Firms and regional innovation

by

Brian Kuang

Thesis Advisor: Gerald Davis

A thesis submitted in fulfillment of
the requirements of the
Michigan Ross Senior Thesis
Seminar (BA 380), 04/2019.

CONTENTS

Introduction

This senior thesis paper examines the role of firms in driving regional innovation. Building on long-standing theoretical frameworks for understanding regional economic growth and combined with recent empirical work. Expanding on a 2017 paper (Berkes and Gaetani, 2017) which examined the connection between population density and the density of patents within a region, I study the density of firms as another potential driver of regional innovation. In doing so, we will be able to further disentangle the myriad of complex factors that drive the innovation and growth of regions.

The results of this study will carry broad economic and public policy implications. First, it will help expand our understanding of the consequences for American innovation caused by the decline in the number of large, publicly-traded American companies. In addition, the findings of this study can be informative for state-and-local policymakers to better understand optimal strategies to spur greater regional private sector innovation within their jurisdictions.

Research Problem

Question

Broadly speaking, this project seeks to further our understanding of the role firms play in driving innovation, and to study how the geographic clustering of firms can lead specific regions to become more competitive in private sector innovation.

There have been several publications dating to the 1990s that outlined a theoretical framework that the geographic clustering of commercial activity – and specifically advanced, knowledge-intensive industries – leads to a greater frequency of transfer of ideas and expertise between individuals and firms, facilitating greater private sector competitiveness and innovation (Porter, 1990; Saxenian, 1996). There have been further works studying the role of knowledge transfer in economic growth and innovation (Lucas, 1998; Lucas and Moll, 2011), and the role of cities in conglomerating and diffusing knowledge (Glaeser et al., 1992; Black and Henderson, 1999; Glaeser, 1999). A more extensive review of these publications can be found further in this proposals' literature review section.

However, these theories have seen relatively limited empirical study until recently; and of those empirical studies, few have tried to isolate the role of the firm in knowledge diffusion and innovation. One notable recent paper sought to examine the connection between population density and regional innovation, and ultimately created a model to mathematically rationalize a connection between these two variables (Berkes and Gaetani, 2017). However, the potential role of firms as agents of innovation did not receive the same sort of empirical treatment in Berkes and Gaetani's paper. Another study has found that greater inter-firm interaction across geographic boundaries leads to greater commercial innovation in the context of startups and venture capital (Shai et al., 2015), which suggests that the geographic proximity of firms is favorable to innovation by facilitating greater inter-firm contact. This framework has also been projected onto other fields, such as in a 2013 paper that found a significant relationship between the clustering of large corporations and their leadership within metropolitan areas driving the greater emergence of corporate-backed nonprofits (Marquis, Davis, and Glynn, 2013).

Ultimately, my goal will be to empirically scrutinize Berkes and Gaetani's 2017 study that established geographically-bounded units of population as an agent of innovation. I plan to introduce geographically-based firms into their model to isolate the[4] extent to which corporations act as drivers of regional innovation, as opposed to individuals. I will study this question through a variety of empirical and statistical methods, relying primarily on open-source data from sources such as the United States Census Bureau and United States Patent and Trademark Office.

Why it matters

There are several reasons why it is pertinent to study the role of firms in regional innovation. The number of large publicly-traded corporations in the United States has been systemically declining for decades despite robust GDP growth, contributing to income inequality as these more concentrated employers have used technology to gain bargaining power over their remaining workforces (Davis, 2016). It is plausible that if the number of competitors within a region and industry concentrates, individual firms may feel less market pressure to pursue game-changing innovation, after controlling for other factors. Furthermore, the interconnectedness of board leadership across corporations has declined (Chu and Davis, 2016). This decline in interfirm connectedness may indicate that in addition to fewer major corporations, there are fewer exchanges of knowledge between the remaining businesses.

There has also been a sweeping transformation to the United States' economic geography as a result of these trends, as both firms and innovations become confined to an increasingly small number of geographies (Kaufmann Foundation, 2016; Berkes and Gaetani, 2017).This project will expand the body of knowledge on not only the drivers of regional economic innovation, but also investigate the economic consequences of the increased concentration in the American private sector.

Furthermore, this project will offer insight for how state-and-local policymakers can better spur innovation within regions. In practice, this often leads to both elected officials and economic development bureaucrats either trying to attract high-profile corporate relocations, or trying to develop a higher-skilled local workforce. The strategy of attracting corporate relocations – such as Foxconn's tentative new plant in Wisconsin, and the bidding war over Amazon's second headquarters – is often controversial as it entails corporate tax breaks and subsidies. In studying the power of firms in driving regional innovation, my project may offer insight as to whether such public policies are beneficial for the communities that pursue them.

Theory and Hypothesis

Hypothesis and Project Summary

This project will empirically test the hypothesis that interactions between geographically proximate firms will lead to a measurable increase in regional innovation that is distinct from the impact caused by person-to-person interactions.

My key hypotheses for this study are as follows:

1. Firm density will have a positive, statistically significant relationship with patent density across a broad sample of geographies over 2004-2014.
2. This statistically significant relationship will persist after applying a fixed effects model to control for individual geographies' baseline propensity to patent.

I plan to study this by recreating the based model in Berkes and Gaetani's 2017 paper – which tied population density to patent innovation by county subdivision (CSD) – and adding the density of large corporations and the connectedness of local corporate boards as dependent variables. These firm-level measures will be applied on top of the preceding paper's population measures, and then a fixed-effects model over 2004 to 2014 will be utilized to control for other factors that are not included in this model.

After creating this new model, I believe I will find that the statistical significance of the preceding paper's population measures in predicting a CSD's level of patent innovation will have decreased. This will suggest that both firm-to-firm and individual-to-individual interactions have separate, measurable effects on a CSD's level of patent innovation, and therefore firms themselves act as agents of innovation akin to individuals.

The findings of this project may also offer causal evidence for one of Berkes and Gaetani's subsequent hypotheses from their 2017 paper. In their findings, it was found that patent density actually peaked in mid-to-high density areas and declined in the most dense areas; but when controlling for patent conventionality using the aforementioned methodology it was found that the most innovative patents came from the densest regions while more sustaining patents came from mid-density areas. The authors theorized that this was because corporate campuses tended to be in mid-density suburban areas, thus driving the bump in patent density, whereas patents in the densest of cities were created moreso by individuals. By finding empirical evidence to suggest that corporations themselves can behave as individuals in driving innovations, I would offer evidence to support their latter hypothesis.

Theoretical Framework

There exists a body of theoretical work that suggests mechanisms by which population and industry clustering can lead to more innovative regions. I will first examine several of the most critical works in this space and their implications for my project, before delving into my own theoretical framework.

The geographic clustering of knowledge-based, globalized industries has been described as one of the greatest paradoxes of the modern economy (Porter, 1990). Unlike in the preceding industrial revolutions, where raw material inputs and workforces needed to be geographically proximate with industrial hubs, globalization and information technology would have theoretically enabled the spatial dispersion of economic activity instead. In a 1990 article, Michael Porter agreed that under a globalized world with instant communication, firms would be able to outsource any good or service from almost anywhere in the world at any time. He further argued that this meant firms' competitive advantage would become tied to knowledge and skills that can only be created by the clustering of industries in a certain region. By being geographically proximate to other actors in their industry value chain, this would lead to greater interaction between firms and their skilled workers, and this would enable more frequent and cheaper knowledge and resource sharing. Ultimately, this would support greater knowledge sharing, the exchange of innovative ideas, and appetite for risk-taking. There have subsequent papers examining the role of cities specifically in conglomerating and diffusing industry knowledge (Glaeser et al., 1992; Black and Henderson, 1999; Glaeser, 1999).

