Heterogeneity in knowledge flows across regions: Investigating patterns and mechanisms

Ashwin Iyenggar, Sai Yayavaram

January 6, 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

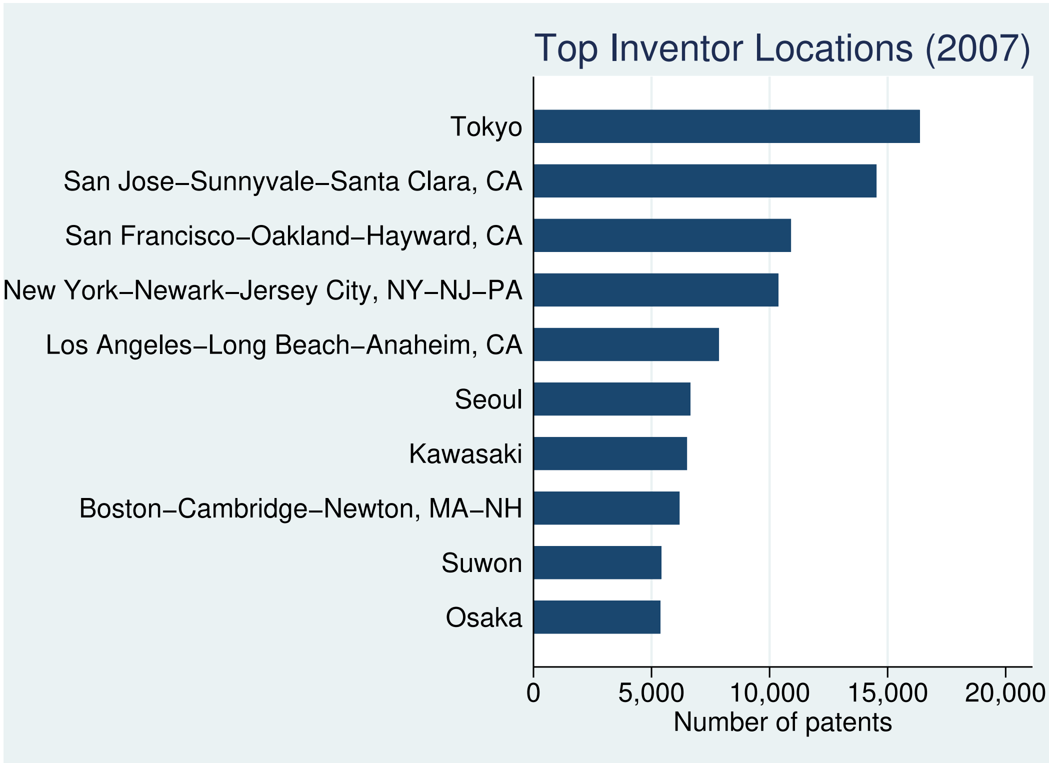

Figure 1: Top Inventor Locations

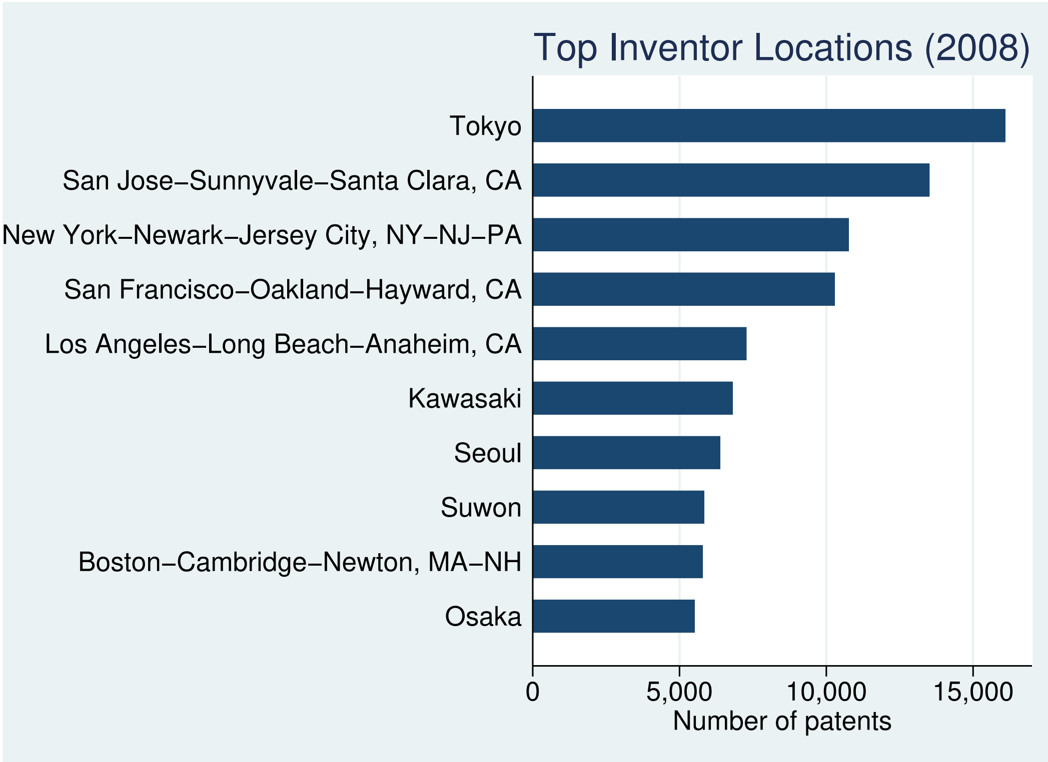

Figure 2: Top Inventor Locations

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

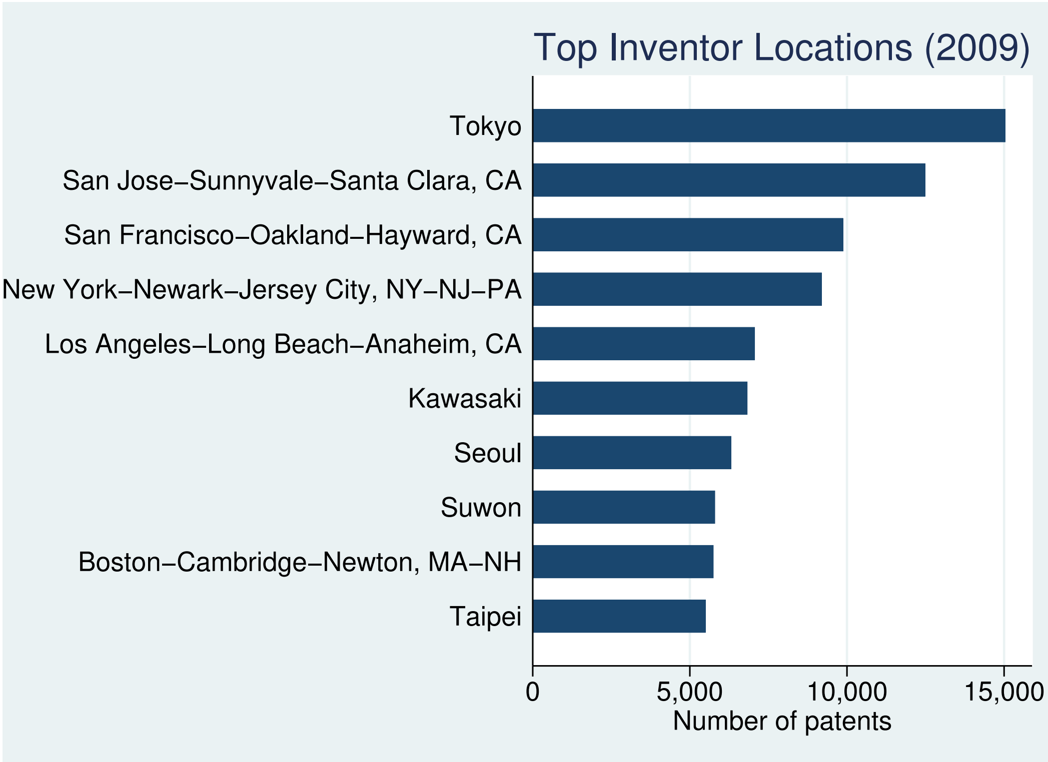

Figure 3: Top Inventor Locations

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 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

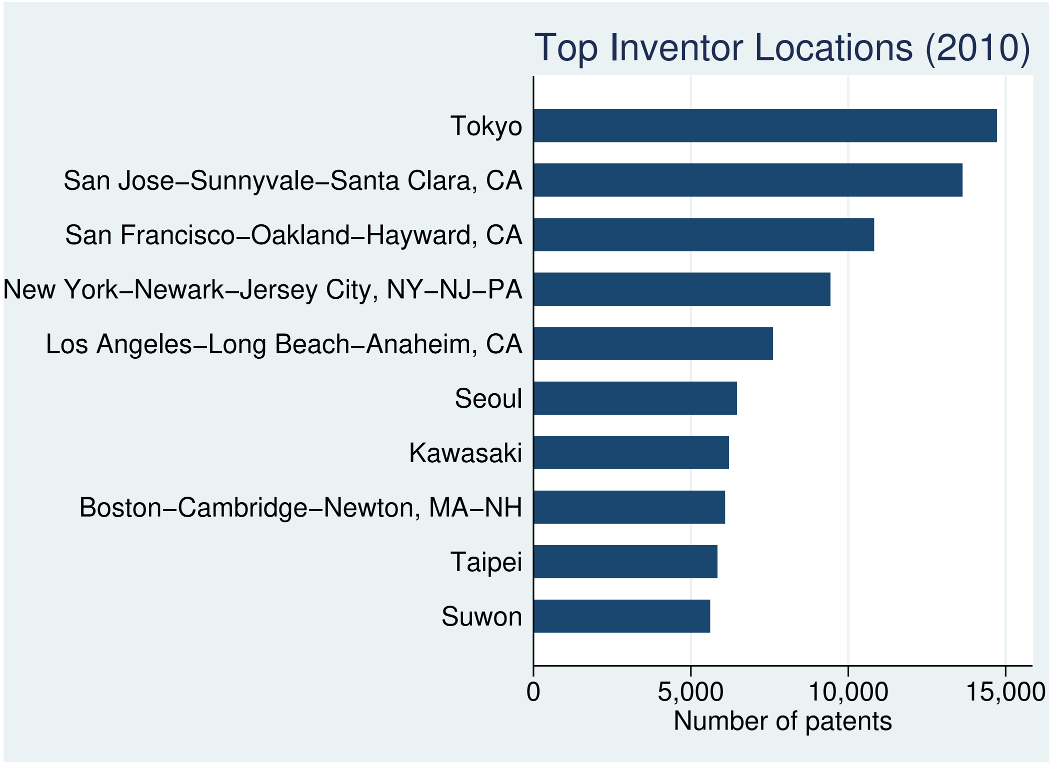

Figure 4: Top Inventor Locations

Table 1: Categories of Knowledge Flows

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

knowledge flows in four mutually exclusive but collectively exhaustive categories as illustrated in Table 1

In Table 1, 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.

