Heterogeneity in knowledge spillovers across regions:
The effects of endogeneity, complexity and IPR environment

Ashwin Iyenggar (1521001)
ashwin.iyenggar15@iimb.ernet.in

January 8, 2017

Abstract

I line up empirical evidence to demonstrate the heterogeneity in the geographical distribution of knowledge spillovers across various regions. I then explore three potential mechanisms that may help explain this heterogeneity: endogenous aspects of the regions themselves, complexity of work, and the intellectual property rights environment of the location. While the empirical results are yet inconclusive and incomplete, the current work extends prior work on geographic spillovers of knowledge by integrating three hitherto alternative explanations.

Keywords: Knowledge Spillovers, Endogeniety, Complexity, IPR

1 Introduction

There has been a long and illustrious scholarly tradition highlighting the agglomeration characteristics of economic regions, going back at least as far as Marshall (2009), whose original work was published in 1890. More recently, scholars over the last three decades have demonstrated the paper trail of these knowledge spillovers through the study of patent citations (e.g., Almeida and Kogut (1999); Jaffe et al. (1993)). This tradition of scholarship has further shaped our theoretical understanding of knowledge spillovers through mechanisms such as the effects of inventor mobility (e.g., Almeida and Kogut (1999)), differential Intellectual Property Rights environments across locations (e.g., Zhao (2006)) and of the role of international geography (e.g., Singh (2007)). The nature and extent of the geographical distribution of knowledge spillovers observed in practice is so highly heterogenous across locations, firms and legal environments, that the understanding of the causal mechanisms leading to knowledge spillovers continues to intrigue the best of scholars. While this is, in no way dismissive of the enormous theoretical strides so far, the question is assumes greater significance in the environment surrounding the second machine age as some scholars have begun to highlight (McAfee and Brynjolfsson, 2014)

Motivated by empirical evidence surrounding the heterogeneity in the nature of knowledge flows across the various regions, I intend to explore the three mechanisms ostensibly influencing knowledge spillovers. Complexity of patents invented as a potential mechanism influencing the extent of local knowledge spillovers. This approach is not to be construed as yet another mechanical departure from the current theory on spillovers. I argue so with the following reasons. First, from a human capital perspective, it is valuable to understand the impact of MNCs that dominate much of the cutting- and bleeding-edge innovation in emerging markets on the development 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? Second, a specific flavor of this question is the investigation of the spillover effects of the innovation process in emerging countries, or those known to have weaker IPR regimes. Specifically, do multinational firms that develop patentable technologies in emerging (or weaker IPR) countries create spillover effects in the host country talent pool? Or do the benefits remain localized to within multinational companies (MNCs) and their home country employees? Finally, the wide disparity in the extent of knowledge spillovers across locations, across firms and across IPR regimes is intriguing to the researcher and calls attention toward a creative response. a researcher to find the mechanisms that may lie behind such a phenomenon. Patents data allows us to ask these questions and to have them answered as has been in the tradition of Jaffe et al. (1993).

The choice of the three explanatory mechanisms is not arbitrary. Indeed there has been a tradition of scholarly work in each of them[1]. First, several studies including Almeida and Kogut (1999) have conclusively demonstrated that the kind and extent of knowledge flows between firms in Silicon Valley is unparalleled in the rest of the world. Indeed, our analysis on a select chosen locations demonstrates this adequately in Figure 1 where the lines in purple and red stand out (Silicon Valley, or the broader San Francisco Bay Area is classified into two Metropolitan Statistical Areas (MSAs) by the United States census). Second, as evidenced by the respective scholarly traditions of Baldwin and Henkel (2015), Ethiraj and Levinthal (2004) and Yayavaram and Ahuja (2008), the complexity, intellectual property and organizational implications have been addressed