These earlier theories have faced some empirical scrutiny in subsequent decades. One notable study found that geographically agglomerated firms benefit from measurably higher productivity (Combes, et al., 2012). The authors of the 2012 study attempted to decompose the causes of this productivity benefit into "firm selection effects" – or increased competitive pressure pushing out less productive neighboring competitors – and "agglomeration effects"; which accounted for productivity-enhancing interactions that occurred due to the increased local density. This supports the contention of earlier authors that the geographic proximity of firms creates tangible benefits to all parties involved through more frequent interactions and knowledge transfer between firm agents. However, while the 2012 paper looked at the broader economic competitiveness of firms, the framework developed by its authors was not applied specifically to patent innovation.

Ten years after Porter's article, a book by AnnaLee Saxenian offered a more complete theory of what these value-adding interactions between firms would look like (Saxenian, 2000). In her book, she noted that both Silicon Valley and the greater Boston area Metro were in the cutting edge of software and computer hardware during the 1970s and 1980s. However, in the following two decades Silicon Valley's software industry would vastly surpass Boston's in both economic output and innovativeness.

Saxenian argued this was because Boston's software industry was dominated by several large firms that stifled the movement of human capital and knowledge with strict non-compete agreements and tight control of intellectual property, boxing in their skilled labor forces and limiting their ability to move between firms or establish new companies altogether. On the other hand, the unenforceability of non-compete agreements in California, combined with a unique entrepreneurial risk-taking culture in the Bay Area led to the constant movement of talent and into newer, more innovative companies that disrupted the most recent disruptors. Silicon Valley's ability to become the global leader in high-tech innovation was dependent on cultural qualities specific to its regions which enabled greater mobility of skilled labor and knowledge between established and upstart firms.

Under Saxenian's argument, skilled individuals who are typically tied to a larger parent institution act as agents of innovation. To bring their ideas to market, they will need to rely on either support within their home institution or interactions with those outside of their firm to bring their ideas to market. In this case, the strict walls between the workforces of Boston's major software and computer companies precluded the second form of individual interactions from occurring. By testing the role of individual units of population and population density in driving innovation within a specific geography, Berkes and Gaetani's 2017 paper empirically validated many of Saxenian's arguments about population-driven innovation.

My project builds on the preceding work in this field to argue that geographic clusters of innovation stem from both interactions among skilled individuals within a population, *and* interactions between firms. By measuring both firm and population variables' relationship with patent innovation within specific geographies, I will be able to isolate two separate affects due to firm agglomeration and population agglomeration. Not only would individual skilled professionals be exchanging and developing innovations with other individuals, but firms would be behaving as individual agents in exchanging knowledge with others as well.

Furthermore, my theory presumes that greater interconnectedness within a local business community leads to a business culture that spurs greater private sector innovation. This would occur because more frequent and incidental interactions between local industry stakeholders leads to more frequent exchanges of ideas and expertise. Finally, geographic proximity facilitates more regular exchanges of ideas and expertise between employees at all levels of business – from product development to the C-suite – due to regular interactions and the transfer of talent between firms, and into altogether new startups.

The relationship between the density of a local business community – as measured by corporate density and the connectedness of local corporate boards – and the formation of corporate-backed nonprofits has been examined in another study (Marquis and Davis, 2013). This study applied a theoretical framework similar to mine in arguing that greater inter-firm interaction proliferated support for nonprofit in a local corporate community.

In addition to these interactions between individuals as a catalyst for the creation of new innovations, I theorize that the interactions between geographically proximate firms – separate from interactions between individuals who might be employed by these firms – also catalyzes the development and implementation of new ideas. For example, the geographic proximity of local competitors could also spur greater local innovation, as firms would become more quickly cognizant of their competitors' innovations and seek to quickly emulate or out-compete them. In addition, when producers are located more closely to their customers – therefore allowing them to serve a greater volume of buyers with more direct communication – they would likely become more attuned to customer's needs and will more quickly develop products and features to better serve them. Geographical proximity could also simply facilitate more ordinary business transactions between firms than would otherwise occur, creating an additional vector for the transfer of knowledge between these organizations.

Data and Methodology

Methodological Framework

To empirically test this theoretical framework, I am presented with two immediate problems: how to meaningfully measure innovation, and how to meaningfully measure the density and interconnectedness of a city's business community. In this section, I will discuss the methodological challenges posed by this project, and the methods used in previous studies that offer me potential solutions. Then, I will develop and justify my own methodological framework.

The challenge in measuring innovation is that it is expressed in numerous ways, and it is difficult to objectively test how novel or groundbreaking a new invention or scientific development can be. Patent filings and awards can offer a partial picture, as inventions must go through a rigorous vetting process to verify their novelty. However, patents also fail to capture the entirety of new scientific and technological development as many firms and individuals opt to not to register their inventions as patents. That said, there exist techniques to assess the novelty of a patent based on what other patents cite it, and what existing patents this new patent cites (Uzzi, 2013; Berkes and Gaetani, 2017; Shai, et al., 2016).

Another approach has been to measure the amount of inbound venture capital funding as a means to test how commercially viable new innovations are (Waters, 2018; Shai, et al. 2016). However, this causes new issues: certain sectors are more prone to venture capital investment at a given point in time, it would take several years for the true novelty and commercial viability of a VC-backed venture to succeed. Finally, using a VC-based metric would constrain my analysis to smaller and younger firms, when I am trying to understand the innovation dynamics of regional economies in their entirety.

Given my focus on replicating Berkes and Gaetani's model from 2017 and the acknowledged limitations of existing measures of innovation, I will use patent density normalized to their level of novelty as my measure of innovation. Admittedly, this is an imperfect metric, but by replicating existing literature as closely as possible I will be able to better discern the impacts of changes in my dependent variables, and patent density offers an arguably broader definition of innovation. To control for other extraneous factors impacting patent rates in a given region beyond those being tested, I plan to also apply a fixed effects model to control for these complicating factors.

As mentioned earlier, there is an additional challenge of reliably measuring how clustered firms are within a given geography, and how frequently they would interact. While it would be relatively straightforward to determine the density of publicly traded corporations within specified geographies, such a measure alone is insufficient in determining the less apparent qualities of inter-firm interactions. However, it is difficult to find quantifiable data that can reliably measure something as qualitative as the business culture within a geography. For example, while survey research can be tremendously useful in discerning the cultural inclinations of a large population, the business community of a specific city would be too narrow to reliably survey given the resources available in this project, for example. Fortunately, other publications offer some potential solutions.