Figure 5: Top Inventor Locations

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.

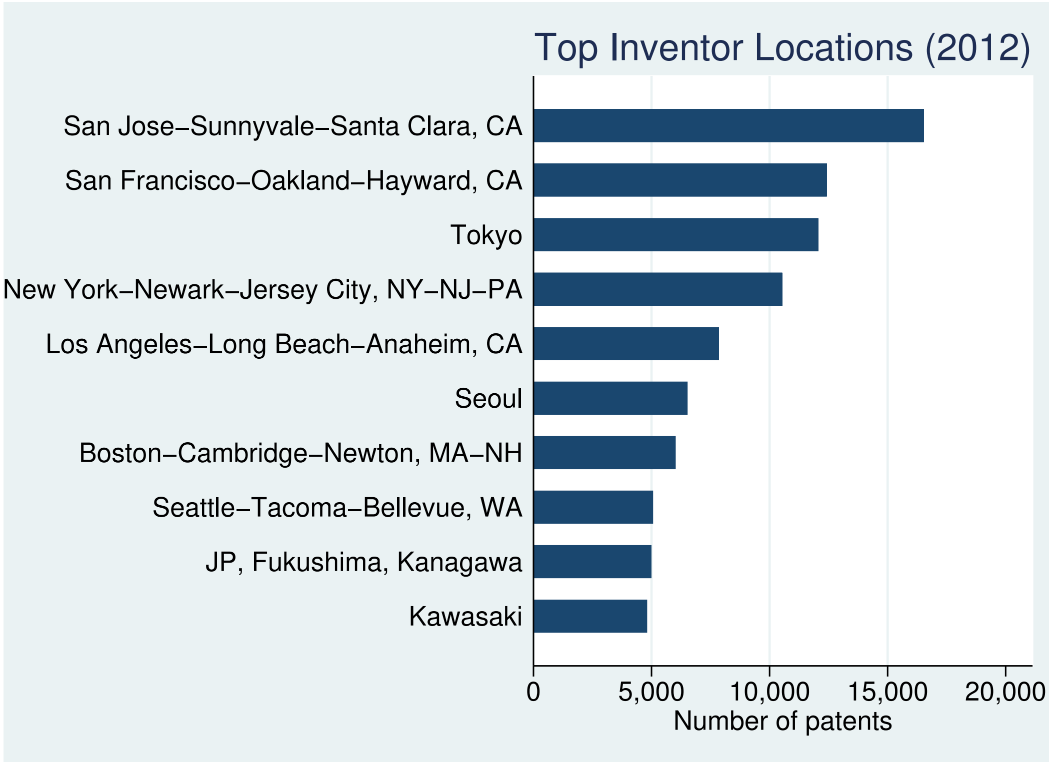

Figure 6: Top Inventor Locations

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 re-

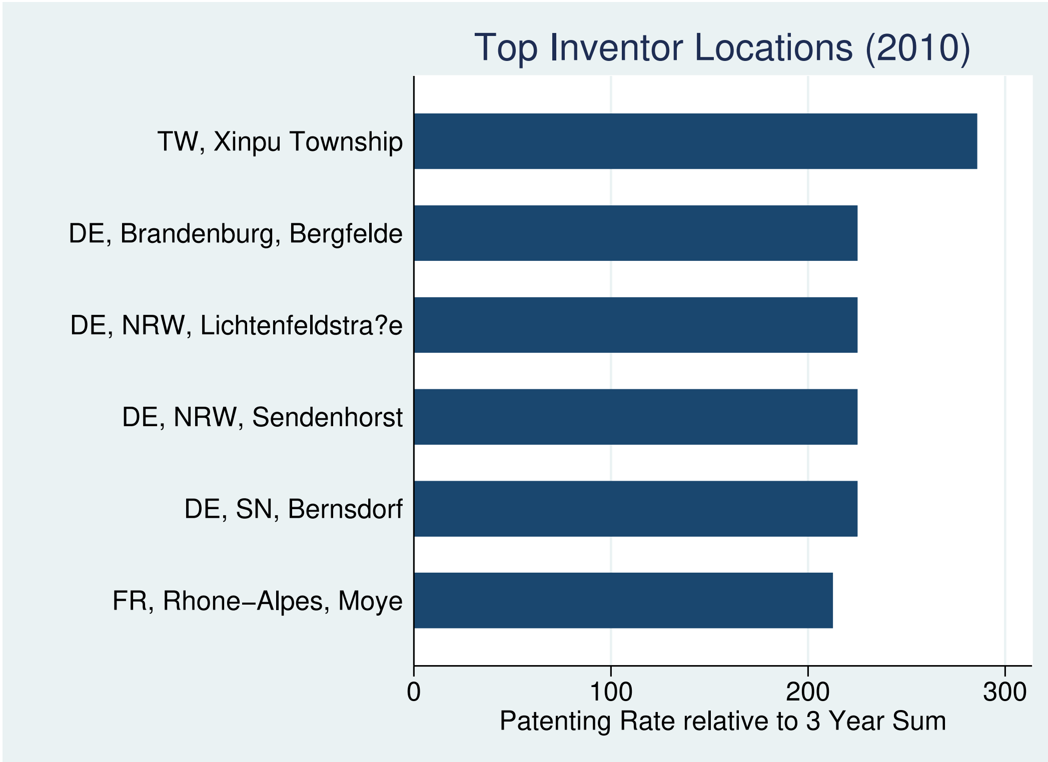

Figure 7: Top Inventor Locations

lated 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 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

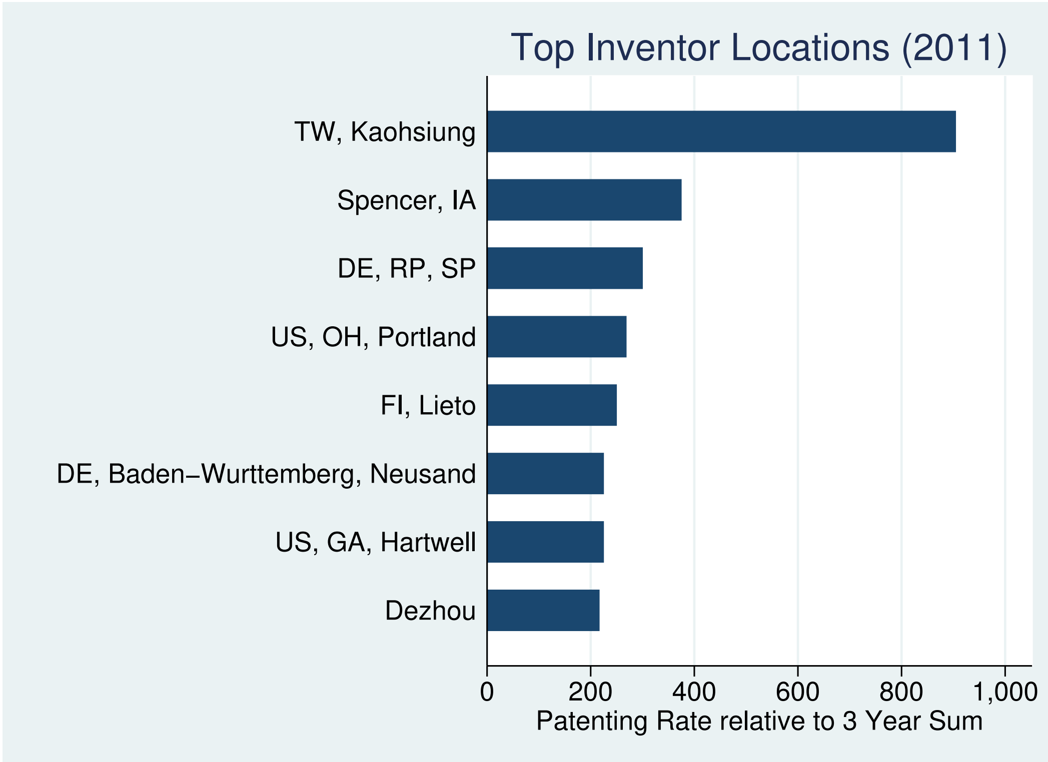

Figure 8: Top Inventor Locations

Table 2: 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

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

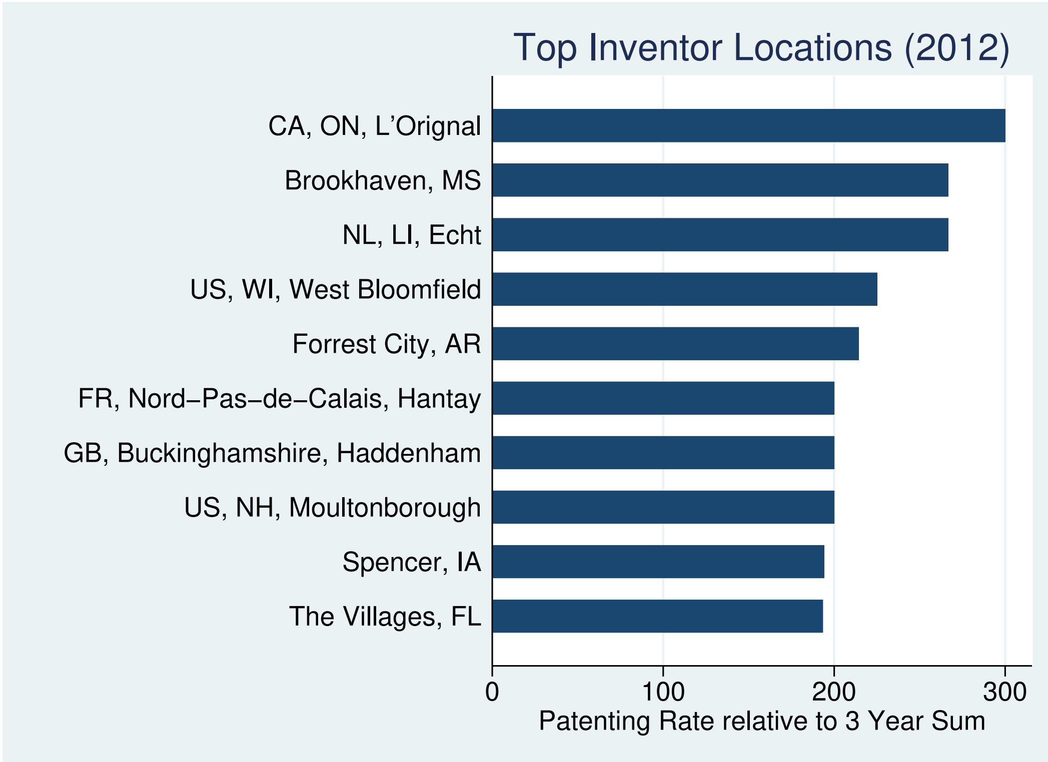

Figure 9: Top Inventor Locations

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 2 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