[1] I am however, unable to go in much of those details in the current article

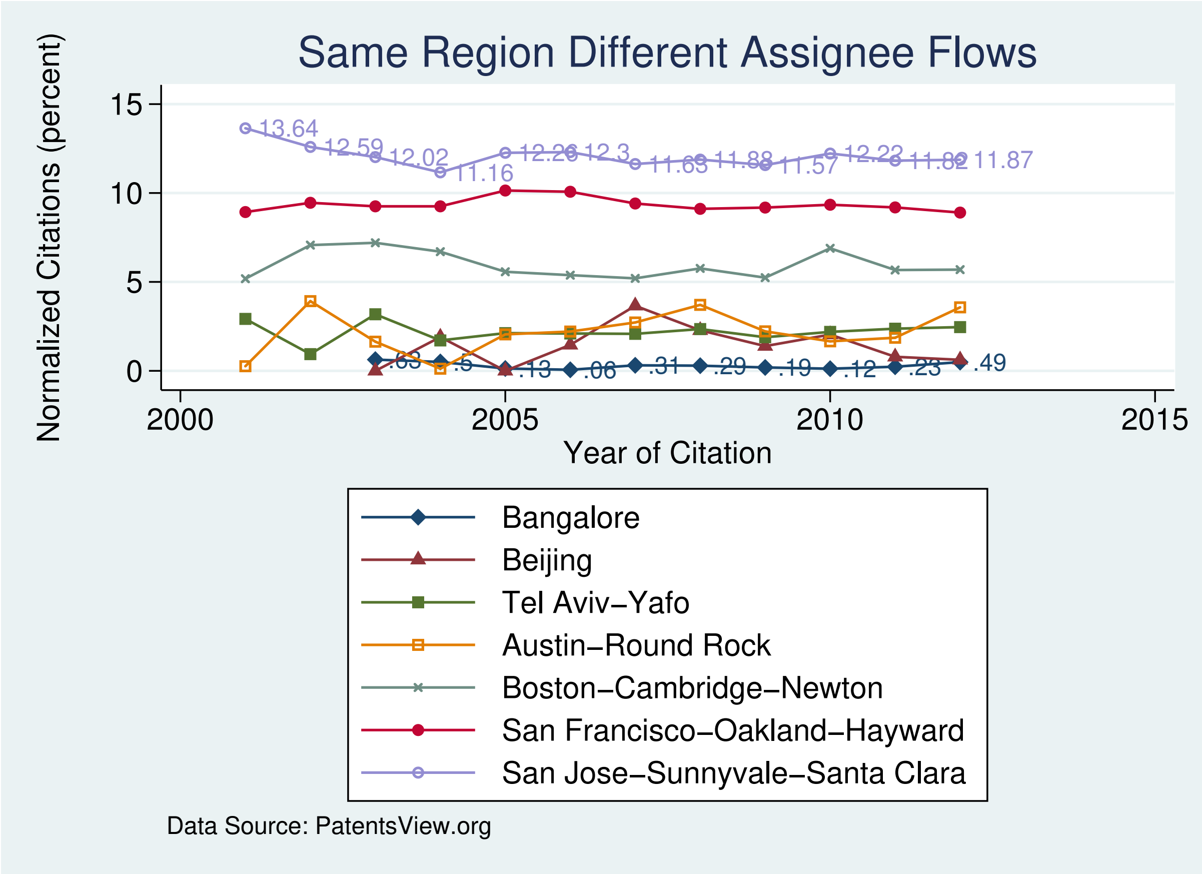

Figure 1: Local and External Flows by Region

by scholars in the context of patenting. In the spirit of Ethiraj and Levinthal (2004), I propose a definition of complexity that is rooted in the question of knowledge spillovers. Specifically, I suggest that complexity may be seen as either an attribute of usage, or as an attribute of invention. A patent that is used (cited) by several patents belonging to distinct and different patent technology classes maybe seen as modular by virtue of it being able to be plugged into multiple, diverse applications. Alternatively, a patent that is constructed with few dependencies may also be seen as being modular by virtue of its capacity to be developed standalone, or with minimal intervention from other modules. For the purposes of this study, I use a definition of Complexity that captures both the effects above. Finally, the scholarly tradition in the international business area has extensively analyzed the relationship between economic geography Singh (2007), intellectual property environments (Zhao, 2006) and political geography (Singh and Marx, 2013).

The current work is placed at the confluence of these three traditions, with the focus on implications for beneficial knowledge spillovers. A second objective of the current work is to understand the local impact of inventing activity by multinationals in emerging nations. I attempt to answer the following questions. First, how does the nature of the geographic distribution of citations made by inventions from a region affect the quantum of citations received. Second, how do complexity of inventions and cross border differences in intellectual property environments affect the previous relationship.

The benefits of understanding geographic and multinational collaboration in invention is that we may seek to inform both managers and firms about the potential opportunities of tapping into or creating spillover effects in the host country talent pool. Does a significant group of local inventors who develop due to spillovers? Do they then move around to cross-pollinate to other firms? How do domestic firms integrate and appropriate rents in this context. There are all interesting and valuable directions spawned by the current approach.