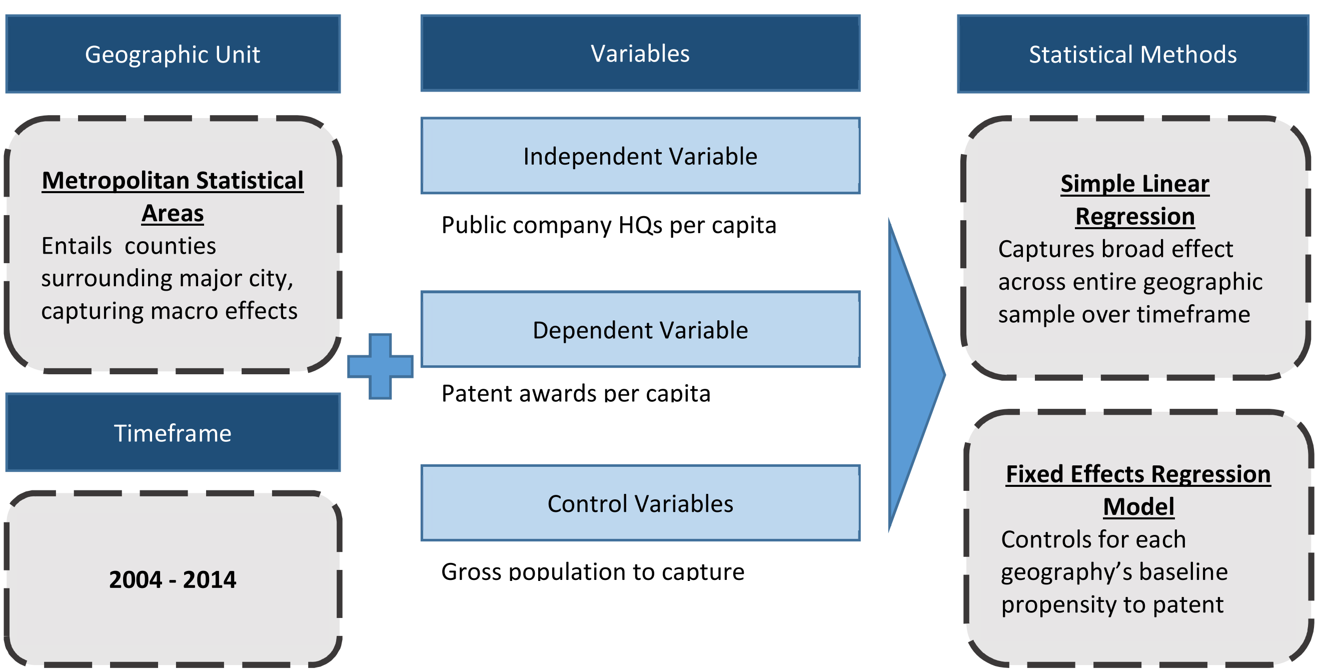
One 2013 paper which studied the relationship between local corporate communities and nonprofit creation offers a potential methodology to quantify a metropolitan area's business culture (Marquis, et al., 2013). The authors of the paper sought to determine whether a region's corporate culture can be objectively measured. The paper identified corporate density, social infrastructure, and cultural infrastructure as significant independent variables for understanding the local business community's connectedness. Corporate density was measured as the number of publicly-traded firms headquartered in a given metropolitan area, and social infrastructure was measured by calculating the network cohesion of corporate board directors within the same metro area. This paper also controlled for other variables by applying a fixed effects model that examined changes within individual geographies over a predetermined period of time.

While this methodology does not fully capture the extent of business density or interconnectedness within a community – as it would exclude smaller or privately-owned firms – it is still a robust approach for my own project. This is because it is able to apply uniform criteria across a large number of firms and geographies, with readily available data over many years. In addition, by using a narrower criteria for firms by limiting it to larger publicly-traded companies, I am able to filter out smaller service-based companies that would be outside of the scope of my hypothesis relatively cleanly. Furthermore, while this paper's definition of "social infrastructure" only factors in the interconnectedness of corporate directors and not other high-knowledge individuals within my theoretical framework, it would still function as an effective proxy for how much interaction occurs between employees of different firms within a community. Finally, I will be able to control for other intangible cultural and economic variances between the geographies in my sample by applying a fixed effects model over a set period of time.

Study Design

This study will use metropolitan statistical areas (MSAs) as its geographic unit of analysis, and will seek to establish a relationship between corporate density (as measured by the number of public company headquarters per capita) and patent density, while controlling for gross population. The time frame of study will range from 2004 to 2014, and will apply both a simple linear regression model across each year-MSA observation, and a fixed effects model to control for differences between each MSA that are not captured by the control variables. To move to a more granular geographic unit of analysis, the same statistical models will be replicated at the county statistical division (CSD) level following the MSA-level models.

Figure 1: Study Design



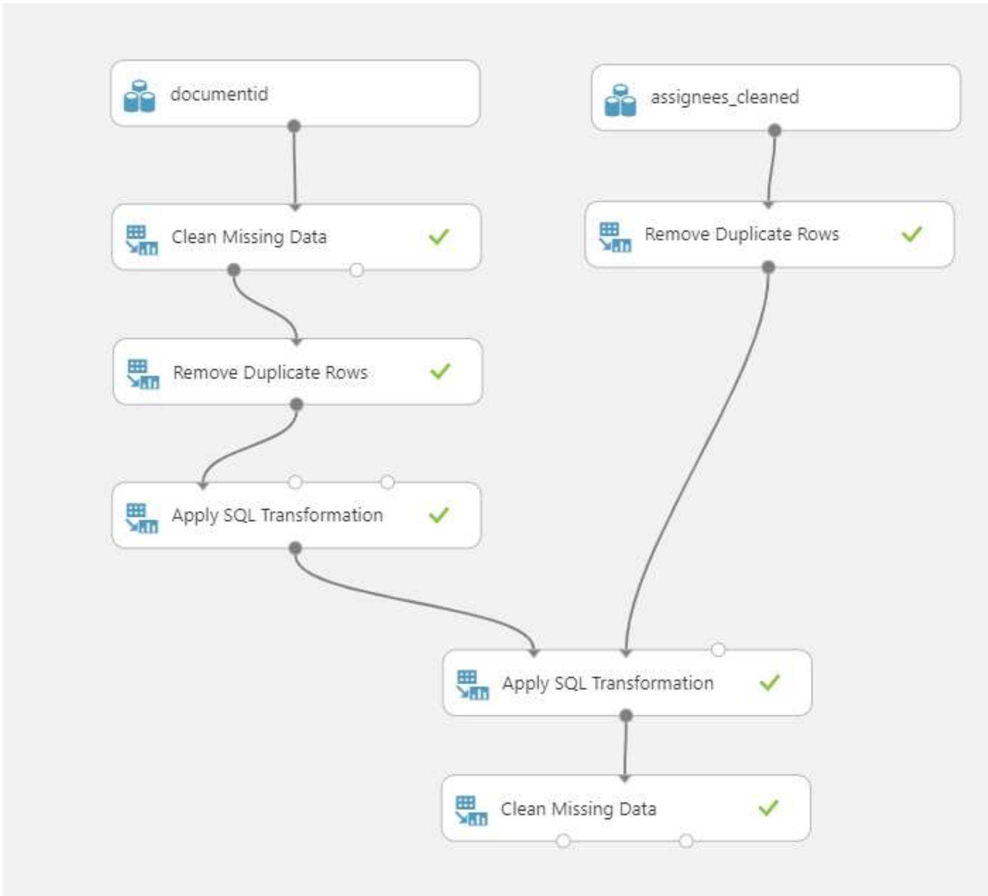
Data Sources and Cleaning

The data used for analysis came from a mix of publicly available federal and academic datasets.

Patent data came from the US Patent and Trademark office's open-source patent assignment database (Marco, et al. 2015). The database was filtered for applications that were awarded a patent, where the application occurred between 2004 and 2014, and where the first assignee is located in the United States. This left 1,125,726 applicable patents for examination. Based on existing literature, this method assumes that the chronologically first patent assignee would be roughly co-located with whoever be considered the "inventor". As a result, patents were matched to MSAs/CSDs by the ZIP code of their respective first assignee.

This cleaning was done in Microsoft Azure using SQL to merge the patent-specific data in documented and merge it with the cleaned set of initial patent assignees. A visualization of this workflow is seen in the following exhibit.

