

Does inventor mobility affect productivity?

Methods for Causal Inference course term paper

Ashwin Iyengar (1521001)
ashwin.iyengar15@iimb.ernet.in

March 3, 2017

Abstract

Using patent inventions data, I attempt to investigate if inventor mobility influences the productivity of those inventors. I leverage a database of worldwide urban regions obtained from remote sensing to demonstrate a significant positive correlation between both inter-region mobility and inter-country mobility and future invention productivity of inventors. I discuss two additional approaches to investigate if a causal explanation exists for this relationship but leave empirical analysis for future work. While the empirical results are incomplete, the current work extends prior work on mobility of inventors by attempting to quantify the direct effects of mobility on inventor productivity.

Keywords: Inventor productivity, Inventor mobility, Economic geography

1 Introduction

Scholars have suggested that inventors carry knowledge with them when they move ([Almeida and Kogut, 1999](#)). While the literature on knowledge flows has investigated the localized nature of knowledge flow driven by the inter-firm

mobility of inventors (Alcácer and Gittelman, 2006; Almeida and Kogut, 1999; Jaffe et al., 1993), we understand very little about the effects of broad mobility of inventors on future inventive productivity. Since firms, regions and countries seek to organize themselves so as to attract the best inventors so as to produce the highest level of innovative output, an understanding of the mobility effects of inventive productivity becomes an important aspect innovation policy. In this paper I ask if the variation in inventor mobility can explain the variation in invention productivity?

Agglomeration economies have been suggested as one reason for localized movement of inventors. These agglomeration economies arise due to labor pooling advantages, economies of specialization of local suppliers, and knowledge spillovers (Krugman, 1991; Porter, 1990). However regions vary in their inventive output (Agrawal et al., 2014) and the nature of knowledge flows in a region may be one source of this variation within regions. In this study therefore, I propose to expand the context to not just localized agglomerations but to all movements of inventors. By doing so, I expect to understand if there is a clear impact of mobility itself on invention productivity. In other words, what is the relationship between the movement of some inventors into or out of a region and the average productivity of inventions from those inventors?

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 mobility of inventors observed in

practice is highly heterogeneous across locations, firms and legal environments. This raises the opportunity to study if a causal effect exists between mobility of inventors and their future productivity. The question is assumes greater significance in the environment surrounding the second machine age (McAfee and Brynjolfsson, 2014) where inventors are expected to influence innovation outcomes in higher proportion.

The innovation policy of emerging countries is influenced with the expectation that the presence of multinational R&D will create value adding spillover effects. Productivity of inventors may provide a richer proxy for value adding innovation. A better understanding of the effects of inventor mobility may therefore help to inform innovation policy. Additionally, work on this line may help to inform managerial decisions about how to organize R&D teams around the world. Current theory seems to suggest both a positive effect due to knowledge spillovers (Almeida and Kogut, 1999) as well as a negative effect due to variation in IPR enforcement ability (Zhao, 2006). I therefore propose the current empirical study to help determine an answer to this question that is not completely explained by theory.

The rest of the paper is organized as follows. The next section defines a framework to classify knowledge flows in a region and motivates our work by demonstrating how different regions fare differently in terms of knowledge flows. We then describe our data and methods in the following section. Our preliminary results are then presented, followed by a discussion of the results. We conclude with next steps and open questions for further research.

2 Motivation

I motivate this study by demonstrating two broad patterns. First, as Figure 1 suggests, there has been increased mobility of inventors across countries over the years. Not only is the incidence of across-country mobility higher, but

there is heterogeneity in the extent across countries and regions. Figure 2 furthers this aspect by depicting the trend in across-region mobility of inventors. Second, Table 1 suggests that about 6% of inventors moved and only about 2% of inventors moved country.

