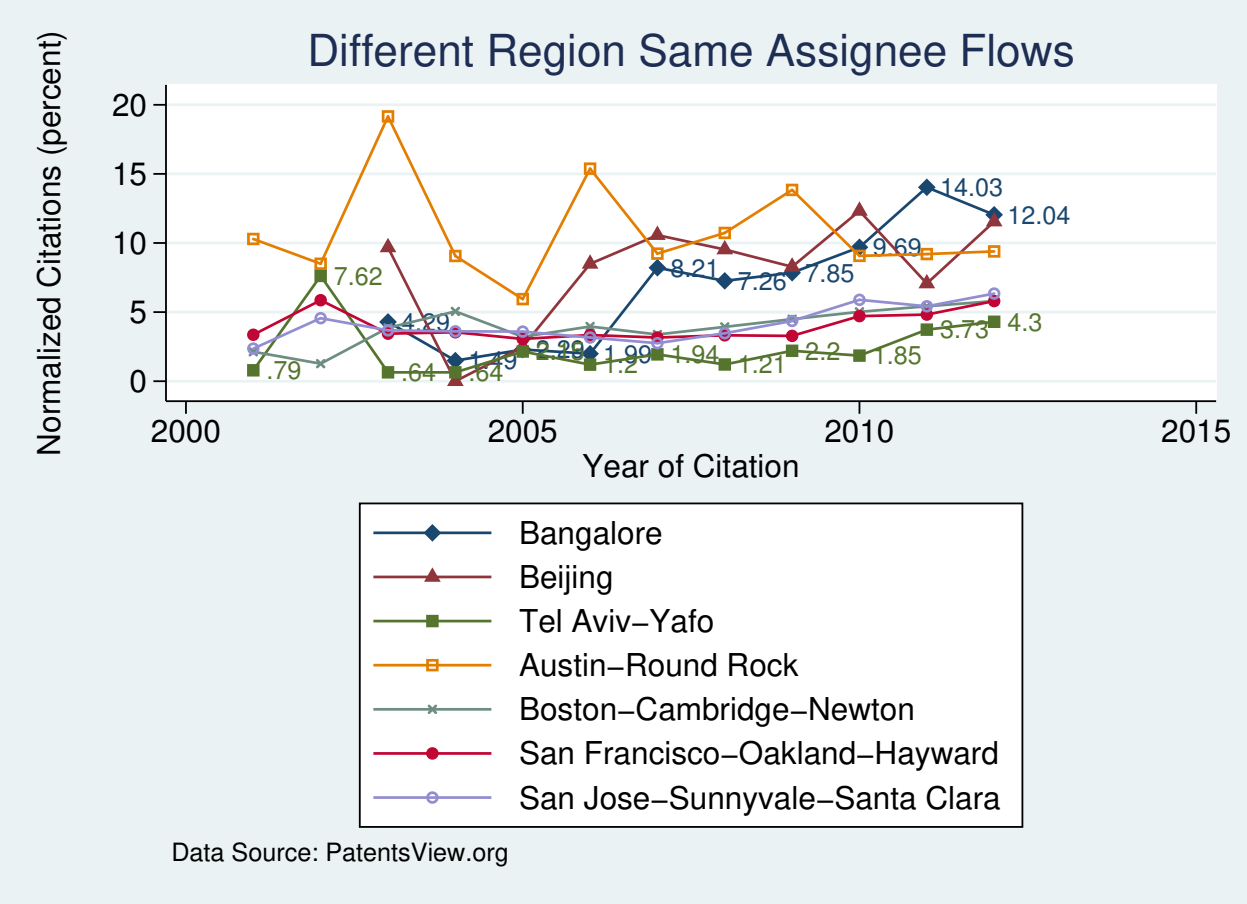

Figure 5: Non-local and Internal Flows by Region

were low to other firms locally. This study throws open opportunities to investigate specific mechanisms that might be leading to such behavior. While scholars have look at IPR regime effects, cross-border effects, we could consider looking at interconnectedness, modularity, and technology dynamics as potential explanatory variables in order to contribute to theory building on knowledge flows across locations.