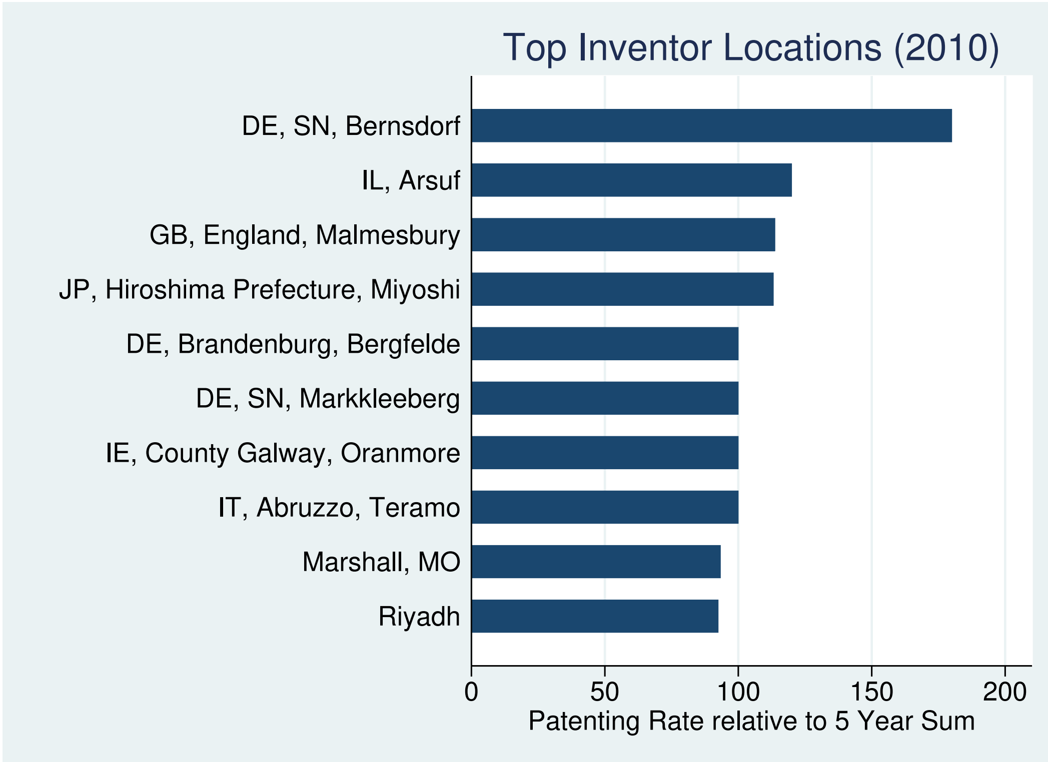

Figure 10: Top Inventor Locations

Table 3: 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

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 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.

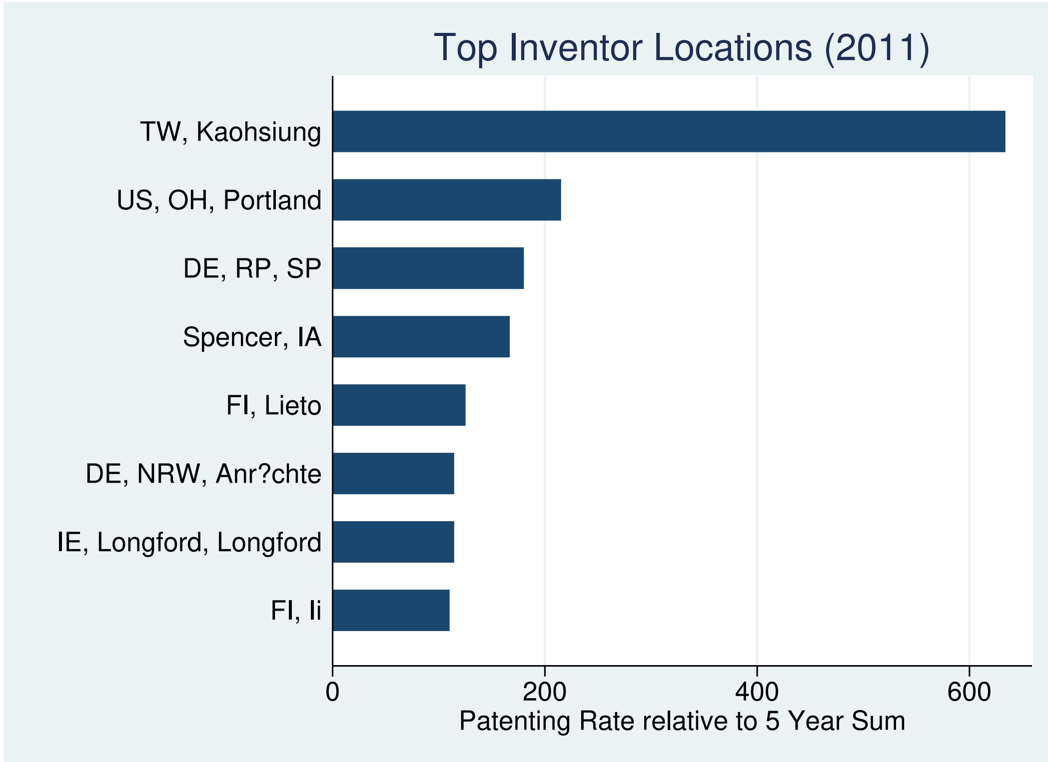

Figure 11: Top Inventor Locations

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 3 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

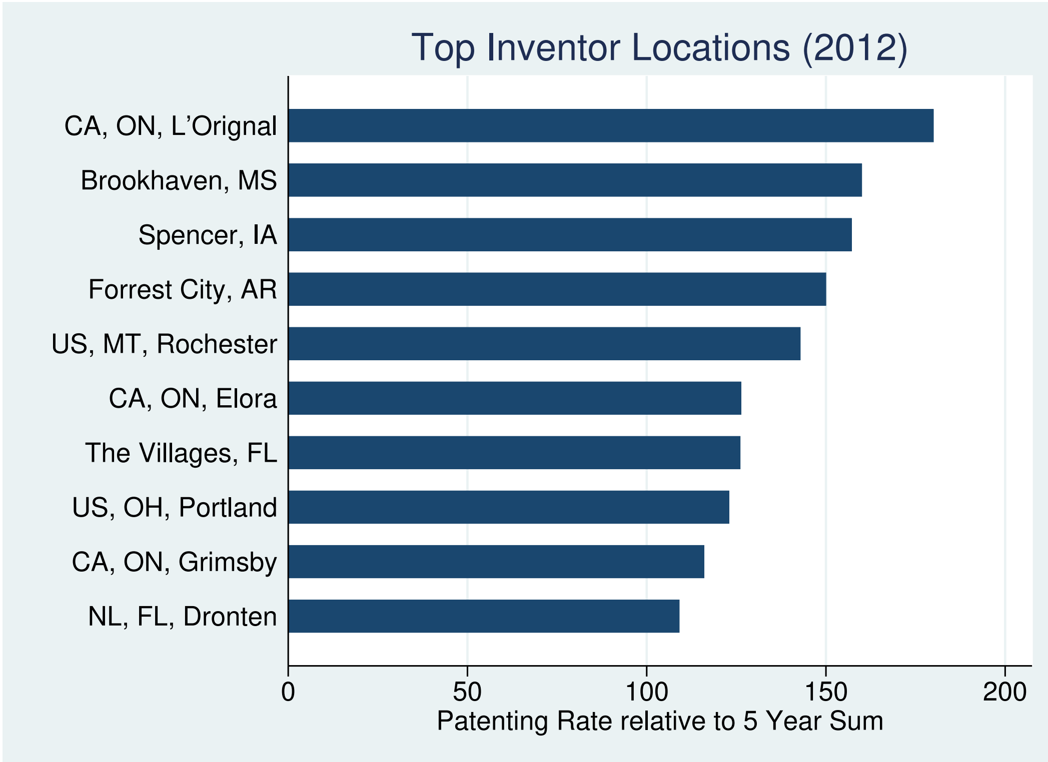

Figure 12: Top Inventor Locations

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 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.

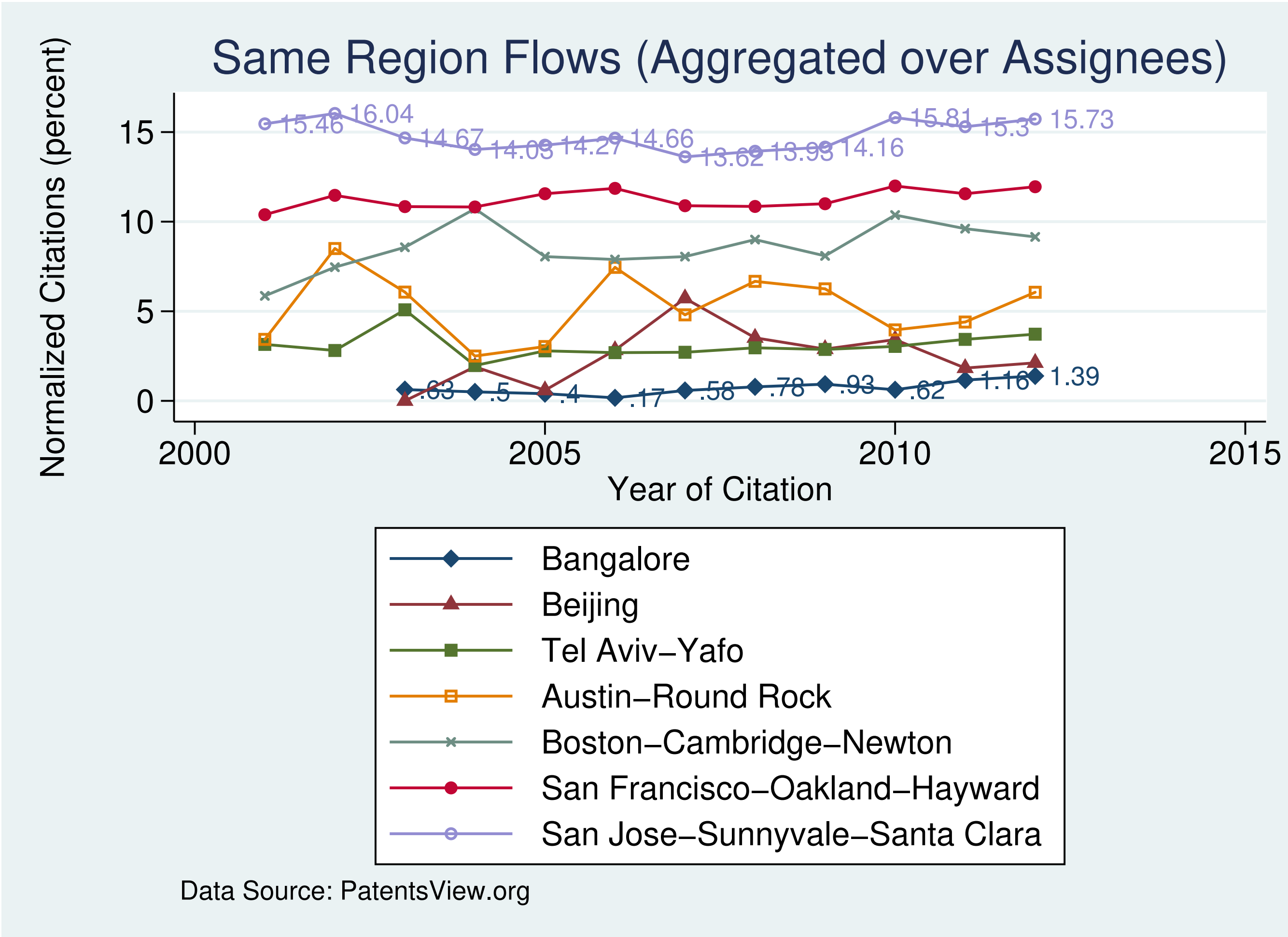

Figure 13: Local Knowledge Flows by Region

5 Preliminary Results

Figure 13 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 14, 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 increase while Beijing has somewhat stalled since 2005.

