The Geography of Unconventional Innovation

Enrico Berkes Ruben Gaetani*

November 19, 2015

(Download the most recent version)

Abstract

This paper studies the role of population density as a driver of innovation. Using a newly assembled dataset of georeferenced patents in the U.S., we show that overall innovation activity is not concentrated in high-density areas as commonly believed. However, when we restrict attention to unconventional innovations - innovations based on unusual combinations of existing knowledge - we show that these are indeed more prevalent in high-density areas. To interpret this relation, we propose the view that informal interactions in densely populated areas help knowledge flows between distant fields, but are less relevant for flows between close fields. We then provide evidence supportive of this view. We build a model of innovation in a spatial economy that endogenously generates the pattern observed in the data: specialized clusters emerge in low-density areas, whereas high-density cities diversify and produce unconventional ideas.

JEL Classification: O33, O40, R11, R12

*Northwestern University, Department of Economics. 2001 Sheridan Road, Evanston, IL, 60208. Enrico Berkes: enrico.berkes@u.northwestern.edu; Ruben Gaetani: gaetani@u.northwestern.edu. We thank for their comments Treb Allen, David Berger, Lawrence Christiano, Matthias Doepke, Martin Eichenbaum, Benjamin Jones, Lorenz Kueng, Matteo Li Bergolis, Lee Lockwood, Guido Lorenzoni, Kiminori Matsuyama, Nicola Persico, Giorgio Primiceri, Jorg Spenkuch and seminar participants at the 2015 Meeting of the Society for Economic Dynamics in Warsaw, 2015 World Congress of the Econometric Society in Montreal, Federal Reserve Bank of Chicago, the Conference of Swiss Economists Abroad 2014 and Northwestern Macro Lunch. Comments are welcome, errors are ours.

1 Introduction

The idea that informal interactions among people are central to innovation and knowledge diffusion has become a cornerstone of recent theories of economic growth (Lucas 1988). If true, the idea implies that economic geography, by controlling the extent of those interactions, should play a first-order role in determining the creation and diffusion of ideas. Specifically, we would expect that innovation should cluster in high density areas, and that cities should be a key engine of technological progress. There exists a sizeable literature on the role of cities and agglomeration for growth that builds on this intuition (Glaeser et al. 1992, Black and Henderson 1999, Glaeser 1999).

In this paper, we examine the link between density and innovation empirically, using narrowly georeferenced information on patenting in the United States. Our geographically disaggregated data show that the advantage of cities in producing innovation is more nuanced than commonly believed. While suburban areas are responsible for a substantial share of overall innovation activity, high-density places disproportionately generate innovation with a high degree of unconventionality. This pattern, which to the best of our knowledge has never been documented, reconciles the intuition that density fosters creativity with the observation that the origin of innovation in the U.S. is dispersed. We then propose a spatial theory of a knowledge-based economy that is consistent with our findings. The theory highlights a novel rationale for why economic activity agglomerates in cities of different sizes and degrees of diversification. This rationale is grounded in the process of knowledge creation and reconciles the tension between returns to local specialization (Marshall 1890) and returns to diversity (Jacobs 1969), without relying on the presence of ex-ante heterogeneous agents. We use our model for policy analysis, and find that that a system of place-based subsidies can have a significant impact on aggregate welfare by changing both the intensity and composition of innovation activity.

The role of geography in shaping innovation outcomes has been receiving growing attention in recent years. As argued by Moretti (2012), this emphasis reflects the fact that knowledge creation has gained increasing importance for the economic success of advanced countries. The theoretical link between innovation and geography relies on the intuition that spatial proximity is a key channel for the diffusion of ideas. Although a broad empirical literature has shed light on several aspects related to this intuition (Jaffe et al. 1993, Audretsch and Feldman 1996), direct assessments of its importance in simultaneously affecting technological change and economic geography are still relatively scarce. Previous research has relied either on a high level of geographical aggregation like MSAs (Carlino et al. 2007) or on indirect evidence such as differences in skill premia and skill mix (Davis and Dingel 2013).

In this paper, we look at variation in patenting across narrowly defined geographical units in the United States. We start by collecting the full-text record of all the patents granted by the USPTO in the years 2002-2014, that we then geo-reference at the County Sub-Division (CSD henceforth) level. We document four novel facts on the spatial distribution of innovation in the United States.

First, the role of high-density regions as engines of innovation is smaller than commonly thought. Over 40% of the patents in our sample originate from units with density of population below $1,000/\text{km}^2$.[1] Unsurprisingly, density is positively related to innovativeness along the extensive margin: more densely populated places are more likely to host permanent innovative activity. However, among continuously innovative CSDs - defined as units that produced at least one patent per year - innovation *intensity* (measured as patents per capita) and population density display a weak relationship that, if anything, points towards a negative slope. Density and patenting have a zero correlation across places with density above $500/\text{km}^2$ and are negatively related across areas with density above $1,000/\text{km}^2$.

Second, innovation produced in densely populated areas is more likely to be built upon unconventional combinations of prior knowledge. To show this fact, we formulate a notion of technological distance that proxies for the intensity of idea flows between fields, based on the observed network of patent citations. We implement an algorithm in the spirit of Uzzi et al. (2013) to evaluate the atypicality of references listed in each patent. Our measure compares the observed frequency of each pairwise combination of citations with the frequency one would expect if references were distributed at random. This procedure defines an index of conventionality (*c-score*) for each citation pair: combinations are unconventional if their empirical frequency is small compared to their random frequency. The c-score ranks inventions along a dimension that we argue to be economically meaningful: first, unconventional patents are significantly more likely to be highly cited compared to conventional ones; second, unconventional patents are significantly less likely to be produced by large, publicly traded firms. We find that unconventional innovations tend to originate disproportionately from densely populated areas. This relationship is statistically and economically significant, emerges both in patent-level and CSD-level regressions and is robust to a wide variety of specifications.

Third, dense cities host a more diversified pool of learning opportunities. Computing the technological distance between two patents produced in each CSD, we show that pairwise combinations of inventions in high-density CSDs are more technically distant than combinations in low-density ones. The implication of this fact is that inventors in dense cities will

[1] As reference points, the density of Manhattan in 2014 was $27,000/\text{km}^2$, Chicago $4,400/\text{km}^2$, Los Angeles $3,200/\text{km}^2$, Austin $1,300/\text{km}^2$, Palo Alto $960/\text{km}^2$ and Armonk (NY) $280/\text{km}^2$.

be more likely to be exposed to ideas from backgrounds distant from their own.

Fourth, the local pool of ideas is a strong predictor of the combinations embedded in new inventions. In other words, inventors exhibit a positive bias toward drawing ideas out of a pool that is locally determined. As a descriptive finding, we observe that for 75% of the class pairs, a patent that references both classes in a given pair $(\mathcal{A}, \mathcal{B})$ is more likely to appear in CSDs with a higher share of patents of class $\mathcal{A}$ and $\mathcal{B}$ produced in the same CSD, and *disproportionately* more likely when the product of the two shares is high (i.e. the coefficient of the interaction is positive). To control for endogenous locational choice, we adopt a difference-in-difference strategy and look at the evolution of patenting of pre-existing firms upon arrival in town of a company from a different sector. We find that the arrival of a firm significantly biases the citation behavior of pre-existing entities toward the field of the arriving firm. To the best of our knowledge, this paper is the first to provide direct evidence of inter-sectoral localized knowledge spillovers operating through this channel.

The facts that we document suggest an alternative interpretations of how technological change interacts with economic geography. Overall, suburban areas play a prominent role in the innovation process: for example, big innovative companies such as IBM or Motorola tend to carry out their research in large office parks located outside main city centers. One view is that these companies can organize knowledge flows efficiently within the organization, and do not have to rely on happenstance interactions in a dense environment. By contrast, informal interactions in dense and diversified areas may become important in generating knowledge flows across technologically distant fields, since specialized *formal* networks (e.g. firms, academic departments and research labs) may not efficiently internalize them. As a result, innovations originating in high-density areas will be built upon more uncommon combinations of prior knwoeldge. The channel of inter-field knowledge spillovers generates an additional benefit from spatial proximity, promoting diversification in dense locations and specialization in sparse ones.

This new set of observations calls for a reassessment of the theoretical link between geography and innovation. In particular, a spatial model of innovation should be able to account for the simultaneous emergence of specialized clusters in suburban areas and diversified hubs in urban centers, while taking the heterogeneity of innovation into account. In the second part of the paper, we propose such a model and use it to perform policy anaysis.

In our setting, innovators are specialized in one of a finite set of scientific fields and choose where to locate balancing rent considerations and innovation opportunities. After developing an idea, innovators can either implement it through an established firm, which keeps the monopoly power over its specific product and increases its productivity, or combine their idea with the knowledge of an inventor of a different field. The second option leads to an unconventional innovation: the inventor can start up a new company and gain leadership over an existing product line, replacing the previous monopolist. Innovators have an incentive to cluster with people of similar background to benefit from intra-field spillovers that increase their ability to develop ideas. However, interactions with inventors from different fields require informal channels and are subject to search frictions. Density reduces frictions across fields but is ineffective in fostering spillovers within fields, that instead occur via formal organizations.

The model reproduces the geographical sorting of innovation activity observed in the data. Conventional and unconventional ideas are complementary. This leads to the emergence of asymmetric sites, both in terms of density and specialization. Densely populated sites diversify and generate unconventional innovation, whereas specialized clusters emerge in low-density areas and produce conventional ideas. The equilibrium implies that geography and both composition and intensity of the innovative activity in the economy are tightly related, and they depend on the parameters of the model in an intuitive way.

This unexplored link opens up novel possibilities for welfare improving place-based transfers. We study optimal policy numerically by calibrating the model using US data. Unconventional ideas are found to be mostly driven by business stealing considerations, as they bring about little technological improvement compared to conventional ones. However, they act as creative destruction events and limit the monopoly power in the economy, which translates into an improved static allocation of labor across firms. The equilibrium conceals a set of externalities that make the outcome inefficient in several dimensions. The welfare analysis reveals that a planner would use place-based policies to increase urbanization and boost creative destruction, at the cost of lowering growth and increasing congestion. The optimal policy of a planner who has the ability to fully affect the urban structure leads to a welfare gain 2 to 3 times larger (in consumption equivalent units) than the one of a planner who can only intervene by reallocating people within the existing urban structure.

This paper contributes to the empirical literature on the estimation of localized knowledge spillovers and the study of their implication for innovation and growth. The importance of localization and geography for the spreading of knowledge, which dates back to Marshall[2] (1890), has been the subject of extensive empirical study in recent years since Lucas (1988), Krugman (1991) and Glaeser et al.'s (1992) seminal papers on economic development and economic geography. Jaffe et al. (1993) analyze the network of patents and

[2] In Marshall's famous words: "When an industry has thus chosen a locality for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighborhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously."

find that citation patterns display a significant bias towards patents that were produced in the same state and metropolitan area. Audretsch and Feldman (1996), Audretsch and Stephan (1996) and Feldman and Audretsch (1999) analyze the geographical concentration of production and innovation activities and find evidence of substantial complementarities between the two. The urban literature has long been interested in the interaction of knowledge spillovers with local specialization and diversity, as in Porter (1990), Florida and Gates (2001), Delgado, Porter and Stern (2014). We contribute to this literature by explicitly considering how local innovation activity affects the *composition* of the knowledge base upon which inventors build new ideas. Our analysis puts particular emphasis on inter-field technology spillovers. Our main finding is broadly consistent with Packalen and Bhattacharya (2015), who find that over the last century newer concepts have been implemented in inventions originating from high-density regions.[3]

This paper also contributes to the theoretical literature on spatial equilibria and knowledge spillovers. Glaeser (1999) proposes one of the first models of knowledge flow in a spatial setting. The dichotomy between specialized and diversified sites in an innovation economy was first introduced by Duranton and Puga (2001). In their model, young firms locate in diversified cities to experiment with different prototypes, while established firms move to specialized sites where intra-field spillovers are stronger. Davis and Dingel (2012) develop a model in which productivity in cities is fostered by informal interactions among people living in a densely populated environment. In their setting, the spatial equilibrium is determined by the comparative advantage of high-skilled individuals in an environment with high learning opportunities. In our setting, individuals are homogeneous in all respects except for the knowledge background they carry: density plays the peculiar role of favoring interactions among people from different fields. As in Berliant and Fujita (2011), knowledge diversity is a key component of growth in our model.

Finally, this paper is related to the literature on endogenous growth and heterogeneous innovation. Akcigit and Kerr (2010) develop a model with heterogeneous firms in the spirit of Klette and Kortum (2004) and explicitly allow for the possibility to carry out exploration R&D to acquire new product lines and exploitation R&D to improve existing ones. We identify exploration R&D with unconventional innovation. This choice is based on the empirical

[3]Packalen and Bhattacharya (2015) find that throughout the last century, patents produced in more densely populated urban areas have made more intense use of newer *concepts*, identified as new sequences of words. Differently from that paper, we look directly at *combinations* of ideas. The pattern of geographical sorting that we document runs through a specific channel, namely, a more hybridized composition of the knowledge base upon which new ideas are built. Packalen and Bhattacharya (2015) also find that the advantage of dense cities is significantly weaker in the part of the sample corresponding to the time period covered by this paper. This suggests that the sorting that we document could be even stronger if an earlier sample of patents were used. This is left for future research.

observation that unconventional patents have a substantially higher technological impact than conventional ones. From a technical point of view, our model closely resembles Peters (2013) and Hanley (2015) in assuming limit pricing and Cobb-Douglas final good aggregator, allowing a simple decomposition of welfare into a static and a dynamic component. We innovate on the existing literature by integrating the idea of heterogeneous innovation in a spatial equilibrium model of a system of cities.

The remainder of the paper is organized as follows: Section 2 introduces the dataset and presents four empirical facts on the geographical organization of innovation activity in the United States. Section 3 introduces the model, characterizes its solution, highlights the mechanism, performs the calibration and studies its implications. Section 4 analyzes optimal place-based policies under fixed and flexible urban structure. Section 5 concludes.

2 Empirical Analysis

Data

The analysis is a based on the universe of patents granted by the US Patent and Trademark Office (USPTO) between January 2002 and August 2014, and filed between January 2000 and December 2010. Table A.1 reports the number of patents by filing year. There are several advantages to focusing on this recent sample. First, the recent digitization of the patent archive has made it easier for authors and reviewers to look for earlier patents to reference. This is reflected in a significantly higher number of citations listed in each patent. Second, by focusing on a short period we minimize long-run changes in the propensity to patent and the technological composition of the sample. Finally, by focusing on the 2000-2010 period, we can reliably link the location information in the patent with socio-economic and demographic characteristics from the Census and the American Community Survey.

Every patent is associated to one of 107 International Patent Classification (IPC) categories.[4] For each grant, we gather information on the identity and location of the original assignee and the inventors and on the full list of referenced patents (up to a maximum of 1,500 citations per patent). Every patent is geolocated following a hierarchical rule: If the patent file reports the name of an institutional assignee (e.g. a company, a research lab or an academic institution) we assign the patent to the geographical coordinates of their location; if the file does not report any assignee or its address is missing or located outside the United

[4]Since each grant is associated with several IPC classes but only one main USPTO class, we build a many-to-one function that maps every USPTO class to a single IPC class based on the association that recurs more often.

States, we attempt to geotag the grant according to the location of its first inventor, otherwise of its second inventor and so on until we are able to assign a location to each patent.[5] Foreign patents are used for the computation of the conventionality score but are discarded in the geographical analysis. We only consider patents that reference at least two citations. The main analysis is performed on a final subset of 1,058,999 patents filed over an 11-year period.

Most of the analysis is conducted at a County Sub-Division (CSD) level. The CSD represents the finest geographical unit that we are able to identify uniquely by intersecting the location information retrievable from the full-text of the patent and the data available from the Census and the American Community Survey.[6] The CSD is much finer than a county, it typically coincides with city boundaries and, in a few cases (e.g. New York City) a city can be partitioned in multiple CSDs. Since demographic data at this level of disaggregation are only available every 10 years, the values of the demographic variables between 2000 and 2010 were interpolated assuming a constant growth rate throughout the years.

2.1 Fact 1: No relationship between density and rate of patenting

The map in Figure 2.1 illustrates the distribution of patenting in the United States between 2000 and 2010 (see also Figure A.2 for close-up maps of the four most denseley populated metropolitan areas).[7] There is a clear tendency for innovative activity to concentrate around main urban centers, highlighting a pattern that most would expect: the East-Coast, the Chicago Area, the Texas Triangle and the Bay Area, among others, are all highly innovative regions. However, two features also emerge. First, a substantial part of patenting activity occurs away from main urban centers, often in low-density areas that are geographically separated from major cities (notably, Armonk, NY and Schenectady, NY). Second, even within major metropolitan areas, a big share of the innovative action takes place in the suburban portion of the latter (e.g. Redmond, WA and Schaumburg, IL). Low density regions seem

[5]Note that we choose to use the location of the assignee, whenever available, instead of the address of the inventor. Most of the literature on the subject, since Jaffe et al. (1993) uses the location of the inventor. Both alternatives raise a number of issues. Since often a patent lists multiple inventors whose locations are too far apart to suggest any interaction through spatial proximity, the address of the institution can represent a more accurate indication of the geographical origin of the invention. Many companies issue patents under several addresses, corresponding to different establishments or research facilities. The main concern with our approach is that the address of the assignee sometimes represents the headquarters of the company instead of the research facility. However, Aghion et al. (2015) report a 92% correlation between the two locations. To attenuate this concern, we run robust checks using only patents assigned to individual inventors. We mention these checks several times throughout the text.

[6]The socio-economic and demographic indicators at the CSD level available at https://nhgis.org.

[7]Note that CSDs are a partition of the US: the empty areas are CSDs where no patents where filed between 2000 and 2010.

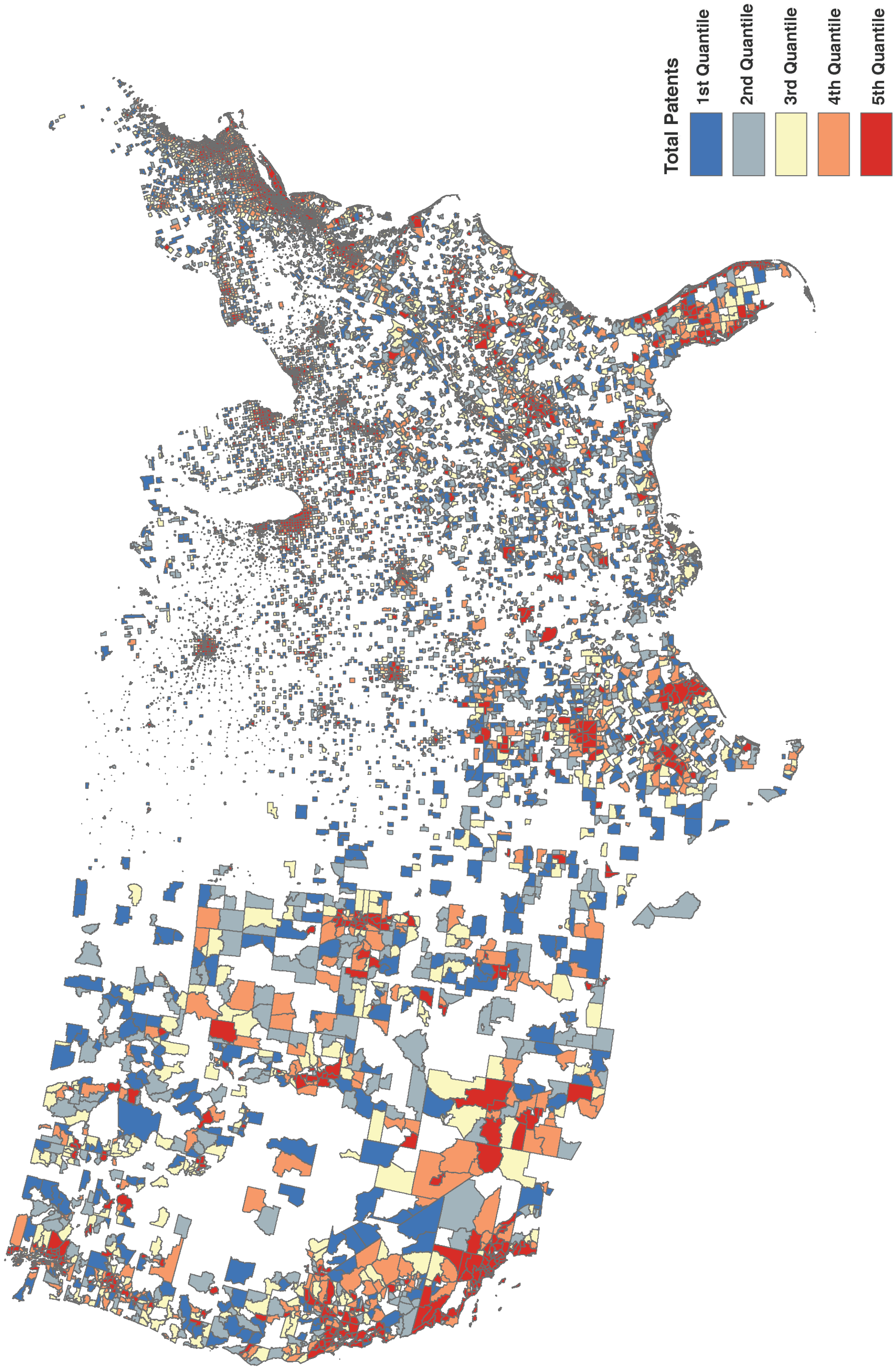


Figure 2.1: The figure shows a map of county sub-divisions in the United States. Each CSD is colored according to the number of patents produced between 2000 and 2010. The more red the higher this value; the more blue, the lower. No patents have been filed in the CSD's that are missing in the map.

to play a key role in the innovation process. About 43.3% of the patents filed between 2000 and 2010 were produced in CSDs with population density below 1,000 residents per square kilometer.[8]

2.1.1 Measurement

In our analysis, we are mostly concerned with characterizing patterns of innovation along the intensive margin. Namely, we restrict the focus to continuously innovative CSDs, defined as units that have filed at least one patent per year between 2000 and 2010. These areas accounted for 53% of the U.S. population and 61% of college graduates in the U.S. labor force in 2010. Roughly 95% of all the patents in our dataset originate from these units. Since we are interested in measuring the extent to which density is related to the flow of knowledge, it is natural to restrict the focus to areas that are continuously involved in innovation activities. The study of areas where innovation does not occur, or occurs only occasionally, is outside the scope of this paper.