Figure 2: Microsoft Azure patent data cleaning workflow



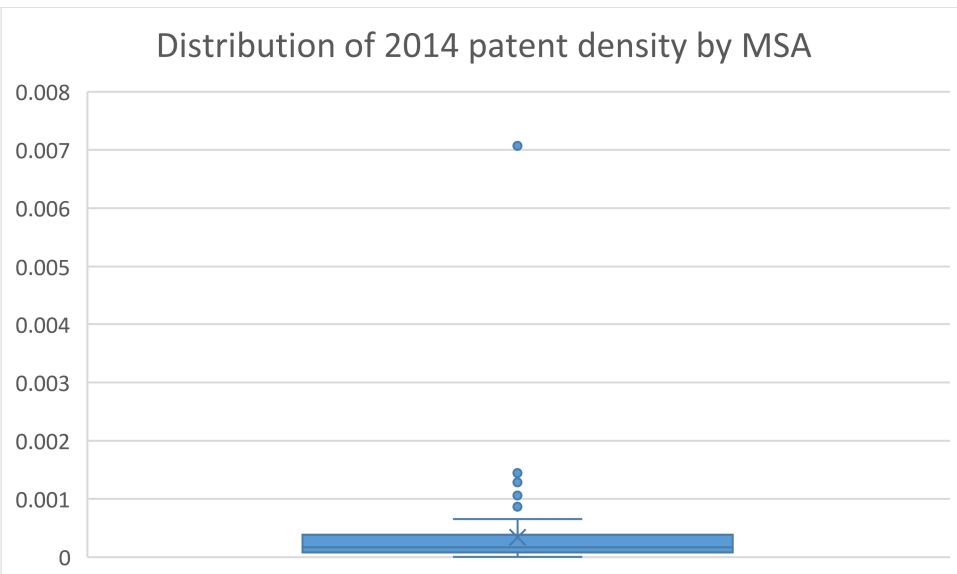
The addresses of all publicly-traded corporate HQs came from the academic WRDS database (Wharton Research Data Services) and was matched to respective MSAs and CSDs by using their ZIP codes. This yielded 175,456 unique corporate headquarter-year observations for analysis, 91,264 of which were located in the top 100 largest MSAs in the United States.

Population figures at the MSA level came from the US Census Bureau's annual community survey, and were matched to their corresponding FIPS codes. In cases where the Census Bureau's boundaries for an MSA expanded to cover a broader area, such as in Honolulu, Hawaii and Youngstown, Ohio, the older population figures were attributed to the most recent boundaries because the effect on the model was limited.

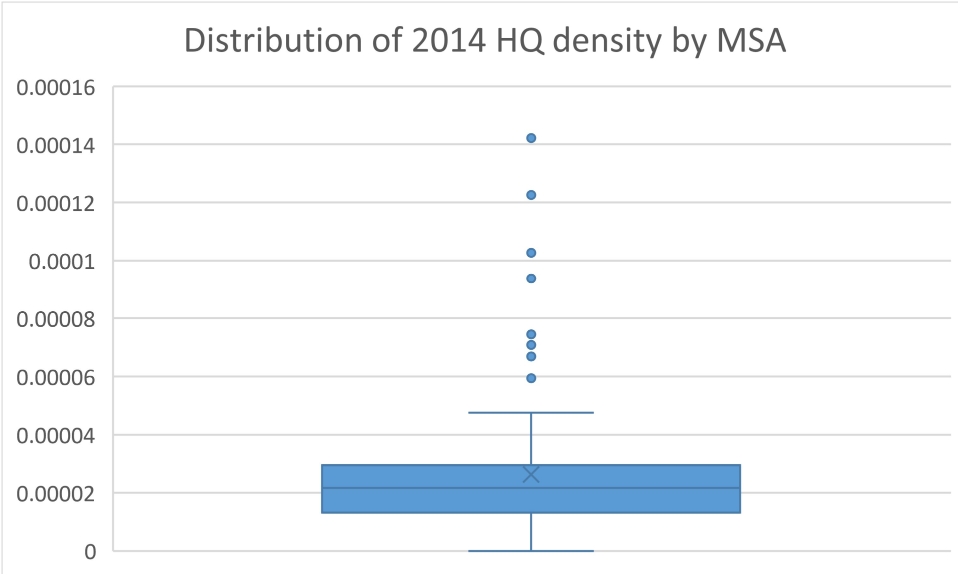
A further complication was found in mapping individual patents, units of population, and corporate headquarters onto their respective CSD. Because County subdivisions cut across multiple ZIP codes, corporate HQs and individual patent assignees cannot be cleanly mapped to a single CSD. However, using ratios of ZIP codes' portion of commercial activity dispersed to each CSD published by the US HUD, one can create a weighted number of corporate HQs and patents by CSD by multiplying the gross number per ZIP code by the portion applicable to each CSD.

Summary of Data

In the dataset used, patent density is dominated by a small number of “rockstar” cities whereas the majority of MSAs are clustered around a tight median. Examining the number of patents per capita across the 100 largest MSAs, the average MSA observed 3.4e-4 patents per capita, with a median of 1.7e-4 in 2014. However, San Jose-Sunnyvale-Santa Clara CA stands out as an extreme outlier, observing 7.1e-3 patents per capita in 2014. San Francisco-Oakland-Hayward CA, Boston-Cambridge-Newton MA, San Diego CA, Bridgeport CT, and Albany-Schenectady-Troy NY stood out as more modest outliers from the mean, with greater than 1.0e-4 patents per capita.



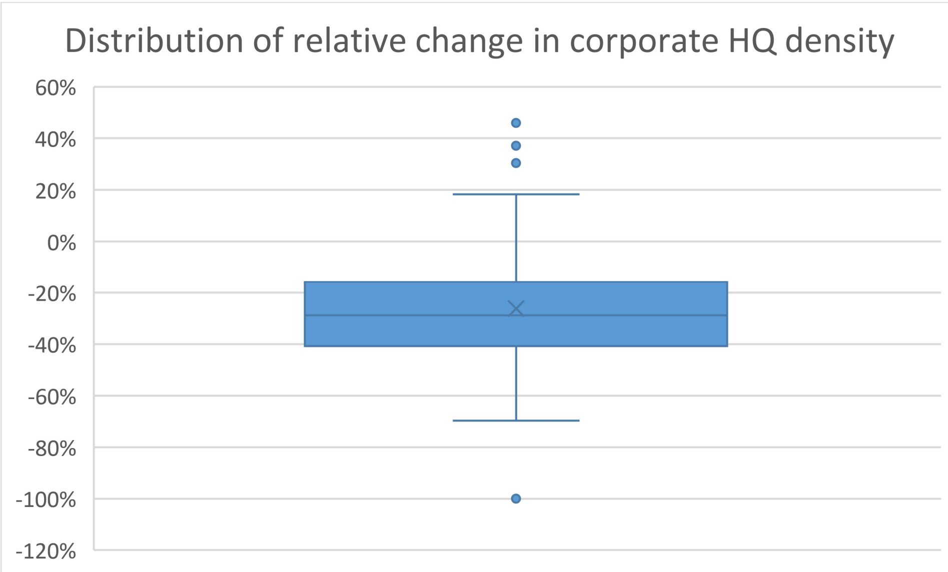
A similar pattern of stratification – albeit less extreme – is also observable when comparing the corporate density of MSAs. The average MSA had 2.6e-5 public company headquarters per capita and a median of 2.2e-5 in 2014. Outliers were relatively closer to the median than was the case for patent density, with San Francisco-Oakland CA, Boston-Cambridge-Newton MA, San Jose-Sunnyvale-Santa Clara CA, and Bridgeport-Stamford-Norwalk CT having the highest densities.



When comparing shifts in the data over time, this pattern of stratified outcomes between MSAs continued to be observed. Over the period of study from 2004 to 2014, sizable shifts in both patent density and corporate headquarter density across the MSAs that were examined. As expected, corporate density contracted in the aggregate as the number of publicly-traded firms has declined. Patent density did not shift as sharply or uniformly. Ultimately, this spread in outcomes creates a rich sample population to analyze the interplay of corporate and patent density.