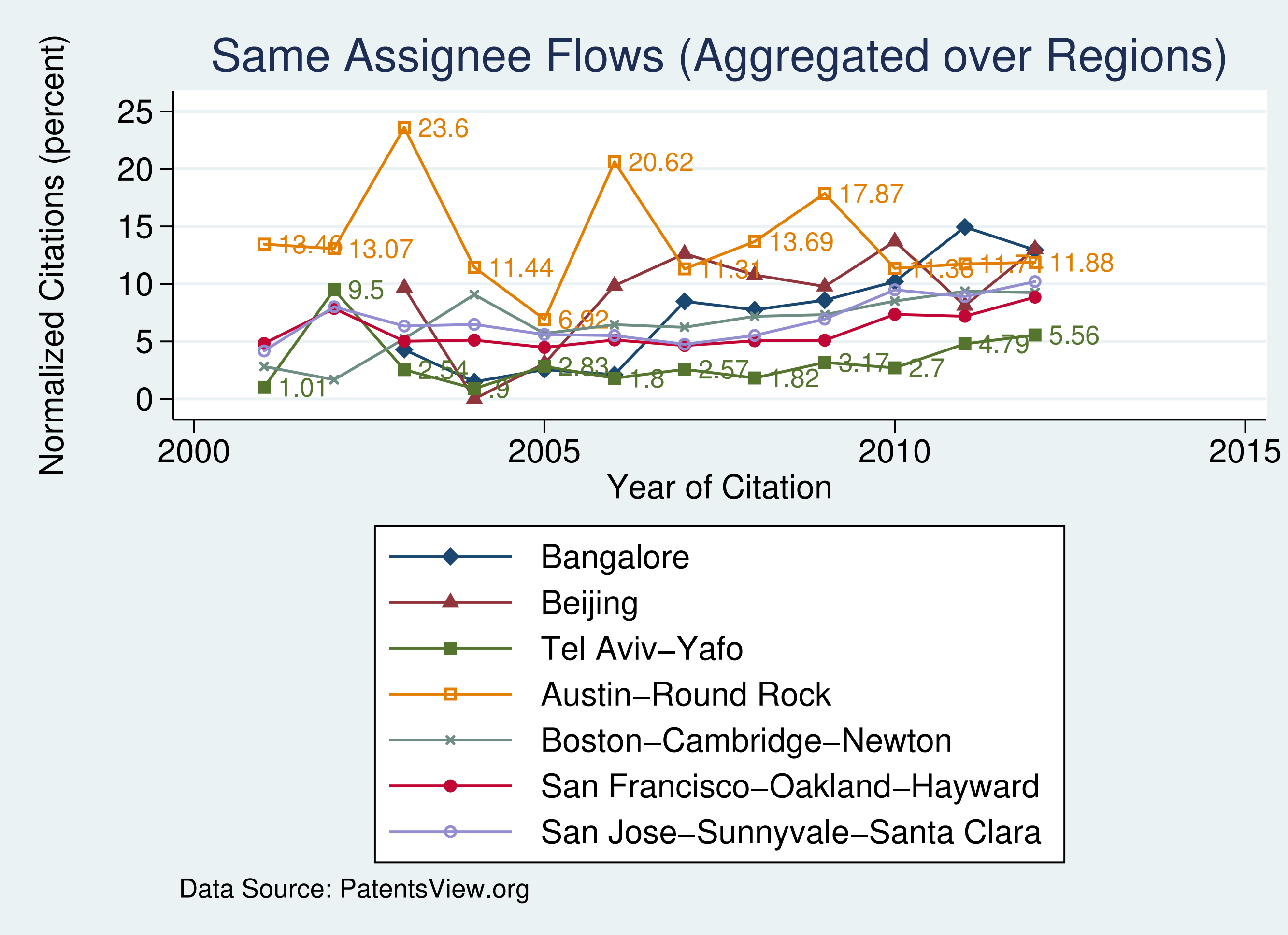

Figure 14: Internal Knowledge Flows by Region

Additionally, Figure 18 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 17 reiterates the Bangalore position as the outsourcing destination for global R&D.

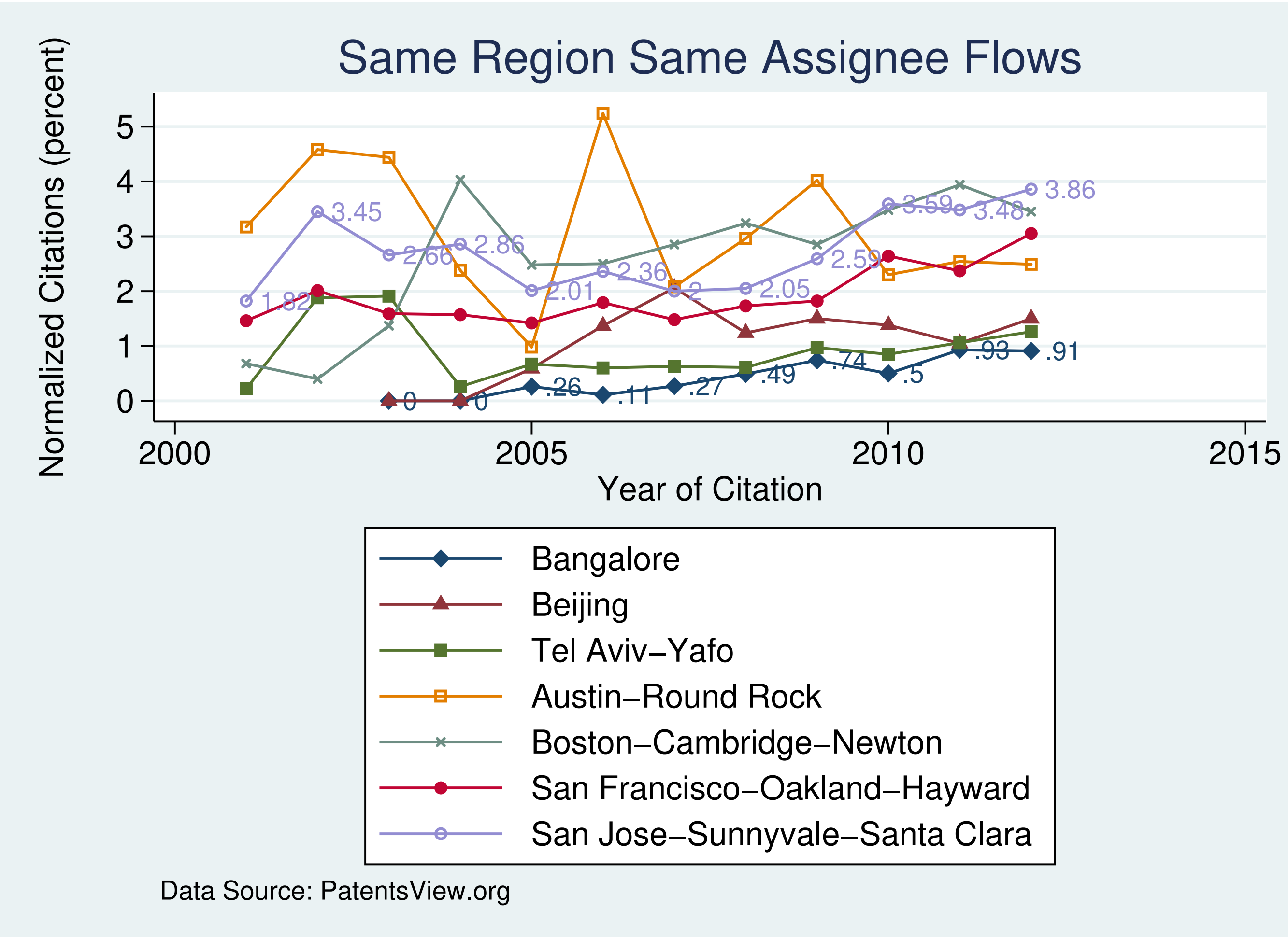

Figure 15: Local and Internal Flows by Region

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 were low to other firms locally. This study throws open opportunities to investigate specific mechanisms that might be leading to such behavior. While scholars