For most of the paper, the units of analysis correspond to CSD-year observations. We use patents per resident as our proxy for innovation intensity. As additional tests, we use patents per worker, patents per college-graduate and the logarithm of patents per capita. In the benchmark results, density is measured as residents per square-kilometer, but we also use density of workers or density of college graduates as additional tests. In this subsection, we weight observations by total population (or, depending on the relation, total workers or total number of college graduates) but the same patterns emerge if observations are unweighted or weighted by number of patents.[9] We control for year fixed effects to account for aggregate trends in density and patenting. Since the panel includes a high number of observations, we illustrate the results using bin-scatter plots: we divide the variable on the x-axis in 20 equally heavy bins and take the mean of the y-variable across the observations falling in each bin. [10]

2.1.2 Finding

Unsurprisingly, in the cross-section of U.S. locations, density is related to innovativeness on the extensive margin: increasing population density by 1% increases the probability of hosting permanent innovation activities by 0.16% (Figure A.3a). As a result, the unconditional

[8] As a reference, the Census defines as urban those areas with a central block of at least $2,500/\text{km}^2$, surrounded by blocks of at least $1,300/\text{km}^2$.

[9] Since the relationship between patenting per capita and density is flat across continuously innovative CSD's, those different weighting methods yield very similar results.

[10] Chetty et al. (2013) show that this methodology is able to graphically capture the correlation between two variables. See http://michaelstepner.com/binscatter/ for a more in depth discussion.

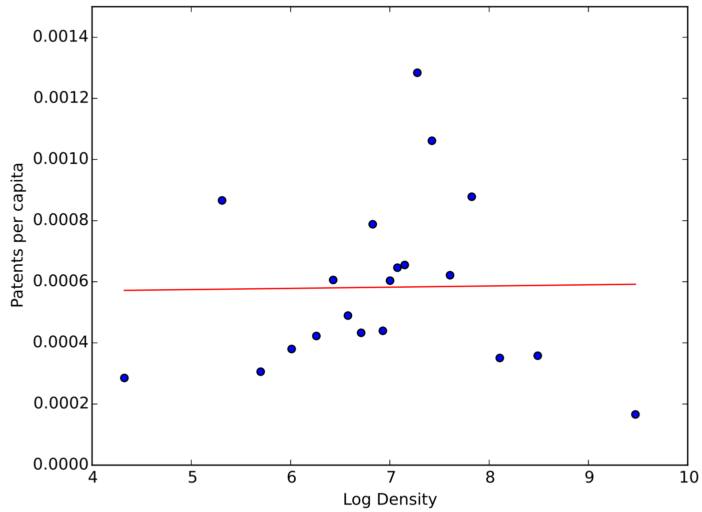


Figure 2.2: Bin-scatter plot of patents per capita and (log) population density in continuously innovative CSDs. The plot is weighted by total population and controls for year fixed effects.

correlation between density and patenting is positive and significant (Figure A.3b). However, this correlation is entirely driven by very low-density places that have zero patenting. The plots in Figure A.3d-f show that the correlation is still positive across places with density above $100/\text{km}^2$ (d), but becomes flat when we condition on density being larger than $500/\text{km}^2$ (e) and even *negative* when we restrict to places with density above $1,000/\text{km}^2$ (f).

We now turn to the focus of this subsection, namely the relation between density and patenting along the *intensive* margin. Figure 2.2 shows a bin-scatter plot of patents per person against (log) population density in the balanced panel of 1,645 continuously innovative CSDs. The relationship between population density and innovation intensity appears to be flat. This finding contrasts with the positive correlation that is observed at the Commuting Zone (CZ) level (Figure A.3c).

Table A.2 in the Appendix presents additional specifications with different weights (e.g. total patents) and fixed effects. In column (4), we control for state fixed effects: the coefficient becomes negative and significant.[11] This fact suggests that although more densely populated states indeed produce more innovation, the low-density portions of those states attract a bigger share of the innovative activity.

Figure A.5 (left-panel) presents an alternative way of visualizing the relationship between density and patenting: we rank CSDs according to their population density and plot the cumulative share of overall patents (horizontal axis) and the cumulative share of overall

[11] A similar result is obtained when controlling for commuting-zone fixed effects (column (7)).

population.[12] As we would expect in the absence of a relationship, the cumulative function largely overlaps with the 45-degree line.

2.1.3 Robustness

It is possible that the relationship in Figure 2.2 is biased downward by the fact that continuously innovative, skill-rich regions tend to be low-density (e.g. college or company towns). In this case, we would be overestimating the relevant interaction opportunities in dense cities and underestimating them for suburban areas. Panel (a) of Figure A.4 shows a similar bin-scatter that captures the partial correlation between density and patenting per capita after controlling for the skill composition (namely, the percentage of college graduates in the population). Panel (b) of the same Figure shows the unconditional correlation, but using density of college graduates and patent per college graduate instead. In both specifications, density and patenting are weakly but negatively correlated.

The choice of a narrow geographical unit of analysis raises the possibility that commuting can confound local population density as a proxy for personal interactions. We address this issue by looking at two extreme cases. In the first case, relevant interactions only occur at the workplace. Since we geolocate the firm whenever possible, we would be correctly assigning the location, but learning opportunities would be mismeasured, as density of workers should be used instead of density of residents. Panel (c) of Figure A.4 shows this relationship for patents assigned to firms or other institutions (i.e. excluding the patents geolocated at the inventor's address). In the opposite case, relevant interactions only take place at the inventor's residence. This time, learning opportunities would be correctly mesured by population density, but the patents issued to institutional assignees would be wrongly geolocated. In Panel (d) of Figure A.4 we plot the relationship only counting patents issued to individual inventors. The relationship is statistically flat in both the polar cases.

By counting the raw number of patents we may be distorting the relationship between density and innovation if inventors and firms locating in low-density places have, other things being equal, a higher propensity to issue low-quality patents. To address this possibility, we weight the number of patents issued by the number of future citations received. In Panel (e) of Figure A.4 we show the partial correlation of population density and citation-weighted patents.[13] Finally, in Panel (f) we plot the partial correlation after controlling for CZ and year fixed effects. The results are largely unchanged. We also run an additional test

[12] The left-panel includes all the CSD-year observations. The right-panel shows the same exercise but drops the observations corresponding to San Jose-Palo Alto.

[13] Although later patents have received on average a lower number of citations, year fixed effects account for this difference.

excluding Delaware and DC from the the sample, as those are classical examples of states for which the address of the assignee and the inventor are more likely to differ.

2.2 Fact 2: Positive relationship between density and unconventionality

The intuition that learning opportunities offered by density should be strong enough to attract the bulk of innovation receives weak support from the data: suburban regions take on a relevant portion of aggregate patenting activity. A possible explanation is that density catalyzes the flow of knowledge across fields that are not connected through established networks, whereas formal organizations are able to internalize knowledge flows efficiently within their own field, without relying on density-driven, informal interactions.

In this subsection, we show that innovations produced in high-density areas tend to be constructed on a more diversified set of prior knowledge. To assess this fact, we need an exact measure of hybridization of the knowledge base of each invention. To construct this measure, we use the distribution of citations across technological classes to infer the intensity of knowledge flows between fields. The fact that a pair of patent classes is recurrently referenced together indicates frequent knowledge flows between the two. Conversely, the fact that the combination of a given pair of categories is atypical denotes the lack of frequent knowledge transmission between the two.

2.2.1 Measurement

We now describe how we measure the degree of interconnection between two technological classes. We adapt the methodology proposed by Uzzi et al. (2013, UMSJ henceforth), who study atypical citation patterns in the universe of academic papers. To the best of our knowledge, this paper is the first to apply a similar algorithm to the universe of patents. The basic idea is to compare the frequency of a bundle of classes in the observed network of references with the frequency one would obtain by assigning citations at random in a replicated network. In this process, the structure of the network is kept constant. In other words, the total number of citations from class $\mathcal{A}$ to class $\mathcal{B}$ is the same in the two networks, but references in the replicated network are reshuffled in a random way, so that the conditional correlation of referenced classes within a patent is zero.[14] The conventionality-score (or *c-score*) of the

[14]While aligning with the basic intuition in UMSJ, we depart from their implementation in two dimensions. First, we do not consider the time dimension explicitely in the replicated network: the total number of citations is kept constant across classes, but not across years. Given that our time window (2000-2010) is relatively short, this simplification is not likely to have a big impact on our estimates. Second, we assume that the number of nodes is big enough such that a law of large numbers hold, which allows us to have an analytical expression for the random frequency. This delivers an exact formula for the c-score, that can be computed without simulating the replicas. See the discussion in the Appendix for the details on the computation.

pair $(\mathcal{A}, \mathcal{B})$ is then defined as the ratio between the observed frequency and the random frequency:

$$c(\mathcal{A}, \mathcal{B}) = \frac{\text{FREQ}_{\text{OBS}}(\mathcal{A}, \mathcal{B})}{\text{FREQ}_{\text{RAND}}(\mathcal{A}, \mathcal{B})}.$$

The interpretation of the c-score is straightforward: a high value of c means that we observe classes $\mathcal{A}$ and $\mathcal{B}$ cited together relatively more often in the data than if citations were assigned pseudo-randomly. We will refer to this citation pair as "conventional" and infer that idea flows between $\mathcal{A}$ and $\mathcal{B}$ are relatively frequent. On the other hand, a low ratio indicates that $\mathcal{A}$ and $\mathcal{B}$ are observed in the data relatively less often than expected. In this case, the combination would be defined as "unconventional". For expositional convenience, in some cases we will refer to $u(\mathcal{A}, \mathcal{B}) = 1 - c(\mathcal{A}, \mathcal{B})$ as unconventionality-score (or *u-score*). The details on the algorithm are provided in Appendix.

Figure A.7 shows a heat-map of the symmetric c-score matrix: each pixel represents a citation pair and it is colored based on its c-score. For example, the pixels on the diagonal represent the score of citation pairs of the form $(\mathcal{A}, \mathcal{A})$. We use a chromatic scale in which brighter pixels denote more conventional pairs and darker pixels denote more unconventional pairs. The figure highlights two patterns that supports the validity of the measure. First, combinations on the diagonal tend to be more conventional than other citation pairs. This is exactly what we would expect: once a patent cites a certain class, it is likely that is going to cite the same class again, since that class is likely to play a central role in the patent development. Second, around the diagonal we observe some "clusters" of conventionality. This happens because the IPC classification system tends to assign close labels to classes that are technologically close. For example, classes in the top-left cluster group all the patents related to human necessities. It is not surprising that a citation that falls in that group is likely to appear with another citation in the same group. However, the c-score identifies technological proximity also between classes that belong to different IPC clusters. The following are some significant examples: Food (belonging to the Human Necessities cluster) and Sugar (belonging to the Chemistry cluster) have a c-score of 1.17; Butchery (Human Necessity) and Weapons (Metallurgy) have a c-score of 1.14; Decorative Arts (Printing) and Photography (Instruments) have a c-score of 1.15; Knitting (Textiles) and Brushware (Human Necessity) have a c-score of 1.84.

We assign to each patent an entire distribution of c-scores, one for each pairwise combination of references (hence, a grant with N references will be assigned $\binom{N}{2}$ possibly identical scores). Two statistics of the distribution are of particular interest. The 10th percentile (or "tail-conventionality") proxies for the most unconventional bundle of classes listed by the

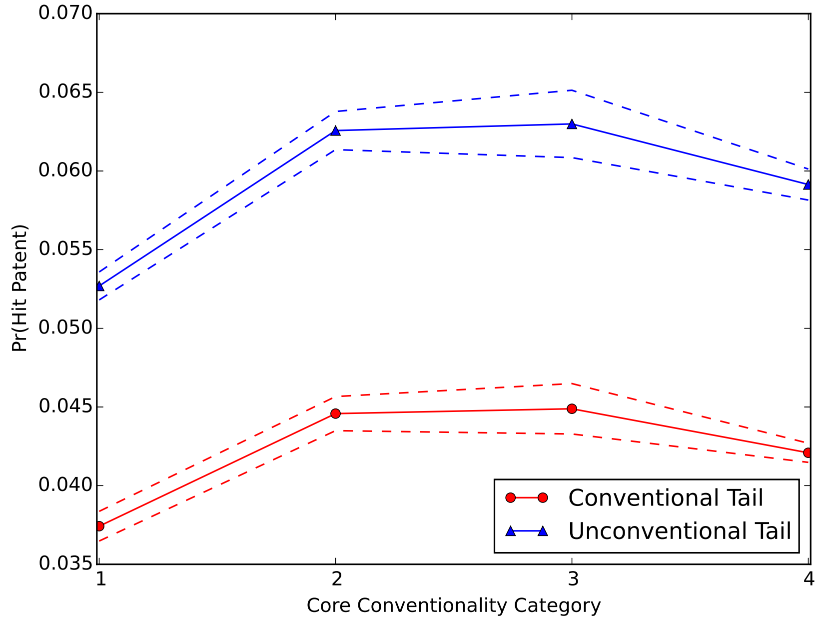


Figure 2.3: Marginal effect of having a conventional tail and being in a certain core-conventionality category on the probability of being a hit patent.

patent. [15] The median c-score (or "core-conventionality") proxies for how tightly grounded the patent is in prior knowledge. Figure A.6 plots the cdf of the core and tail-conventionality in our final sampe. Consistently with the findings in UMSJ, it shows that the median patent is highly conventional at the core (its core-conventionality is well above one).

Next, we show that having an unconventional tail is a powerful predictor of technological impact. To show this, in the spirit of UMSJ, we define a hit patent as an invention that received more citations than 95% of the other grants issued in the same year and belonging to the same class. We estimate a logit model of the form:

$$\text{logit}\left(\text{Hit}_i\right) = \alpha + \delta_c + \delta_t + \beta \times \text{Unconv.Tail}_i + \gamma \times \text{Core Cat.}_i \tag{2.1}$$

where Hit_i is an dummy that takes value 1 if grant i is a hit patent, Unconv.Tail_i is a dummy that takes value 1 if the tail-conventionality is below the median of class c in year t, Core Cat._i is a set of 4 indicators denoting the core-conventionality quartile (in class c and year t), δ_c and δ_t are class and time fixed-effects respectively.[16]

[15] In this paper we follow USMJ and use the 10th percentile for tail-conventionality, but our results are robust to using the minimum. We winsorize the c-score at the 1% level.

[16] We include time and class fixed effects to account for the fact that discreteness in defining the top 5% of the citation distribution leads some classes/years to have a mechanically higher share of hit patents. A linear probability model yields very similar results.

Figure 2.3 shows the joint marginal effects of the two variables on the probability of becoming a hit patent. The conditional probability ranges from 3.7% of a patent with a conventional tail and an unconventional core to 6.2% of a patent with an unconventional tail and a somewhat conventional core. By construction, the unconditional probability is 5%. Having an unconventional tail increases this probability by about 1.7 percentage points. On the other hand, the core seems to have a smaller impact and having an unconventional core *decreases* the chances of being a hit patent. Our result is very similar to the one obtained by UMSJ on the sample of academic papers: scientific research with the highest impact appears strongly rooted in existing knowledge and at the same time displays the intrusion of novel combinations. This surprising similarity seems to suggest that the process of innovation, no matter if academic or applied, follows universal patterns.[17]

This strong correlation between unconventionality and technological impact suggests that the c-score is ranking patents along a meaningful dimension. Motivated by this discussion, in what follows we will use tail-conventionality as our reference measure. We will interpret this measure as capturing the most distant pieces of knowledge assembled in a given invention.

2.2.2 Finding

One hypothesis is that density plays the decisive role of catalyzing knowledge diffusion across unrelated fields. If this intuition is correct, we should observe that patents from high-density regions display more unconventional references. By facilitating interactions, density allows people to gain insights they cannot acquire through their formal network. This translates into new ideas obtained by assembling a more hybridized set of prior knowledge.

Table 2.1 and Figure 2.4 show several CSD-level correlations between (log) density of population (or college educated workers) and the tail-unconventionality (defined as one minus tail-conventionality) of the median patent filed in a given CSD/Year observation. In all the specifications, increasing density of population has a positive and significant impact on the tail-unconventionality of the median patent. In the baseline specification, an increase in density of population equal to the weighted inter-quartile range increases tail-unconventionality by 36% of its weighted inter-quartile range.

To study this relationship more in depth, we add to the specification various CSD specific

[17] The fact that high-impact research is novel and, at the same time, tightly grounded, is explained at least in part by UMSJ by the necessity to efficiently deliver an idea to an inertial audience. They bring the example of Charles Darwin's *On The Origin of Species*, one of the pieces of research with the highest impact in human history, where the groundbreaking idea of natural selection is not addressed until the second part of the work, the first part being entirely dedicated to a much more uncontroversial subject, the selective breeding of cattle and dogs.

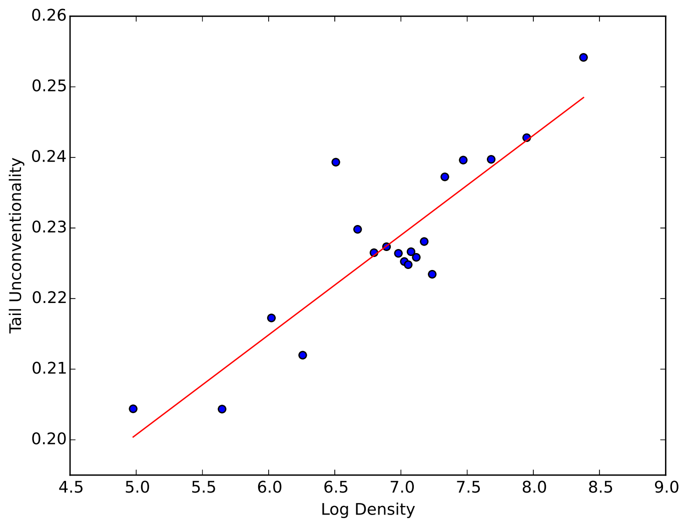


Figure 2.4: The dependent variable is defined as one minus the tail-conventionality of the median patent in the CSD-year observation. The bin-scatter plot is weighted by the total number of patents filed in the CSD/Year observation and controls for State and Year fixed effects.

controls, including (log) median income, the percentage of people with a college degree, inequality (measured by the Gini index). The results are reported in Table 2.2. The effect of density on tail-unconventionality stays positive and statistically significant. The coefficient on median income is always positive and statistically significant. This is probably driven by company or college towns that are typically not the places where hybridized innovation is produced. The share of college graduates and the degree of inequality (Gini index) both have a positive effect, but the coefficients are not statistically significant.

Table 2.3 reports the marginal effects of a patent-level logit regression of (log) density on the probability that the patent has an unconventional tail. Consistently with the CSD-level results, the coefficient is positive and significant. This patent-level regression allows to control for whether the patent is produced by a publicly traded firm. Traded firms tend to produce conventional innovation, consistently with the interpration of unconventional innovation as creative destruction events. This is an interesting fact per se and would deserve further research. We leave this for future work.

2.2.3 Robustness

Table A.8 in Appendix shows that these results are not driven by any of the four most densely populated urban centers (New York City, Boston, San Francisco and Chicago). The bin-scatter plots in Figure A.9 repeat most of the robustness checks mentioned for Fact 1. Panel (a) controls for the share of college graduates, Panel (b) uses density of college grad-

	Median tail-unconventionality					
	(1)	(2)	(3)	(4)	(5)	(5)
Log population density	0.012***	0.013***			0.0075***	0.0097***
	(0.0039)	(0.0032)			(0.0012)	(0.0014)
Log college-graduate density			0.010***	0.010***		
			(0.0035)	(0.0031)		
st/y f.e.	no	yes	no	yes	no	no
Weighted	Pat	Pat	Pat	Pat	no	Pop
N. Obs	18,095	18,095	18,095	18,095	18,095	18,095
R^2	0.023	0.14	0.016	0.13	0.003	0.01

Table 2.1: The dependent variable is defined as one minus the tail-conventionality of the median patent in the CSD-year observation. All regressions, except for (5) and (6), are weighted by the total number of patents filed in the CSD/Year observation. Standard errors in all the regressions are clustered at the CSD level. All variables are winsorized by year at the 1% level.

uates as independent variable. Panel (c) and (d) show what happens when, respectively, only patents with institutional assignee and individual inventor are considered (and plotted against respectively density of workers and density of residents). Finally, Panel (e) and (f) plot the unweighted and the unconditional correlation respectively. All these alternative specifications yield similar results.