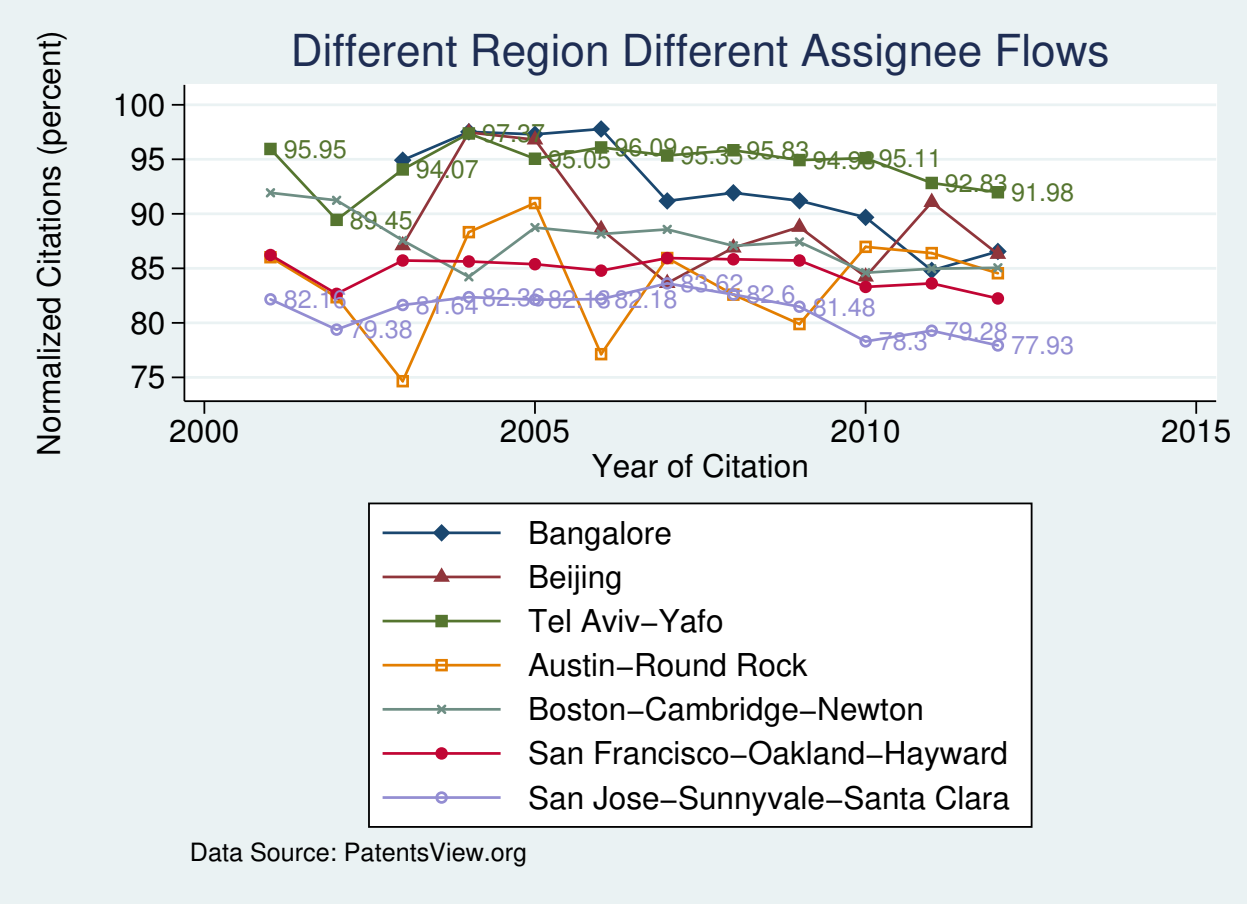

Figure 6: Non-local and External Flows by Region

8 Conclusion

We started this article attempting to understand some of the factors that may explain the heterogeneity in knowledge flows across regions. While we have yet to get there, we have in the process demonstrated strong support for both the existence of the heterogeneity in knowledge flows across regions, as well as identifying some patterns relative to certain locations. We found that the San Francisco and San Jose MSA regions have the highest proportion of local knowledge flows, while Tel Aviv has the least. We also found that most of the flows for Tel Aviv patents are from different firms in different locations. Bangalore flows were higher to same firms in other locations, and were low to other firms locally. This study throws open opportunities to investigate specific mechanisms that might be leading to such behavior. While scholars have look at IPR regime effects, crossborder

effects, we could consider looking at interconnectedness, modularity, and technology dynamics as potential explanatory variables in order to contribute to theory building on knowledge flows across locations.

References

Almeida, P. (1996). Knowledge sourcing by foreign multinationals: Patent citation analysis in the u.s. semiconductor industry. *Strategic Management Journal*, 17(S2):155–165.

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

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

Lecocq, C. and Van Looy, B. (2016). What differentiates top regions in the field of biotechnology? an empirical study of the texture characteristics of biotech regions in north america, europe, and asia-pacific. *Industrial and Corporate Change*.