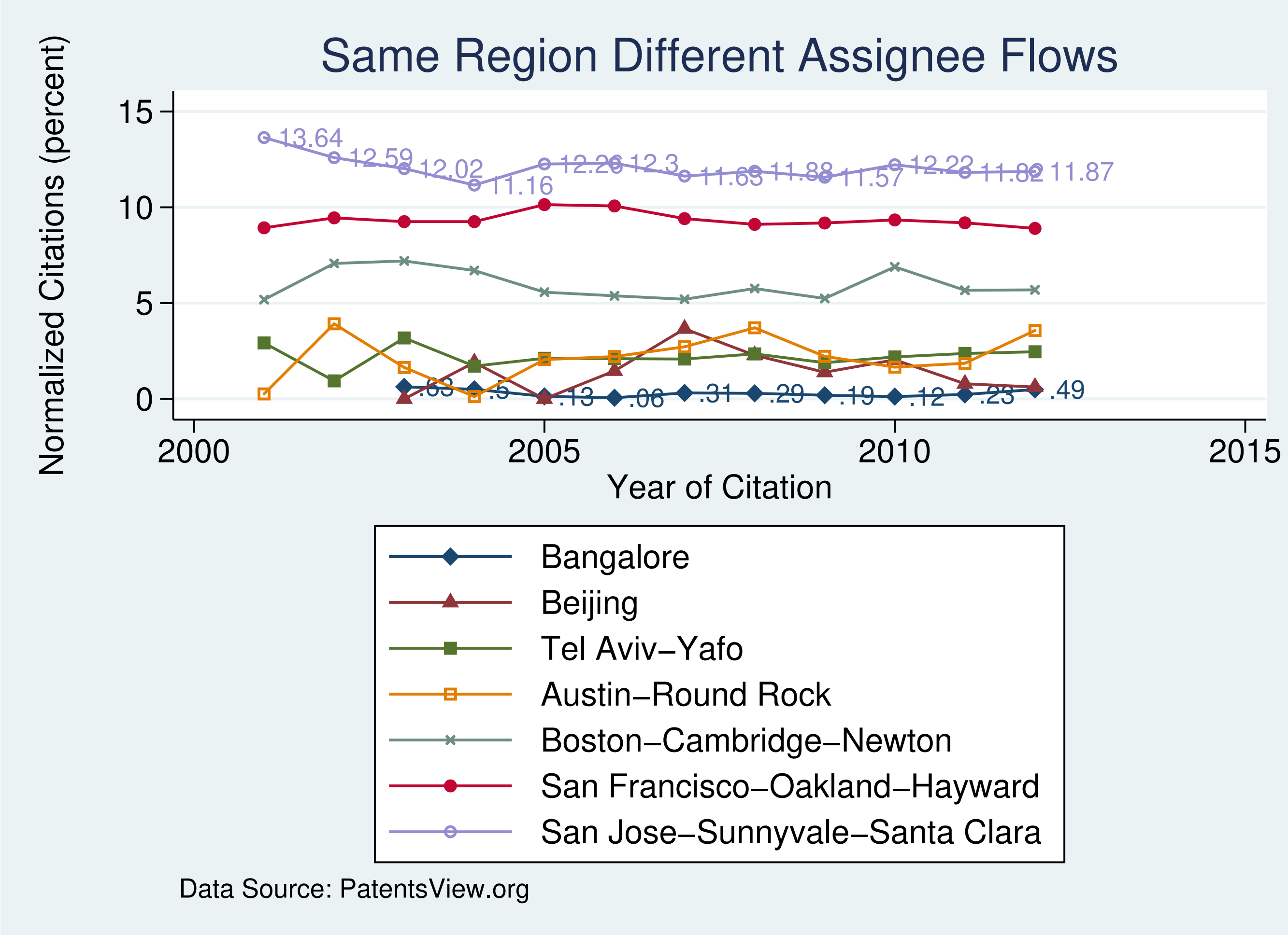

Figure 16: Local and External Flows by Region

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.

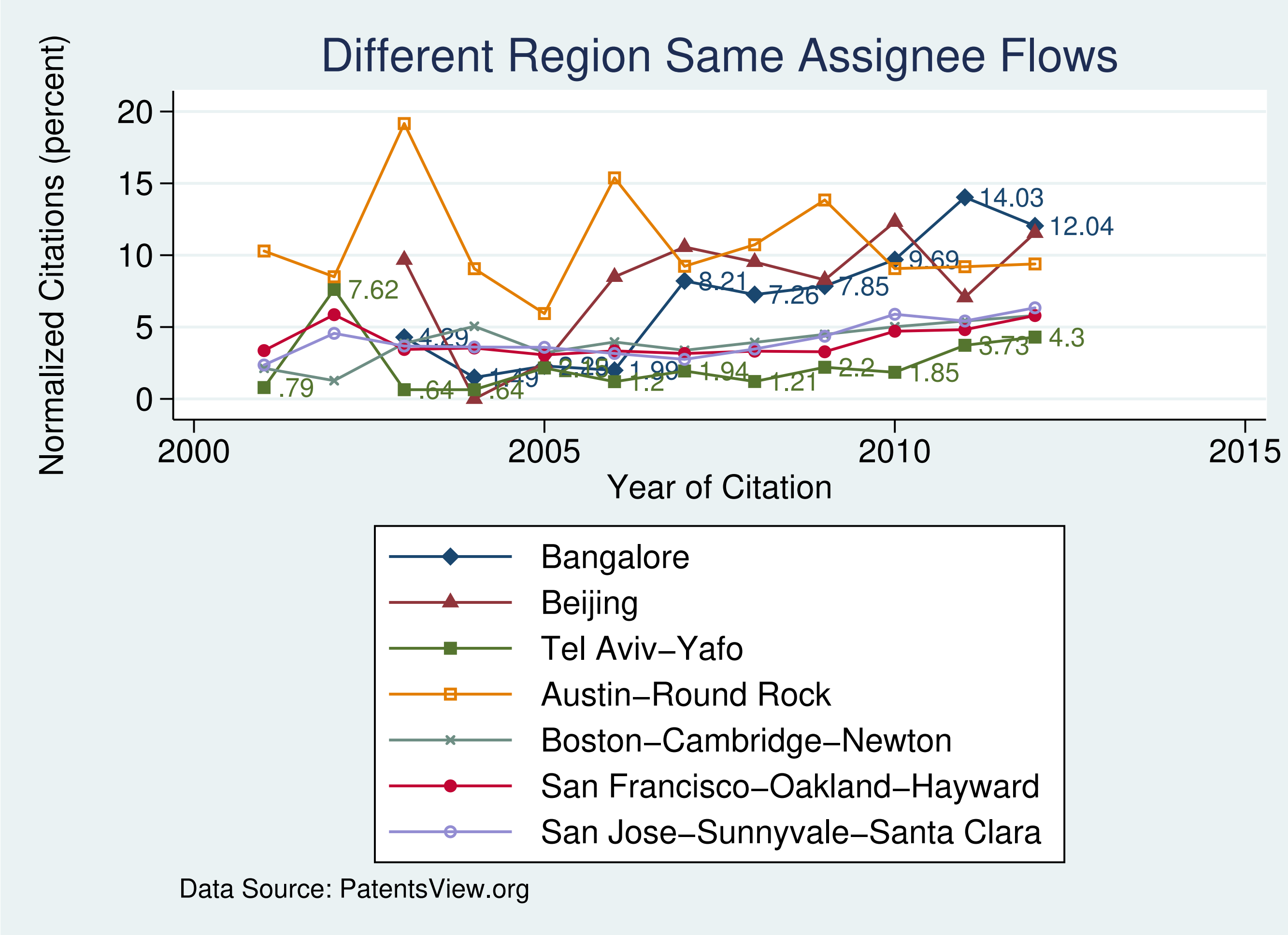

Figure 17: Non-local and Internal Flows by Region

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

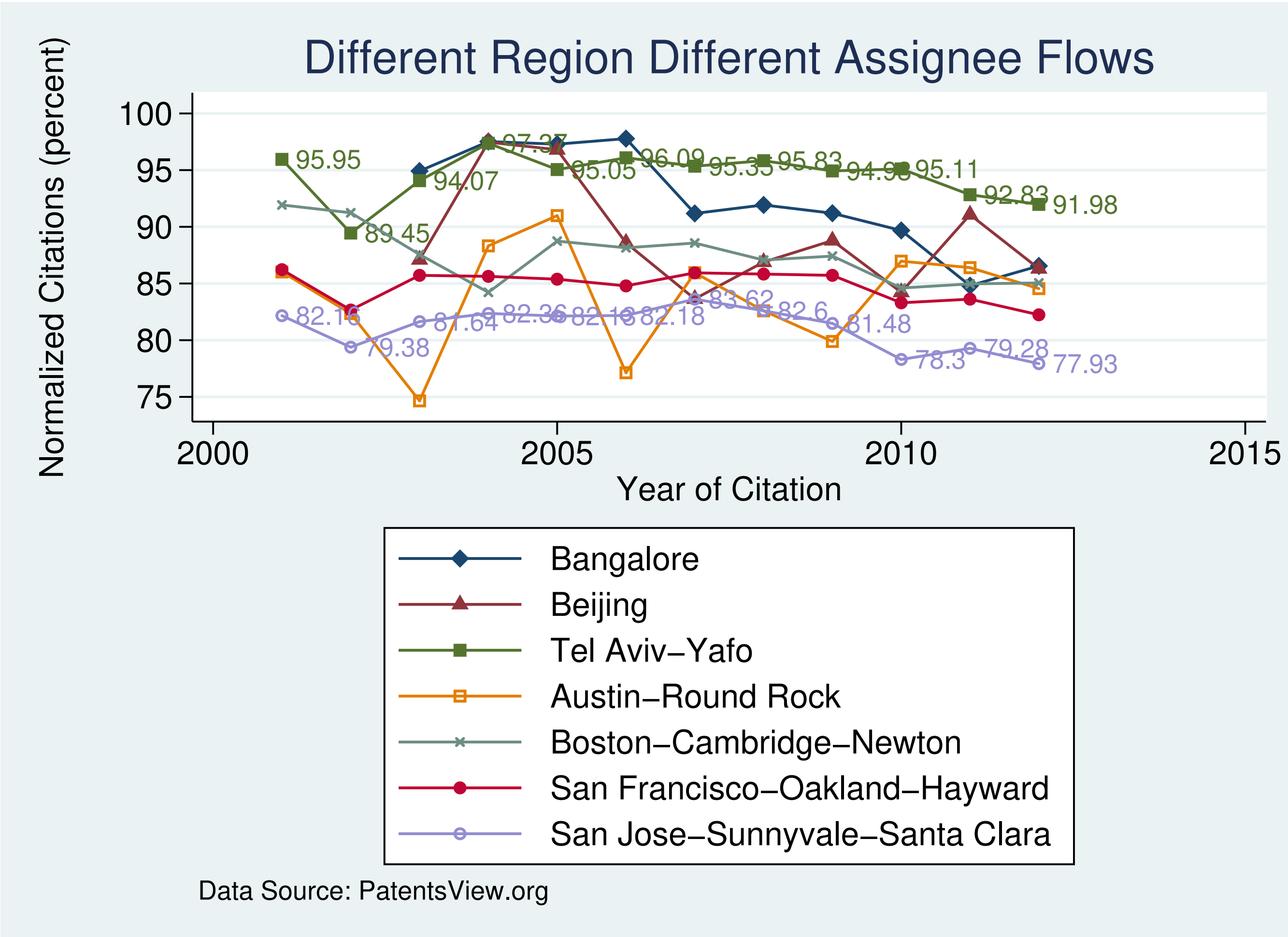

Figure 18: Non-local and External Flows by Region

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.

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.