2.3 Fact 3: Positive relationship between density and diversification

Figure 2.2 and 2.4 show that density of population and innovation are indeed tightly related. In particular, density seems to affect the type, rather than the rate, of local innovation activies. This pattern of geographical sorting runs through a previously unexplored channel, namely, a more hybridized composition of the knowledge base upon which new ideas are built. In the remainder of this section, we show that (1) dense cities offer a more diversified pool of interaction opportunities and (2) those interactions can be inferred by looking at innovation outcomes. These two findings suggest that the geographical sorting that we documented can be explained as a result of the local interactions available in densely populated areas.

In this subsection, we show that dense cities tend to be more diverse in their innovation output. In particular, we use the concept of the u-score to show that dense cities host a diversified range of innovation activities spanning technologically disconnected fields, whereas low-density areas are markedly specialized in a set of technologically close fields.

	Median Tail Conventionality			
	(1)	(2)	(3)	(4)
Log population density	0.012***	0.0095***	0.0093***	0.0073***
	(0.0032)	(0.0028)	(0.0029)	(0.0029)
Log median income		-0.0193***	-0.0277***	-0.0199***
		(0.0068)	(0.0100)	(0.0090)
% College Graduates			0.0322	0.0227
			(0.0220)	(0.0281)
Gini				0.1147
				(0.0824)
st/y f.e.	yes	yes	yes	yes
Weighted	Pat	Pat	Pat	Pat
N. Obs	18,095	18,095	18,095	17,995
R^2	0.14	0.15	0.15	0.15

Table 2.2: The dependent variable is defined as one minus the tail-conventionality of the median patent in the CSD-year observation. All regressions are weighted by the total number of patents filed in the CSD/Year observation. Standard errors in all the regressions are clustered at the CSD level.

	Unconventional Tail	
	(1)	(2)
Log population density	0.0069**	0.0074**
	(0.0035)	(0.0033)
Publicly Traded		-0.0161**
		(0.0068)
st/y/class f.e.	yes	yes
N. Obs	706,469	706,469
Pseudo R^2	0.007	0.007

Table 2.3: Marginal effects of a logit regression. Dependent variable is a dummy that takes value 1 if the Tail Conventionality of the patent is below the median of its year-class bin. Standard errors in all the regressions are clustered at the CSD level.

2.3.1 Measurement

In addition to assessing the degree of unconventionality of a single patent, the concept of the u-score can also be useful for evaluating the technological diversification of a given subset of inventions: a group of patents is highly diversified if two items drawn at random from the group are likely to belong to technologically distant fields. This idea can be applied to evaluate the degree of technological diversification of a given region over a certain period.

Specifically, we consider all the pairwise combinations of patents filed in each CSD/Year bin. Each of these combinations is assigned the u-score corresponding to the pair of patent classes to which the two grants belong. For example, a CSD that has produced N patents in

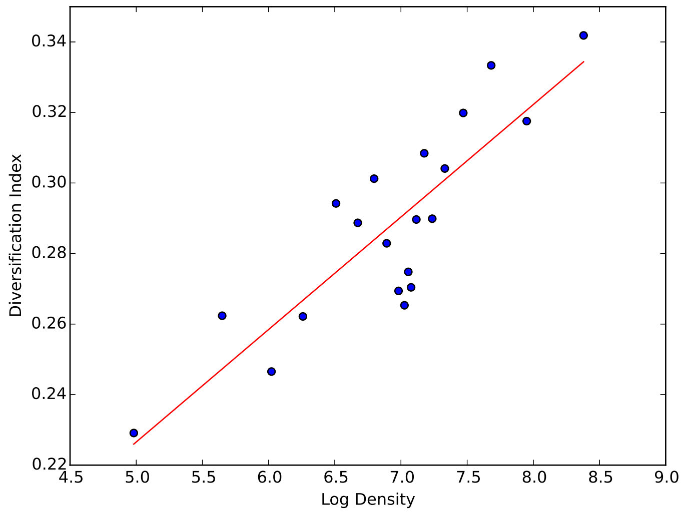


Figure 2.5: Diversification of innovation output and log density of population. The bin-scatter plot is weighted by the total number of patents filed in the CSD/Year observation.

a given year will be assigned $\binom{N}{2}$ u-scores.[18] We then compute the median u-score of those combinations. This procedure delivers an index of diversification for County Sub-Division CSD in year t defined as:

$$\text{Diversification}\left(CSD_t\right) \equiv \text{median}\left(\left\{u\left(CLASS_i, CLASS_j\right) \mid (i, j) \in CSD_t\right\}\right). \tag{2.2}$$

2.3.2 Finding

The bin-scatter plot in Figure 2.5 shows the correlation between density of population and the diversification index defined in (2.2). High-density regions are significantly more diversified than low-density ones. The magnitude of this effect is economically meaningful: a regression of log-density on the diversification index yields a coefficient of 0.03, which implies that an increase in density of population equal to the weighted inter-quartile range decreases diversification by 42% of its weighted inter-quartile range.

2.3.3 Robustness

Since the measure in (2.2) computes the median of a set whose cardinality grows at a binomial rate with the number of local patents, a possible concern is that CSDs with a higher number of patents (as it is typically the case with dense cities) will have a mechanically

[18]To clarify, in this case we are not evaluating the set of references of a given patent, but rather the technological distance of the innovation output itself.

high index of diversification. To address this possibility, we conduct a placebo experiment in which we generate 50 datasets identical to the original one in terms of total number of patents assigned to each CSD/Year bin and to each technology class, but reshuffling the geographical allocation of individual patents at random. We then run 50 regressions of log-density on the simulated indexes of diversification. The resulting coefficients are plotted in Figure A.11. Although the distribution of coefficients for the simulated datasets has a slightly positive average, proving that the index in (2.2) has indeed a dimensionality bias, the estimated coefficients range between $-.0004$ and .00099 (with a mean of .00037), two orders of magnitude smaller than the estimated coefficient on the original sample.

2.4 Fact 4: The local pool of ideas predicts local inventions

The key implication of Figure 2.2 is that, if local interactions matter, people in densely populated regions will have a more diversified pool of possible ideas to draw from. In the extreme case in which local interactions are the only source of ideas, having access to a local pool of innovators from remote fields will be a necessary condition for generating unconventional patents. In this subsection, we show that the local pool of ideas indeed matters.

Inter-field spillovers should be a key component of the benefits from geographical proximity in the production of innovation. As ideas can flow almost freely within interconnected subjects but can hardly spill over across remote fields, spatial proximity should be essential for assembling unconventional combinations of knowledge.[19] In this section, we perform a series of empirical exercises to shed light on the existence and the strength of such spillovers.

2.4.1 A descriptive analysis

As a first step, we check whether our data suggest a correlation between the local technological mix and the citation patterns of locally produced patents. In particular, we ask whether a patent that cites *both* class $\mathcal{A}$ and $\mathcal{B}$ is disproportionately more likely to originate from a CSD with a high share of patents of class $\mathcal{A}$ *and* $\mathcal{B}$. This can be accomplished by running a set of patent-level logit regressions of the form:

$$\text{logit}\left(\mathbb{1}_{\{\mathcal{A}\wedge\mathcal{B}\}}\right) = \alpha + \beta^{\mathcal{A}}\, S^{\mathcal{A}} + \beta^{\mathcal{B}}\, S^{\mathcal{B}} + \beta^{\mathcal{A}\mathcal{B}}\, S^{\mathcal{A}} S^{\mathcal{B}} + \epsilon \tag{2.3}$$

for any pair of technology classes $(\mathcal{A}, \mathcal{B})$. This implies running $\binom{107}{2}$ regressions, one for each combination of classes. In (2.3), $\mathbb{1}_{\{\mathcal{A}\wedge\mathcal{B}\}}$ is a dummy variable that takes value 1 if the

[19]Consistently with this idea, Inoue et al. (2015) find that in Japanese patent applications spatial proximity is more relevant in inter-field collaborations than in intra-field collaboration.

patent cites, at the same time, items from class $\mathcal{A}$ and $\mathcal{B}$, while $S^{\mathcal{A}}$ and $S^{\mathcal{B}}$ denote the share of patents produced in the same CSD/Year belonging to class $\mathcal{A}$ and $\mathcal{B}$ respectively. The logit regression tries to predict whether a patent will display the combination of references $(\mathcal{A}, \mathcal{B})$ based on the frequency of $\mathcal{A}$ and $\mathcal{B}$ in the local innovation pool, including the interaction between the two frequencies. A convenient way of interpreting this regression is looking at two polar cases. Given that patents of class $\mathcal{A}$ are more likely to cite other items from $\mathcal{A}$ (and similarly for $\mathcal{B}$), if the local pool is completely irrelevant, we should observe $\beta^{\mathcal{A}}$ and $\beta^{\mathcal{B}}$ to be positive and the coefficient of the interaction $\beta^{\mathcal{AB}}$ to be zero. At the other extreme, if exposure to local spillovers is the only channel through which $\mathcal{A}$ and $\mathcal{B}$ can be combined, we should observe $\beta^{\mathcal{A}}$ and $\beta^{\mathcal{B}}$ to be zero and the coefficient of the interaction to be positive. Figure A.12 in Appendix is a graphical representation of our results. Every pixel in the heat-map is colored according to the sign of $\beta^{\mathcal{AB}}$, blue if negative, red if positive. The estimate of $\beta^{\mathcal{AB}}$ is positive in more than 75% of the regressions. This appears clearly from visual inspection of the heat-map. Surprisingly, red and blue pixels appear to be evenly distributed over the map, and are not concentrated along the diagonal.

2.4.2 Predicting combinations from the arrival of new firms

Figure A.12 suggests that the local technological mix is reflected in the citation behavior of inventors. However, from this descriptive analysis it is unclear whether this fact reflects local knowledge spillovers or it simply results from endogenous locational choice. Places that produce (or are expected to produce) significant knowledge flows between two fields can be endogenously populated by firms belonging to those fields. For example, a company that aims to produce high-tech shoes, might find it optimal to locate in a town hosting strong CPU and footware sectors.

To control for this possibility, we adopt a difference-in-difference approach and follow the evolution of the citation behavior of pre-existing firms upon arrival in their location of a company from a different industry. The assumption is that the location of pre-existing firms is uncorrelated with the locational choice of incoming firms. Pre-existing firms are all the companies that patent at least once in a given CSD at the beginning of the sample (year 2000). Incoming firms are all the companies that file the first patent in a given CSD in some year after 2000 (we run a robustness exercise considering only firms entering from 2005 onwards). Each incoming firm is assigned to the technology class corresponding to the most recurring class among its patents. Then, for each class/CSD/time observation, we construct an arrival shock as:

$$A_{cdt} = \frac{\sum_{\tau=2001}^{t} R_{cd\tau}}{P_{d,2000}} \tag{2.4}$$

	Share of citations to class $\mathcal{A}$			
	(1)	(2)	(3)	(4)
Arrival of new firm of class $\mathcal{A}$	0.0049*** (0.0002)	0.0043*** (0.0002)	0.0092*** (0.0011)	0.0065*** (0.0002)
Class/CSD & f.e.	yes	yes	yes	yes
Class/Time f.e.	no	yes	no	yes
Shock arrival year	2001	2001	2005	2005
Average $\bar{S}$	0.0043	0.0043	0.0043	0.0043
N. Obs	682,116	682,116	682,116	682,116
Adj. R^2	0.0352	0.0091	0.0109	0.005

Table 2.4: This table reports the coefficients of a regression of the share of citations received by patent class $\mathcal{A}$ from patents of classes other than $\mathcal{A}$ in a given CSD at a given time on time/class and class/CSD fixed effects and the cumulative normalized arrival of new firms of class $\mathcal{A}$ in that CSD. Columns 2 and 4 include time/class fixed effects. Columns 3 and 4 only include incoming firms on or after 2005. Standard errors clustered at the CSD/class level are reported in parenthesis.

where $R_{cd\tau}$ is the number of patents filed in year τ by incoming firms of class c in CSD d and $P_{d,2000}$ is the total number of patents filed in 2000 by pre-existing firms in the same CSD. In other words, the numerator of A_{cdt} proxies for the cumulative inflow of patents of class c, while the denominator normalizes by the size of potentially affected firms. As dependent variable, we use the share of citations that class c receives in patents filed by pre-existing firms of any class different than c.[20] We denote this share by S_{cdt}. The unconditional average of S_{cdt} is 0.0043.[21]

The specification of the regression is the following:

$$S_{cdt} = \alpha + \delta_{ct} + \delta_{dc} + \beta A_{cdt} \tag{2.5}$$

where δ_{ct} and δ_{dc} are class/time and CSD/class fixed effects respectively. To estimate the parameter of interest, β, we exploit the variation in the increase in the propensity to cite class c that results from a higher relative inflow of firms of class c. The identifying assumption is that the citation shares display parallel trends within the same class, across different CSDs. To see this formally, consider the diff-in-diff representation of (2.5) between year t and year

[20]For example, how frequently patents that belong to any class different from CPU reference items in CPU.

[21]Given that we have 107 classes, if citations were distributed at random, every class should receive a share of citations from other classes equal to $\frac{1}{106} = 0.0094$ on average. The fact that the unconditional average is about half that number is simply telling us that on average half of the citations go to items in the same class of the citing patent itself.

$t + r$ for class c in places d_1 and d_2:

$$\left(S_{cd_1(t+r)} - S_{cd_1t}\right) - \left(S_{cd_2(t+r)} - S_{cd_2t}\right) = \beta\left[\frac{\sum_{\tau=t+1}^{t+r} R_{cd_1\tau}}{P_{d_1,2000}} - \frac{\sum_{\tau=t+1}^{t+r} R_{cd_2\tau}}{P_{d_2,2000}}\right].$$

If $\beta > 0$, pre-existing firms producing, say, laptops in a town that has received a higher inflow (compared to its size) of, say, apparel firms, have disproportionately shifted their citation behavior towards apparel. The results are shown in Table 2.4. The estimates of β are always positive and statistically significant, as well as economically meaningful: the arrival of a firm producing exactly as many patents as $P_{d,2000}$ results in an increase in S_{cdt} equal in size to its unconditional mean (column 3). We also report results where we only consider incoming firms that arrive in or after 2005 (column 4): the results are robust and larger in magnitude.[22]

Taking Stock

Our empirical findings can be summarized as follows: (1) the relationship between density and patenting is flat across continuously innovative locations and negative when we consider variation within larger areas (e.g. states or CZ); (2) innovations originating from densely populated areas are built on a more unconventional bedrock of prior knowledge; (3) higher population density is associated with higher diversification of the innovation output; (4) the local technological mix predicts the composition of the knowledge background upon which new inventions are built. In the next section, we will embed these findings into an endogenous growth model of a spatial economy to study how they can help redefine the link between economic geography and innovation, and its implications for macroeconomic outcomes and growth.

3 Model

In this section, we explore the interaction between economic geography and composition of innovation in a fully-specified, endogenous growth model of a spatial economy, in which the heterogeneity in innovation is explicitly taken into account. In its positive implications, the model rationalizes the observed geographical patterns: specialized clusters emerge in low-density areas and produce conventional innovation, while high-density cities become diversified hubs and generate unconventional ideas. The theory provides a novel rationale

[22] The fact that the estimated coefficient is larger in magnitude suggests, as one would expect, the presence of a positive correlation between the class of firms arriving before and after 2005.

for the coexistence of asymmetric cities (both in terms of size and degree of diversification) without assuming ex-ante heterogeneous agents, differentiated products or intrinsic productivity differences across different locations.

The key assumption of the model is that conventional and unconventional ideas are qualitatively different: while the former is crucial to the improvement of existing products and processes, the latter is the foundation for creating new products or disruptively entering a market by displacing existing producers. This assumption is supported by the fact that unconventionality is a strong predictor of a patent's success, as showed in Section 2.2. In the model, the urban structure and the conventionality mix are jointly determined. This fact highlights a novel channel through which place-based policies can have an impact on growth and other macro aggregates. In the next section, we study the normative implications and discuss the main policy trade-offs at play.

3.1 Setting

Consider a continuous time environment in which a representative consumer has access to a homogeneous final good which is valued according to:

$$W_t = \int_t^\infty e^{-\rho(s-t)} \log\left(c_t\right) \mathrm{d}s. \tag{3.1}$$

where $\rho > 0$ is the time discount rate.

The final good Y_t is produced by a competitive firm that aggregates a continuum of intermediate varieties in the interval $[0,1]$ through a Cobb-Douglas production function:

$$\log\left(Y_t\right) = \int_0^1 \log\left(y_{it}\right) \mathrm{d}i. \tag{3.2}$$

The final good producer takes prices of the intermediate varieties as given. Normalizing the price of the final good to $P_t = 1$, profit maximization implies:

$$Y_t = p_{it}\, y_{it}.$$

The form of the demand function of each variety reveals that the revenues of intermediate producers only depend on aggregate output. Hence, intermediate varieties are only produced by the most efficient intermediate firm that charges the highest possible price in order to minimize total production costs.

3.1.1 Intermediate Producers

The most efficient producer (the "leader", denoted by a superscript L) of each variety i employs unskilled labor l_{it} at wage w_t to produce output y_{it}, according to a linear production function:

$$y_{it} = a_{it}^L \, l_{it}$$

where a_{it}^L denotes the labor productivity of the leader. We follow the recent literature on Schumpeterian growth with limit pricing[23] and assume that each intermediate variety i at time t can be identified by a leader-follower distance $\Delta_{it} \geq 0$, such that:

$$a_{it}^L = (1+\lambda_0)\,(1+\lambda_1)^{\Delta_{it}}\, a_{it}^F \tag{3.3}$$

where a_{it}^F is the labor productivity of the second most efficient producer (the "follower"), $(1+\lambda_0)$ is the jump factor by which the *previous* leader's productivity is improved upon losing leadership and $(1+\lambda_1)$ is the factor by which the *current* leader's productivity is improved after receiving a conventional innovation. The leader maximizes current profits by setting a price that is equal to the follower's marginal cost:

$$p_{it} = \frac{w_t}{a_{it}^F}.$$

This results in a markup over its own marginal cost equal to:

$$\mu_{it} \equiv \mu\left(\Delta_{it}\right) = \frac{a_{it}^L}{a_{it}^F} = (1+\lambda_0)\,(1+\lambda_1)^{\Delta_{it}}. \tag{3.4}$$

Profits can be written as:

$$\pi_{it} \equiv \pi_t\left(\Delta_{it}\right) = p_{it}\, y_{it} - \frac{w_t}{a_{it}^L}\, y_{it} = Y_t\left(1-\mu_{it}^{-1}\right).$$

It is easy to see that, given aggregate output Y_t, profits are an increasing and concave function of Δ_{it} that converges to Y_t as Δ_{it} grows to infinity. Substituting the optimal intermediate firm's decisions into (3.2), the expression for aggregate output becomes:

$$\log\left(Y_t\right) = \log\left(L^F\right) + \int_0^1 \log\left(a_{it}^L\right)\, \mathrm{d}i + \int_0^1 \log\left(\mu_{it}^{-1}\right)\, \mathrm{d}i - \log\left(E\left[\left[\mu\left(\Delta\right)\right]^{-1}\right]\right) \tag{3.5}$$

[23] Peters (2013) and Hanley (2015) generalize the original Schumpeterian growth model in Klette and Kortum (2004) by allowing for the possibility of heterogeneous markups.

where $L^F = \int_0^1 l_i \, \mathrm{d}i$ is the total amount of unskilled labor employed by intermediate producers.

Expression (3.5) decomposes aggregate output into an "aggregate input" term, $\log\left(L^F\right)$, an "aggregate technology" term, $\int_0^1 \log\left(a_{it}^L\right) \mathrm{d}i$, and a "static distortion" term:

$$\int_0^1 \log\left(\mu_{it}^{-1}\right) \, \mathrm{d}i - \log\left(E\left[\mu_{it}^{-1}\right]\right) \tag{3.6}$$

which reflects the misallocation of labor resulting from heterogeneous markups. To see why the third term represents a static loss from resource misallocation, note that, by Jensen's inequality, it is always weakly negative, and is equal to zero only if almost every intermediate producer charges the same markup.

3.1.2 The Leader

Consider the leader in product line i who currently holds an advantage on the follower of size Δ_{it}. Two types of idiosyncratic events can hit the leader: a *conventional* innovation that improves her productivity, and an *unconventional* innovation that pushes her out of the market. For now, we take the frequency of these shocks as exogenous and endogenize it in Section 3.2.