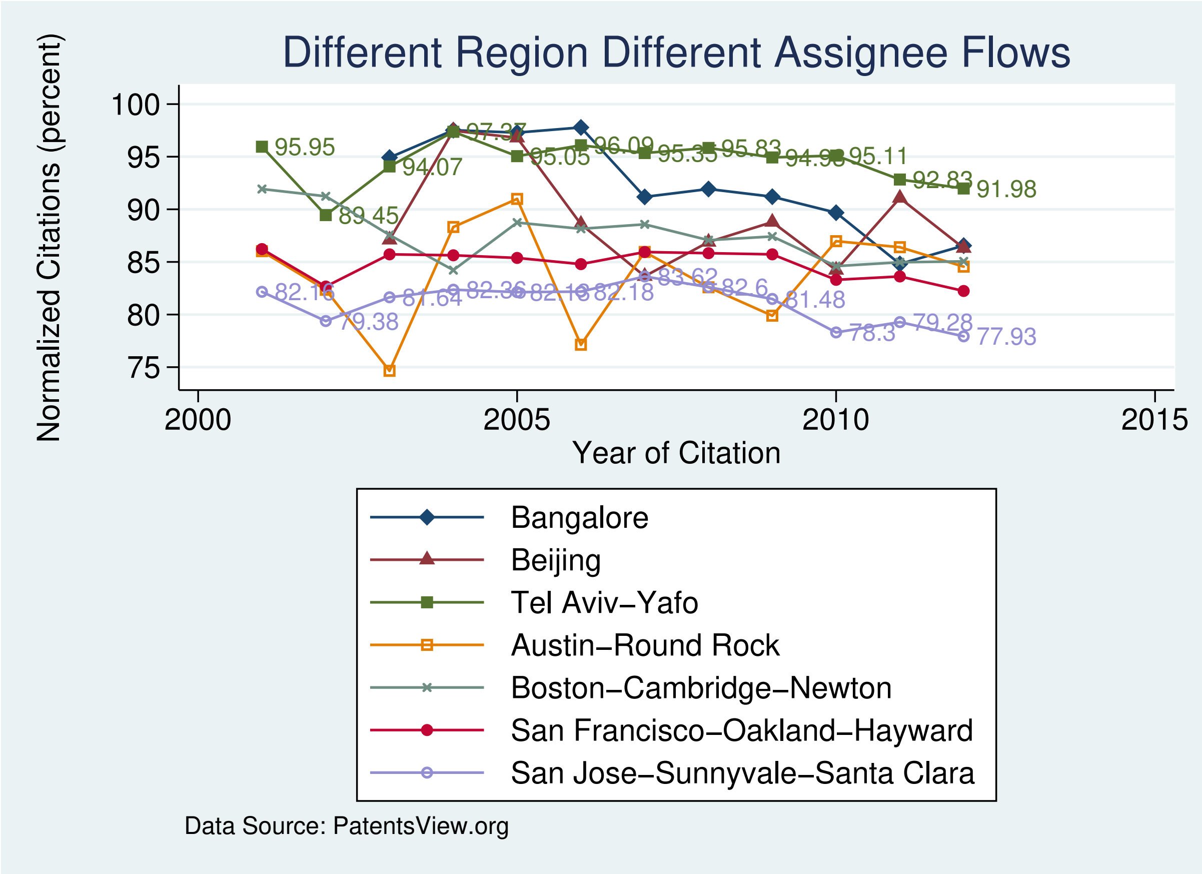

Figure 2: Non-local and External Flows by Region

2 Motivation

I motivate this study with a small sample analysis of the knowledge spillover characteristics of a selected sample of seven regions across continents, and technological specialization and IPR strength. In each of Figure 1, Figure 2, Figure 3 and Figure 4, citation counts of regions are expressed as a normalized percentage number (so as to be able to fairly compare across regions with vastly different pools of knowledge and inventors). Figure 1 already demonstrated that the northern California regions stood out in terms of the extent of local knowledge spillovers to other firms. Scholars have explained this using the mechanism of employee mobility (Almeida and Kogut, 1999). Figure 2, on the other hand demonstrates that Tel Aviv-Yafo in Israel stands out as accounting of the highest proportion of flows to external firms in external locations. The employee mobility explanation may not be able explain the phenomenon here. Figure 3 suggests that the Bangalore region sees very little local knowledge spillovers at all, while Figure 4 suggests that the Bangalore region flows are dominated by those to the multinational parent location. This rather wide disparity between a selection of inventing locations, provides us with the context to dive into the understanding the mechanisms that underlie this divergence in knowledge spillover patters across regions.