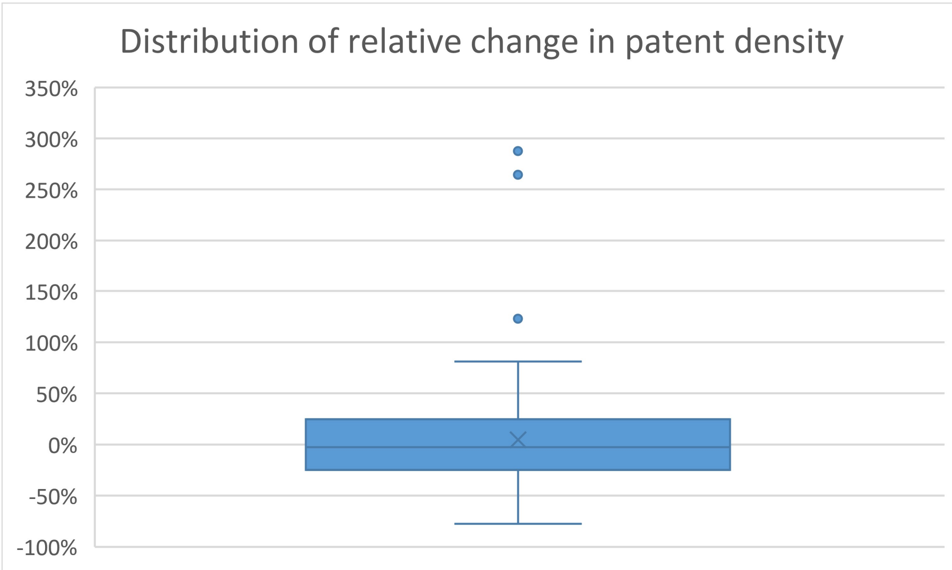
The median MSA saw a 29% contraction in the number of corporate headquarters per capita, with an average decrease of 26%. Only 11 of the 100 MSAs saw any increase in headquarter density. Unsurprisingly, several high-growth outliers overwhelmingly bucked this trend, including San Francisco-Oakland-Hayward (40% growth), Chicago-Naperville-Elgin (37% growth), and Washington-Arlington-Alexandria (30%). Many MSAs experienced an extreme decline in corporate density in excess of 50%, including Baltimore-Columbia-Towson MD (-53%), Riverside-San Bernardino CA

(-58%), Birmingham-Hoover AL (-50%), Albuquerque NM (-68%), Syracuse NY (-58%), and Lancaster PA (-51%).



The endogenous shock to the density of public corporate headquarters observed during this timeframe in numerous MSAs provides an ample natural experiment to examine whether the rapid shift in corporate density leads to a measurable effect on patent density as well.

No analogous uniform swing was observed with patent density. The median shift by MSA in patent density during 2004 to 2014 was -3%, with an average 5% growth in patent density. Several notable high outliers exist, including Atlanta-Sandy Springs Roswell GA (70% growth), San Diego-Carlsbad CA (81% growth), Denver-Aurora-Lakewood CO (64%), McAllen TX (288%), and Lancaster PA (264%). Notably large declines were observed in Fresno CA (-57%), El Paso, TX (-67%), Allenton PA (-65%), Boise ID (-78%), and Lakeland-Winter Haven FL (-71%).

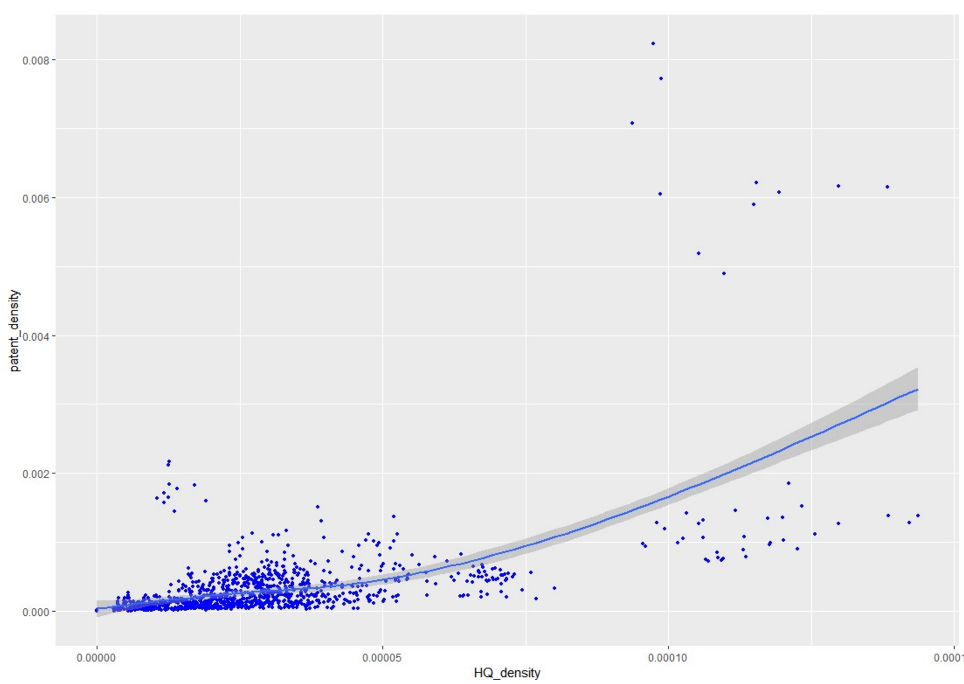

Distribution of relative change in patent density
350%
300%
250%
200%
150%
100%
50%
0%
-50%
-100%

Results

MSA-level results

A statistically significant and positive correlation was found between patents per capita and corporate headquarters per capita when plotting every year-observation for the 100 largest MSAs across 2004-2014. This is initially supportive of the theory that greater firm agglomeration leads to a more innovative environment within a geography. However, this simple regression is insufficient to reach any rigorous conclusions, and further analysis is necessary.

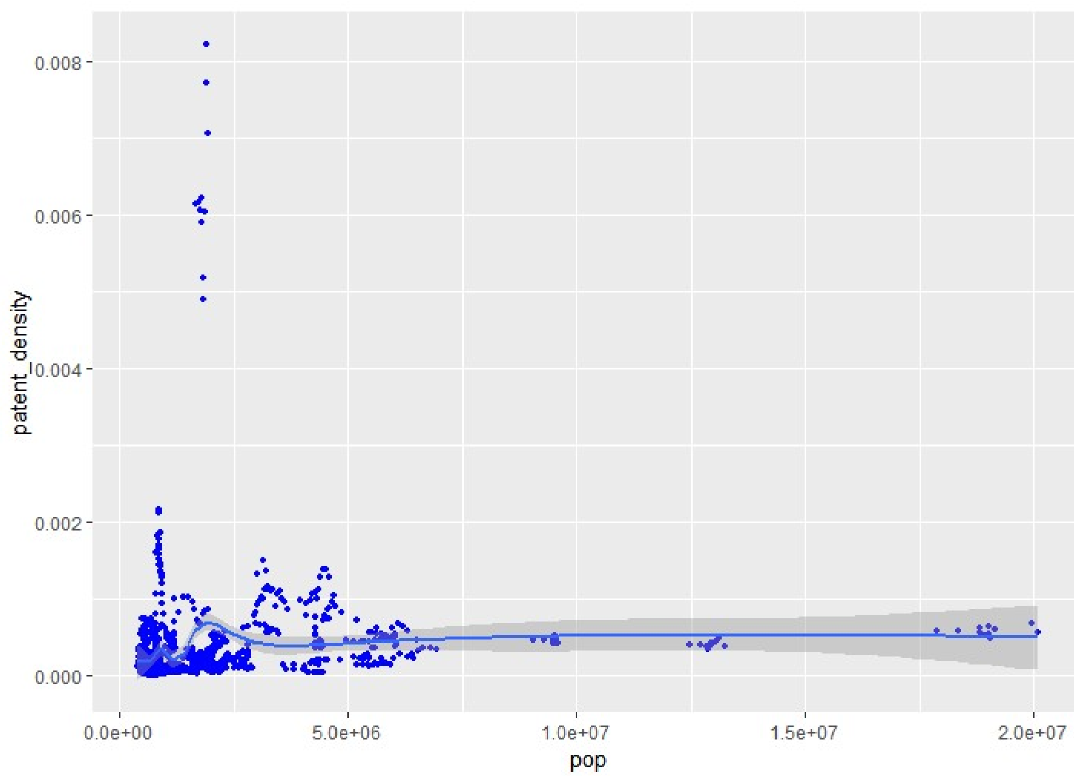
Figure 3: Smoothed curve of patent density vs. HQ density by MSA-year