1. At Poisson rate $\psi > 0$, the leader is contacted by an innovator who offers her a conventional technological improvement that increases her productivity by a factor $(1 + \lambda_1)$. We assume that conventional innovators always find it optimal to contact the current leader. As a result, the productivity of followers is stagnant. Patent protection of previous underlying technologies prevents the innovator from making any alternative use of the idea. Denoting by $V_t\left(\Delta_{it}\right)$ the value of the leader at Δ_{it}, the resulting surplus is:

   $$S_t\left(\Delta_{it}\right) = V_t\left(\Delta_{it} + 1\right) - V_t\left(\Delta_{it}\right).$$

   If a conventional innovator contacts the leader, a bargaining process begins and a fraction $b \in (0, 1)$ of the resulting surplus is paid by the firm to the innovator. The incremental innovator receives a payment equal to:

   $$\beta_t\left(\Delta_{it}\right) = b\, S_t\left(\Delta_{it}\right).$$

2. At Poisson rate $\zeta > 0$, an inventor develops an unconventional innovation that improves the productivity of the current leader by a factor $(1 + \lambda_0)$. However, while conventional ideas rely on underlying technologies for which the leader enjoys patent

protection, unconventional ideas can be implemented without infringing the leader's intellectual property. The inventor starts up a new firm and becomes the new leader, and the previous leader becomes the current follower. This event resets the technological lead in product line i to $\Delta_{it} = 0$.

In what follows, whenever the time subscript is dropped, we refer to the corresponding variable in balanced growth path (BGP). For all non-stationary variables, we impose stationarity by dividing the corresponding quantity by Y_t.[24] The stationary value function for a leader with technological lead Δ is therefore:

$$(\rho + \psi + \zeta)\ V(\Delta) = \pi(\Delta) + \psi\ [V(\Delta + 1) - \beta(\Delta)] . \tag{3.7}$$

Equation (3.7) makes use of the fact that, along the balanced growth path, the interest rate is constant and equal to:

$$r = \rho + g.$$

The analytical expression for the stationary value function is found by guessing and verifying the following form:

$$V(\Delta) = A - B\ [\mu(\Delta)]^{-1} . \tag{3.8}$$

Matching coefficients for A and B delivers:

$$A = \frac{1}{\rho + \zeta} \qquad B = \frac{(1 + \lambda_1)}{(1 + \lambda_1)\ [\rho + \zeta] + \psi\ (1 - b)\ \lambda_1}.$$

This gives the value of a conventional innovation to a product line with technological lead Δ:

$$\begin{aligned} \beta(\Delta) &= b\,B \left\{ [\mu(\Delta)]^{-1} - [\mu(\Delta + 1)]^{-1} \right\} \\ &= \frac{b\,B\,\lambda_1}{(1 + \lambda_1)}\ [\mu(\Delta)]^{-1} \end{aligned} \tag{3.9}$$

It is easy to see that $\beta(\Delta)$ is decreasing in Δ, reflecting endogenous decreasing returns from conventional improvements.

[24]For example, we let $V(\Delta) = \frac{V_t(\Delta)}{Y_t}$ in BGP. Also, by definition, $Y = 1$.

3.1.3 Stationary Distributions and Balanced Growth Path

Let $\nu(\Delta)$ denote the stationary mass of product lines with technological lead equal to Δ. It can be computed as the solution of the following recursive system:

$$\begin{cases} \zeta\ [1-\nu(\Delta)] = \psi\,\nu(\Delta) & \Delta = 0 \\ \psi\,\nu(\Delta-1) = (\zeta+\psi)\ \nu(\Delta) & \Delta \geq 1 \end{cases}$$

This system has the following solution:

$$\nu(\Delta) = \frac{\zeta}{\zeta+\psi}\left(\frac{\psi}{\zeta+\psi}\right)^{\Delta}.$$

The stationary distribution of technological leads is geometric with an intercept that negatively depends on the ratio of conventional and unconventional innovation $\frac{\psi}{\zeta}$ (or "incrementalism").

From (3.5), we can see that along the BGP, the growth rate of output is simply given by the average growth rate of productivity of the intermediate varieties:

$$g = \int_0^1 \frac{\dot{a}_i^L}{a_i^L}\,\mathrm{d}i = \lambda_0\,\zeta + \lambda_1\,\psi. \tag{3.10}$$

3.2 Economic Geography

Up to this point, the innovation rates ζ and ψ have been treated as exogenous. We now endogenize them by assuming that innovation takes place in a system of cities. For expositional simplicity, assume that all the intermediate varieties in the economy are high-tech devices (e.g. smartphones) that are obtained by combining a software component ($\mathcal{S}$) with a design blueprint ($\mathcal{D}$). The model easily generalizes to the case of multiple components or multiple sectors.[25]

3.2.1 Agents, Cities and Housing

The world is populated by a measure L of unskilled workers and a measure N of skilled innovators. Each innovator is born either as a programmer ($\mathcal{S}$) or a designer ($\mathcal{D}$). For simplicity, we focus on the symmetric case in which the mass of designers is equal to the mass

[25] The extension simply requires an additional equilibrium condition that pins down the optimal degree of diversification of diversified cities.

of programmers:

$$N_{\mathcal{S}} = N_{\mathcal{D}} = \frac{N}{2}.$$

Skilled workers choose where to live and are fully mobile. Unskilled workers live in rural areas (close to production facilities or in the outskirts of cities) where there are no congestion costs and their rent is normalized to zero. There is a large mass of potential sites of area 1, which implies that we can think of local population and local density interchangeably. These sites are owned by absentee competitive landlords, and governed by city developers,[26] who have the ability to tax and provide subsidies to the local economy. They have three options for how to utilize their own site:[27]

1. They can establish a *company town* that provides research facilities for innovators to implement their ideas. Innovators living in a company town can only interact with agents of their own type (e.g. at the workplace), but cannot interact with innovators of the other type.

2. They can establish a *generic town* that does not provide research facilities directly but allows people of different types to potentially interact together.

3. They can leave their site deserted.

In order to attract innovators, the developers can commit to provide type-specific subsidies ($\tau_{\mathcal{S}}$ and $\tau_{\mathcal{D}}$) to the research activity of local inventors. The subsidies are financed by taxing the rent paid by the residents to the landlords. City developers act to maximize profits (taxes minus subsidies) and since option 3 leads to zero profits, a free-entry condition can be used to pin down the active mass of sites of type 1 and 2. We denote by N^k the skilled population in town k and L^k the local unskilled labor input. Each skilled individual demands one unit of housing and, since the area of each site is equal to one, we impose the additional constraint $N^k \leq 1$. Housing services are provided by competitive landlords, who face a local housing production function:

$$N^k = q \left(L^k\right)^{\alpha} \tag{3.11}$$

where the paremeters $\frac{1}{\alpha} > 1$ and $q > 0$ control the strength of the congestion force. The rent

[26] As in Becker and Henderson (2000).

[27] The choice between option 1 and 2 is introduced to simplify the definition and the analytical characterization of the equilibrium. Eliminating the ex-ante distinction between option 1 and 2 would require to introduce an additional condition on whether innovators would choose to interact with people from the other field (option 2) or interact only with people of the same field (option 1). As shown in the Appendix, imposing this condition under the baseline calibration would induce exactly the same equilibrium.

paid by residents of city k is equal to the marginal cost of producing housing services:

$$R^k = \frac{w}{q^{\frac{1}{\alpha}}\,\alpha}\left(N^k\right)^{\frac{1-\alpha}{\alpha}}. \tag{3.12}$$

The entire rent is taxed by the local developer, whose revenue is equal to $N^k R^k$. To clarify, city developers are large agents at the local level but are small from the point of view of the macroeconomy: they can affect local rents but take all aggregate quantities and prices as given.

3.2.2 Innovation

Skilled agents are fully mobile and choose to live in the town that offers them the best combination of rent and innovation opportunities, taking into account the subsidies provided by city developers. Innovation opportunities are determined by two orders of factors.

1. Individuals receive *intra-field spillovers* by agents of the same type that live in the same location. Being surrounded by a high number of "colleagues" increases the Poisson arrival rate of ideas.[28] Specifically, an agent of type $\mathcal{S}$ living in a city with $N^k_{\mathcal{S}}$ innovators (note, of her *same type)* will receive ideas at Poisson rate $d\,\left(N^k_{\mathcal{S}}\right)^{\phi}$, where $d > 0$ and $\phi \in (0, 1)$ control the extent of the learning externalities. Similarly, inventors of type $\mathcal{D}$ living in a city with $N^k_{\mathcal{D}}$ peers receive ideas at Poisson rate $d\,\left(N^k_{\mathcal{D}}\right)^{\phi}$.

2. Upon receipt of an idea, the agent must either be matched with an existing company to which the idea will be sold, or meet an *innovator of the other type* to start up a new firm. Specifically, a programmer (designer) with an idea can either look for an existing firm whose software (design) can be improved upon, or combine it with a design blueprint (software component) to create a new product. The first option is only available to agents living in company towns through the local formal network: in this case, the agent draws a product line $i \in [0, 1]$ at random and sells her conventional improvement to the current leader of line i, receiving a payoff of $\beta\,(\Delta_i)$. The second option is only available to agents living in diversified towns: the programmer (or designer) starts a search process in which he randomly draws a point in the city and finds an innovator of the opposite type with probability $z\,N^k_{\mathcal{D}}$ (or $z\,N^k_{\mathcal{S}}$), where $z \in (0, 1)$ controls the efficiency of the search process.[29] If search is unsuccessful, the idea is lost.

[28] This source of agglomeration externality is akin to the one considered by Duranton and Puga (2001) in that it only affects agents of the same industry.

[29] Since $N^k \leq 1$, this probability is always well defined.

To save on notation, in what follows we conjecture that company towns will be *fully specialized* (i.e. they will host innovators of only one type). This conjecture will be proven formally in Proposition 3.3. Let $\mathcal{K}^G$ denote the set of generic cities and $\mathcal{K}^C_{\mathcal{S}}$ (or $\mathcal{K}^C_{\mathcal{D}}$) denote the set of $\mathcal{S}$-specialized (or $\mathcal{D}$-specialized) company towns. The stationary period utility of an inventor of type $\mathcal{S}$ living in city k can be written as (the one for type $\mathcal{D}$ is analogous, but with inverted indexes):

$$U^k_{\mathcal{S}} = \begin{cases} \left(1+\tau^k_{\mathcal{S}}\right) z\,d\, \left(N^k_{\mathcal{S}}\right)^{\phi} N^k_{\mathcal{D}} V(0) - R^k & k \in \mathcal{K}^G \\ \left(1+\tau^k_{\mathcal{S}}\right) d\, \left(N^k_{\mathcal{S}}\right)^{\phi} \mathrm{E}_{\Delta}\left[\beta(\Delta)\right] - R^k & k \in \mathcal{K}^C_{\mathcal{S}} \end{cases} \tag{3.13}$$

where $V(0)$ is the value of starting up a new firm and $E_{\Delta}\left[\beta(\Delta)\right]$ is the expected return from a conventional innovation.

Once the spatial distribution of innovators is determined, the innovation rates can be derived as:

$$\zeta = \int_{\mathcal{K}^G} z\,d \left[\left(N^k_{\mathcal{S}}\right)^{\phi+1} N^k_{\mathcal{D}} + \left(N^k_{\mathcal{D}}\right)^{\phi+1} N^k_{\mathcal{S}}\right] \mathrm{d}k \tag{3.14}$$

$$\psi = \int_{\mathcal{K}^C_{\mathcal{S}}} d \left(N^k_{\mathcal{S}}\right)^{\phi+1} \mathrm{d}k + \int_{\mathcal{K}^C_{\mathcal{D}}} d \left(N^k_{\mathcal{D}}\right)^{\phi+1} \mathrm{d}k. \tag{3.15}$$

In (3.14), the aggregate rate of unconventional innovation is given by the integral over all the generic locations of the Poisson rate of arrival of ideas for $\mathcal{S}$-type innovators ($d \left(N^k_{\mathcal{S}}\right)^{\phi}$) multiplied by the mass of $\mathcal{S}$-type innovators in city k ($N^k_{\mathcal{S}}$) and multiplied by the probability that the search for a $\mathcal{D}$-type innovator is successful ($z\,N^k_{\mathcal{D}}$), plus the same product for $\mathcal{D}$-type innovators. In (3.15), the aggregate rate of conventional innovation is given by the Poisson rate of arrival of ideas for $\mathcal{S}$-type innovators in $\mathcal{S}$-specialized company towns, plus the same rate for $\mathcal{D}$-type company towns.

The following assumption, that will be maintained throughout, is necessary to insure that agglomeration externalities are not sufficiently strong to perpetually dominate the congestion force:

Assumption A: $\frac{1}{\alpha} > 2 + \phi$.

3.2.3 Equilibrium

In spatial equilibrium, agents of the same type must be indifferent across active locations:

$$U^k_{\mathcal{S}} = U^{k'}_{\mathcal{S}} \qquad \forall k, k' \in \mathcal{K}^G \cup \mathcal{K}^C_{\mathcal{S}}$$

$$U^k_{\mathcal{D}} = U^{k'}_{\mathcal{D}} \qquad \forall k, k' \in \mathcal{K}^G \cup \mathcal{K}^C_{\mathcal{D}}.$$

In what follows, we will focus on symmetric equilibria in which the contribution to aggregate growth of the two types of innovators is the same. This simply requires ex-ante utility to be equalized also across types:

$$U_{\mathcal{D}}^{k} = U_{\mathcal{S}}^{k'} \qquad \forall k \in \mathcal{K}^{G} \cup \mathcal{K}_{\mathcal{D}}^{C} \qquad k' \in \mathcal{K}^{G} \cup \mathcal{K}_{\mathcal{S}}^{C}.$$

A local developer's revenues are equal to the total rent paid by inventors to the competitive landlord:

$$\mathrm{Rev}^{k} = R^{k} \, N^{k} = \frac{w \, (1-\alpha)}{q^{\frac{1}{\alpha}} \, \alpha} \left(N^{k}\right)^{\frac{1}{\alpha}}.$$

Its expenses are equal to the total subsidies paid to the innovators:

$$\mathrm{Exp}^{k} = \begin{cases} z \, d \left[\tau_{\mathcal{S}}^{k} \left(N_{\mathcal{S}}^{k}\right)^{\phi} N_{\mathcal{D}}^{k} + \tau_{\mathcal{D}}^{k} \left(N_{\mathcal{D}}^{k}\right)^{\phi} N_{\mathcal{S}}^{k}\right] V(0) & k \in \mathcal{K}^{G} \\ \tau_{\mathcal{S}}^{k} \, d \left(N_{\mathcal{S}}^{k}\right)^{\phi} \mathrm{E}_{\Delta}\left[\beta(\Delta)\right] & k \in \mathcal{K}_{\mathcal{S}}^{C} \\ \tau_{\mathcal{D}}^{k} \, d \left(N_{\mathcal{D}}^{k}\right)^{\phi} \mathrm{E}_{\Delta}\left[\beta(\Delta)\right] & k \in \mathcal{K}_{\mathcal{D}}^{C} \end{cases}.$$

In equilibrium, free entry of city developers will drive their profits to zero:

$$\mathrm{Rev}^{k} = \mathrm{Exp}^{k} \qquad \forall k \in \mathcal{K}^{G} \cup \mathcal{K}^{C}.$$

To save on notation, in deriving the equilibrium, we will normalize the returns on unconventional innovation and the unskilled wage rate by the expected returns on conventional innovation (hence, the reurns on conventional ideas will be normalized to 1). Once the relative prices and the aggregate rates of innovation are found, the nominal returns can be backed up through (3.8) and (3.9). The derivation of the resulting relative prices can be found in the Appendix:

$$\mathcal{V} \equiv \frac{V(0)}{E_{\Delta}\left[\beta(\Delta)\right]} = \frac{\left[(1+\lambda_0)\,\lambda_1 \psi\,(1-b) + \lambda_0\,(1+\lambda_1)\,[\rho+\zeta]\right]\left[(1+\lambda_1)\,\zeta + \lambda_1 \psi\right]}{[\rho+\zeta]\, b \zeta \lambda_1\,(1+\lambda_1)} \tag{3.16}$$

$$\mathcal{W} \equiv \frac{w}{E_{\Delta}\left[\beta(\Delta)\right]} = \frac{(1+\lambda_1)\,[\rho+\zeta] + \psi\,(1-b)\,\lambda_1}{b \lambda_1 L^{F}}. \tag{3.17}$$

In (3.17), L^F is the total amount of unskilled labor employed by intermediate producers (that is, unskilled labor that is not emloyed in the housing sector):

$$L^{F} = \int_0^1 l_i \, \mathrm{d}i = L - \int_{\mathcal{K}^{G} \cup \mathcal{K}^{C}} L^{k} \, \mathrm{d}k. \tag{3.18}$$

The following proposition highlights a complementarity that will be key for finding the equilibrium.

Proposition 3.1. *The relative return on unconventional ideas, $\mathcal{V}$, is increasing in the aggregate rate of conventional innovation ψ and decreasing in the aggregate rate of unconventional innovation ζ.*

Proof. Taking derivatives of (3.16) with respect to ψ and ζ immediately yields the result. □

Proposition (3.1) can be strengthened by noticing that the relative returns are infinity if the rate of unconventional innovation is zero ($\zeta = 0$). Intuitively, developing a new product is convenient if there is high availability of agents who are able to incrementally improve upon it, and low probability of someone else taking over the product line.

We now have all the ingredients to provide a definition of a symmetric Balanced Growth Path equilibrium for this economy.

Definition 3.2. A symmetric BGP equilibrium is a set of company towns and generic cities $\mathcal{K} = \left\{\mathcal{K}^C, \mathcal{K}^G\right\}$ a utility level U, aggregate innovation rates ζ and ψ, relative prices $\mathcal{V}$ and $\mathcal{W}$, subsidies $\left\{\tau^k_{\mathcal{S}}, \tau^k_{\mathcal{D}}\right\}_{k\in\mathcal{K}}$, local populations $\left\{N^k_{\mathcal{S}}, N^k_{\mathcal{D}}\right\}_{k\in\mathcal{K}}$, local rent $\left\{R^k\right\}_{k\in\mathcal{K}}$, local unskilled labor $\left\{L^k\right\}_{k\in\mathcal{K}}$ and unskilled labor employed in production L^F such that:

1. City developers optimally choose $\tau^k_{\mathcal{S}}$ and $\tau^k_{\mathcal{D}}$ and make zero profits
2. ζ and ψ are defined as in (3.14) and (3.15)
3. U is defined as in (3.13) and is equal across types and active sites
4. $\mathcal{V}$ and $\mathcal{W}$ are defined as in (3.16) and (3.17)
5. L^k and R^k are defined as in (3.11) and (3.12)
6. Labor markets clear: $\int_{\mathcal{K}} N^k_{\mathcal{S}} + N^k_{\mathcal{D}}\,\mathrm{d}k = N$ and $L^F = L - \int_{\mathcal{K}} L^k\,\mathrm{d}k$.

3.3 Characterization

We start by solving the city developer's problem of determining the type, size and composition of its location and the optimal subsidies. We can solve the problem of a developer who aims to found a company and a generic town separately. The free-entry condition will drive profits to zero and make the developer indifferent between establishing any of the two categories of locations (and leave the site deserted).

The problem of a city developer who chooses to establish a company town can be written as:

$$\max_{N_{\mathcal{S}}^k,\tau_{\mathcal{S}}^k,N_{\mathcal{D}}^k,\tau_{\mathcal{D}}^k} \frac{\mathcal{W}\ (1-\alpha)}{q^{\frac{1}{\alpha}}\ \alpha} \left(N_{\mathcal{S}}^k + N_{\mathcal{D}}^k\right)^{\frac{1}{\alpha}} - \tau_{\mathcal{S}}^k\, d\, \left(N_{\mathcal{S}}^k\right)^{\phi+1} - \tau_{\mathcal{D}}^k\, d\, \left(N_{\mathcal{D}}^k\right)^{\phi+1}$$

$$\begin{aligned} \text{subject to :} \qquad & \left(1+\tau_{\mathcal{S}}^k\right)\ d\ \left(N_{\mathcal{S}}^k\right)^{\phi} - \frac{\mathcal{W}}{q^{\frac{1}{\alpha}}\ \alpha} \left(N_{\mathcal{S}}^k + N_{\mathcal{D}}^k\right)^{\frac{1-\alpha}{\alpha}} && \geq\ \mathcal{U} \\ & \left(1+\tau_{\mathcal{D}}^k\right)\ d\ \left(N_{\mathcal{D}}^k\right)^{\phi} - \frac{\mathcal{W}}{q^{\frac{1}{\alpha}}\ \alpha} \left(N_{\mathcal{S}}^k + N_{\mathcal{D}}^k\right)^{\frac{1-\alpha}{\alpha}} && \geq\ \mathcal{U} \end{aligned}$$

In this problem, the maximand represents the developer's profits, while the constraint represents the level of utility the developer must guarantee to the inventors to convince them to join the location.[30]

The maximization of a city developer choosing to establish a diversified city is

$$\max_{N_{\mathcal{S}}^k,\tau_{\mathcal{S}}^k,N_{\mathcal{D}}^k,\tau_{\mathcal{D}}^k} \frac{\mathcal{W}\ (1-\alpha)}{q^{\frac{1}{\alpha}}\ \alpha} \left(N_{\mathcal{S}}^k + N_{\mathcal{D}}^k\right)^{\frac{1}{\alpha}} - \tau_{\mathcal{S}}^k z\, d\, \left(N_{\mathcal{S}}^k\right)^{\phi+1} N_{\mathcal{D}}^k \mathcal{V} - \tau_{\mathcal{D}}^k z\, d\, \left(N_{\mathcal{D}}^k\right)^{\phi+1} N_{\mathcal{S}}^k \mathcal{V}$$

$$\begin{aligned} \text{subject to :} \qquad & \left(1+\tau_{\mathcal{S}}^k\right)\ z\, d\ \left(N_{\mathcal{S}}^k\right)^{\phi}\ N_{\mathcal{D}}^k\, \mathcal{V} - \frac{\mathcal{W}}{q^{\frac{1}{\alpha}}\ \alpha} \left(N_{\mathcal{S}}^k + N_{\mathcal{D}}^k\right)^{\frac{1-\alpha}{\alpha}} && \geq\ \mathcal{U} \\ & \left(1+\tau_{\mathcal{D}}^k\right)\ z\, d\ \left(N_{\mathcal{D}}^k\right)^{\phi}\ N_{\mathcal{S}}^k\, \mathcal{V} - \frac{\mathcal{W}}{q^{\frac{1}{\alpha}}\ \alpha} \left(N_{\mathcal{S}}^k + N_{\mathcal{D}}^k\right)^{\frac{1-\alpha}{\alpha}} && \geq\ \mathcal{U} \end{aligned}$$

The following proposition characterizes the solution to the developer's problem and the equilibrium system of cities, given relative prices $\mathcal{V}$ and $\mathcal{W}$.