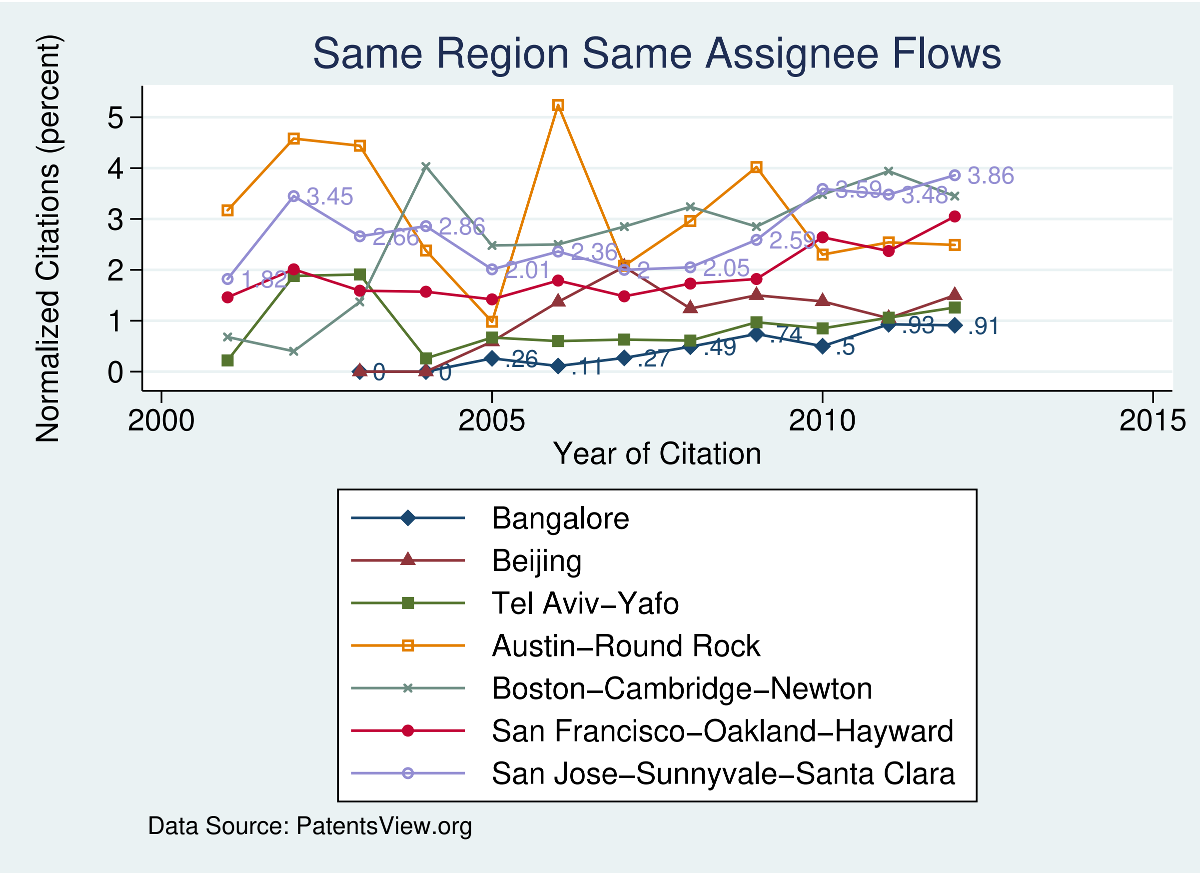

Figure 3: Local and Internal Flows by Region

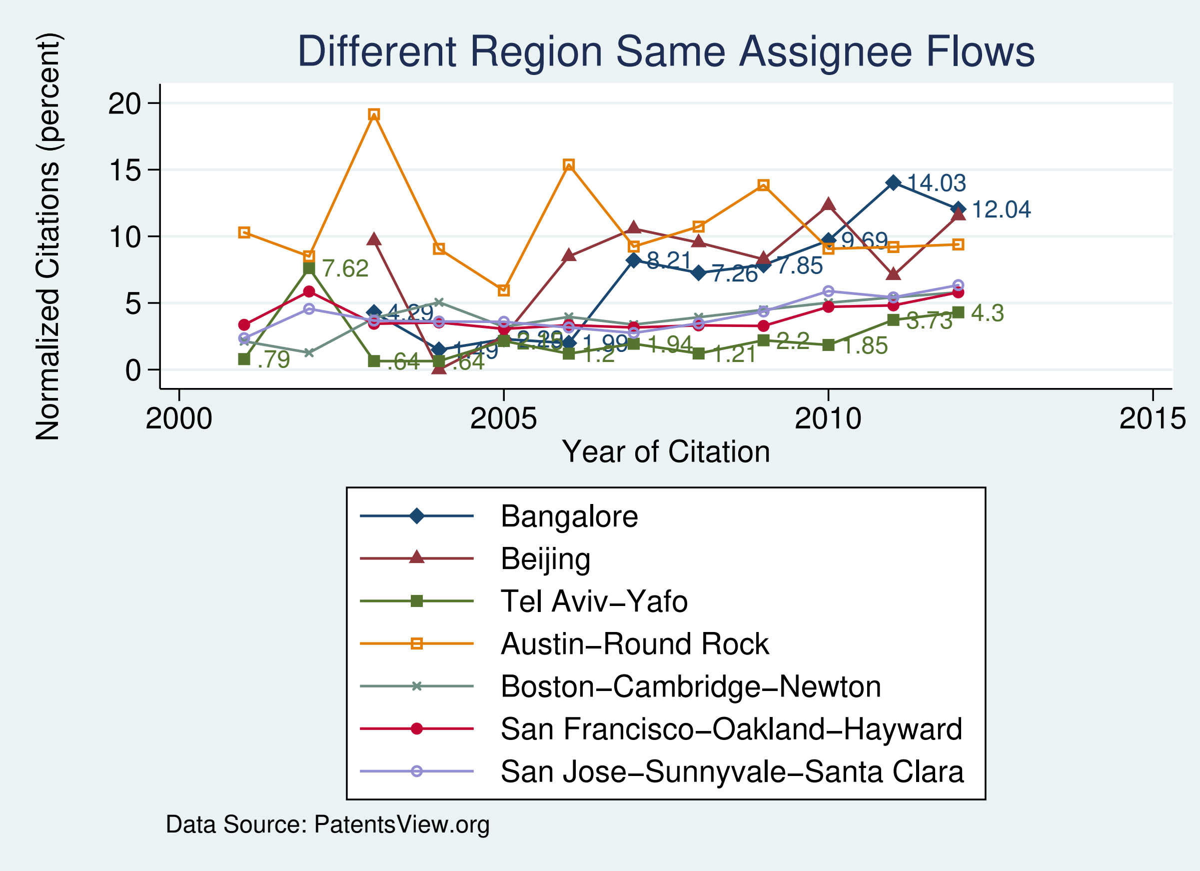

Figure 4: Non-local and Internal Flows by Region

3 Theory

Lay the hypotheses here

Hypothesis 1: I hope I can have something to go

4 Research Design

4.1 Complexity

I construct my measure of Complexity based interactions between the different patent sub-classes. Since each of the interactions between patent sub-classes may introduce a new interaction, I model interactions on a binomial function. Specifically, when `subclass` represents the number of distinct patent sub-classes, I define `interaction(subclass)` as follows:

$$interaction(subclass) = \begin{cases} 1 & : subclass \leq 2 \\ \binom{subclass}{2} & : subclass > 2 \end{cases}$$

I would expect, from a user perspective that the more number of contexts in which the patent is valuable, the higher should be the Complexity. If `Complexity` represents my measure of the Complexity of the patent, and `usage contexts` represents the number of distinct contexts where the patent is found valuable, I should expect the following relationship to hold:

$$\texttt{Complexity} \propto \frac{1}{usage\ contexts}$$

Similarly, from an inventor perspective, the more the number of contexts that the patent is built on, the lower should be the Complexity. A patent that is developed without citing any other patents is an extreme case of highest Complexity, while one that requires to be built upon several `source contexts` is properly understood as being less modular. The relationship between `source contexts` and `Complexity` is therefore an inverse one as depicted below.

$$\texttt{Complexity} \propto sourcecontexts$$