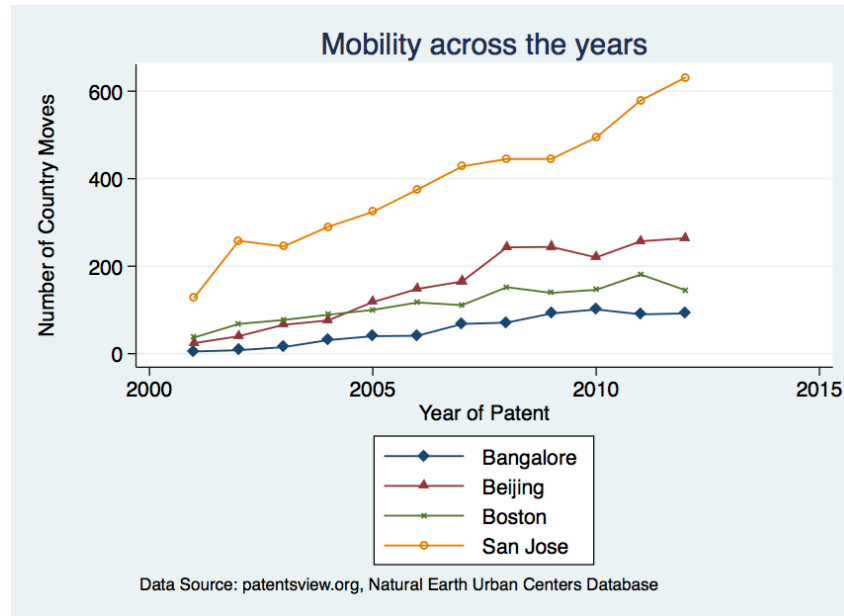


Figure 1: Country moves by year

The two broad patterns together provide us both the motivation as well as the variation in the data to be able ask the question of whether the mobility of inventors affects invention productivity. It must be noted at this point that while we have data on inventors who moved and patented in their new location, and data on inventors who did not move and continued to patent at their existing location, the patents data itself does not provide us data on those inventors who moved but did not patent, and those who did not move and did not patent. This may cause us to either overestimate the effect of mobility or underestimate the effect of not moving.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	N
Moved Region (MR)	0.06	0.238	4368062
Moved Country (MC)	0.021	0.143	4368062
Productivity	1.79	2.647	4368062
Prior Patents of Inventor (PPI)	4.716	16.718	4368062
Prior Patents of Team (PPT)	24.557	112.7	3721064

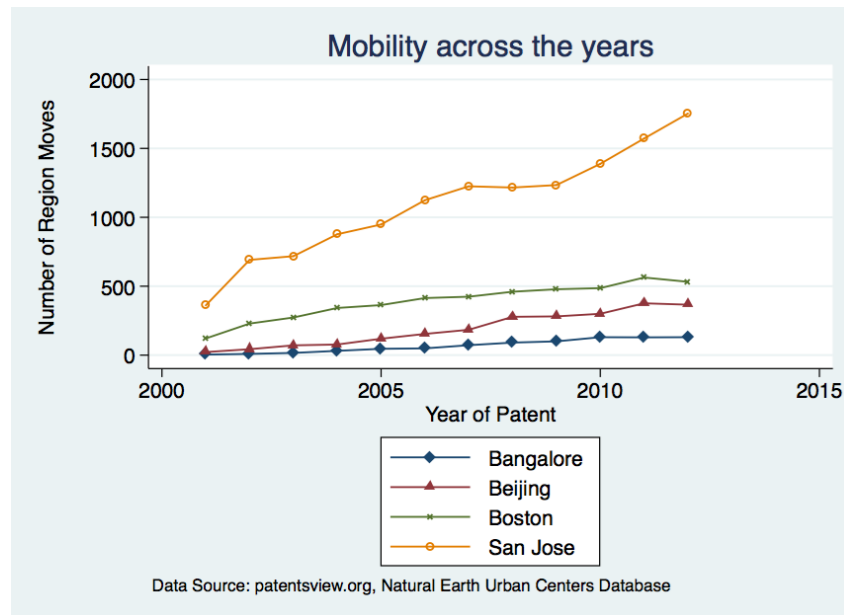


Figure 2: Region moves by year

3 Theory

In this section, I develop three hypothesis building off of prior literature on inventor mobility and knowledge spillovers.