Table 4: Countries and their IPR scores (Lesser, 2010)

Country	**IPR Score**
Afghanistan	
Albania	4.7682
Algeria	2.7608
Angola	1.8734
Anguilla	
Antigua and Barbuda	
Argentina	5.4684
Armenia	4.4032
Aruba	
Australia	11.1872
Austria	9.4024
Azerbaijan	3.1358
Bahamas	
Bahrain	5.7736
Bangladesh	2.3664
Barbados	
Belarus	3.2344
Belgium	9.6096
Belize	

Country	**IPR Score**
Benin	
Bermuda	
Bhutan	4.9300
Bolivia	4.2752
Bosnia and Herzegovina	2.9580
Botswana	6.2666
Brazil	5.2612
British Virgin Islands	
Brunei Darussalam	5.4230
Bulgaria	5.3598
Burkina Faso	3.5496
Burma	
Cambodia	1.9720
Cameroon	2.1692
Canada	11.1872
Cayman Islands	
Central African Republic	1.9720
Chad	1.5776
Chile	9.2152
China	6.1586
Colombia	6.2572
Congo	1.8734
Cook Islands	
Costa Rica	6.8388
Cote d'Ivoire	2.0706
Croatia	5.8528
Cuba	
Cyprus	7.2526
Czech Republic	6.4444
Democratic Republic of the Congo	3.8260
Denmark	11.7788
Djibouti	
Dominica	
Dominican Republic	
Ecuador	3.7822
Egypt	2.7608
El Salvador	3.3524
Equatorial Guinea	
Estonia	9.1166
Ethiopia	2.6622
Fiji	
Finland	11.3844
France	10.3984
French Guiana	10.3984
Gabon	2.8594

Country	IPR Score
Gambia	2.8594
Georgia	4.9106
Germany	10.4970
Ghana	4.5904
Greece	5.4878
Greenland	
Guadeloupe	
Guam	
Guatemala	3.3524
Guernsey	
Guinea	1.7748
Guinea-Bissau	
Guyana	2.5636
Haiti	1.7748
Honduras	3.2100
Hong Kong	8.0852
Hungary	7.6376
Iceland	10.1912
India	4.0974
Indonesia	4.5018
Iran (Islamic Republic of)	1.7748
Iraq	1.4790
Ireland	9.6290
Isle of Man	
Israel	8.6236
Italy	6.8488
Jamaica	2.9580
Japan	10.2012
Jersey	
Jordan	6.5430
Kazakhstan	2.6622
Kenya	3.7822
Korea, Democratic Republic of	
Korea, Republic of	7.1640
Kuwait	4.0426
Kyrgyzstan	3.4864
Lao People's Democratic Republic	1.9720
Latvia	6.0500
Lebanon	2.4650
Lesotho	
Liberia	3.0566
Libyan Arab Jamahiriya	
Liechtenstein	
Lithuania	7.4404
Luxembourg	8.8302

Country	IPR Score
Macau	
Madagascar	2.9580
Malawi	3.2538
Malaysia	5.1820
Mali	2.7608
Malta	
Mauritania	2.4650
Mauritius	5.3244
Mexico	4.8668
Monaco	
Mongolia	3.4072
Montenegro	
Morocco	5.8628
Mozambique	2.4650
Namibia	4.4370
Nepal	2.2678
Netherlands Antilles	11.3844
Netherlands	11.3844
New Caledonia	
New Zealand	11.8774
Nicaragua	5.0740
Niger	2.8594
Nigeria	3.2100
Northern Mariana Islands	
Norway	10.1912
Oman	7.0360
Pakistan	4.1074
Palau	
Palestine	
Panama	5.2164
Papua New Guinea	2.0706
Paraguay	3.6836
Peru	5.3892
Philippines	4.1074
Poland	7.5390
Portugal	8.3278
Puerto Rico	
Qatar	7.6470
Republic of Moldova	4.1218
Reunion	
Romania	6.3558
Russia	4.0332
Saint Barthelemy	
Saint Kitts and Nevis	
Saint Lucia	

Country	IPR Score
Saint Pierre and Miquelon	
San Marino	
Saudi Arabia	4.2398
Senegal	2.9580
Serbia	4.4470
Seychelles	
Sierra Leone	2.1692
Singapore	11.6802
Slovakia	7.0460
Slovenia	8.3716
Solomon Islands	
South Africa	7.2432
Spain	8.6236
Sri Lanka	3.0566
Sudan	1.4790
Suriname	3.6482
Svalbard	
Swaziland	4.2946
Sweden	11.6802
Switzerland	11.4830
Syrian Arab Republic	3.5596
Taiwan	7.2626
Tajikistan	1.9720
Thailand	4.0974
The former Yugoslav Republic of Macedonia	
Togo	2.7608
Trinidad and Tobago	5.1626
Tunisia	5.8528
Turkey	6.9474
Turkmenistan	
Turks and Caicos Islands	
Uganda	2.4650
Ukraine	3.7822
United Arab Emirates	6.4090
United Kingdom	10.2012
United Republic of Tanzania	2.5636
United States Virgin Islands	10.0040
United States	10.0040
Uruguay	8.2192
Uzbekistan	3.6388
Venezuela	3.6144
Vietnam	4.2752
Yemen	2.0706
Zambia	2.9580
Zimbabwe	2.9142