As a robustness check to ensure corporate HQ density isn't just serving as a proxy for the absolute size of a given metro area, a multiple linear regression model was applied to the same dataset adding absolute population as an additional independent variable. Under this robustness check, a statistically significant but slightly negative relationship was found for absolute population size. A statistically significant and positive collinearity was found between gross population and headquarters density, suggesting larger metro areas are simply able to sustain greater corporate agglomeration.

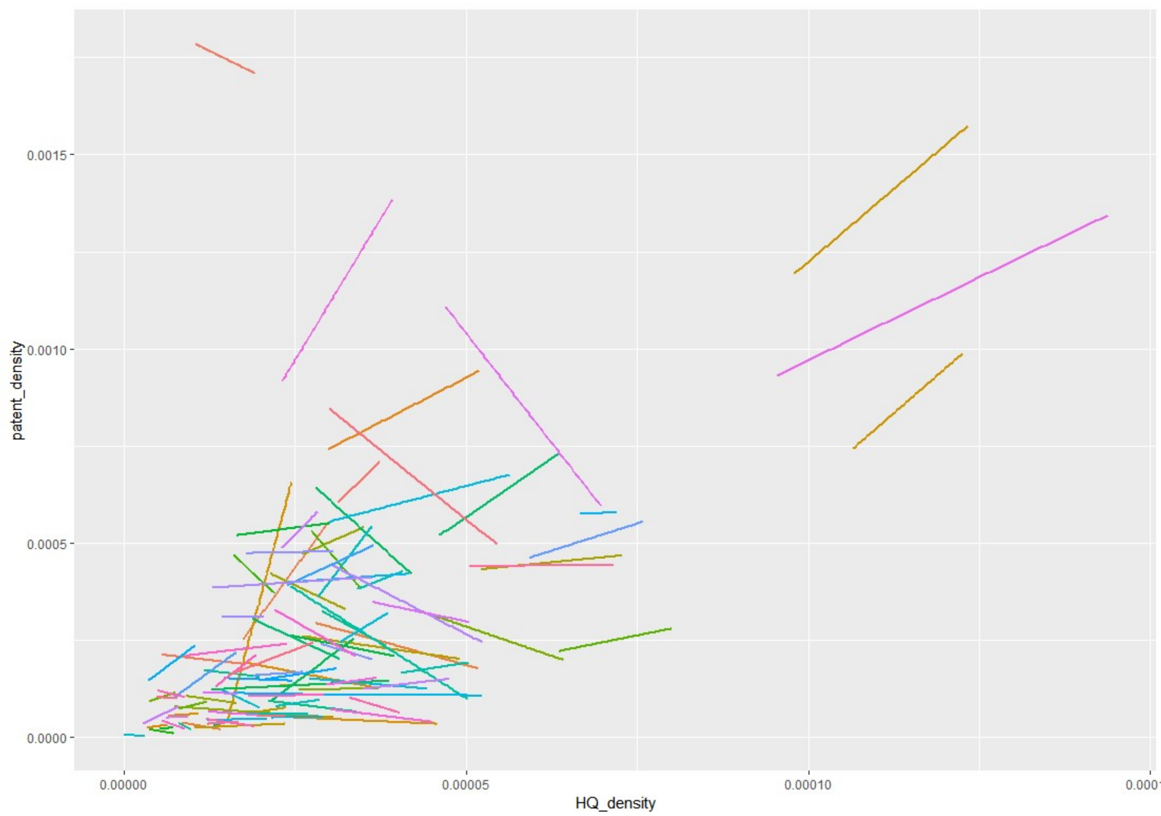
In addition, when plotting gross population against patent density, gross population size alone was much weaker than corporate density in explaining variations in patent density across the dataset.

Figure 4. Smoothed curve of patent density vs. gross population by MSA-year



However, this statistical significance found in the initial regression model disappears when switching to a fixed-effects model for each individual MSA over the same timeframe. In addition, the coefficient for corporate density became negative, running contrary to my theoretical framework.

Figure 5: Visualization of fixed effects regression by MSA



When running the same fixed effects model as a multiple regression with gross population, the statistical significance of corporate density faded even further, although the coefficient for corporate density became more positive.

This loss of significance is despite many of the MSAs experiencing an exogenous shock by losing a significant portion of their major corporate headquarters during the time period of study. These findings suggest that while differences between metro areas can be driven by differences in firm agglomeration, these effects are often Furthermore, they are indicative that MSAs might be too large of a unit of analysis to determine meaningful relationships between these variables as there is too much additional complexity within each metropolitan area that can't fully be captured in these regressions.

Finally, while the findings of the fixed effects do not support my hypothesis, they do not outright disprove it either. When compared to the findings of the linear model across all MSAs, it suggests that firm agglomeration within metro areas increases a local propensity for innovation, and that this tendency is resilient against short-term shocks to the number of firms in the region through the time period of we examined. To reach a further level of confidence, however, it is necessary to move to CSDs as a smaller unit of analysis.

Table 1. Regression coefficient outputs of MSA-level models

Variable	Model 1	Model 2	Fixed Effects Model 1	Fixed Effects Model 2
Intercept	-1.33E-04 ***	-1.08E-04 ***		
Firm Density	1.63E+01 ***	1.80E+01 ***	-8.08E-01	-4.52E-01
Gross Population		-3.72E-11 ***		3.52E-11
R-squared	0.3040	0.3214	0.9759	0.9759

Signif. codes: 0 *** 0.001 ** 0.01 *0.05 . 0.1 ‘ ’ 1

"Rockstar" cities and innovation

Unexpectedly, the fixed effects regression model identified 34 "star" MSAs out of the sample which exhibited a consistent overperformance of their expected baseline patent density, and this overperformance was highly statistically significant, with a p-value less than 0.001. The 34 MSAs were as follows:

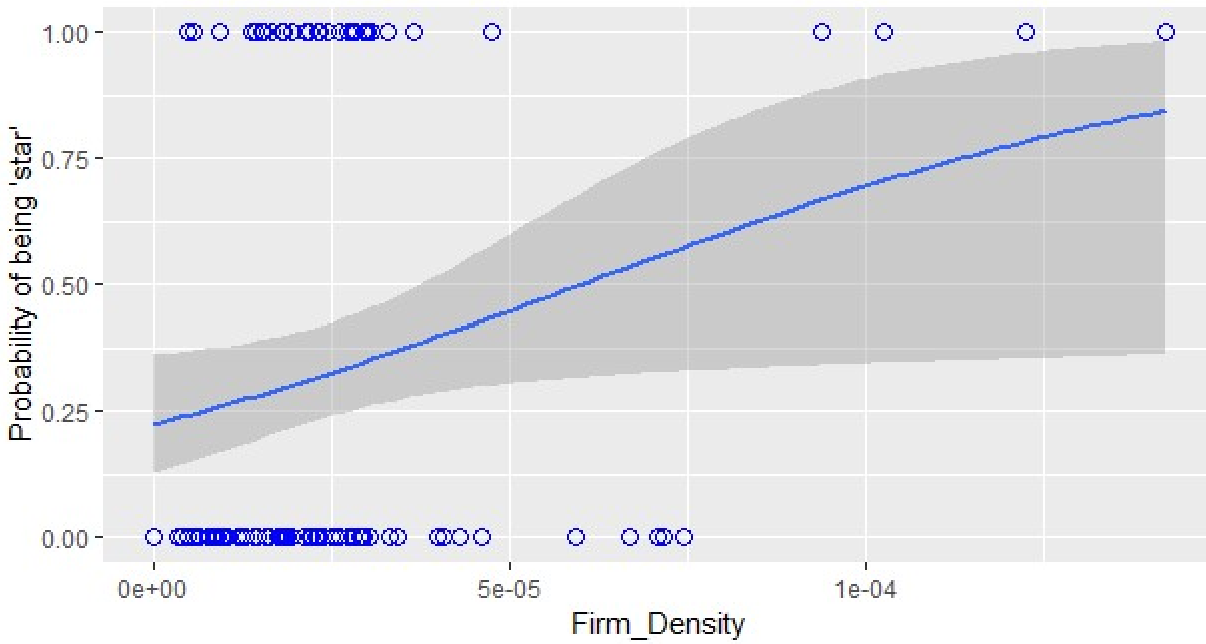
Table 2. "Star" cities

1. San Jose, CA	8. Boston, MA	15. Des Moines, IA	22. Boise, ID	29. Greensboro, NC
2. Albany-Schenectady, NY	9. Worcester, MA	16. Madison, WI	23. Raleigh, NC	30.Syracuse, NY
3. Bridgeport, CT	10. Akron, OH	17. Cleveland, OH	24. Allentown, PA	31. Winston-Salem, NC
4. San Francisco, CA	11. Hartford, CT	18. Oxnard, CA	25. Cincinnati, OH	32. Albuquerque, NM
5. Seattle, WA	12. Grand Rapids, MI	19. Portland, OR	26. Salt Lake City, UT	33. Toledo, OH
6. San Diego, CA	13. Minneapolis, MN	20. New Haven, CT	27. Lancaster, PA	34. Ogden, UT
7. Austin, TX	14. Rochester, NY	21. Provo, UT	28. Providence, RI	

The findings of this model suggest that many of these metro areas possess unique traits outside of this scope's study which allow them to so consistently outperform the patent density that is expected of them based on their firm density and population. While many of these MSAs are not shocking, including San Jose, San Francisco, and Austin, TX, some of the MSAs included are somewhat more surprising, such as Toledo, OH and Lancaster, PA. As a result, a new question emerges: what do all these "star" MSAs have in common that allows them to outperform their peers so consistently?

To study this question, a logistic regression is applied to all MSAs in the sample, to determine how strongly firm density and population predict whether a given MSA is classified as a "star". Based on the regression outputs, illustrated in the following chart, firm density was highly predictive – though not decisive – in determining whether a given MSA would be a consistent outperformer. A further logistic regression controlling for population would also find that firm density's predictive qualities persisted.

Figure 6: Logistic regression of P(star) vs. firm density



While these findings do not conclusively prove a causal link between whether a metropolitan area would be an overperformer in per capita patent innovation, these results are still highly informative. First, they give further credibility to the theory that while firm density itself may not act as a primary driver for immediate growth in patent density, greater firm density could lead to the development of a broader infrastructure and ecosystem that is conducive to greater local innovation.

Table 3. Logistic regression coefficient outputs

Variable	Logit Model 1	Logit Model 2	Logit Model 3
Intercept	-1.2365 ***	-2.91E-01	-9.29E-01 *
Firm Density	20624 *		4.96E+04 **
Gross Population		-2.12E-07	-5.96E-07 **

Signif. codes: 0 *** 0.001 ** 0.01 *0.05 . 0.1 ' ' 1

CSD-level results

While the original study design called for the preceding analyses used on MSA-level data to be applied to CSDs, computing and data constraints were encountered which prevented this from being completed. With over 40,000 individual CSDs within the sample, and demographic, corporate, and patent data syndicated from a broad set of sources, individual corrupt data points would trigger fatal errors in the computing software, and the large datasets made it obstructively time consuming to troubleshoot the root cause. Under a future study, it would be ideal to re-attempt this analysis to better discern local-level interactions between corporate and patent density.

Discussion

MSA-level results

This paper's analysis has found a statistically significant relationship between firm density and per capita patent density at the MSA-level that cannot easily be explained away by gross population. This finding suggests that there is indeed a positive effect to regional innovation caused by firm agglomeration that is separate and measurable from the effect of population agglomeration.

However, this is complicated by the findings in the fixed-effects MSA model. In finding that the correlation for corporate density and patent density becomes negative under a fixed effects model, and with a p-value well beyond statistical significance, a clear immediate connection between corporate density and patent density can’t be drawn. Nonetheless, this also does not entirely invalidate my hypothesis that firm agglomeration still acts as a driver for innovation.

When reconciling that a regression across all MSA-year observations found a statistically significant relationship between firm density and patent density with the findings of the fixed effects, this suggests several possibilities that are not mutually exclusive with the earlier theory. First, these results suggest that firm agglomeration within a region may lead to a regional propensity for innovation that is inflexible to short-term shocks to firm density, at least the decade-long timeframe being examined. This phenomenon suggests that firm agglomeration may support factors that drive regional innovation, such as the clustering of skilled human capital, which can outlast the firms themselves. A second possibility is that there is still a direct, measurable effect of firm agglomeration regional innovation, but when limiting the unit of analysis to individual

MSAs under a fixed effects model for 10 years, there are too many exogenous variables over too small a sample to find a statistically significant connection.

Furthermore, in identifying “star” MSAs which consistently outperformed their firm-density-implied patent baseline, a new avenue of interaction between firm agglomeration and regional innovation is raised as a possibility. The finding that a specific subset of MSAs consistently outperform their expected patent density suggests that regional innovation is heavily driven by factors that were not tested in this study, such as population education, the presence of major universities, or the presence of specific industries. However, by identifying that greater firm density was predictive of whether a given city would be a star performer, these findings suggest that firm agglomeration could be conducive to building this necessary innovation infrastructure. Once this infrastructure is present patent density does not immediately move in conjunction with changes in firm density. That said, it must be noted that this finding also does not establish a causal link, and further study will need to be conducted to prove this theory.

Ultimately, while there are advantages to studying the interaction of these variables at the metropolitan level, this unit of geographic analysis is so large that it is difficult to disentangle the many economic and demographic factors that could affect patent density aside from firm density. As a result, a more granular analysis of patent density at the CSD level will be able to offer more nuanced findings.

Limitations

Despite the findings of this paper, the limitations of the methodology used must also be mentioned to contextualize potential shortcomings in its conclusions.

First, the units of geographic analysis used in this study suffer from both a lack and excess of precision. MSAs, while encompassing almost the entirety of a metropolitan area's economic activity, are too large and cumbersome as divisions to segment local demographic and economic ecosystems. As a result, regression analysis using MSAs is able to capture the entirety of a region's innovativeness and business concentration but fails to offer a more nuanced and granular decomposition of local systems that may contribute to its patent density.