Proposition 3.3. *City developers in company towns (C) and generic towns (G) set the optimal subsidy to:*

$$\tau^C = \phi \qquad \tau^G = 1+\phi. \tag{3.19}$$

The optimal population in the two types of locations is:

$$\begin{aligned} N^C &= \left[\frac{d\,\phi\,\alpha\, q^{\frac{1}{\alpha}}}{1-\alpha}\right]^{\frac{\alpha}{1-\alpha\,(\phi+1)}} \left[\frac{1}{\mathcal{W}}\right]^{\frac{\alpha}{1-\alpha\,(\phi+1)}} \\ N^G &= \left[\frac{2^{-(\phi+1)}\, z\, d\,(1+\phi)\,\alpha\, q^{\frac{1}{\alpha}}}{1-\alpha}\right]^{\frac{\alpha}{1-\alpha\,(\phi+2)}} \left[\frac{\mathcal{V}}{\mathcal{W}}\right]^{\frac{\alpha}{1-\alpha\,(\phi+2)}} \end{aligned} \tag{3.20}$$

Company towns are perfectly specialized. Generic towns are perfectly diversified ($N_{\mathcal{S}}^G = N_{\mathcal{D}}^G = \frac{N^G}{2}$) and are more densely populated than company towns.

Proof. See Appendix. □

[30]Notice that we have written the maximization normalizing everything by $E_{\Delta}\left[\beta\left(\Delta\right)\right]$, including $\mathcal{U} = \frac{U}{E_{\Delta}[\beta(\Delta)]}$. Hence, the returns on conventional ideas that enter the developer's cost and the inventors utility is normalized to one.

Proposition 3.3 represents the model counterpart to Figures 2.4 and 2.5, that show the empirical correlation between density and unconventionality of patenting and diversification of the knowledge pool, respectively. The developer's optimal strategy maximizes the value of local output per person, given the relative prices $\mathcal{V}$ and $\mathcal{W}$,[31] but the equilibrium is in general constrained-inefficient. The intuition behind Proposition 3.3 is that agents perceive an additional benefit from agglomerating in diversified cities compared to specialized clusters, and this induces them to trade off additional congestion costs and lower intra-field spillovers for the opportunity of having a higher exposure to inter-field interactions. To see this, compare the elasticity of the local externalities in a specialized company town with the elasticity in a diversified city. In the former case, it is simply ϕ, i.e. the elasticity of intra-field spillovers, whereas in the latter case it is $\phi + 1$, where the $+1$ results from the fact that joining a diversified town increases the matching probability for the inventors of the other field. The developer internalizes this additional externality and, as a result, diversified towns are more densely populated than specialized ones.

The equilibrium solution laid out in Proposition 3.3 also sheds light on the relationship between density of population and patenting per capita (Figure 2.2 and Table A.2), as detailed in the following:

Corollary 3.4. *In equilibrium, patenting per capita is larger in specialized sites if and only if the following inequality is satisfied:*

$$z \left(\frac{N^G}{2} \right)^{\phi+1} < \left(N^C \right)^{\phi} \tag{3.21}$$

A sufficient condition for (3.21) to hold is:

$$\frac{[1 - \alpha \ (\phi + 1)] \ (\phi + 1)}{[1 - \alpha \ (\phi + 2)] \ \phi} z^{\frac{1}{\phi}} < 2^{\frac{\phi+1}{\phi}} \tag{3.22}$$

Proof. See Appendix. □

Condition (3.22) requires congestion forces and intra-field spillovers not to be too small. Figure A.14 (in Appendix) shows some comparative statics with respect to α and ϕ and argues that the Condition is indeed satisfied in all empirically relevant cases. In particular, Condition (3.22) is satisfied at our calibrated parameters.

Once the optimal strategy of city developers has been characterized, it can be combined analytically with the equilibrium definitions, as shown in the Appendix. This leads to a system of two equations in two unkowns (the relative return on innovation $\mathcal{V}$ and the equi-

[31] This property was named Henry George Theorem by Stiglitz (1977).

librium mass of diversified cities $\mathcal{I}^G$)[32] that can be easily solved numerically:

$$\begin{cases} \mathcal{V} = \gamma\left(\mathcal{I}^G\right) \\ \mathcal{V} = \chi\left(\mathcal{I}^G\right) \end{cases} . \tag{3.23}$$

The Appendix describes in details the derivation of $\gamma(\cdot)$ and $\chi(\cdot)$ and the conditions that guarantee existence of the equilibrium with asymmetric locations. Once a solution to (3.23) is found, all the other endogenous variables can be backed up analytically.

3.4 Calibration

In this section we perform a calibration to study the combination of forces that control the geographical organization of innovation and to discipline the parameters for the normative analysis in the next Section. The model has only 11 structural parameters that need to be assigned or estimated. We set 5 of them externally and calibrate the remaining 6 to match relevant moments extracted from the dataset.

The time discount rate ρ is set to 0.05 to obtain an annual discount factor of roughly 0.95%. The mass of unskilled labor is normalized to 3 and the mass of skilled labor is set to 1.74 , so that 36% of the labor force is skilled. Since the estimation of intra-sectoral local technological spillovers is beyond the purposes of this paper, we rely on the extensive empirical literature on the subject to calibrate the elasticity of patenting with respect to neighbor innovators, ϕ. Matray (2014) estimates the local elasticity of patenting activity of small firms to patenting of geographically close listed firms to range between 0.17 and 0.24. In the main calibration, we set ϕ to 0.2 and we experiment with different values of ϕ to see how this choice affects our results. The three parameters controlling growth (d, λ_0 and λ_1) cannot be separately identified. Hence, we normalize $\lambda_0 = 0.02$ and calibrate the other two.

We are left with 7 parameters (λ_1, q, α, d, b, z and p),[33] that we set in such a way as to minimize the distance between some observed moments and their model generated counterpart, as shown in Table 3.2, according to the following metric:

$$D = \sum_j \frac{\left|\text{model}_j - \text{data}_j\right|}{\left|\text{data}_j\right|}.$$

The resulting parameter values are listed in Table 3.1.

[32] Namely, $\mathcal{I}^G = \int_{\mathcal{K}^G} \mathrm{d}k$.

[33] The parameter p is the step factor of the patent citation distribution and is introduced exclusively to match the average of the observed citation distribution. See the "Expected citations" section in the Appendix for details on the assumed citation structure and the role played by the parameter p.

Parameter	Value	Source/Target
Assigned Parameters		
ρ	0.05	Annual discount factor ~ 0.95
λ_0	0.02	Normalized
L	3	Normalized
N	1.74	36% of skilled workers
ϕ	0.2	Spillover elasticity
Calibrated Parameters		
λ_1	0.039	Step factor conv.
q	0.83	Housing
α	0.41	
d	0.77	Poisson arrival of ideas
b	0.22	Appropriability of conv.
z	0.42	Efficiency of search
p	0.49	Patent citations step factor

Table 3.1: Parameter Values

We target an annual growth rate of output of 2% per year. The relative supply of conventional and unconventional innovation depends crucially on two parameters that are novel to the literature, namely, the appropriability of conventional improvements, b, and the efficiency of the inter-sectoral search process, z. These parameters can be identified by the labor share (that has a well known analytical expression) and the average and standard deviation of the citation distribution. Although the average is not per se an interesting moment to match (it depends on the assumed citation structure and requires to introduce an ad-hoc patameter, p), for given average, the dispersion of citations is an important indicator of the degree of incrementalism in the economy. In the "Expected citations" section in the Appendix, we explain the details of the assumed citation distribution, formally derive an analytical expression for the latter and discuss the relationship between the dispersion of citations and the composition of innovation.

The city size distribution is approximated by a Pareto-tail parameter computed using the share of people living in the top 10% of most populated CSDs that lives in the top 1%. In 2000, this share is 38%. This gives a Pareto-tail index equal to:

$$\eta = \frac{-\log(0.1)}{[\log(0.38) - \log(0.1)]} \simeq 1.72.$$

The model counterpart identifies the Pareto-tail index through the share of agents living in the top $\frac{\mathcal{I}^H}{\mathcal{I}^H + \mathcal{I}^L}$ most populated locations. To identify the housing parameters, we target the share of expenditure from individuals with a college degree or higher that are devoted to

Moment	Data	Model
Aggregate growth rate	0.02	0.02
Share of expenditure to housing and non-tradebles	0.48	0.47
Labor Share	0.58	0.58
Skill premium	1.8	1.8
Pareto tail - city distribution	1.72	1.56
Std. Dev. Citation distr. in $T = 12$	28.8	28.8
Ave. Citation distr. in $T = 12$	16.5	16.4

Table 3.2: Moments: Data and Model

housing and non-tradables.[34]

3.5 Mechanism

Table 3.3 reports some statistics on the equilibrium quantities and prices at the calibrated parameters. The model generates the geographical sorting of innovation illustrated in Figure A.13. The equilibrium is characterized by the symultaneous emergence of a large mass of low-density specialized clusters and a smaller mass of high-density diversified cities.

This endogenous geography completely pins down the innovation variables. The degree of incrementalism is $\frac{\psi}{\zeta} = 18.3$: every unconventional idea receives on average 18.3 incremental improvements before being replaced by another unconventional idea. At the estimated parameters, this results in a relative price of unconventional innovation of $\mathcal{V} = 42.53$. The estimated parameter for the efficiency of the search process z is 0.42. This relatively low value implies that informal interactions are a risky activity and must be compensated by a relatively high price of unconventional innovation. Most of the surplus generated by conventional improvements goes to the current leader in the product line, as reflected by the estimate of $b = 0.22$. The calibration further reveals that the step factor of conventional ideas, λ_1,[35] is substantially larger than the step factor assumed for unconventional innovation, λ_0. An important implication of this result is that, other things being equal, there is a trade-off between growth and creative destruction. We will analyze this tradeoff more in detail when studying the planner's problem.

Figure A.17 shows how different values for the elasticity of intra-sectoral spillovers, ϕ, and for the appropriability of conventional innovation, b, affect the degree of incrementalism, $\frac{\psi}{\zeta}$ and the growth rate of output. Higher ϕ increases the relative benefit from agglomer-

[34]This value is obtained from the 2014 Survey of Consumer Expenditure by adding the share of expenditure of households with maximum educational attainment of college or higher that goes to Food away from home, Housing, Public transportation, Entartainment, Apparel services, Vehicle maintainance and Medical services.

[35]This estimated value is in line with the estimated value in Peters (2013).

Moment	Equilibrium Value
Mass of diversified cities $\mathcal{I}^G$	1.05
Mass of specialized clusters $\mathcal{I}^C$	35.43
Unocnventional innovation rate ζ	0.027
Conventional innovation rate ψ	0.49
Density of diversified cities N^G	0.45
Density of specialized clusters N^C	0.03
Relative returns to unconventional innovation $\mathcal{V}$	42.53
Relative wage $\mathcal{W}$	3.96

Table 3.3: Equilibrium: Statistics

ating with innovators from the same field. This implies higher incrementalism, but mechanically it reduces the arrival rate of ideas and lowers aggregate growth. The appropriability of conventional innovation, b, unambigously increases the propensity towards specialization. This results in higher incrementalism and higher growth, as it increases the incentives of innovators with the same background to cluster in the same location. Note that changes in the value of b affect the equilibrium allocation but by construction cannot affect the planner's solution. Figure A.16 shows the relationship between the relative returns to unconventional ideas $\mathcal{V}$ and other equilibrium outcomes. Higher $\mathcal{V}$ increases the share of innovators living in diversified cities and reduces the degree of incrementalism.

To illustrate the key relationship between geography and macroeconomic outcomes, in Figure A.19 we fix the urban structure $\mathcal{K} = \{\mathcal{K}^G, \mathcal{K}^C\}$ at the equilibrium of the baseline calibration and we then artificially impose a different distribution of inventors across company and generic towns (this exercise will be formalized in Section 4.1). In the top-left panel, we show that increasing the share of innovators in diversified towns reduces the aggregate growth rate of output. There are two reasons behind this: First, as discussed in Corollary 3.4, reducing the share of innovators in specialized towns reduces the aggregate rate of invention. Second, the composition of innovation shifts towards more unconventionality, whose step factor $(1+\lambda_0)$ is estimated to be lower than the one of conventional ideas, $(1+\lambda_1)$. In the top-right panel, we show that increasing density in diversified cities reduces the static misallocation of resources. The reason is that lower incrementalism, $\frac{\psi}{\zeta}$ is associated to a more compressed distribution of markups across firms. However, since $N^G > N^C$ and the congestion function is convex, this requires diverting unskilled labor from production of the final good to the non-tradeable sector, as shown in the bottom-left panel.[36]

[36]It is also interesting to see how a change in the share of inventors in diversified towns is reflected in the dispersion of patent citations. For low values, higher rate of unconventionality increases overall variation in technological leads and produces an increase in the coefficient of variation of the citation distribution. For high values, increasing unconventionality further compresses the distribution of technological leads and decreases

4 Welfare and Policy

We now turn to the problem of optimality of the equilibrium. Although city developers internalize knowledge externalities at the local level and the associated congestion costs, they cannot internalize the impact of local acivities on aggregate growth and other macroeconomic outcomes. In general, the equilibrium will be suboptimal. In this section, we analyze the optimal local policy of a constrained planner who can tax the innovators and provide local-based subsidies. The planner can affect the locational choice of innovators, but cannot effectively control pricing and production operations carried out by monopolistic firms. In our setting, aggregate welfare has a well known form that corresponds to the balanced growth version of (3.1):

$$W_0 = \int_0^\infty e^{-\rho t} \left[\log \left(c_0 e^{gt} \right) \right] \, dt \tag{4.1}$$

where c_0 represents consumption at time 0 and includes all the static factors that determine equilibrium output for a given level of technology, while g represents the endogenous growth rate of technology. We now analyze these two components separately.

Initial consumption c_0 is determined by the congestion costs and by the static monopolistic distortion in (3.6). Congestion costs are captured by the amount of unskilled labor that is employed in the production of housing: given the convex cost function for landlords, the higher the geographical concentration the higher the amount of labor employed in housing, which reduces c_0. To see this, we can look at the amount of unskilled labor employed in production:

$$L^F = L - \int_{\mathcal{K}^G \cup \mathcal{K}^C} \frac{N_k}{q}^{\frac{1}{\alpha}}.$$

Since $\alpha \in (0,1)$, a more concentrated geography results in a lower value of L^F. Notice that congestion costs can be fully internalized by city developers, but cannot be traded off for higher growth or a different conventionality mix in a decentralized equilibrium. The monopolistic distortion is an increasing function of the rate of incrementalism $\frac{\psi}{\zeta}$. Substituting in (3.6) the relevant expressions, one gets:

$$\int_0^1 \log \left(\mu_{i,t}^{-1} \right) \, di - \log \left(E \left[\left[\mu \left(\Delta \right) \right]^{-1} \right] \right) = - \frac{\log \left(1 + \lambda_1 \right) \psi}{\zeta} - \log \left(\frac{\zeta}{\zeta \left(1 + \lambda_1 \right) + \psi \lambda_1} \right).$$

It is easy to see that the above expression is decreasing in $\frac{\psi}{\zeta}$. Note also that, by Jensen's inequality, the above expresion is always weakly negative, and is equal to zero only if the

the coefficient of variation. This delivers the inverted-U shape of the line in the bottom-right panel of Figure A.19.

dispersion of mark-ups is zero. This can only be the case if the rate of conventional innovation ψ is itself zero.

The growth rate of technology, g, is determined by the aggregate rates of conventional and unconventional innovation:

$$g = \zeta\lambda_0 + \psi\lambda_1.$$

A planner can be interested both in increasing overall agglomeration, which would increase both ζ and ψ via its impact on localized spillovers, or in changing the composition of innovation. The decentralized equilibrium might have too little growth and too much unconventional innovation (ζ high but ψ low): this can result from the fact that business stealing incentives from unconventional ideas are too strong. Conversely, the equilibrium might have too much growth and too little creative destruction (ζ low and ψ high), which results in a high monopolistic distortion.

Following the above discussion, by integrating (4.1), welfare can be rewritten as:

$$W^P \propto F + \underbrace{\frac{\lambda_1\psi + \lambda_0\zeta}{\rho}}_{\text{Growth}} + \underbrace{\log\left(L^F\right)}_{\text{Congestion}} - \underbrace{\left[\frac{\log\left(1+\lambda_1\right)\psi}{\zeta} + \log\left(\frac{\zeta}{\zeta\left(1+\lambda_1\right)+\psi\lambda_1}\right)\right]}_{\text{Static Misallocation}}. \tag{4.2}$$

where F is a constant that only depends on strucutral parameters and the current level of technology $\int_0^1 a_i \, \mathrm{d}i$, which we normalize to one without loss of generality. This expression decomposes welfare into its three main components: growth, congestion costs and monopolistic distortions.

As we will see, depending on the policy tools the planner has at its disposal, some of these objectives will trade-off each other or will be achievable simultaneously. We will consider two extreme types of constrained planners. First, we consider a planner who takes the urban structure ($\mathcal{K}^H$ and $\mathcal{K}^L$) as given and designs a system of transfers that can affect the locational choice of innovators. Second, we consider a planner that has full flexibility in affecting the urban structure.

4.1 Fixed urban structure

We first consider the extreme case of an urban structure that is fixed as prescribed by its decentralized equilibrium. Existing sites can neither be withdrawn by their respective developers nor can their nature of diversified/specialized location be changed. Moreover, new locations cannot be created. In this case, the zero profit condition of city developers does not need to hold. The mass of locations $\mathcal{I}^G$ and $\mathcal{I}^C$ is fixed. The planner can simply reallocate skilled workers across the different sites by designing a simple system of lump-sum transfers

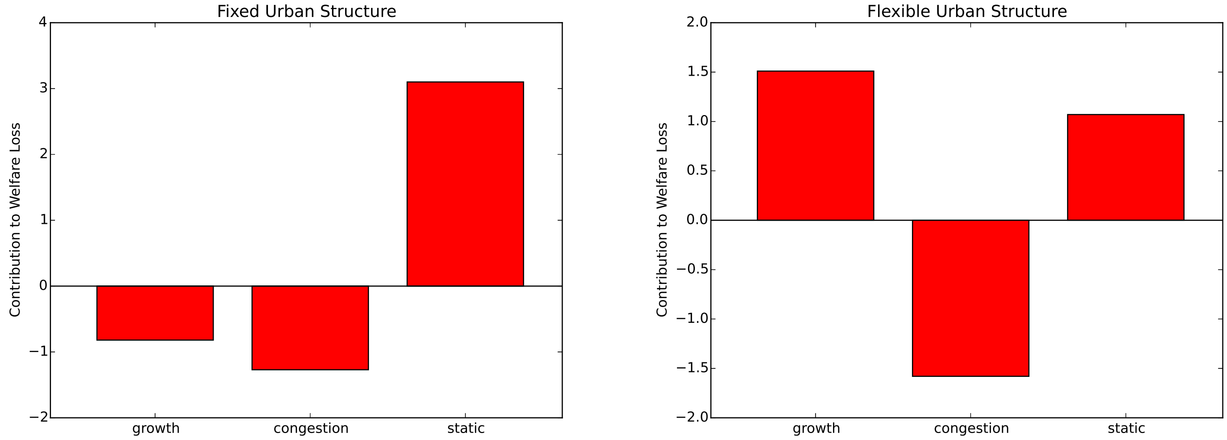


Figure 4.1: The figure shows the contribution to the welfare loss in equilibrium compared to the planner allocation. By construction, the sum of the three columns is one.

$\left\{T^k_{\mathcal{S}}, T^k_{\mathcal{D}}\right\}_{k\in\mathcal{K}}$ that are technology and site specific, with the objective of shifting innovation activity away or towards a given type of location.