Lesser, W. (2010). Measuring intellectual property strength and effects: An assessment of patent scoring systems and casuality. *J. Bus. Entrepreneurship & L.*, 4:345.

Marshall, A. (2009). *Principles of Economics: Unabridged Eighth Edition*. Cosimo, Inc.

Porter, M. (2003). The economic performance of regions. *Regional Studies*, 37(6-7):549–578.

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

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

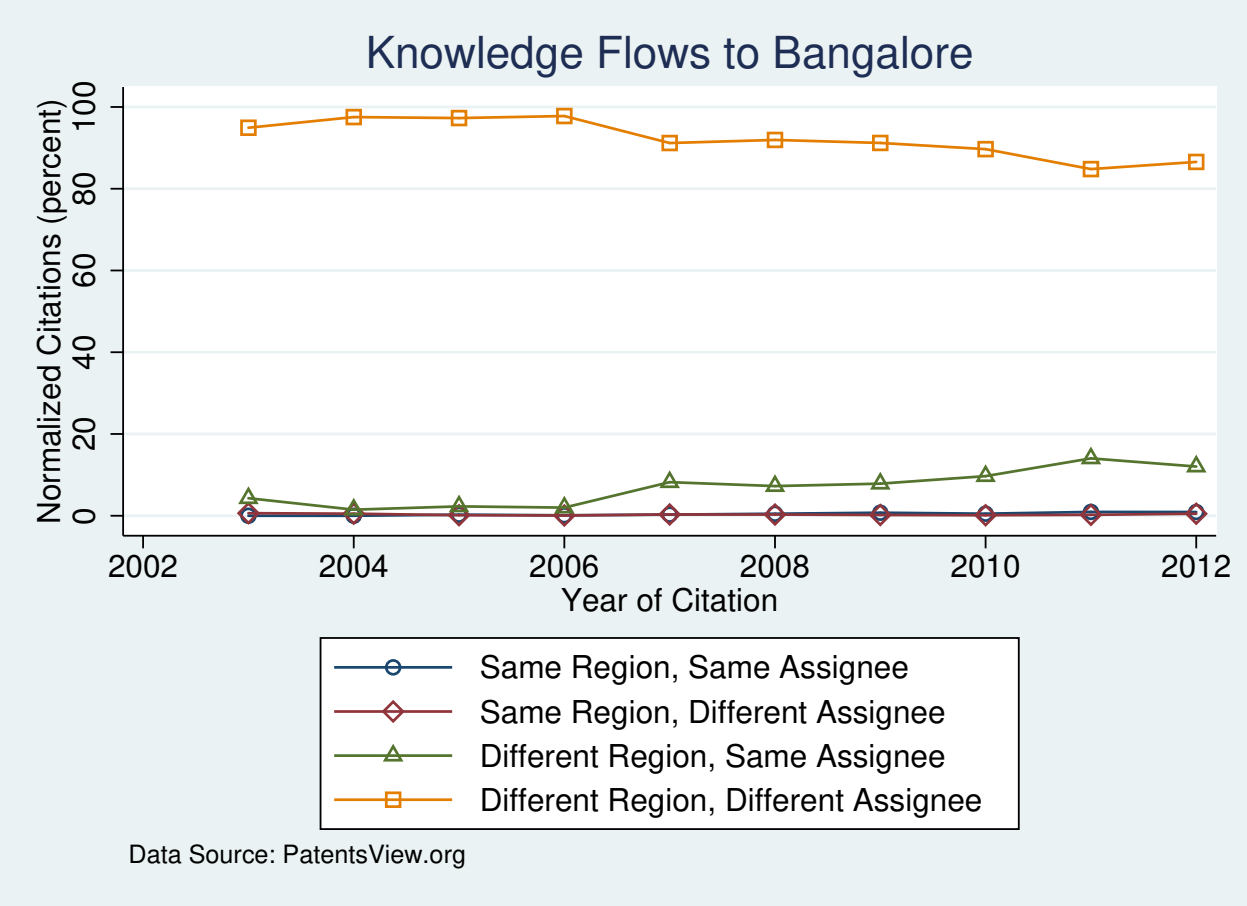

Figure 7: Relative Flows by Region : Bangalore

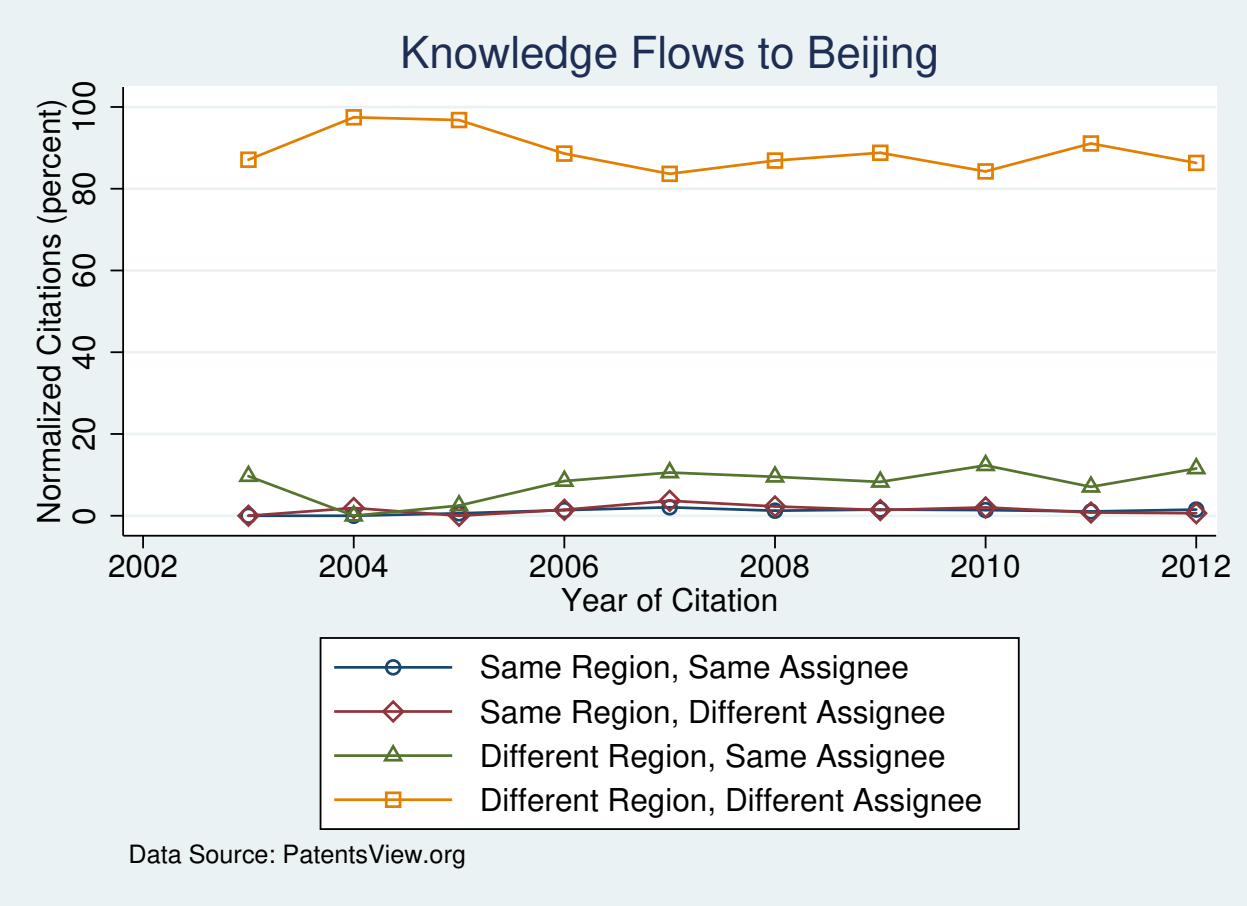