Using the principles above, I therefore develop the following definition of Complexity.

$$\texttt{Complexity} = \frac{interaction(subclass_{\text{cited}})}{interaction(subclass_{\text{patent}})}$$

By the definition above, a patent that cites no patents (and hence has $subclass_{\text{cited}} = 0$) but is itself assigned to 4 sub-classes (and hence has $subclass_{\text{patent}} = 4$) will have a raw Complexity score of $\frac{\binom{4}{2}}{1} = 6$. If the patent itself had been assigned onto to 2 sub-classes, the raw Complexity score would have been just 1. Therefore, the more the number of patent sub-classes a patent is assigned to, the higher its Complexity score (by a square term). A similar but inverse relationship would hold for sub-classes arising out of cited patents. Here, I take a set union of patent sub-classes assigned to each cited patent, and use that count to determine the value of the `interaction` function.

4.2 IPR Classification

A review of the academic literature surrounding the construction of IPR indexes indicated that there were several, as was also evident in Zhao (2006) constructing a composite measure for the purposes of her article. Lesser (2010) provides an alternative, composite scoring system that includes the following components: protectable subject matter, membership in convention, enforcement, administration and duration of protection. I have therefore used the scores generated by Lesser (2010) for the purposes of this study. The extensive table of IPR scores in presented in Table 4. The listing has several countries for which scores have not been provided. However none of the top patenting nations were among them, and I therefore chose to go along with this scale.