The planner's problem in a symmetric BGP reduces to the choice of a share $\eta \in [0,1]$ representing the fraction of innovators living in diversified cities. Hence:

$$N^G_{\mathcal{S}} = N^G_{\mathcal{D}} = \left(\frac{\eta N}{2\mathcal{I}^G}\right)$$

$$N^C_{\mathcal{S}} = N^C_{\mathcal{D}} = \frac{(1-\eta)\,N}{\mathcal{I}^C}.$$

This geographical allocation can be easily implemented by a pair of lump-sum transfers $\left\{T^G, T^C\right\}$ that, given the resulting prices and quantities, make innovators indifferent in their locational choice. This constrained planner's problem can be written as:

$$\max_{\eta\in(0,1)} \frac{\lambda_1\psi + \lambda_0\zeta}{\rho} + \log\left(L^F\right) - \left[\frac{\log\left(1+\lambda_1\right)\psi}{\zeta} + \log\left(\frac{\zeta}{\zeta\left(1+\lambda_1\right)+\psi\lambda_1}\right)\right]$$

$$\begin{aligned} \text{subject to}: \quad \zeta &= \left(2\mathcal{I}^G\right)^{-(\phi+1)} zd\left(\eta N\right)^{\phi+2} & \psi &= d\left(\mathcal{I}^C\right)^{\phi}\left(\left(1-\eta\right)N\right)^{\phi+1} \\ L^F &= L - \mathcal{I}^G\left(\frac{\eta N}{q\mathcal{I}^G}\right)^{\frac{1}{\alpha}} - \mathcal{I}^C\left(\frac{(1-\eta)N}{q\mathcal{I}^C}\right)^{\frac{1}{\alpha}} & & \mathcal{I}^G, \mathcal{I}^C \text{ given} \end{aligned}$$

Figure A.20 shows the value of welfare as a function of η, while the horizontal line represents the decentralized equilibrium. Column 2 of Table 4.1 shows the solution to the planner's problem under fixed urban structure compared with the corresponding decentralized equilibrium. The planner can achieve an increase in welfare that corresponds to 3.05% annu-

	Equilibrium	Fixed Geo	Flexible Geo
Growth	0.02	0.019	0.024
N^G	0.458	0.521	0.61
N^C	0.035	0.033	0.12
$\mathcal{I}^G$	1.05	1.05	0.89
$\mathcal{I}^C$	35.43	35.43	9.76
Average Lead	18.26	12.95	13.97
Labor in production L^F	2.73	2.64	2.48
Static Distortion	-0.140	-0.062	-0.075
Labor Share	58%	65.8%	64.1%
Cons. equiv.	0	+3.05%	+7.28%

Table 4.1: Equilibrium and Planner: Statistics

alized consumption equivalent units. In the left panel of Figure 4.1 we show the composition of the welfare gain. The three columns in the histogram sum up to one by construction. With fixed cities, under the calibrated parameters, the planner chooses to sacrifice growth and increase congestion by moving innovators towards the diversified cities. This results in higher rate of creative destruction and a lower static misallocation of labor.

At the estimated parameter values, the decentralized equilibrium produces too much conventional innovation and too little creative destruction. Hence, welfare is maximized by reallocating inventors where inter-field spillovers are stronger. In this case, improving the static allocation and increasing growth are two contrasting objectives that cannot be achieved at the same time. The increase in congestion costs comes as a pure byproduct of this policy: since diversified cities are more dense, reallocating inventors from specialized to diversified locations increases the spread between the two, and average congestion costs increase because of the convexity of the housing cost function.

4.2 Flexible urban structure

We now turn to the opposite extreme case. We consider a planner that has full flexibility in affecting the urban structure. In this case, the system of cities is not predetermined and the zero profit condition of the developers must hold. To preserve tractability, we assume that the planner's policy tool consists of class/location specific transfers that multiplicatively subsidize innovation outcomes, financed through a lump-sum tax. The planner chooses $\{T^k_{\mathcal{S}}, T^k_{\mathcal{D}}\}_{k\in\mathcal{K}}$ between -1 and $+\infty$ and pays to successful innovators $(1+T)$ times the effective value of the innovation.

Assuming symmetry in the planner's solution, the optimal system of transfers reduces to two numbers, $\{T^G, T^C\}$ for diversified and specialized locations respectively. Given this policy choice, the resulting equilibrium can be found as in Proposition 3.3, with the exception that the inventor's income is now incremented by a multiplicative factor of $(1+T)$. The resulting geography has an analytical solution:

$$\begin{aligned} N^G &= \left[\frac{1+T^G}{1+T^C}\right]^{-\frac{\alpha\phi}{1-\alpha}} C^G_{pl}\mathcal{V}^{-1} \\ N^C &= \left[\frac{1+T^G}{1+T^C}\right]^{-\frac{\alpha(\phi+1)}{1-\alpha}} C^C_{pl}\mathcal{V}^{-1} \end{aligned} \tag{4.3}$$

where C^G_{pl} and C^C_{pl} are constants to be determined in equilibrium and the value of $\mathcal{V}$ can be found by solving a system of two equations in two unknowns as in (3.23). The optimal policy can be found numerically.

The third column of Table 4.1 shows the solution to the planner's problem under the flexible urban structure. In this case, the planner can achieve an increase in welfare that corresponds to 7.28% annualized consumption equivalent units, which is substantially larger than in the case of fixed urban structure. The right panel of Figure 4.1 shows the composition of the welfare gain. Again, the three columns add up to one by construction. In this case, the objectives of higher growth and smaller static distortions are no longer in contrast: they can both be achieved by increasing the geographical concentration of both types of innovation. This is reflected in higher density N^G and N^C and a smaller mass of active locations $\mathcal{I}^G$ and $\mathcal{I}^C$. The intuition behind this is that only part of the benefits from increasing growth, and none of the benefits from reducing the static distortion, are internalized by the city developers. On the contrary, congestion costs are fully internalized by them. Hence, the equilibrium tends naturally towards too little congestion. The planner can trade off part of the congestion cost to increase local knowledge spillovers.

5 Conclusion

Understanding the process through which creative ideas are generated is crucial to fully exploit the comparative advantage of advanced economies in today's world. In this paper, we explore a specific aspect of this process, namely how the economic geography shapes the creative content of innovation. We show that high-density regions have an advantage in producing unconventional ideas. We do this by assembling a new dataset of georeferenced patents and by assigning a measure of creativity that is novel to the macro literature on innovation. Our empirical analysis reveals that the combination of ideas embedded into inventions is determined by the local technology mix. This supports the hypothesis that knowledge spillovers across fields resulting from informal interactions are a key component of the innovation process. High-density areas promote diversification and facilitate informal interactions, resulting in a higher degree of unconventionality in innovation. Our analysis reconciles the fact that a big portion of innovative activity takes place outside cities with the common wisdom, rooted in the literature, that density is an important catalyzer of knowledge diffusion.

We integrate these findings into an endogenous growth model with spatial sorting and heterogeneous innovation. In our setting, the choice between producing conventional and unconventional ideas depends on their relative price and, crucially, on the local degree of density and diversification. In equilibrium, low-density specialized cities coexist with high-density diversified ones. This asymmetry is dictated by the complementarity of unconventional and conventional ideas in the innovation process and does not depend on the existence of agents with ex-ante heterogeneous abilities. The composition of innovation determines the balance between growth, static allocation of resources and congestion costs, which in equilibrium is suboptimal. Our analysis reveals that a constrained planner would sacrifice growth and congestion costs, increase urbanization, promote the creation of unconventional ideas and reduce the monopolistic distortions. Whether a planner has some flexibility in adjusting the urban structure makes a big difference in determining the welfare benefits from place-based policies.

Further research should be devoted to understanding the locational choice of large firms and whether these are subject to the same type of incentives as small, typically more unconventional firms. Another promising avenue would be exploring what are the cultural and economic factors that can be held responsible for the "urban revolution" that is rapidly reshaping the geography and the path of technical change in advanced economies.

References

[1] Akcigit, Ufuk, and William R. Kerr. 2010. "Growth Through Heterogeneous Innovations." NBER Working Paper No. 16443.

[2] Acemoglu, Daron, Ufuk Akcigit, and Murat A. Celik. 2014. "Young, Restless and Creative: Openness to Disruption and Creative Innovations." NBER Working Paper No. 19894.

[3] Aghion, Philippe, Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323-351

[4] Aghion, Philippe, Ufuk Akcigit, Bergeaud, Antonin, Blundell,Richard and Hémous, David. 2015. "Innovation and Top Income Inequality" NBER Working Paper No. 21247.

[5] Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review*, 86(3): 630-40.

[6] Audretsch, David B., and Paula E. Stephan. 1996. "Company-Scientist Locational Links: The Case of Biotechnology," *American Economic Review*, 86(3): 641-52.

[7] Berliant, Marcus, and Masahisa Fujita. 2011. "The Dynamics of Knowledge Diversity and Economic Growth." *Southern Economic Journal*, 77(4): 856-84.

[8] Becker, Randy and Vernon Henderson. 2000. "Intra-Industry Specialization and Urban Development." in "Economics of Cities" edited by Jean-Marie Huriot and Jacques-Francois Thisse. Cambridge University Press. 138-166.

[9] Black, Duncan and Vernon Henderson. 1999. "A Theory of Urban Growth." *Journal of Political Economics*, 107(2): 252-85.

[10] Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593-2632.

[11] 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

[12] Davis, Donald R., and Jonathan I. Dingel. 2014. "The Comparative Advantage of Cities." NBER Working Paper No. 20602.

[13] Davis, Donald R., and Jonathan I. Dingel. 2012. "A Spatial Knowledge Economy." NBER Working Paper No. 18188.

[14] Delgado, Mercedes, Michael Porter and Scott Stern. 2014. “Clusters, convergence, and economic performance.” *Research Policy,* 43(10): 1785-99

[15] Duranton, Gilles, and Puga, Diego. 2000. ”Diversity and Specialisation in Cities: Why, Where and When Does it Matter? .” *Urban Studies*, 37(3): 533-555.

[16] Duranton, Gilles, and Puga, Diego. 2001. ”Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products.” *American Economic Review*, 91(5): 1454-1477.

[17] Feldman, Maryann P., and David B. Audretsch. 1999. "Innovation in Cities: Science-based Diversity, Specialization and Localized Competition." *European Economic Review*, 43(2): 409-29.

[18] Florida, Richard, and Gary Gates. 2001. “Technology and Tolerance: The Importance of Diversity to High-Technology Growth.” The Brookings Institution Survey Series, June.

[19] Gabaix, Xavier. 1999. “Zipf's Law for Cities: An Explanation.” *Quarterly Journal of Economics*, 114 (3): 739-767.

[20] Gaubert, Cecile. 2014. “Firm Sorting and Agglomeration”, Mimeo

[21] Glaeser, Edward L., Hedi Kallal, Jose' Scheinkman and Andrei Schleifer. 1992. “Growth in Cities." *Journal of Political Economics*, 100(6): 1126-52.

[22] Glaeser, Edward L. 1999. "Learning in Cities." *Journal of Urban Economics*, 46(2): 254-77.

[23] Glaeser, Edward L. and Joshua D. Gottlieb. 2009 ”The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States.” *Journal of Economic Literature*, 47(4): 983-1028.

[24] Griliches, Zvi. 1990. “Patent Statistics as Economic Indicators: A Survey.” *Journal of Economic Literature*, 28(4): 1661-1707.

[25] Hall, Bornwyn H., Aadam B. Jaffe, and Manuel Trajtenberg. 2001. “The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools.” NBER Working Paper No. 8498.

[26] Hanley, Douglas. 2015. “Innovation, Technological Interdependence, and Economic Growth.”, Mimeo

[27] Imbs, Jean, and Romain Wacziarg. 2003. "Stages of Diversification ." *American Economic Review*, 93(1): 63-86.

[28] Inoue, Hiroyasu, Kentaro Nakajima and Yukiko U. Saito. 2015. “Localization of Collaborations in Knowledge Creation.”, Mimeo

[29] Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics*, 108(3): 577-98.

[30] Jovanovic, Boyan. 2015. “Learning and Recombination.” Mimeo

[31] Katz, Bruce and Julie Wagner. 2014. “The Rise of Innovation Districts: A New Geography of Innovation in America.” Metropolitan Policy Program at Brookings, May

[32] Klette, Tor J., and Samuel Kortum. 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy*, 112(5): 986-1018.

[33] Kogan, Leonid, Dimitris Papanikolaou, and Noah Stoffman. 2013. “Technological Innovation: Winners and Losers.” NBER Working Paper No. 18671.

[34] Kogan, Leonid, Dimitris Papanikolaou, Amit Seru and Noah Stoffman., 2012, "Technological Innovation, Resource Allocation, and Growth." NBER Working Paper No. 17769.

[35] Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy*, 99(3): 483-99.

[36] Lucas, Robert Jr. 1988. ”On the Mechanics of Economic Development.” *Journal of Monetary Economics*, 22(1): 3-42.

[37] Lucas, Robert Jr. and Moll, Benjamin. 2014. “Knowledge Growth and the Allocation of Time.” *Journal of Political Economy*, 112(1): 1-51.

[38] Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan and Co., Ltd.

[39] Matray, Adrien. 2014. ”The Local Innovation Spillovers of Listed Firms.” https://sites.google.com/site/adrienmatray/research.

[40] Moretti, Enrico. 2004. “Workers’ Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions.” *American Economic Review*, 94(3): 656-90.

[41] Moretti, Enrico. 2012. “The New Geography of Jobs”. Houghton Mifflin Harcourt.

[42] Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. “German Jewish Émigrés and US Invention.” *American Economic Review*, 104(10): 3222-55.

[43] Packalen, Mikko, and Jay Bhattacharya. 2015. “Cities and Ideas.” NBER Working Paper No. 20921.

[44] Peters, Michael. 2013. “Heterogeneous Mark-Ups, Growth and Endogenous Misallocation.” http://mipeters.weebly.com/research.html.

[45] Porter, Micheal E. 1998. *The Competitive Advantage of Nations*. New York: Free Press.

[46] Rossi-Hansberg, Esteban and Wright, Mark (2007). “Urban Structure and Growth” *Review of Economic Studies*, 74(2): 597-624.

[47] Saxenian, Annalee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and route 128*. Cambridge, MA: Harvard University Press.

[48] Stiglitz, Joseph. 1977. “The Theory of Local Public Goods.” in *The Economics Of Public Services.* New York, NY: Macmillan

[49] Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Benjamin Jones. 2013. “Atypical Combinations and Scientific Impact.” *Science*, 342(6157): 468-472.

[50] Weitzman, Martin. 1998. “Recombinant Growth.” *The Quarterly Journal of Economics*, 113(2): 331-60.

Appendix

C-Score: Details and Example

The c-score of the class pair $(\mathcal{A},\mathcal{B})$ is calculated according to the following algorithm:

1. The frequency of the citation pair $(\mathcal{A},\mathcal{B})$ in the dataset is computed. To avoid that our results are disproportionately driven by patents that give a large number of citations, we weight every occurrence by the number of possible pair combinations in a certain patent. Mathematically,

$$\text{FREQ}_{\text{OBS}}(\mathcal{A},\mathcal{B}) = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{C_n-1}\sum_{l=m+1}^{C_n}\frac{1}{\binom{C_n}{2}}\mathbb{1}_{\{c_m=\mathcal{A},c_l=\mathcal{B}\vee c_m=\mathcal{B},c_l=\mathcal{A}\}}$$

   where N is the total number of patents in the dataset, C_n is the total number of citations in patent n, c_k and c_l are the k-th and l-th citation of patent n, respectively. It is easy to see that $\text{FREQ}_{\text{OBS}}(\mathcal{A},\mathcal{B})$ is a symmetric function.

2. The theoretical frequency of the citation pair $(\mathcal{A},\mathcal{B})$ is computed. This is the frequency with which one would expect $(\mathcal{A},\mathcal{B})$ to occur if the number of citations from and to a certain class were to be respected. We weight the contribution of each patent by its total number of citations given. Formally,

$$\text{FREQ}_{\text{RAND}}(\mathcal{A},\mathcal{B}) = \begin{cases} \sum_{h=1}^{H}\frac{N_h}{N}\,2\left(\frac{1}{N_h}\sum_{g\in\mathcal{P}_h}\sum_{k=1}^{C_g}\frac{\mathbb{1}_{\{c_k=\mathcal{A}\}}}{C_g}\right)\left(\frac{1}{N_h}\sum_{g\in\mathcal{P}_h}\sum_{k=1}^{C_g}\frac{\mathbb{1}_{\{c_k=\mathcal{B}\}}}{C_g}\right) & if\ \mathcal{A}\neq\mathcal{B} \\ \sum_{h=1}^{H}\frac{N_h}{N}\left(\frac{1}{N_h}\sum_{g\in\mathcal{P}_h}\sum_{k=1}^{C_g}\frac{\mathbb{1}_{\{c_k=\mathcal{A}\}}}{C_g}\right)^2 & if\ \mathcal{A}=\mathcal{B} \end{cases}$$

   where H is the total number of classes, $\mathcal{P}_h$ is the set of patents of class h, C_g the number of citations of patent g patent, and c_k is the k-th citation of patent g. The first term in parenthesis in the first expression is the (weighted) empirical probability that a patent of class i is cited in class h if we took a citation at random from the pool of all the citations of class h. The second term is the (weighted) empirical probability that a patent of class j is cited in class h if we took a citation at random from the pool of all the citations of class h. The multiplication of these two terms is therefore the probability of observing a citation pair $(\mathcal{A},\mathcal{B})$ if two citations were taken at random from the pool keeping the network of citations from class to class constant. This expression is multiplied by two for symmetry reasons. Finally, these probabilities are weighted by

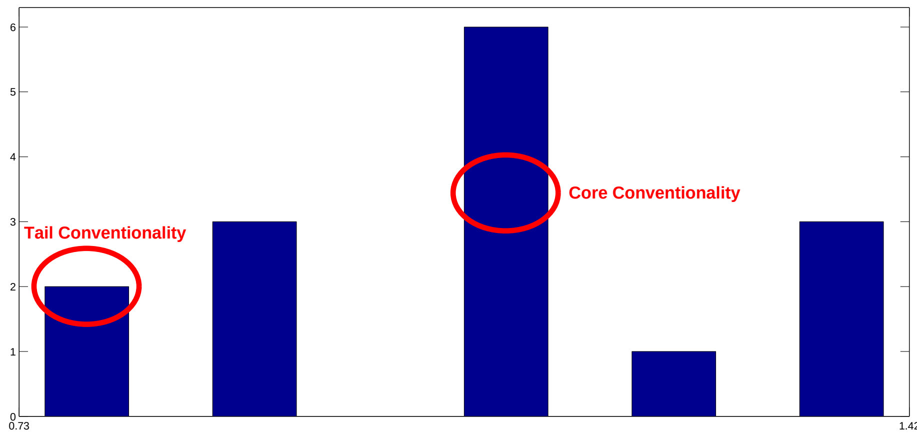


Figure A.1: Example of c-score distribution for a patent. Tail-conventionality corresponds to the 10th percentile of the distribution, core-conventionality corresponds to the median. Similarly, we define tail-unconventionality as one minus tail-conventionality.

the frequency of each class in the universe of patents.

The second expression implements the same idea in the case $\mathcal{A} = \mathcal{B}$.

3. The c-score of each citation pair is calculated as follows:

$$c\left(\mathcal{A},\mathcal{B}\right) = \frac{\text{FREQ}_{\text{OBS}}\left(\mathcal{A},\mathcal{B}\right)}{\text{FREQ}_{\text{RAND}}\left(\mathcal{A},\mathcal{B}\right)}$$

when the c-score is smaller than 1, the pair $\left(\mathcal{A},\mathcal{B}\right)$ is observed in the data less often than what one would expect by taking the some paper in a pseudo-random fashion. We consider this a sign of novelty. On the contrary, when the c-score is bigger than 1, the pair is observed more frequently than the pseudo-random distribution. We consider this a sign of commonality. For expositional convenience, we refer to unconventionality-score (u-score) as:

$$u\left(\mathcal{A},\mathcal{B}\right) = 1 - c\left(\mathcal{A},\mathcal{B}\right)$$

4. Each of the $\begin{pmatrix} C_n \\ 2 \end{pmatrix}$ different citation pairs of each patent is assigned its corresponding c-score. This gives the distribution of c-scores for each patent.

The following is an example of how a patent is assigned a distribution of c-scores. Consider a patent that cites 6 patents of 3 different classes ($CPU \times 3$, $MONITOR \times 2$, $SHOES \times 1$):

$$\{CPU, CPU, CPU, MONITOR, MONITOR, SHOES\}.$$

Take all pairwise combinations of citations and assign each of these combinations the corresponding c-score:

$$\underbrace{(CPU,\, CPU)}_{c=1.4} \times 3 \quad \underbrace{(MON,\, MON)}_{c=1.25} \times 1 \quad \underbrace{(CPU,\, MON)}_{c=1.1} \times 6 \quad \underbrace{(CPU,\, SH)}_{c=0.9} \times 3 \quad \underbrace{(SH,\, MON)}_{c=0.75} \times 2$$

This generates a distribution of c-score for this specific patent (Figure A.1) from which we can extract the 10th percentile (tail-conventionality) and its median (core-conventionality).