Figure 8: Relative Flows by Region : Beijing

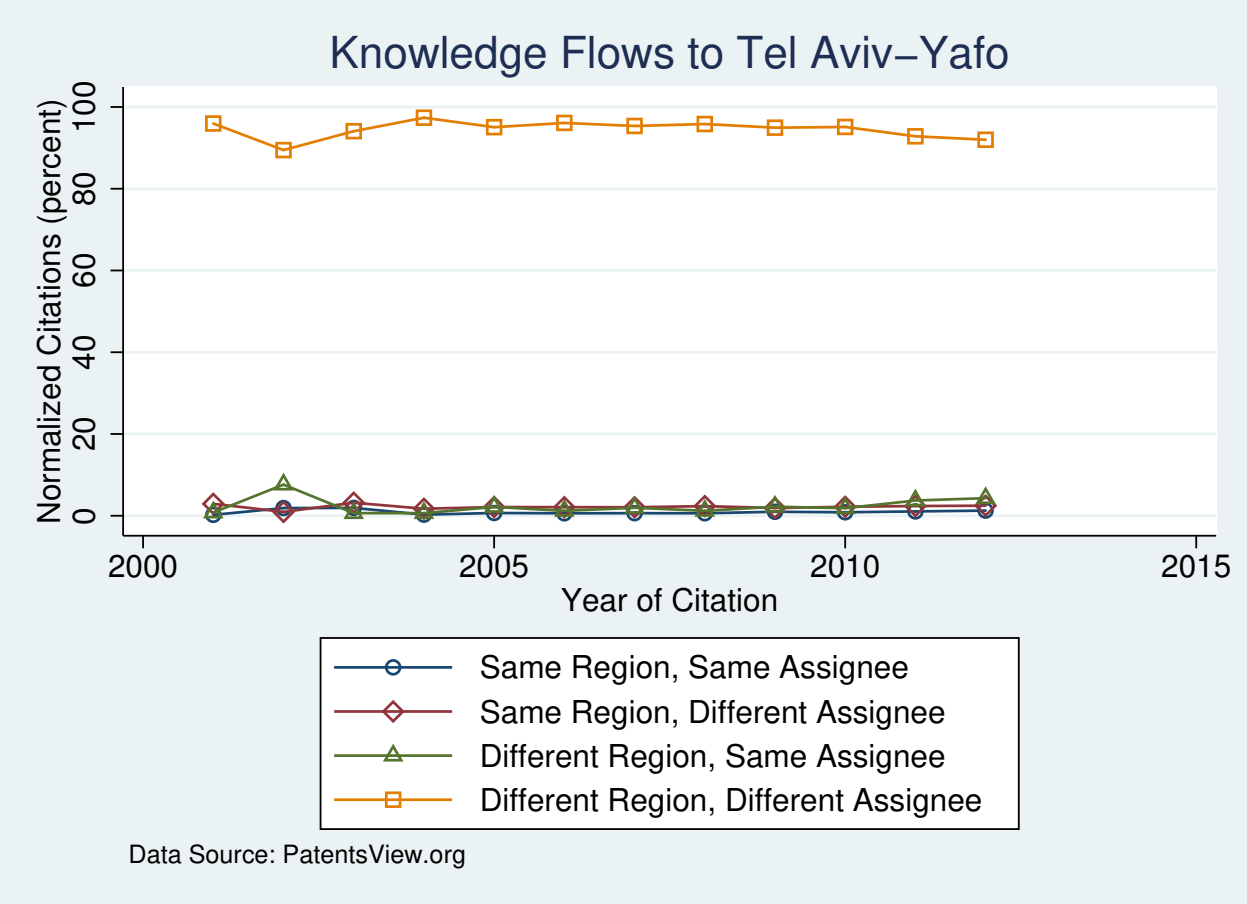

Figure 9: Relative Flows by Region : Tel Aviv-Yafo

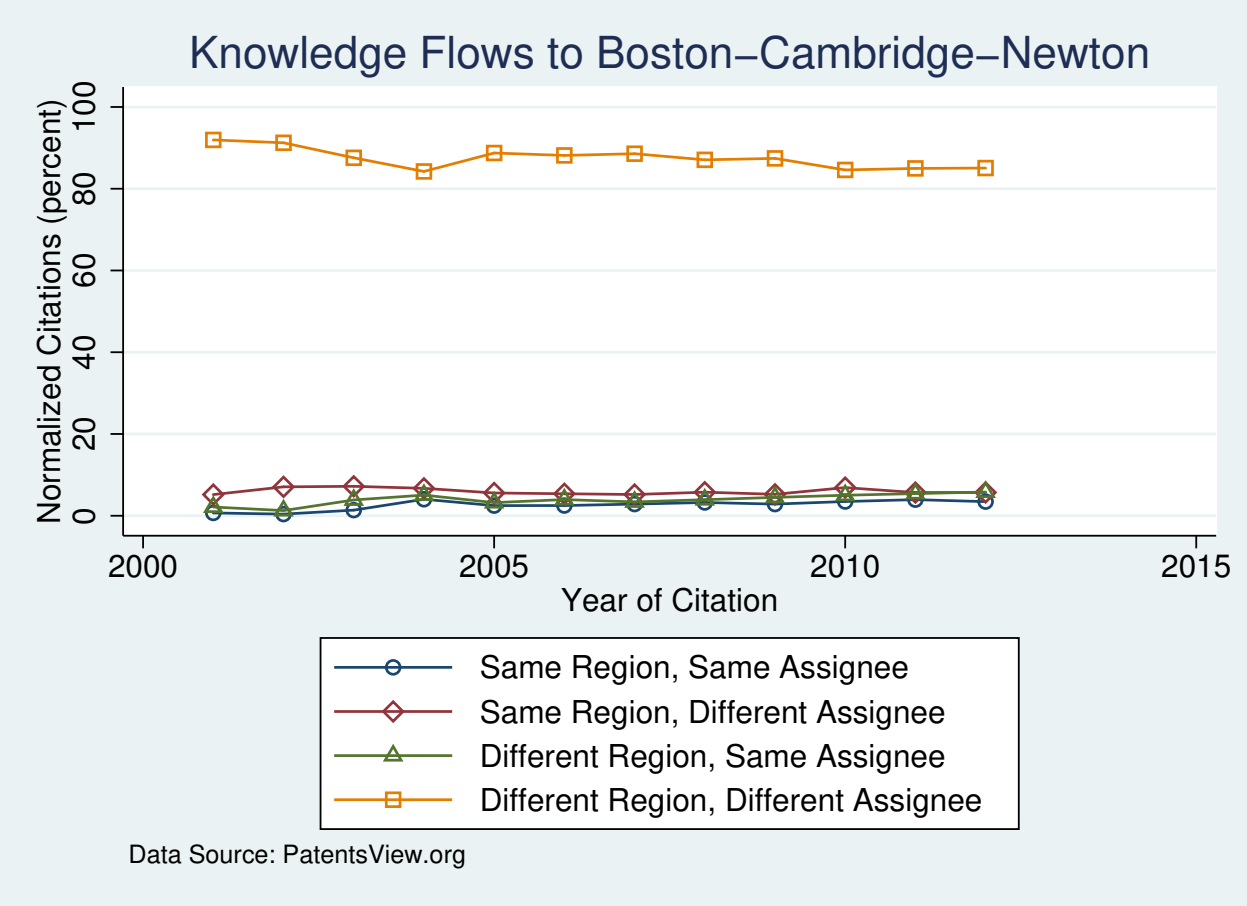