The literature on agglomeration economies and knowledge spillovers suggest that when inventors move across regions, that newer combinations of knowledge now become possible Almeida and Kogut (1999); Jaffe et al. (1993). Literature would therefore predict better and more numerous inventions arising out of the movement of inventors. I therefore propose hypothesis 1 as follows:

Hypothesis 1: An increase in the average mobility of inventors in a region increases the average productivity of the inventor

Singh (2007) suggests that inventors who were highly productive previously are likely to carry that after the move. I therefore propose hypothesis 2 as follows:

Hypothesis 2: The effect in Hypothesis 1 is moderated positively by the strength

of the prior pool of inventions by the inventor

Finally, the literature of weak and strong IPR suggests that teams within weak IPR locations are likely to integrate their knowledge within global organizations, thus leading to a lower standalone inventive output. I therefore propose hypothesis 3 as follows:

Hypothesis 3: The effect in Hypothesis 1 is moderated negatively by the strength of the prior pool of inventions by the inventing team

4 Data and Measures

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 country definitions, I use the resources provided by [Thematic Mapping](#). To map location data of inventors to regions, I use urban centers data for worldwide locations from [Natural Earth Data](#) that uses remote sensing data to determine urban agglomerations (a process developed in [Schneider et al. \(2003\)](#)). While it has been common practice to use Metropolitan Statistical Areas (MSA) for analyses related to economic geography in the U.S., an equivalent measure is unavailable for the rest of the world. For comparability and consistency, I choose to use the urban centers definitions from [Natural Earth Data](#) for all regions both within U.S. and outside U.S. Sample region definitions are depicted in [Figure 3](#) and [4](#) in the appendix.

4.1 Unit of Analysis

The unit of analysis for this study is the `inventor - year`. For each inventor-year, I capture if the inventor moved in that year. As was previously discussed,

the nature of our data is that we do not have data for years in which inventors did not patent. However for the years for which patents were applied for, we identify if there was a movement or not.

4.2 Dependent Variable

My primary dependent variable is the productivity of an inventor in a year. I measure productivity by the number of patents invented by the inventor in the year and in two years succeeding that year. The dependent variable may be zero, or 1 or a value higher depending on the inventive productivity of the inventor.

4.3 Explanatory Variables

I use two primary explanatory variables - between-region mobility, and between-country mobility. For each inventor-year, I determine mobility for the inventor in that year by comparing the region of the previous patent to that of the current patent. If the region has not changed, I mark between-region mobility to zero for that inventor-year. Else between-region mobility is marked as one for that inventor-year.

Similarly, to determine between-country mobility for an inventor-year, I compare the country of the previous patent to that of the current patent. If the country has not changed, I mark between-country mobility to zero for that inventor-year. Else between-country mobility is marked as one for that inventor-year.

I define two additional moderating variables. Prior patents of inventor (PPI) is intended to capture the inventor specific aspects and is computed by adding up the number of patents granted to the inventor up to but not including the patents granted in the current year. Prior patents of team (PPT) is intended to capture the team specific aspects affecting inventor productivity. It is computed by

summing up the number of patents granted to the most prolific of the inventors within the team (not including the focal inventor).

4.4 Control Variables and Fixed Effects

I use the technology classification defined by ? to control for various technology subcategories. All models also include year dummies so as to account for any year specific effects. Depending on the model run, we cluster standard errors at the region level, the country level or the technology subclass level.

4.4.1 Citations Made

My primary explanatory variables are the four counts of Citations Made along the two dimensions of same/different region and same/different assignee. 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 Control Variables

4.5.1 Size of the patent pool, and Number of patents generated in a region - year

Since the priors of the region may themselves explain the extent of citations received, I control for both the number of patents generated that year, as well as the aggregate pool of patents invented within that region. The reason both variables are logged is because of the exponential nature of the estimation method used.

5 Results

The preliminary results from our analysis are presented in Table ???. Model 1 reports the results for all regions worldwide, while Model 2 and Model 3 report results for U.S. locations and non-U.S. locations respectively. Across the three models we find that local knowledge flows within the firm do not seem to have a significant impact on the quality of inventions as measured by the number of citations received. This result is robust to using non-self citations as the dependent variable as presented in Table ???.

5.1 Spillover Effects

Table 2 presents the preliminary results from my regression. The first model uses year dummies but does not include region fixed effects, while the second includes both. As evident, the effects do not seem stable and much empirical work will be needed to determine the right identification method. A couple of approaches that are planned to be tried are to look at controlling for technology class, as regions may vary widely on this aspect. As indicated at the outset, the empirical aspect of this article remains work in progress.