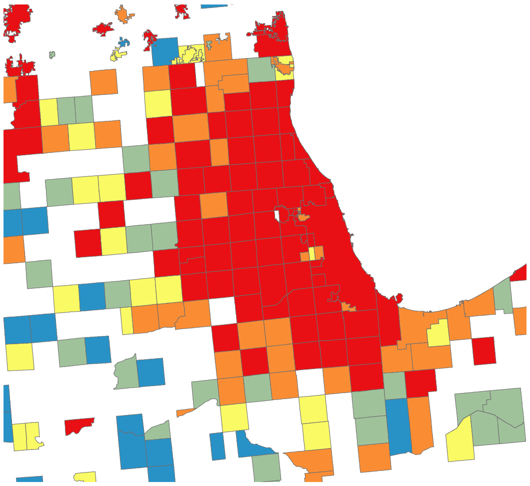
Figures and Tables

Filing Year	# Patent Grants	Filing Year	# Patent Grants
2000	161,388	2006	202,601
2001	209,259	2007	204,957
2002	209,957	2008	199,802
2003	199,752	2009	180,558
2004	198,383	2010	166,985
2005	200,204	Total	2,155,901

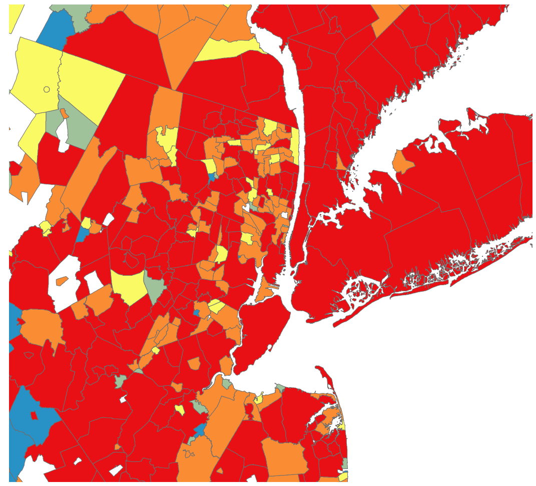
Table A.1: This table reports the number of patents issued from January 2002 to August 2014 and re-arranged by filing year.

	Patents per capita						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log population density	-0.0003 (0.0007)	0.00006 (0.00004)	0.00004 (0.0005)	-0.033** (0.0012)	0.00002 (0.00006)	-0.00009 (0.00004)	-0.034*** (0.0121)
st/y f.e.	No	No	No	Yes	Yes	Yes	CZ/y
Weighted	N.Pat	Pop	No	N.Pat	Pop	No	N.Pat
Winsor	Yes	Yes	Yes	No	No	No	No
N. Obs	18,095	18,095	18,095	18,095	18,095	18,095	18,095
R^2	0.003	0.001	0.001	0.59	0.59	0.01	0.66

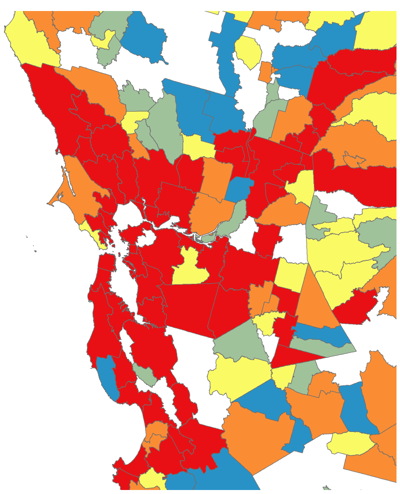
Table A.2: The dependent variable is patents per capita in the CSD/year observation. Standard errors in all the regressions are clustered at the CSD level. Winsorization is by year at the 1% level.



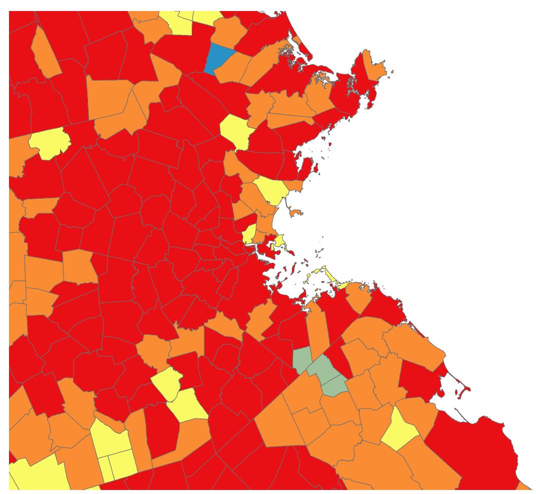
(a) Chicago



(b) New York

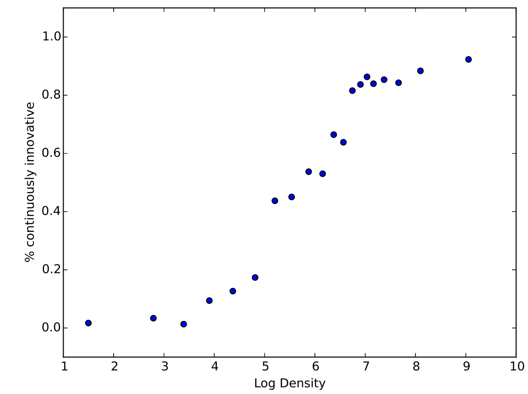


(c) San Francisco

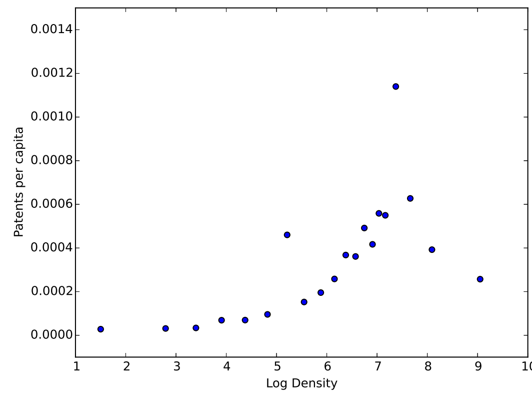


(d) Boston

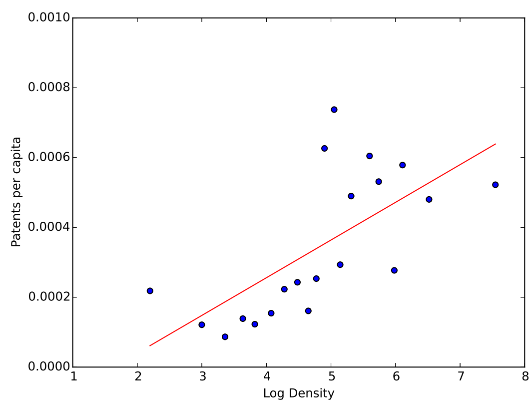
Figure A.2: The figure shows a map of county sub-divisions in the United States. Each CSD is colored according to the number of patents produced between 2000 and 2010. The more red the higher this value; the more blue, the lower. No patents have been filed in the CSD's that are missing in the map.



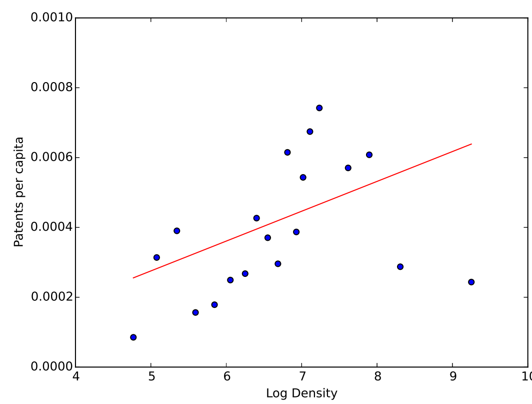
(a) Share of continuously innovative CSD's



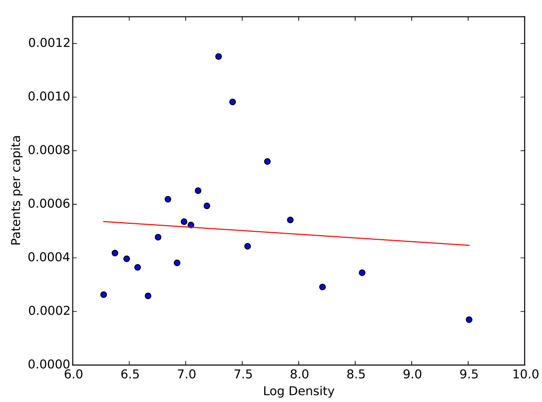
(b) Patenting per capita and density: all CSD's



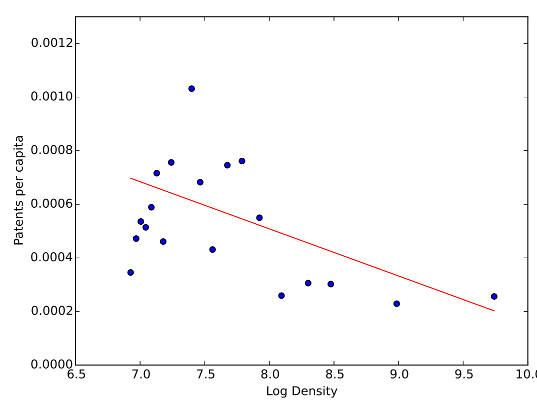
(c) Continuously innovative CZs: patents per capita and density



(d) Patents per capita and density: *dens* $> 100/\text{km}^2$

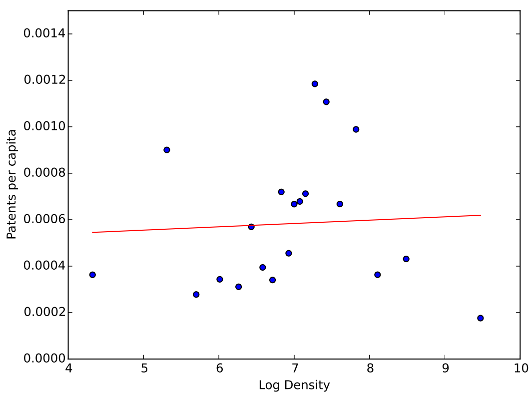


(e) Patents per capita and density: *dens* $> 500/\text{km}^2$

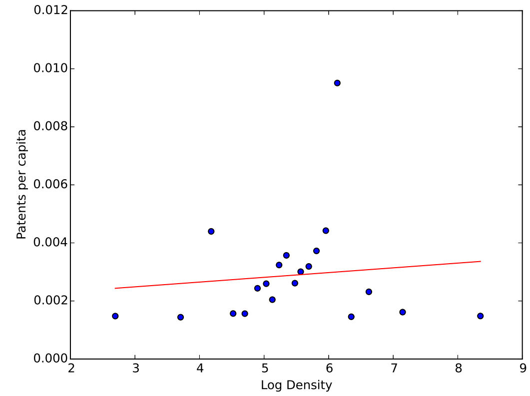


(f) Patents per capita and density: *dens* $> 1000/\text{km}^2$

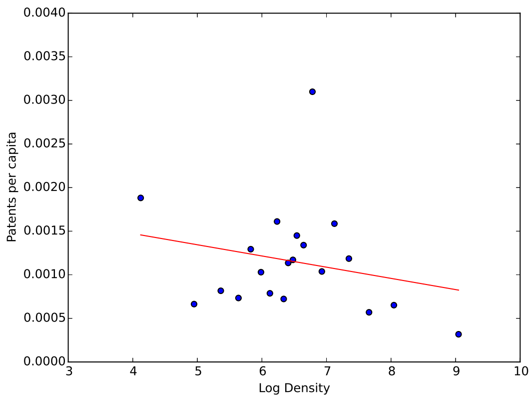
Figure A.3: All the bin-scatter plots are weighted by total population and control for year fixed effects (except for (a)).



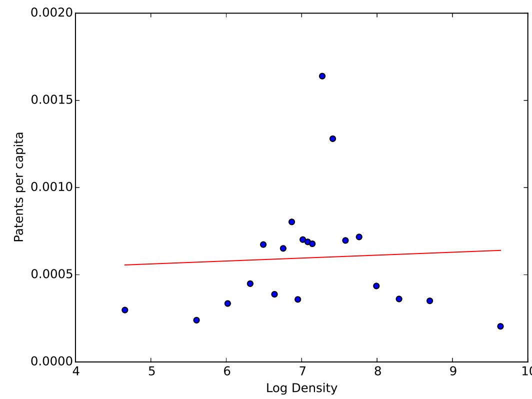
(a) Controlling for share of college graduates



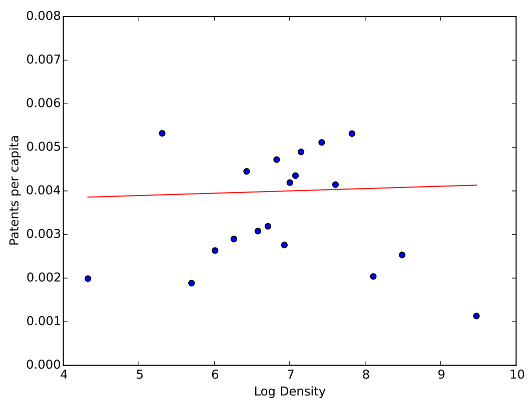
(b) Density of College Graduates



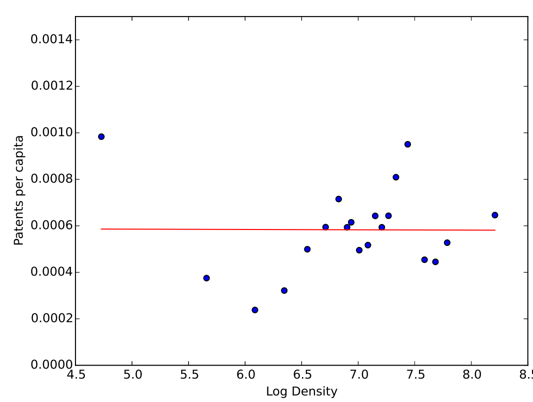
(c) Patents with assignee only and density of workers



(d) Patents with no assignee only



(e) Patents weighted by citation received (year fixed effects included)



(f) Controlling for year and state fixed effects

Figure A.4: Density of population and patents per capita: alternative specifications. All bin-scatter plots except for (b) and (c) weight CSDs by total population and control for year fixed effects. Panel (b) is weighted by total college graduates. Panel (c) is weighted by total workers.

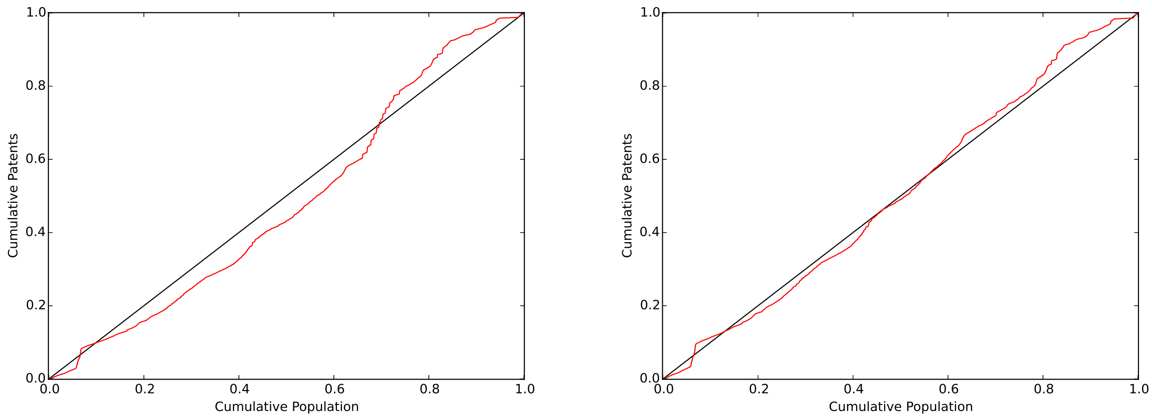


Figure A.5: Cumulative distribution of total population and total patents. In the graphs, CSDs are ranked by their population density. The left-panel includes all the CSDs, the right-panel drops the observations of the San Jose-Palo Alto CSD. The magenta line represents the 45 degree line.

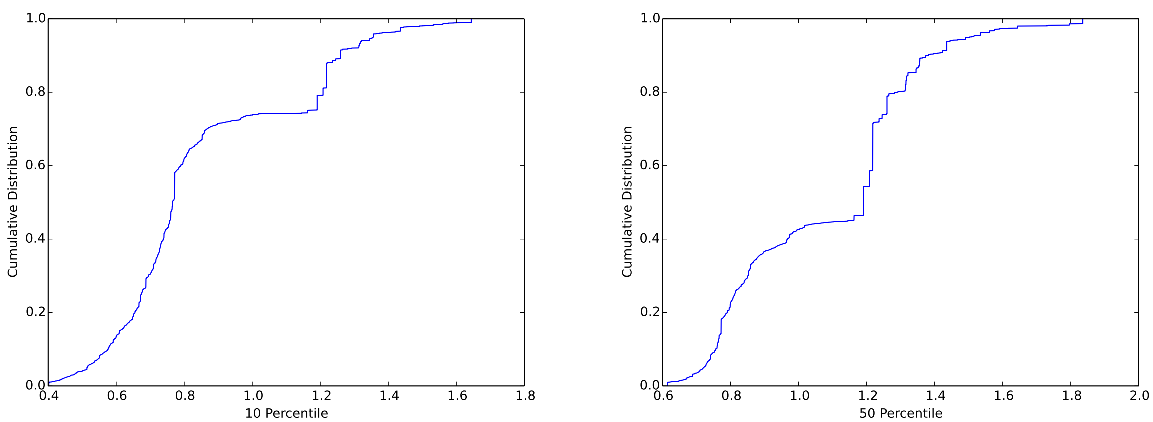


Figure A.6: Cumulative distribution functions of tail conventionality (left) and core conventionality (right) in the universe of patents.

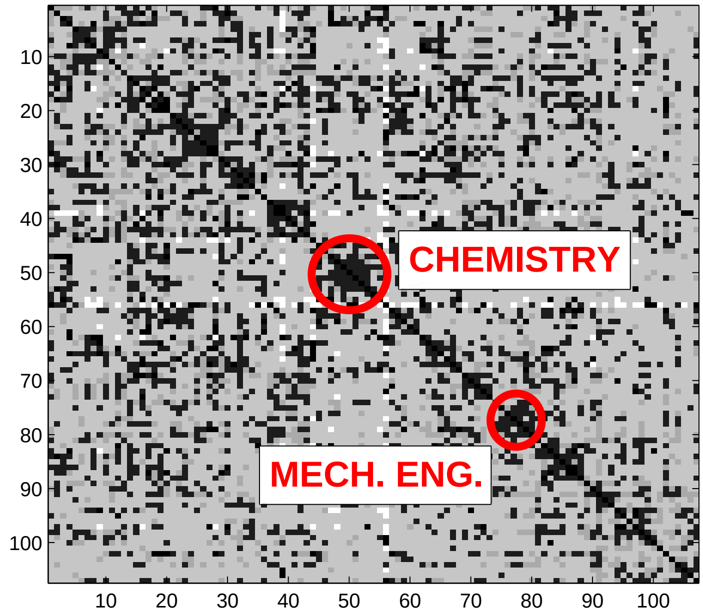
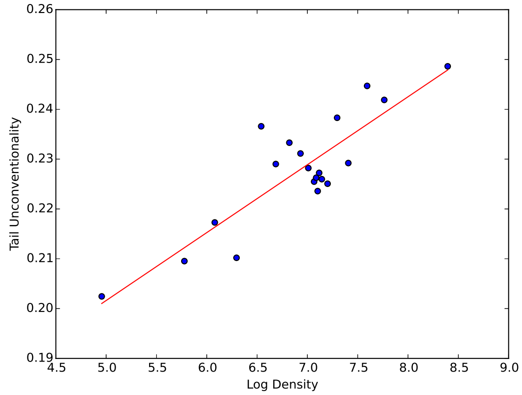


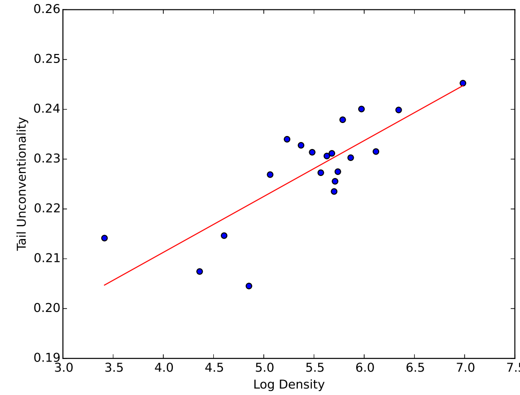
Figure A.7: Every pixel in the matrix indicates a patent class pair. The darker the pixel the higher the c-score assigned to that class pair, the lighter the lower the c-score. Diagonal elements of the matrix show a clear red tendency compared to the rest of the matrix. The "class-clusters" of Chemistry and Mechanical Engineering, among the others, are clearly visible around the diagonal.

	Median Tail Conventionality	
Log population density	0.0131***	0.0136***
	(0.0032)	(0.0035)
Chicago		-0.0113
		(0.0100)
Boston		0.0348**
		(0.0172)
New York		-0.0118
		(0.0097)
San Francisco		-0.0145***
		(0.0055)
st/y f.e.	yes	yes
N. Obs	18,095	18,095
R^2	0.14	0.14

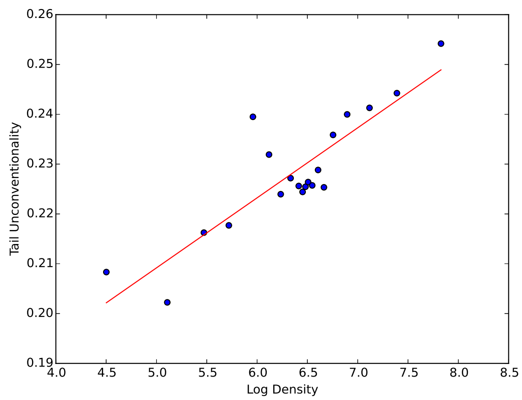
Figure A.8: All regressions are weighted by the total number of patents filed in the CSD/Year observation. Standard errors in all the regressions are clustered at the CSD level.