CSDs, on the other hand, are much smaller and therefore are better able to capture block-level dynamics within a metropolitan area that contribute to innovation, and are small enough for population density to become a meaningful measure. However, these units are often more precise than the geographic information available for pinpointing the "inventor" of patents. There is a level of imprecision that is expected with the filing address of patent assignees which we use to approximate the location where a patent was "invented". For example, a piece of intellectual property may be developed in a corporate lab in Pittsburgh that falls within one CSD, but the assignee applies from an address in an adjacent CSD in Pittsburgh. As a result, the patent under this study's methodology would be recorded in the "wrong" CSD while still being recorded in the correct MSA.

Ultimately, these geographic units of analysis cannot be easily replaced with an alternative as almost all data can only be referenced under these boundaries. However, in relying on both macro-level MSAs and granular CSDs, and applying robust statistical methods across a sufficiently large time sample, these imperfections can be mitigated to an extent.

A second key limitation of this study is its potentially incomplete measure of firm agglomeration within a region. While solely measuring the number of publicly-traded company headquarters within each geography of analysis applies a minimum filter for scale and an easy reference to its location, this fails to capture the large number of privately-held mature companies and startups across the country. In addition, by solely looking for corporate headquarters, this overlooks the major operations of firms that are located outside of their headquarters region. While existing literature has found that many companies' headquarters and key research facilities tend to be located near each other (e.g., Ford Motor Company's Dearborn research center), there are numerous examples where this is not the case. By examining only headquarters, many of these satellite operations – where significant R&D and product development may occur – are not measured.

However, this is partially mitigated by a number of factors. First, corporate headquarter density can be viewed as a proxy for how many companies have major operations within a region – even if a major company has a large research campus in one metro area separate from its headquarters, this research hub could still draw other firms to the region, which would be captured under this study's methodology. In addition, the filtering based on publicly traded company headquarters ensures a uniform criteria of location assignment and places a rough minimum on the scale of the firms being measured.

The third key limitation of this study is the time scale being examined. By limiting the time series to 2004 through 2014, we can minimize the effect of changes in norms around corporations filing patents, and issues with the availability and consistency of patent data. However, the findings from the regression and fixed effect regression across the MSA-year observations suggests that while firm density is a statistically significant driver of patent density, it takes well over 10 years for a statistically discernable impact to emerge on patent density following an exogenous shock to corporate density.

The final limitation of this study's methodology is the imperfection of using patents as a measure of innovation. While patents are subjected to a standard vetting process to ensure their newness over other existing documents, not all innovations are filed as patents (certain industries prefer to maintain trade secrets instead, for example), and some granted despite having a limited actual impact. That said, the application of a fixed effects model is intended to control for variances in regional industries' propensity to file their innovations as patents. In addition, the vetting process for awarded patents ensures a minimum level of improvement over existing knowledge has been achieved. Further inquiry should seek to find a means to normalize patents on their innovativeness based on the pattern of preceding patents they cite and subsequent patents that cite it.

Areas of future inquiry

Most immediately, the findings of this study raises the question of what distinguishes "star" MSAs from median and laggard-performers in patent density. While initial analysis suggests that higher firm density is likely in creating a "star" performer in regional innovation, no causal link has yet been established, and there are likely other factors beyond firm density and population that are predictive of whether an MSA becomes a "star". A future inquiry should design a study which builds a more comprehensive model to calculate the baseline expected patent density using an expanded set of infrastructure and demographic variables (e.g., the presence of major universities, the presence of major labs, and the portion of the population holding a college degree). Using the baseline established in an expanded model, a future study could help better isolate what factors could be driving whether a given metropolitan area becomes a “star” innovator.

In addition, future inquiries should seek a natural experiment which can better illuminate whether a causal link can be established within the theoretical framework laid out in this paper. While this study sought to study how shifts in firm density over the 10-year study period could shift patent density, this was an inherently endogenous shift and therefore any statistically significant relationship established between these variables could not prove such a link was causal. To allay this issue, a future study should seek distinct exogenous events where shifts (or the lack thereof) in patent density can be measured. One potential avenue of exploration in this would be to study a sample of metropolitan areas experiencing major corporate relocations (either away from the region, or into the region) due to exogenous economic shocks such as the 2007 financial crisis, and following the patent density after the shift.

A final potential avenue for future inquiry is to replicate this study with a firm-based unit of analysis. Specifically, it would be worthwhile to break down firms within a region by year-cohorts of how long they have operated within a given MSA, and examine how their propensity to patent shifts over time as the local firm density moves. By taking this firm-cohort based methodology, one can better establish a baseline for local patent density, and examine how the endogenous shift in local firm density will affect long-standing incumbent firms’ tendency to patent.

Conclusion

Ultimately, the findings of this study suggest that the agglomeration of firms within a geography leads to a discernable impact on the firm's innovativeness that can be separated from the impact of population factors. The identification of specific MSAs which consistently (and statistically significantly) outperform their expected patent density, and the linkage of this "star" status to greater patent density, further lends credence to the theory that greater firm density is conducive to greater regional innovation. However, the fixed effects model across the time series raises questions about how immediate this connection is, and how sensitive it is to short-and-medium-term shocks.

While this study failed to establish a causal link between firm and patent density, it raises the possibility that an denser agglomeration of firms within a region supports the development of local conditions and infrastructure that is conducive to more intense local innovation. For example, the clustering of firms within a region could support a deeper pool of human capital, as well as strong local research, educational, and entrepreneurial infrastructure to drive local innovations. This new theory would also explain why firm and patent density would not move in conjunction under the 10-year fixed effects regression model.

Finally, it must be acknowledged that innovation and its underlying drivers are complex, and impossible to fully capture in purely quantitative measures. Especially when analyzing innovation at a regional level, so many factors – economic, demographic, and perhaps freak chance – interplay that it is immensely challenging to isolate any singular root causes. That said, the findings of this study – particular the identification of “star” cities and laggards – opens up a rich new set of questions to be examined in further research.

References

1. Berkes, E., & Gaetani, R. (2017, June 21). The Geography of Unconventional Innovation.
2. BERNSTEIN, S. , GIROUD, X. and TOWNSEND, R. R. (2016), The Impact of Venture Capital Monitoring. The Journal of Finance, 71: 1591-1622. doi:10.1111/jofi.12370
3. Black, Duncan and Vernon Henderson. 1999. “A Theory of Urban Growth.” Journal of Political Economics, 107(2): 252-85.
4. Chu, Johan and Davis, Gerald. "Who Killed the Inner Circle? The Decline of the American Corporate Interlock Network," American Journal of Sociology 122, no. 3 (November 2016): 714-754.
5. Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga and Sebastien Roux. 2012. “The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection.” Econometrica, 80(6): 2543-2594
6. Glaeser, Edward L., Hedi Kallal, Jose’ Scheinkman and Andrei Schleifer. 1992. “Growth in Cities.” Journal of Political Economics, 100(6): 1126-52.
7. Glaeser, Edward L. 1999. ”Learning in Cities.” Journal of Urban Economics, 46(2): 254-77
8. Marco, Alan C., Graham, Stuart J.H., Myers, Amanda F., D'Agostino, Paul A and Apple, Kirsten, The USPTO Patent Assignment Dataset: Descriptions and Analysis (July 27, 2015)

9. Marquis, C., Davis, G., Glynn, M. (2013) Golfing Alone? Corporations, Elites, and Nonprofit Growth in 100 American Communities. Organization Science 24(1):39-57.
10. Porter, M. (1998) ‘Clusters and the new economics of competition’, Harvard Business Review, November/December, pp. 77–90.
11. Saxenian, A. (2000). Regional advantage: culture and competition in Silicon Valley and Route 128. Cambridge, Mass.: Harvard Univ. Press.
12. Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Benjamin Jones. 2013. “Atypical Combinations and Scientific Impact.” Science, 342(6157): 468-472.
13. Wharton Research Data Services. "Compustat Annual Updates " wrds.wharton.upenn.edu, accessed 02/05/2019.