Table 2: Regression Results

Regional Mobility of Inventors and Productivity of Inventors			
VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Region (MR)	1.3974*** (0.0997)	0.7966*** (0.0531)	0.5571*** (0.0556)
Prior Patents of Inventor (PPI)		0.0852*** (0.0041)	0.0741*** (0.0027)
Prior Patents of Team (PPT)		0.0004*** (0.0001)	0.0005*** (0.0001)
MR x PPI			0.0282*** (0.0054)
MR x PPT			-0.0005*** (0.0002)
Constant	1.7123*** (0.0894)	1.7258*** (0.0468)	1.7404*** (0.0504)
Observations	2,886,438	2,507,963	2,507,963
R^2	0.0282	0.2940	0.3010
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Region	Region	Region

Robust standard errors in parentheses

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

Table 3: Regression Results

Regional Mobility of Inventors and Productivity of Inventors			
VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Region (MR)	1.5681*** (0.0926)	0.9204*** (0.0606)	0.6239*** (0.0654)
Prior Patents of Inventor (PPI)		0.0803*** (0.0138)	0.0683*** (0.0124)
Prior Patents of Team (PPT)		0.0003*** (0.0000)	0.0005*** (0.0000)
MR x PPI			0.0319*** (0.0051)
MR x PPT			-0.0005*** (0.0002)
Constant	1.7227*** (0.0197)	1.7344*** (0.0218)	1.7489*** (0.0207)
Observations	4,120,577	3,563,673	3,563,673
R^2	0.0298	0.2888	0.2984
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Technology	Technology	Technology

Robust standard errors in parentheses

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

Table 4: Regression Results

Country Mobility of Inventors and Productivity of Inventors			
VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Country (MC)	1.7074*** (0.4054)	0.9275*** (0.1655)	0.5005*** (0.1186)
Prior Patents of Inventor (PPI)		0.0875*** (0.0066)	0.0775*** (0.0027)
Prior Patents of Team (PPT)		0.0004** (0.0002)	0.0005*** (0.0002)
MC x PPI			0.0381*** (0.0062)
MC x PPT			-0.0009* (0.0005)
Constant	1.7096*** (0.1609)	1.7260*** (0.0727)	1.7337*** (0.0810)
Observations	3,981,822	3,440,812	3,440,812
R^2	0.0176	0.2962	0.3063
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Country	Country	Country

Robust standard errors in parentheses

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

Table 5: Regression Results

Country Mobility of Inventors and Productivity of Inventors			
VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Country (MC)	1.7541*** (0.1470)	0.9859*** (0.0872)	0.5779*** (0.1018)
Prior Patents of Inventor (PPI)		0.0812*** (0.0139)	0.0719*** (0.0125)
Prior Patents of Team (PPT)		0.0004*** (0.0000)	0.0005*** (0.0000)
MC x PPI			0.0351*** (0.0075)
MC x PPT			-0.0009* (0.0005)
Constant	1.7081*** (0.0213)	1.7267*** (0.0217)	1.7340*** (0.0213)
Observations	4,120,577	3,563,673	3,563,673
R^2	0.0179	0.2845	0.2940
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Technology	Technology	Technology

Robust standard errors in parentheses

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

6 Extensions to identify causality

6.1 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:

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

I would expect, from a user perspective that the more number of contexts in which the patent is valuable, the lower 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:

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

Similarly, from an inventor perspective, the more the number of contexts that the patent is built on, the higher should be the complexity. A patent that is developed without citing any other patents is an extreme case of lowest complexity, while one that requires to be built upon several `source contexts` is properly understood as being more complex.

The relationship between `source contexts` and `complexity` is therefore a normal one as depicted below.

$$\text{complexity} \propto \text{source contexts}$$

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

$$\text{complexity} = \frac{\text{interaction}(\text{subclass}_{\text{cited}})}{\text{interaction}(\text{subclass}_{\text{patent}})}$$

By the definition above, a patent that cites no patents (and hence has $\text{subclass}_{\text{cited}} = 0$) but is itself assigned to 4 sub-classes (and hence has $\text{subclass}_{\text{patent}} = 4$) will have a raw Complexity score of $\frac{1}{\binom{4}{2}} = 0.16$. If the patent itself had been assigned onto 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 lower 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.

6.2 IPR Classification

For the IPR Scores, I rely on the scores generated by Lesser (2010). Zhao (2006) has argued that multinational enterprises may benefit from conducting R&D in countries with weak IPR protection by making up for the weaker IPR protection through better internal organization. Scholars (Baldwin and Henkel, 2015; Yayavaram and Ahuja, 2008), have argued that increased interaction with a larger number of components creates organizational impediments to an increase in reusability of prior work. In the presence of a stronger differential in the IPR environments between inventing locations, Zhao (2006) suggests that organizational mechanisms may stand to counter the treat posed by weaker property rights. In a similar vein, I argue that a differential in the IPR rights environment creates the organizational response to increase complexity of the inventions shared across country and IPR boundaries.