Figure 10: Relative Flows by Region : Boston-Cambridge-Newton

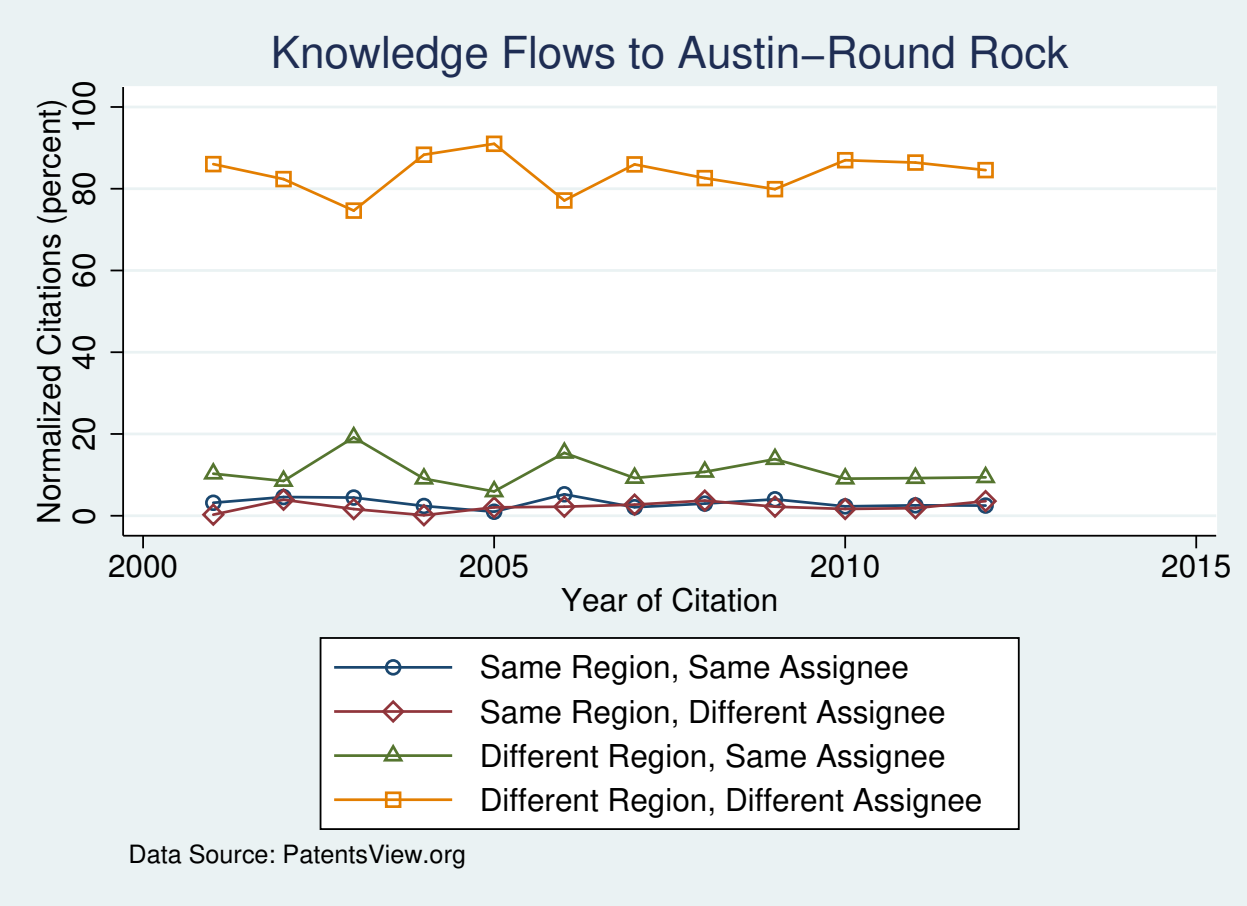

Figure 11: Relative Flows by Region : Austin-Round Rock

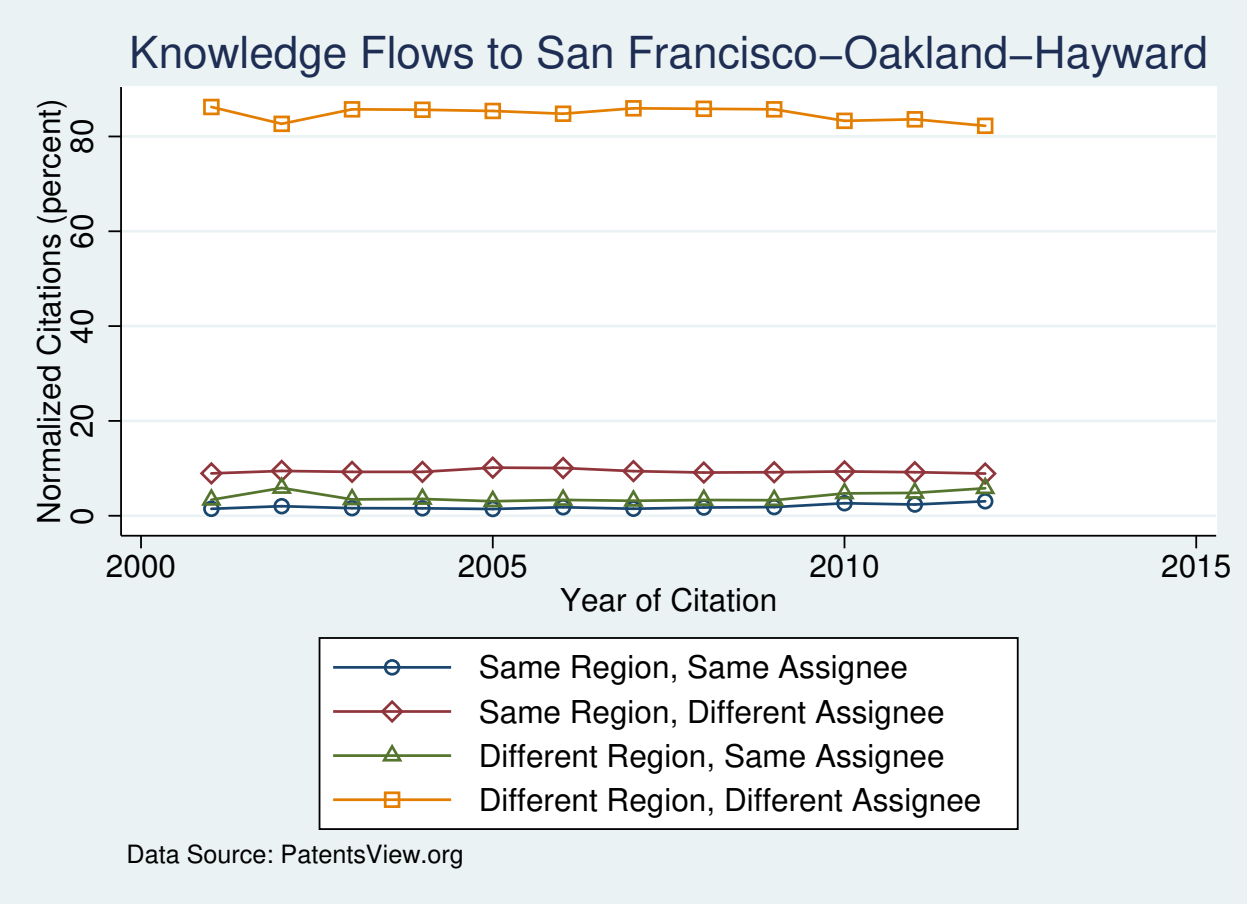