4.3 Data Source

I derive all patents data for this study from patentsview.org. The dataset considered is for all USPTO patents filed in the period 1976 to 2015. For the IPR

Scores, I rely on the scores generated by Lesser (2010). For country definitions, I use the resources provided by Thematic Mapping. To determine if spillovers are local, I use a composite data source as described in the following. For locations in the United States, it has been standard to use Metropolitan Statistical Areas (MSA) for analyses related to economic geography. Such standardized data is unavailable for non-US locations. Urban areas are a close substitute for economic centers, and I therefore determine to use one such definition for non-US locations. My data source for MSA of US locations is the US census and that for urban areas for world wide locations is Natural Earth Data.
This automatically raises conflicting definitions for locations in the United States. So that the MSA definitions take precedence, I eliminated all data pertaining to US locations from the Natural Earth urban centers data and integrated this with the MSA information. With this I generated a single database of location information for economic centers around the world.

4.4 Dependent Variable

For the first two questions, I assess the impact of IPR regime and Cross-border inventions on Complexity of patents. Since I have used a binomial measure for the raw value of Complexity, I define my dependent variable for the first two questions to be the logarithm of the raw Complexity score. The summary statistics for my primary quantitative variables is provided in Table **??**.

For the third question on impact of Complexity on local spillovers, I define as my dependent variable, `local spillover` which is a variable that takes a value between 0 and 1, with 0 representing no local spillover and 1 representing the maximum spillover. Therefore for $patent_i$

$$local\ spillover_i = \frac{Citations\ to\ Patents\ from\ Same\ Country_i}{Total\ Citations_i}$$

4.5 Explanatory Variables

4.5.1 IPR (Inventor Location) and IPR (Assignee Location)

While most patents have multiple inventors, and some patents also have multiple assignees, my question requires us to associate a single location to the inventor of a patent, and a single location for the assignee of the patent. For the inventor location, I tabulate the count of each of the regions that each inventor is a resident of at the time of the filing of the patent application. In doing so, I treat all inventors equally and allocate the most frequently occurring location as the location of the inventor for that patent. In case of a tie, I assign the location of the first inventor (given by the sequence number of the inventor on the patent) as the location of the inventor of the patent.

For the assignee location, I treat multiple assignees as having been granted separate patents. I do this since the number of patents with multiple assignees is small, and so as to not lose potentially valuable information.

4.5.2 Localized Invention

Patents for which the inventor resides in the same country as the assignee are marked as inventions that are localized. This is in contrast to those where the inventor and assignees are located in different countries. I capture this difference to identify the potential variation in Complexity caused on account of cross-border collaborations in inventions.

4.5.3 Crossborder Invention

A patent in which the location country of the inventor differs from the country of the assignee is marked as a crossborder invention.

5 Results

5.1 Localized vs. Cross border Inventions

The preliminary answer to my first question on the impact of the location on log(Complexity) is provided by a t-test in Table 3. I see that localized inventions on an average score a log(Complexity) value of about 0.26 higher than those scored by cross border inventions. This result comes as a bit of a surprise as one might have expected that inventions being developed across borders may have been the ones with the least interaction with others. The result on 4.25 Million patents filed between 1975 and 2015 provides a statistically significant result that it is the cross border inventions that are of lower Complexity (or alternatively, of higher complexity).

Table 1: Main Regression Results

Table 2: 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

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

Table 3: Investigating potential mechanisms driving spillovers

Effect of Modularity on Local Spillovers

VARIABLES	(1) Local Spillover	(2) Local Spillover
Log Modularity	0.0050***	0.0049***
	(0.0013)	(0.0013)
IPR(Inventor Location)		0.0005
		(0.0009)
IPR(Assignee Location)		0.0027
		(0.0017)
Localized Invention		-0.0044
		(0.0486)
IPR(Assignee Location) * Localized Invention		0.0023
		(0.0059)
Constant	0.0487***	-0.0001
	(0.0156)	(0.0156)
Observations	4,252,508	4,239,225
R^2	0.0032	0.0043

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

5.2 Strong vs. Weak IPR location of Assignee

5.3 Regressions on Local Spillovers

In Table 3, I present my results from my initial nvestigation on whether Complexity affects local spillovers. Models 1 and 2 demonstrate that there is indeed a positive effect of log (Complexity) on local spillovers across my sample of 4.2 million patents. When coupled with my findings in the previous section, I have some support for the effect of patenting activity by multinational firms in weak IPR locations on local spillovers. I discuss this in the following section on conclusions.

6 Conclusions

I started this study attempting to understand if I could use the mechanism of patent Complexity to explain the heterogeneity in knowledge spillovers across locations, firms and IPR regimes. My study found that IPR regimes do not by themselves seem to directly affect the Complexity of patents invented in those locations. However, I find strong evidence for the fact that multinationals (or cross border inventions, where the inventor and assignee are located in different countries) on an average file for more complex patents. My third finding was that Complexity of patent work was positively correlated with local spillovers. Putting results two and three suggests therefore, that cross border patenting performed in both weak IPR and strong IPR locations are unlikely to generate high local spillovers for those locations.