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 has not been presented here to adhere to the page restriction, but can be made available on request. 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.

7 Limitations and Looking Ahead

Direction of Causality Underestimation bias of mobility - effects Mechanism by which mobility affects productivity - other explanations Alternative measures of productivity I started this study attempting to understand if I could bring three mechanisms to bear together in explaining the heterogeneity in knowledge flows across regions. While there seems to be theoretical promise to exploring this question, this was a study too big to have been completed within the constraints of a term. Specifically, the endeavor has exposed me to the challenges to demonstrating empirically driven work with the objective of building theory. I intend to continue to pursue this further and integrate the IPR level data and complexity data to the flows. In addition, I plan to explore the prospect of controlling for technology classes, and see if that may lead to a strong result.

Conclusion

While still at a preliminary stage, our analysis seem to suggest that local knowledge flows as measured by patent citations may not have a significant effect on

the quality of inventions produced in a region. This casts a doubt on the widely accepted idea that local knowledge spillovers are an important source of agglomeration economies. A potential extension of the study could be to conduct empirical analysis at the level of firm-year rather than at the region-year. We could additionally look at the additional dimension of technology (within and outside technological domain, [Rosenkopf and Nerkar \(2001\)](#)) in addition to those of within/outside region and within/outside firm. This may provide us with a more nuanced understanding of the factors affecting invention quality. Future studies could potentially examine other measures of invention outcomes such as breakthrough inventions. Finally, while our work suggests that local knowledge spillovers do not affect invention quality, it is not quite as clear why this may be the case. We hope that our current work spurs further research in this direction.

Acknowledgements

I am greatly indebted to Vidhya Soundarajan for having helped me with thinking deeply about my research design, level of analysis and causality. While I might have not done as much justice to the causality instruction, I cannot imagine having made as much empirical progress on this project if not for that guidance. I am also indebted to her for having encouraged me to see the theoretical relevance and contribution of empirical work.

I am also grateful to Sai Yayavaram for having introduced me to the literature on innovation, and for having hand held me with working on the patents data. Indeed many of the skills in understanding the data underlying this article owe their origin to him. All mistakes though, remain entirely mine.

References

- Agrawal, A., Cockburn, I., Galasso, A., and Oettl, A. (2014). Why are some regions more innovative than others? the role of small firms in the presence of large labs. *Journal of Urban Economics*, 81:149 – 165.
- Alcácer, J. and Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics*, 88(4):774–779.
- Almeida, P. 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.
- Ethiraj, S. K. and Levinthal, D. (2004). Modularity and innovation in complex systems. *Management Science*, 50(2):159–173.
- Ge, C., Huang, K.-W., and Png, I. P. L. (2016). Engineer/scientist careers: Patents, online profiles, and misclassification bias. *Strategic Management Journal*, 37(1):232–253.
- 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.
- Krugman, P. R. (1991). *Geography and trade*. MIT press.
- Lesser, W. (2010). Measuring intellectual property strength and effects: An assessment of patent scoring systems and causality. *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.

- Porter, M. E. (1990). *The Competitive Advantage of Nations*. Free Press New York.
- Rosenkopf, L. and Nerkar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal*, 22(4):287–306.
- Schneider, A., Friedl, M. A., McIver, D. K., and Woodcock, C. E. (2003). Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, 69(12):1377–1386.
- 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.