Figure 12: Relative Flows by Region : San Francisco-Oakland-Hayward

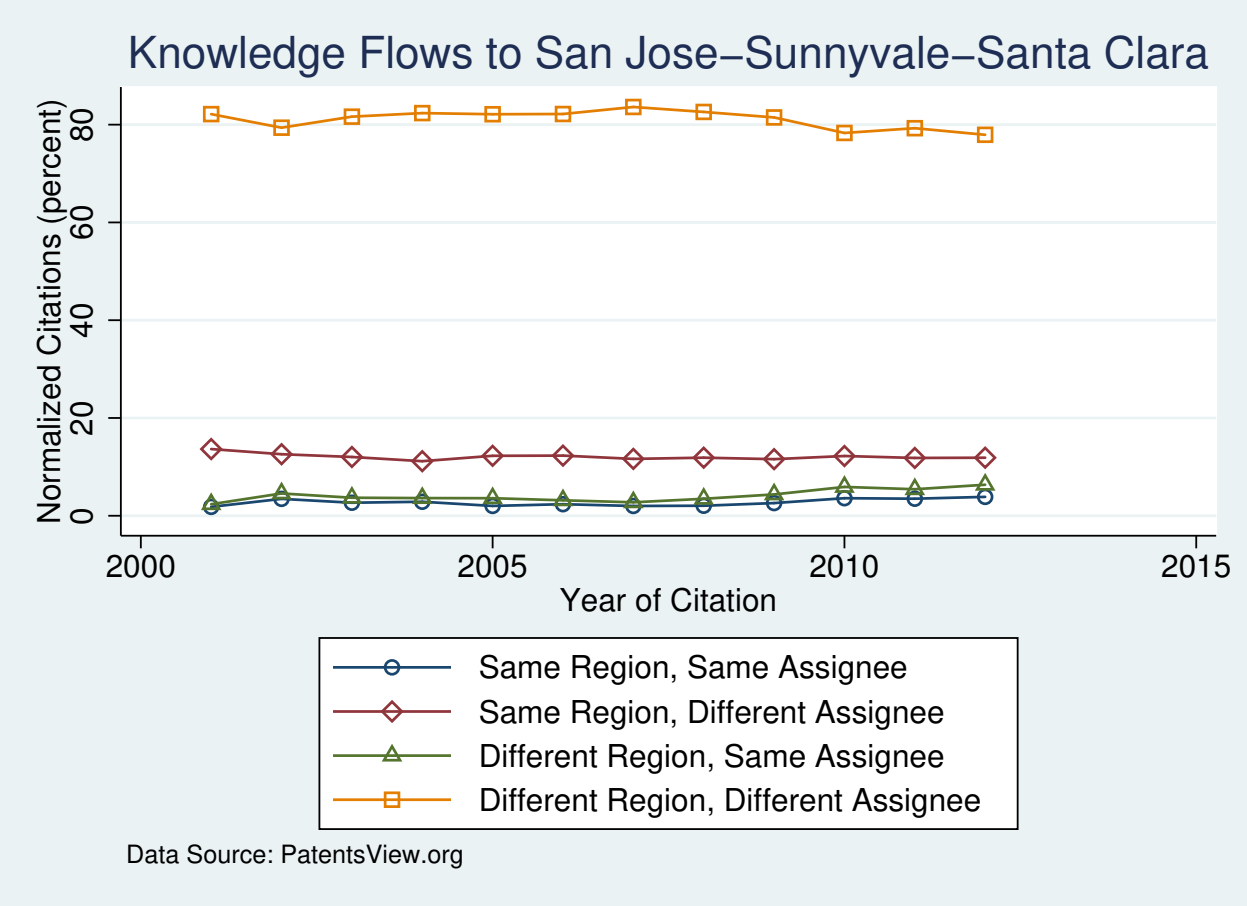

Figure 13: Relative Flows by Region : San Jose-Sunnyvale-Santa Clara