References

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

Baldwin, C. Y. and Henkel, J. (2015). Modularity and intellectual property protection. *Strategic Management Journal*, 36(11):1637–1655.

Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why u.s. manufacturing firms patent (or not). Working Paper 7552, National Bureau of Economic Research.

Ethiraj, S. K. and Levinthal, D. (2004). Modularity and innovation in complex systems. *Management Science*, 50(2):159–173.

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.

Kogut, B. and Zander, U. (1993). Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies*, 24(4):625–645.

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.

McAfee, A. and Brynjolfsson, E. (2014). The second machine age. *NY: WW Norton & Company*.

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

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

Yayavaram, S. and Ahuja, G. (2008). Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. *Administrative Science Quarterly*, 53(2):333–362.

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

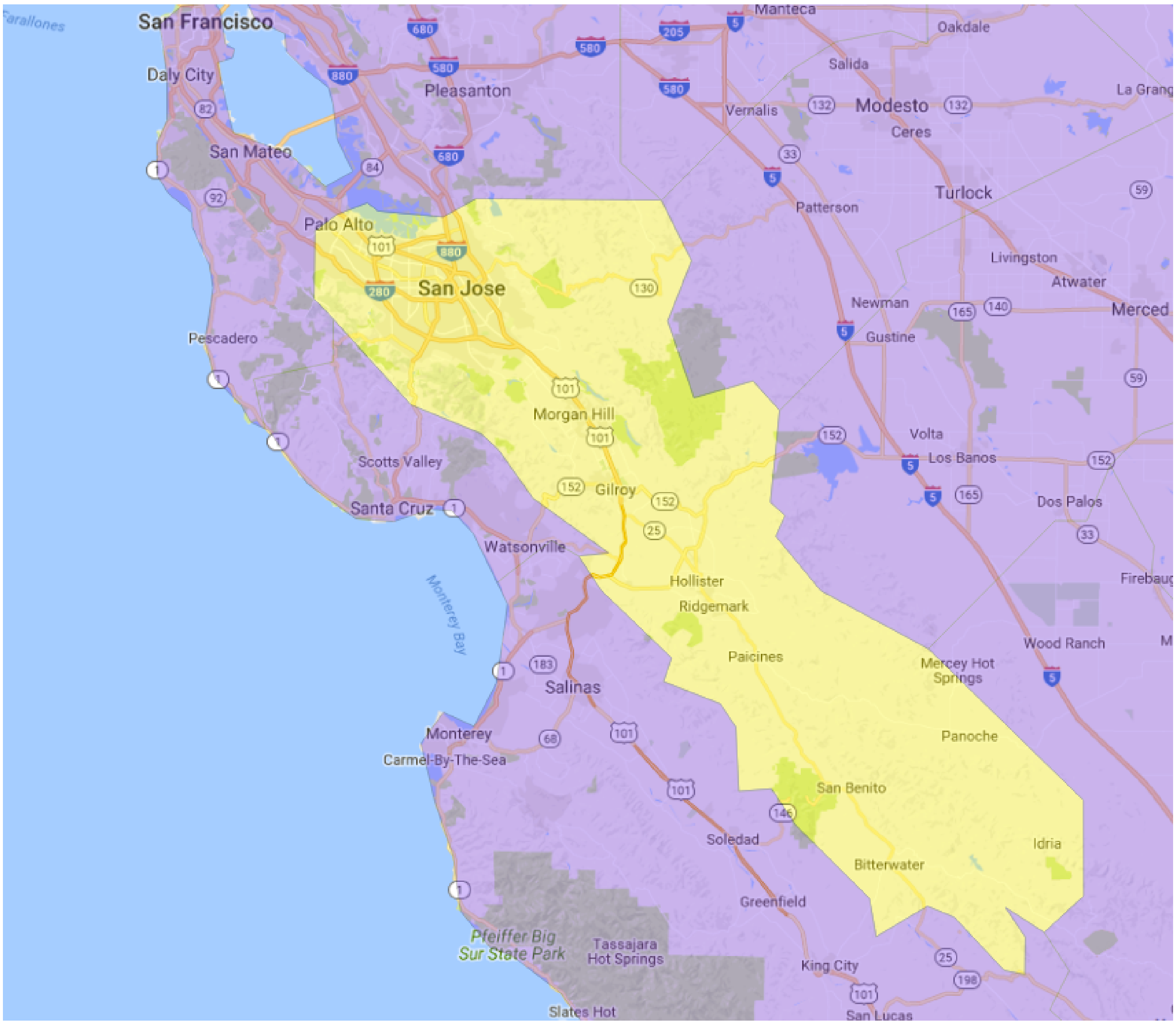

Figure 5: Geographic Definition of San Jose-Sunnyvale-Santa Clara, CA

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	

Country	**IPR Score**
Australia	11.1872
Austria	9.4024
Azerbaijan	3.1358
Bahamas	
Bahrain	5.7736
Bangladesh	2.3664
Barbados	
Belarus	3.2344
Belgium	9.6096
Belize	
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	