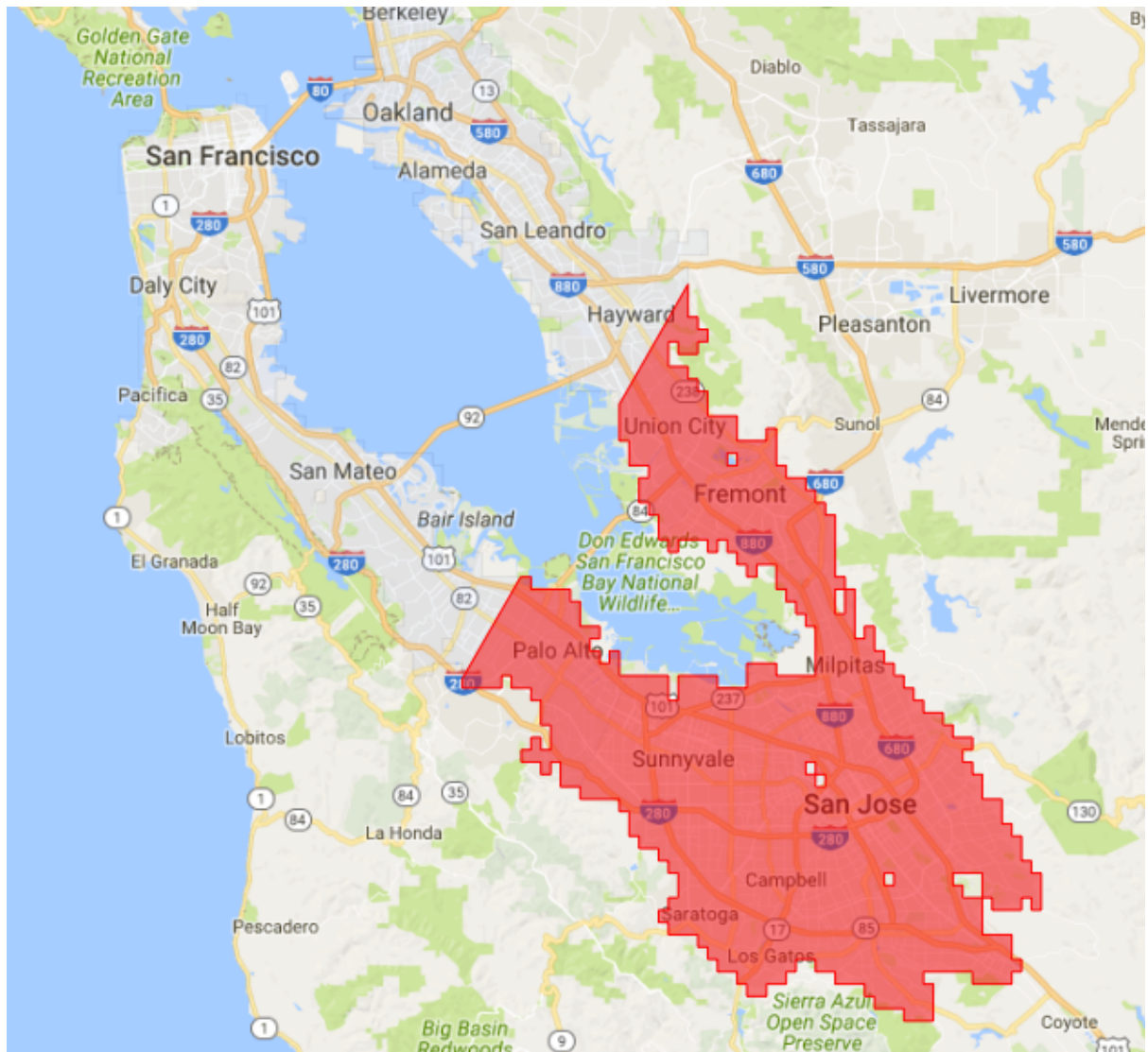


Figure 3: Geographic Definition of San Jose, CA

Table 6: Regression Results

Regional Mobility of Inventors and Productivity of Inventors			
VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Region (MR)	1.5328*** (0.2967)	0.8679*** (0.1278)	0.5668*** (0.0851)
Prior Patents of Inventor (PPI)		0.0866*** (0.0065)	0.0740*** (0.0019)
Prior Patents of Team (PPT)		0.0004* (0.0002)	0.0005*** (0.0002)
MR x PPI			0.0335*** (0.0078)
MR x PPT			-0.0006*** (0.0002)
Constant	1.7242*** (0.1644)	1.7335*** (0.0746)	1.7487*** (0.0848)
Observations	3,981,822	3,440,812	3,440,812
R^2	0.0292	0.3001	0.3096
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Country	Country	Country