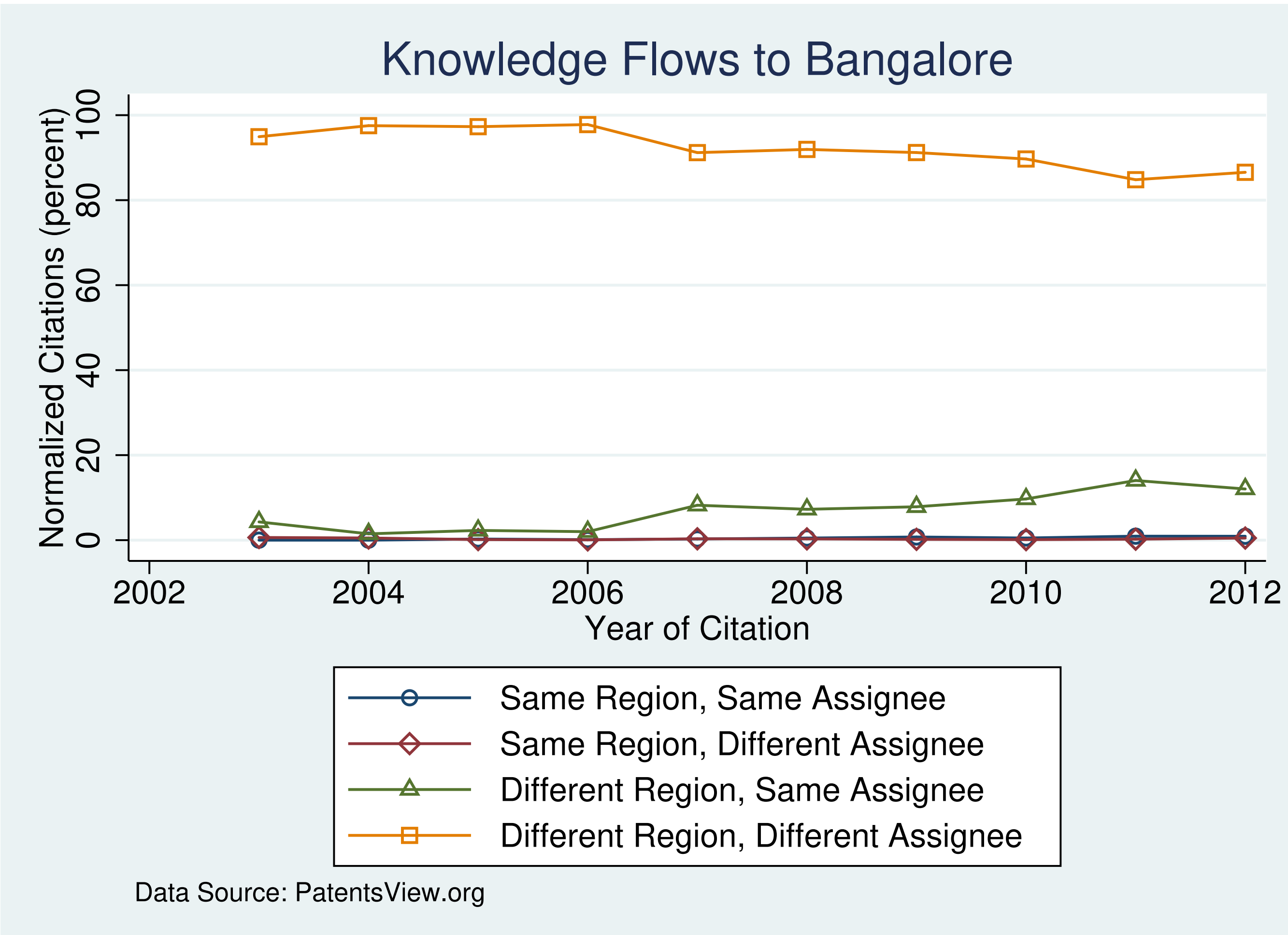

Figure 19: Relative Flows by Region : Bangalore

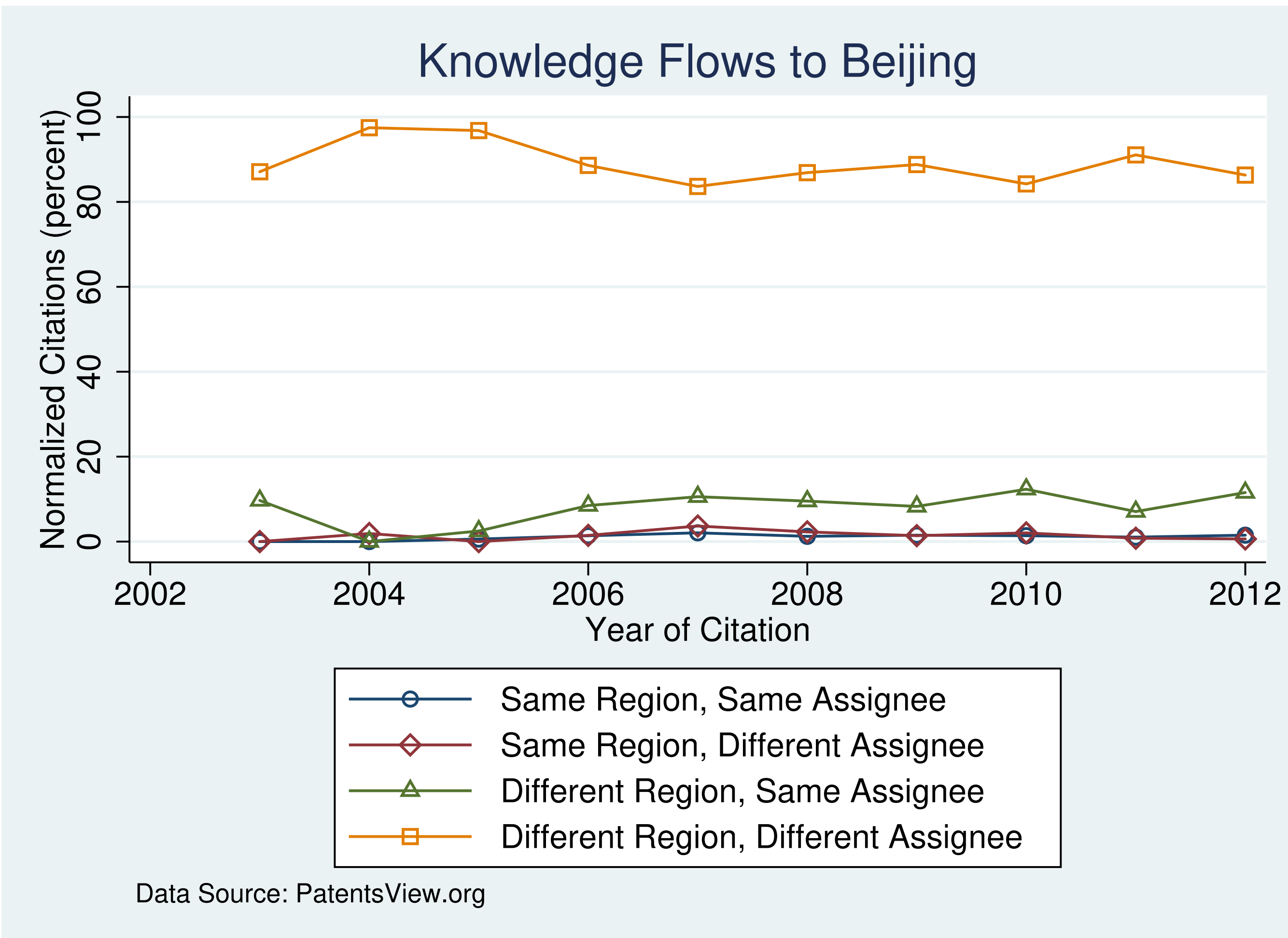

Figure 20: Relative Flows by Region : Beijing

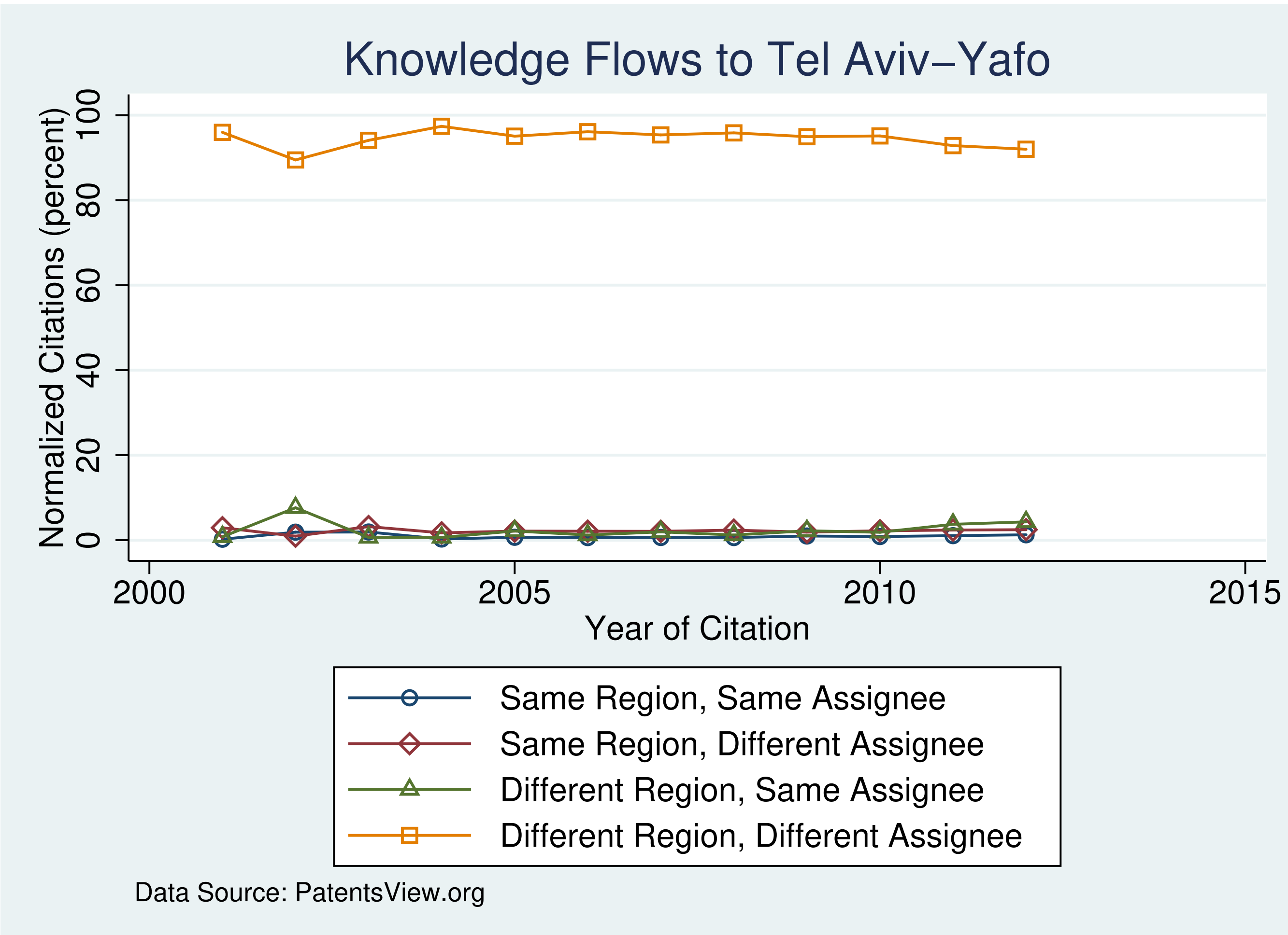

Figure 21: Relative Flows by Region : Tel Aviv-Yafo

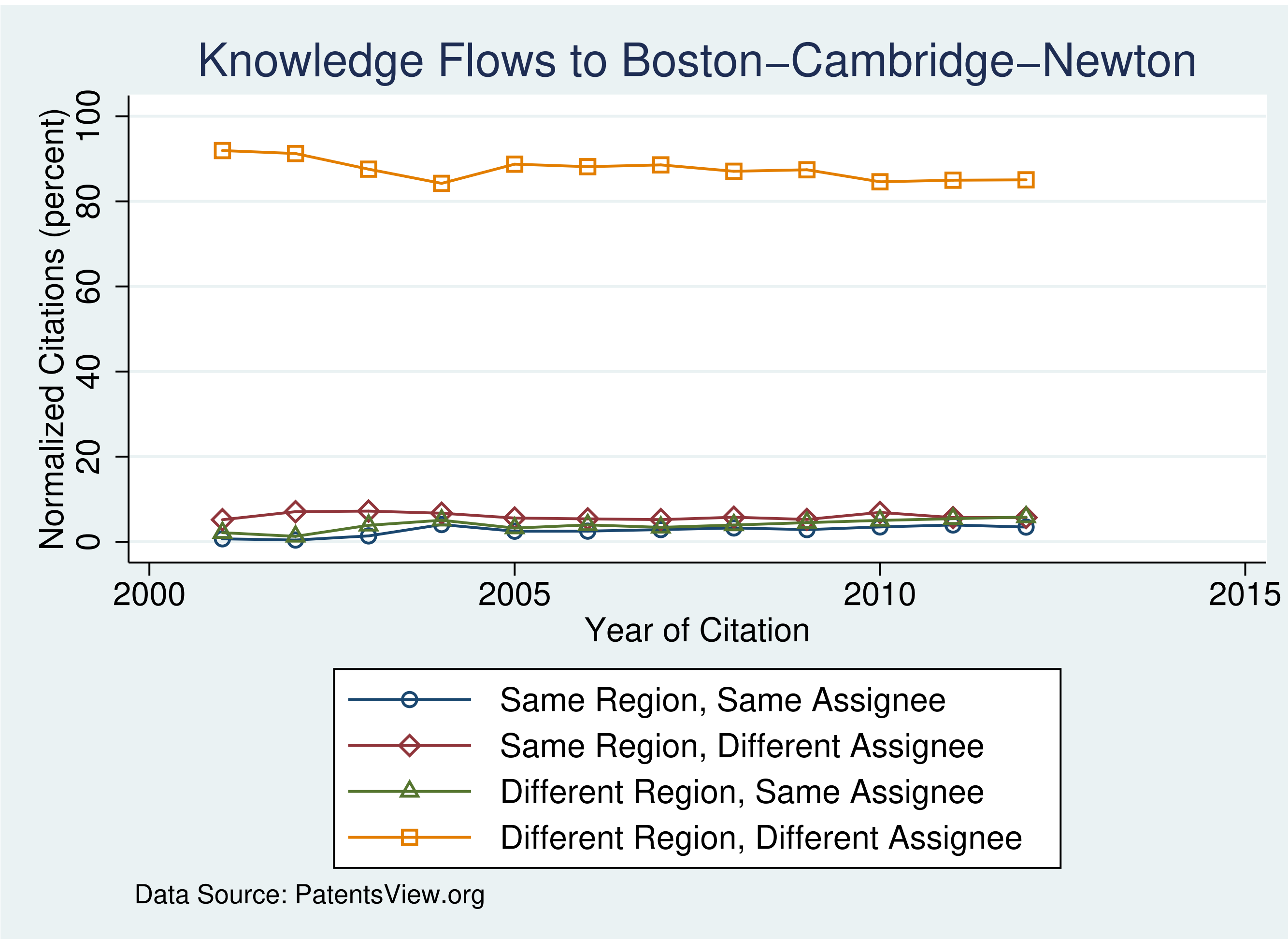

Figure 22: Relative Flows by Region : Boston-Cambridge-Newton