Country	**IPR Score**
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
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

Country	IPR Score
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
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

Country	IPR Score
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	
Saint Pierre and Miquelon	
San Marino	
Saudi Arabia	4.2398

Country	**IPR Score**
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

Country	**IPR Score**
Uzbekistan	3.6388
Venezuela	3.6144
Vietnam	4.2752
Yemen	2.0706
Zambia	2.9580
Zimbabwe	2.9142

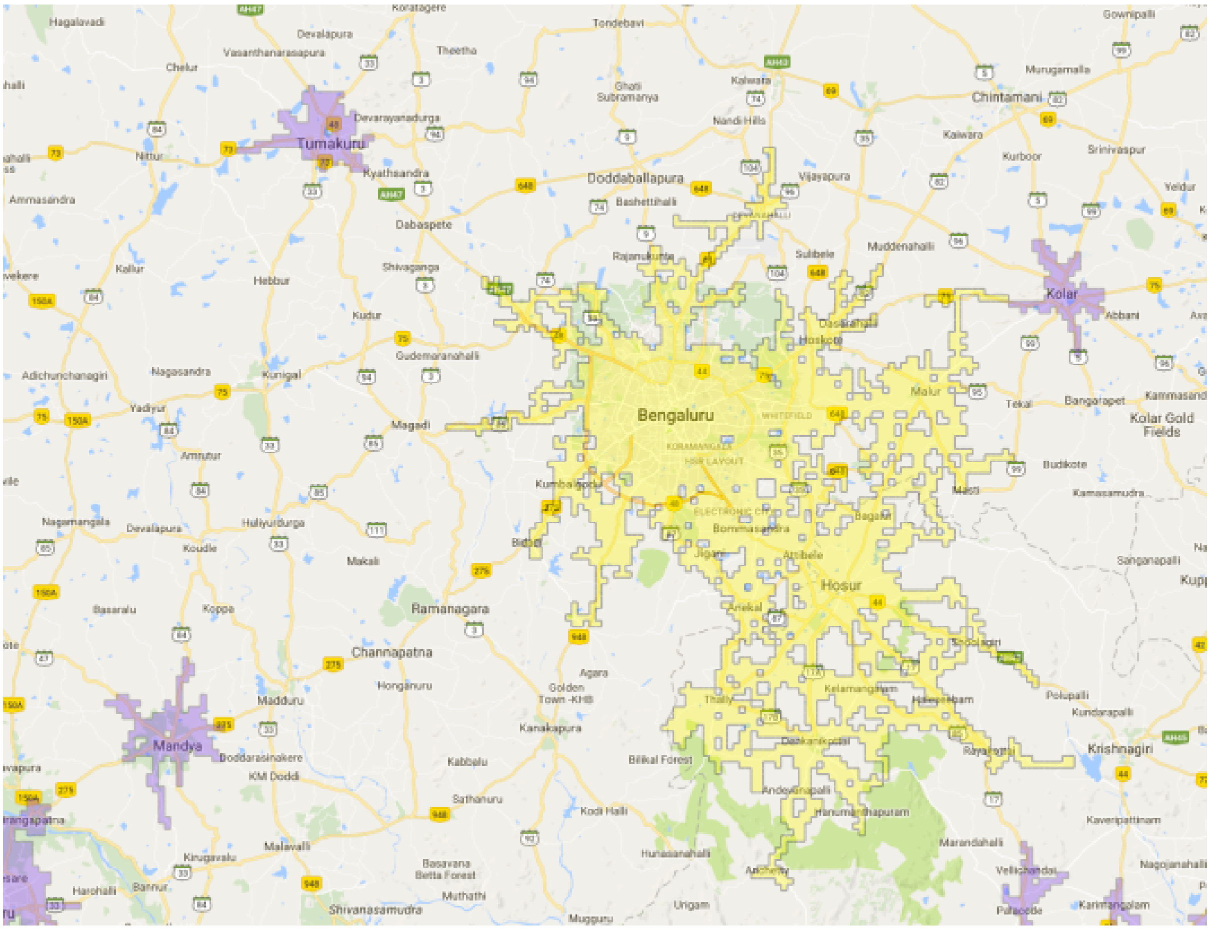

Figure 6: Geographic Definition of Bangalore