Robust standard errors in parentheses

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

Table 7: Regression Results

Country Mobility of Inventors and Productivity of Inventors			
VARIABLES	(1) Productivity	(2) Productivity	(3) Productivity
Moved Country (MC)	1.8086*** (0.1945)	0.9959*** (0.0892)	0.6032*** (0.1072)
Prior Patents of Inventor (PPI)		0.0860*** (0.0041)	0.0774*** (0.0036)
Prior Patents of Team (PPT)		0.0004*** (0.0001)	0.0005*** (0.0001)
MC x PPI			0.0332*** (0.0083)
MC x PPT			-0.0006 (0.0006)
Constant	1.7019*** (0.0871)	1.7212*** (0.0453)	1.7294*** (0.0483)
Observations	2,886,438	2,507,963	2,507,963
R^2	0.0173	0.2903	0.2980
Year FE	Yes	Yes	Yes
Technology FE	Yes	Yes	Yes
Clustered SE	Region	Region	Region

Robust standard errors in parentheses

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

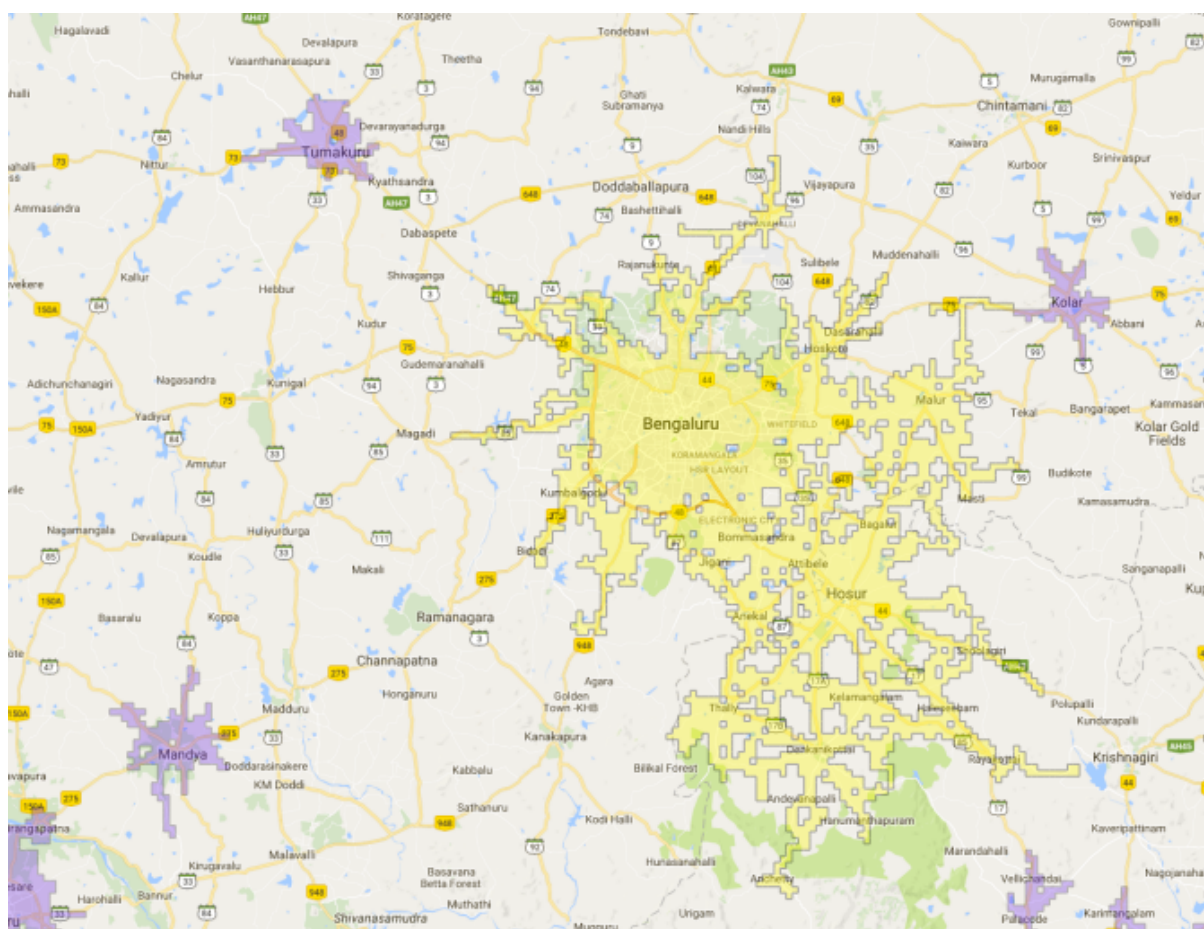


Figure 4: Geographic Definition of Bangalore