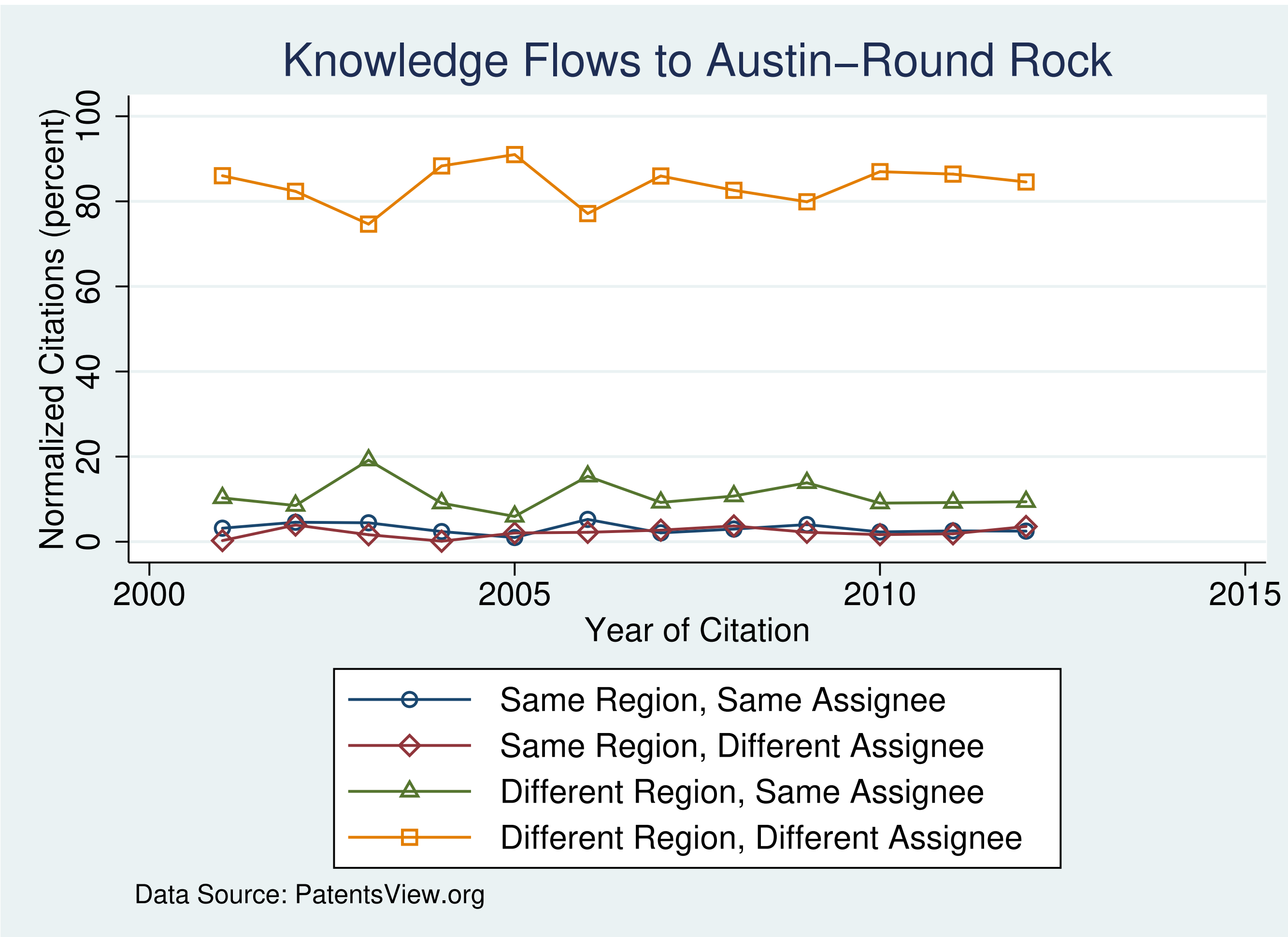

Figure 23: Relative Flows by Region : Austin-Round Rock

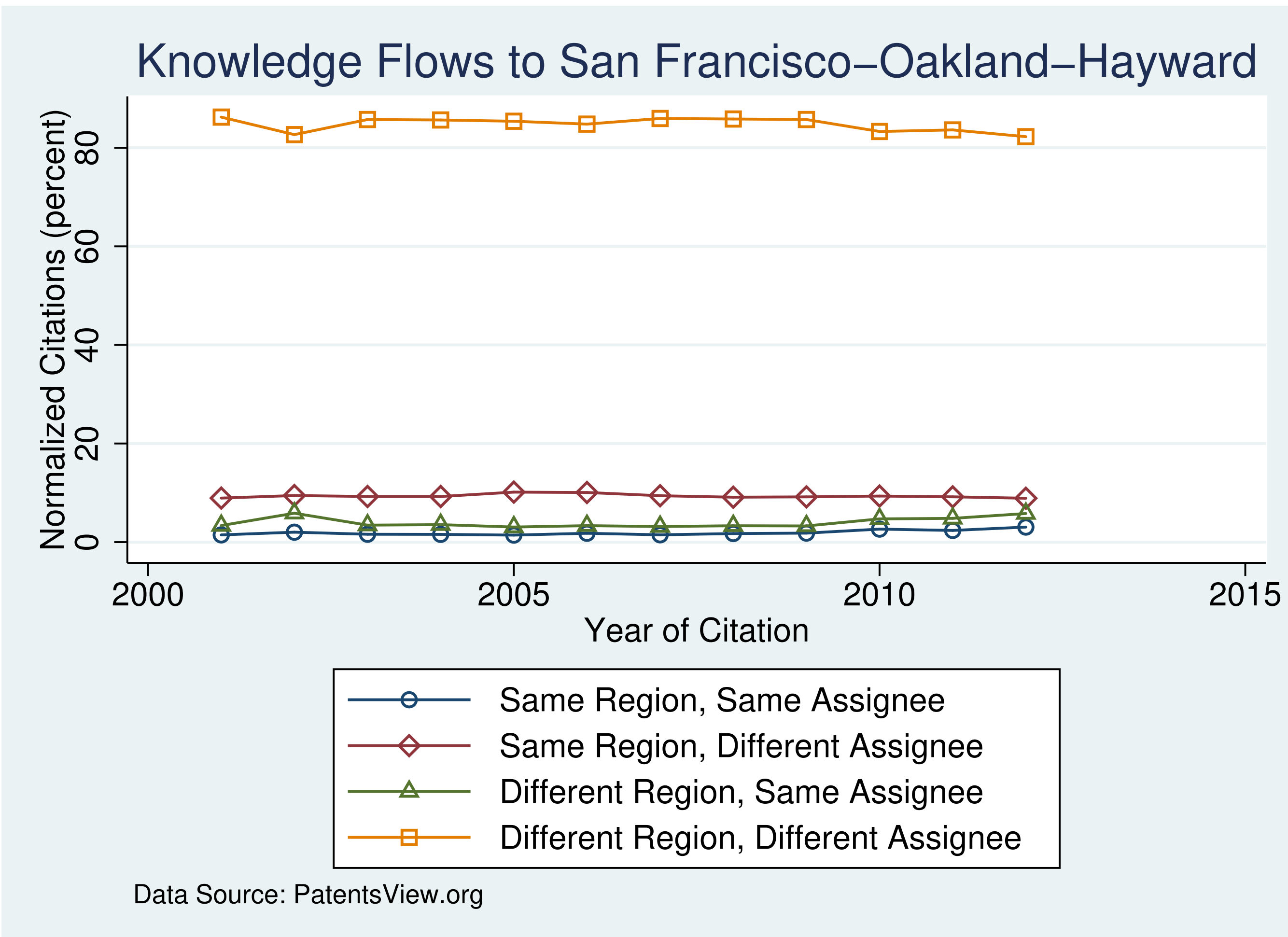

Figure 24: Relative Flows by Region : San Francisco-Oakland-Hayward

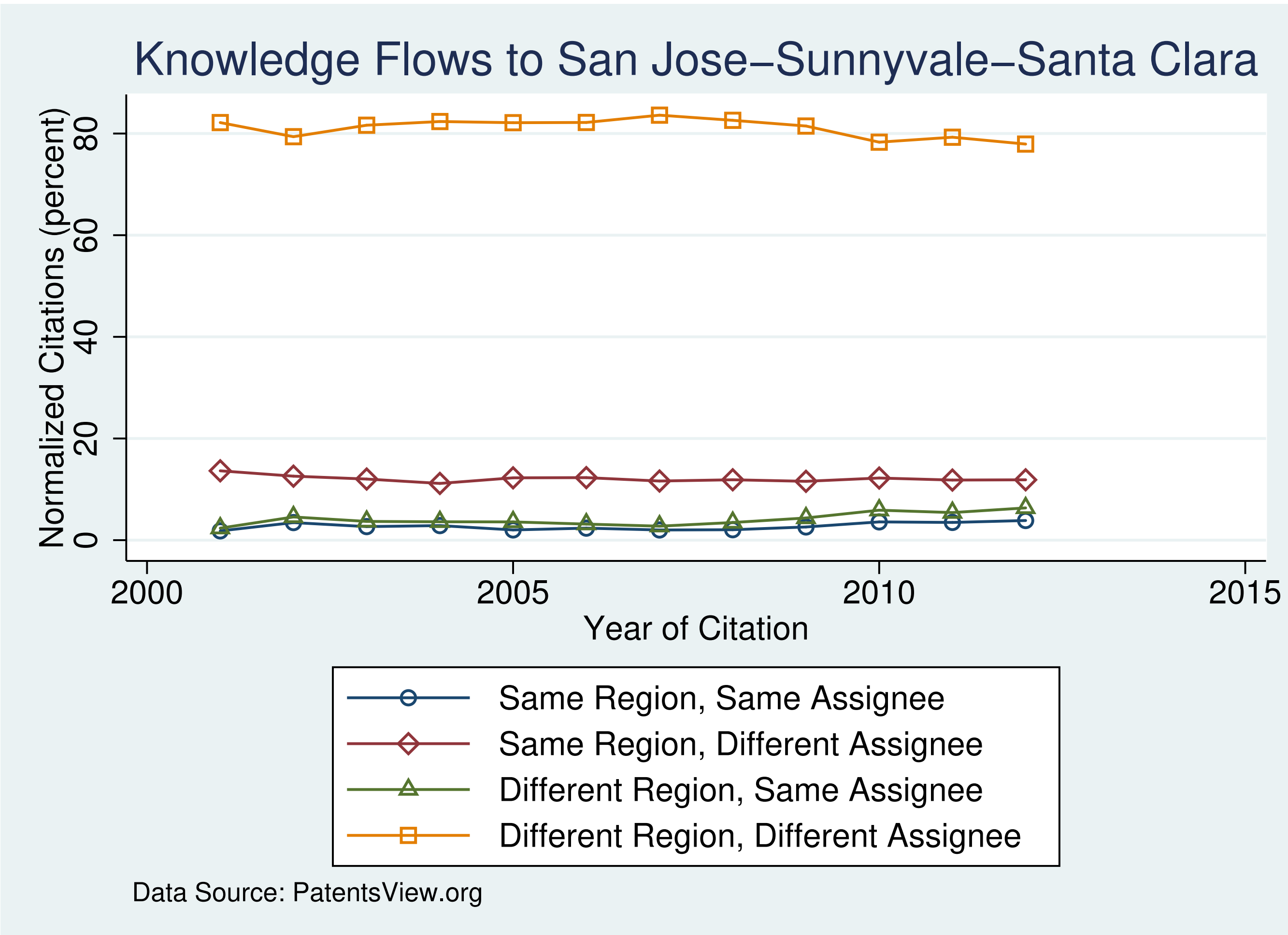

Figure 25: Relative Flows by Region : San Jose-Sunnyvale-Santa Clara