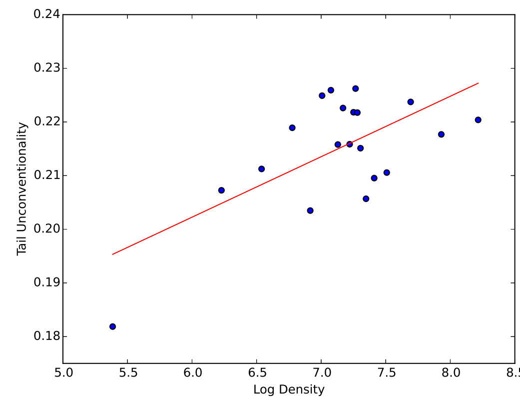
(a) Controlling for share of college graduates



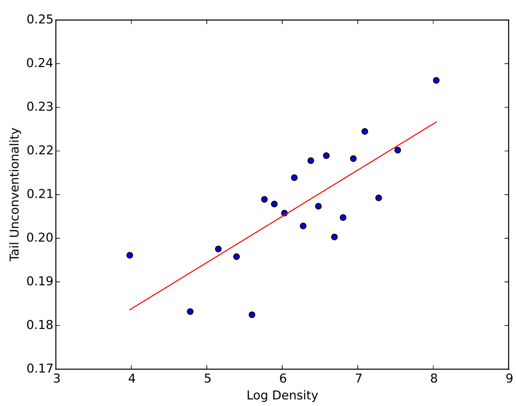
(b) Density of College Graduates



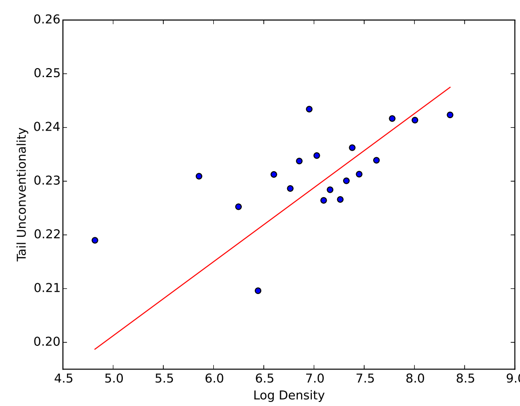
(c) Patents with assignee only and density of workers



(d) Patents with no assignee only

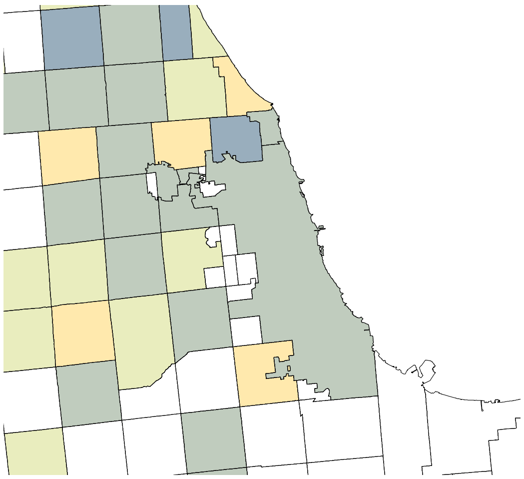


(e) Unweighted

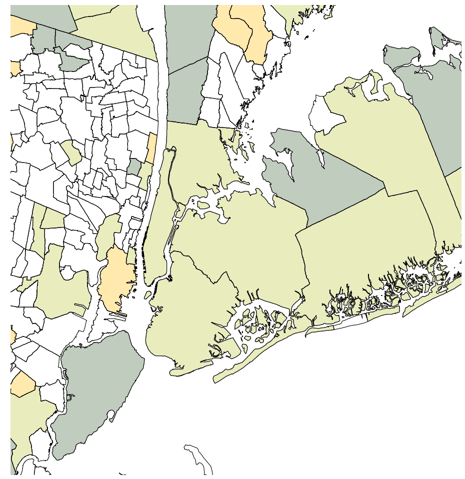


(f) Unconditional

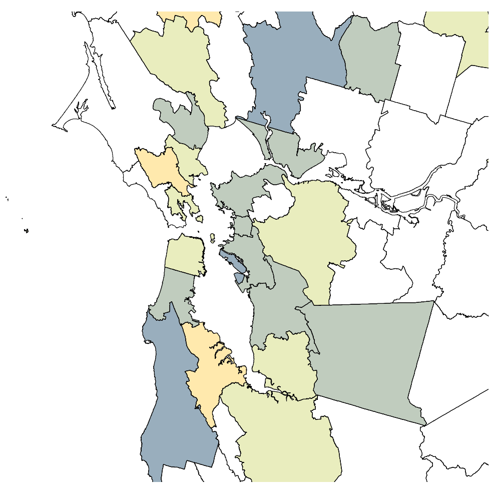
Figure A.9: Density of population and Tail Conventionality of the median patent: alternative specifications. All bin-scatter, except for (f) control for state-year fixed effects. All bin-scatter plots except for (e) weight CSDs by number of patents produced.



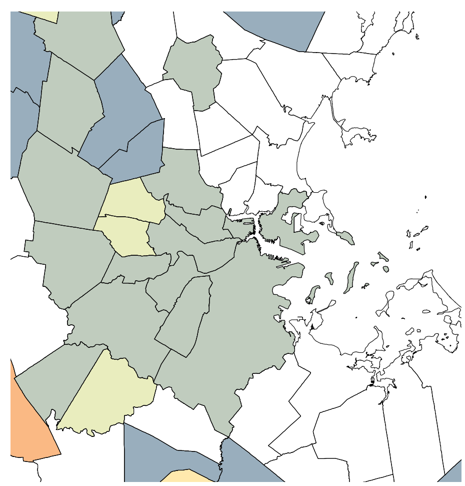
(a) Chicago



(b) New York



(c) San Francisco



(d) Boston

Figure A.10: Four major urban areas divided into County Sub-Divisions. The CSDs colored towards blue produced on average more unconventional patents, whereas CSDs leaning towards red produced more conventional patents according to the associated c-scores. The maps include only CSDs that produced at least 10 patents per year.

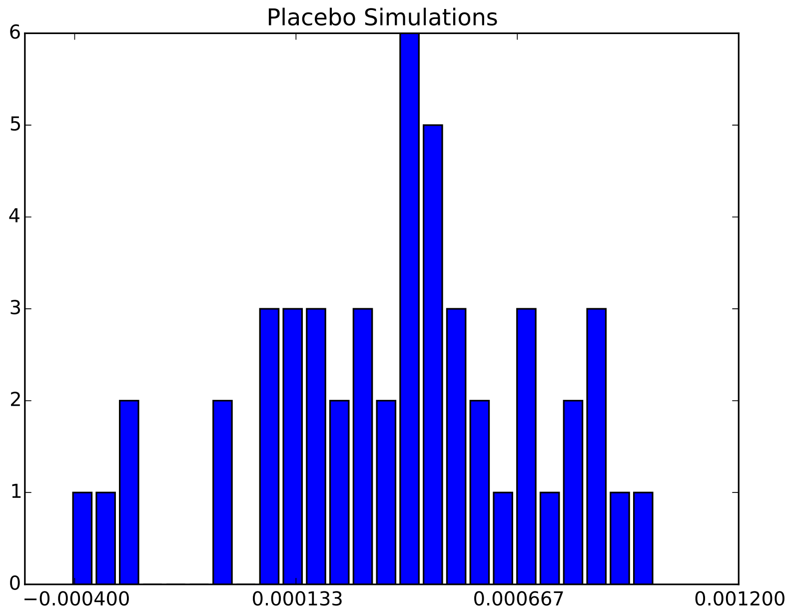


Figure A.11: Placebo experiment: Estimated coefficients from 50 regressions of log-density on concentration index on simulated patent networks.

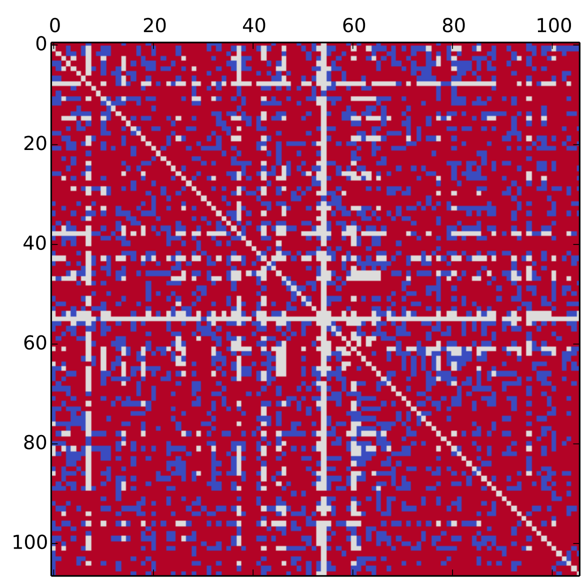


Figure A.12: Sign of the coefficient of the regression in (2.3). Red pixel denotes $\beta_{\mathcal{AB}} > 0$, blue pixel denotes $\beta_{\mathcal{AB}} < 0$. White pixel denotes class pairs for which there are no observations for which $\mathbb{1}_{\{\mathcal{A} \wedge \mathcal{B}\}}$ is true. The matrix is symmetric by construction.

Proofs and Derivations

Derivation of (3.16) and (3.17)

To derive (3.16), simply evaluate $V(\Delta)$ at $\Delta = 0$ and divide by:

$$E_{\Delta}\left[\beta\left(\Delta\right)\right] = \frac{bB\lambda_1}{(1+\lambda_1)}E\left[\mu\left(\Delta\right)^{-1}\right] = \frac{bB\lambda_1\zeta}{(1+\lambda_0)\left[(1+\lambda_1)\zeta+\lambda_1\psi\right]}.$$

To derive (3.17), we need an expression for the normalized wage $w = \frac{w_t}{Y_t}$. The labor demand of an individual firm is:

$$l_{it} = \frac{y_{it}}{a_{it}^L} = \frac{Y_t}{p_{it}a_{it}^L} = \frac{Y_t a_{it}^F}{w_t a_{it}^L} = \frac{Y_t}{w_t}\mu_{it}.$$

In BGP, integrating over all i's:

$$L^F \equiv \int_0^1 l_i \,\mathrm{d}i = \frac{E\left[\mu\left(\Delta\right)^{-1}\right]}{w}$$

and rearranging:

$$w = \frac{\zeta\left(1+\lambda_1\right)}{L^F\left(1+\lambda_0\right)\left[(1+\lambda_1)\zeta+\lambda_1\psi\right]}.$$

Dividing w by $E_{\Delta}\left[\beta\left(\Delta\right)\right]$ yields (3.17).

Proof of Proposition 3.3

We start with the maximization problem of the developer that sets up a company town. We conjecture and verify at the end of the proof that company towns are fully specialized. We focus on the case of a $\mathcal{S}$-specialized location, the one for $\mathcal{D}$-specialized sites is identical. Letting θ^C denote the Lagrange multiplier on its participation constraint, the first order conditions of its problem can be expressed as:

$$\theta^C = N_{\mathcal{S}}$$

$$\tau_{\mathcal{S}}^C = \phi.$$

Plugging this solution in the profit function and imposing the zero profit condition yields:

$$N_{\mathcal{S}}^C = \left[\frac{d\phi\alpha q^{\frac{1}{\alpha}}}{1-\alpha}\right]^{\frac{\alpha}{1-\alpha(\phi+1)}}\left[\frac{1}{\mathcal{W}}\right]^{\frac{\alpha}{1-\alpha(\phi+1)}}.$$

As for the case of a generic town, let $\theta^G_{\mathcal{S}}$ and $\theta^G_{\mathcal{D}}$ denote the Lagrange multipliers on the participation constraints on innovators of type $\mathcal{S}$ and $\mathcal{D}$ respectively. The first order conditions for the developer's maximization problem yield:

$$\theta^G_{\mathcal{S}} = N^G_{\mathcal{S}} \qquad \theta^G_{\mathcal{D}} = N^G_{\mathcal{D}}$$

$$\tau^G_{\mathcal{S}} = \phi + \left(\frac{N^G_{\mathcal{D}}}{N^G_{\mathcal{S}}}\right)^{\phi} \qquad \tau^G_{\mathcal{D}} = \phi + \left(\frac{N^G_{\mathcal{S}}}{N^G_{\mathcal{D}}}\right)^{\phi}$$

while symmetry implies that $\mathcal{U}^G_{\mathcal{S}} = \mathcal{U}^G_{\mathcal{D}}$, which gives:

$$\left(\frac{N^G_{\mathcal{S}}}{N^G_{\mathcal{D}}}\right)^{1-\phi} = \frac{1+\tau_{\mathcal{S}}}{1+\tau_{\mathcal{D}}}.$$

It is easy to see that this problem admits a unique solution in which $N^G_{\mathcal{S}} = N^G_{\mathcal{D}} = \frac{N^G}{2}$ and:

$$\tau^G_{\mathcal{S}} = \phi + 1 \qquad \tau^G_{\mathcal{D}} = \phi + 1.$$

Plugging this solution in the profit function and imposing the zero profit condition gives:

$$N^G = \left[\frac{2^{-(\phi+1)} zd\,(1+\phi)\,\alpha q^{\frac{1}{\alpha}}}{1-\alpha}\right]^{\frac{\alpha}{1-\alpha(\phi+1)}} \left[\frac{\mathcal{V}}{\mathcal{W}}\right]^{\frac{\alpha}{1-\alpha(\phi+2)}}.$$

Finally, we need to show that $N^G > N^C$. Plugging the expressions for N^G and N^C in the utility of the inventor and imposing $\mathcal{U}^G = \mathcal{U}^C$ allows us to write:

$$\mathcal{W} = \left[\frac{2^{-(\phi+1)} zd\,(2+\phi)\left(C^G_1\right)^{\phi+1} - \frac{q^{-\frac{1}{\alpha}}}{\alpha}\left(C^G_1\right)^{\frac{1-\alpha}{\alpha}}}{d\,(1+\phi)\left(C^C_1\right)^{\phi} - \frac{q^{-\frac{1}{\alpha}}}{\alpha}\left(C^C_1\right)^{\frac{1-\alpha}{\alpha}}}\right] \mathcal{V}^{\frac{1-\alpha(\phi+1)}{\alpha}}$$

where

$$C^G_1 \equiv \left[\frac{2^{-(\phi+1)} zd\,(1+\phi)\,\alpha q^{\frac{1}{\alpha}}}{1-\alpha}\right]^{\frac{\alpha}{1-\alpha(\phi+1)}} \qquad C^C_1 = \left[\frac{d\phi\alpha q^{\frac{1}{\alpha}}}{1-\alpha}\right]^{\frac{\alpha}{1-\alpha(\phi+1)}}.$$

Define:

$$C^W \equiv \left[\frac{2^{-(\phi+1)} zd\,(2+\phi)\left(C^G_1\right)^{\phi+1} - \frac{q^{-\frac{1}{\alpha}}}{\alpha}\left(C^G_1\right)^{\frac{1-\alpha}{\alpha}}}{d\,(1+\phi)\left(C^C_1\right)^{\phi} - \frac{q^{-\frac{1}{\alpha}}}{\alpha}\left(C^C_1\right)^{\frac{1-\alpha}{\alpha}}}\right]$$

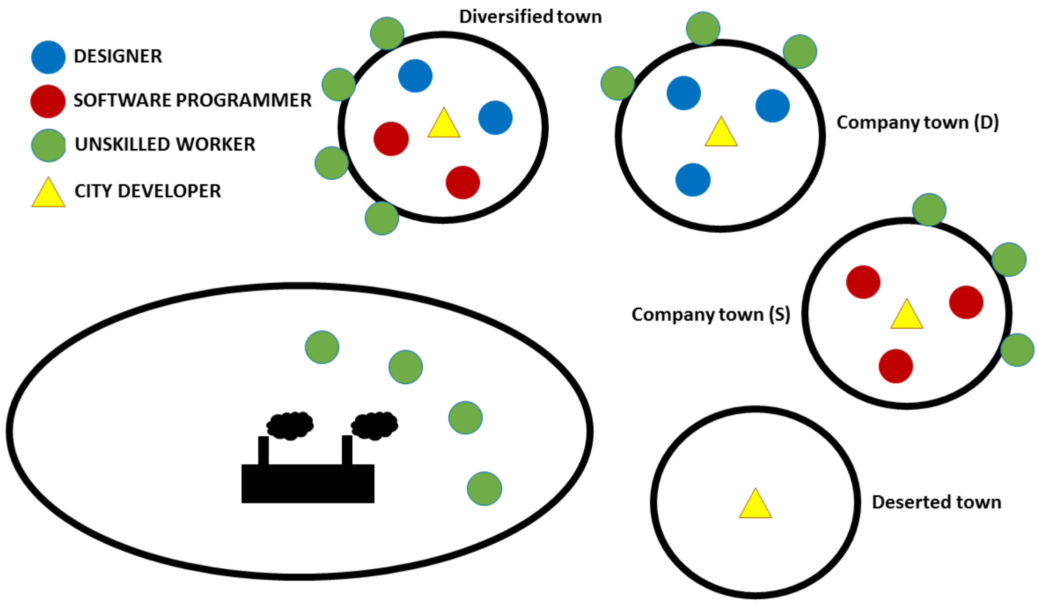


Figure A.13: Spatial economy: Illustration. Innovators from background $\mathcal{S}$ and $\mathcal{D}$ (programmers and designers) sort themselves into the downtown areas of cities. Unskilled labor lives in the outskirts of cities and in the rural areas. Production takes place in rural areas between cities.

then, $N^G > N^C$ if and only if:

$$C_1^G > C_1^C \left(C^W\right)^{\frac{\alpha^2}{[1-\alpha(\phi+2)][1-\alpha(\phi+1)]}}.$$

Working out the expression explicitely, after some algebra, reveals that this is always the case.

It is left to show that company towns are fully specialized. This follows directly from the fact that in a company town, for a given city population, the value of innovation per person is maximized by maximizing intra-field spillovers, i.e. by setting $N^k = N^k_{\mathcal{S}}$ or $N^k = N^k_{\mathcal{D}}$.□

Proof of Corollary 3.4

Patents per capita in diversified cities and specialized clusters can be written respectively as:

$$\begin{aligned} \iota^D &= d\,z \left(\frac{N^G}{2}\right)^{\phi+1} \\ \iota^S &= d \left(N^C\right)^{\phi} \end{aligned}.$$

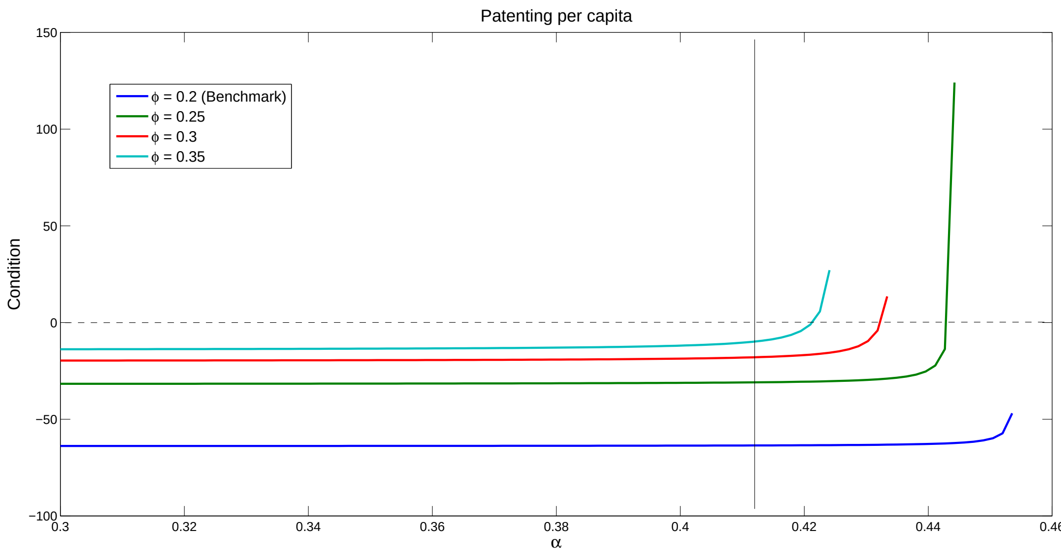


Figure A.14: The two lines represent Condition (3.22) expressed as $LHS - RHS$ for different values of ϕ and α. The baseline parametrization is $\phi = 0.2$. The vertical line denotes the estimated value of $\alpha = 0.41$.

The condition $\iota^S > \iota^D$ is equivalent to (3.21). Moreover, since in equilibrium $N^G < 1$, a sufficient condition for (3.21) to hold is:

$$z\left(\frac{N^G}{N^C}\right)^{\phi} < 2^{\phi+1}.$$

Using the equilibrium expressions for N^G and N^C in Proposition 3.3 and rearranging yields (3.22). □

Derivation of (3.23)

Innovators must earn the same utility from living in a diversified city or in a specialized cluster. Plugging (3.19) and (3.20) into (3.13) and equalizing across locations:

$$\mathcal{U}^G = zd\,(2+\phi)\,C_1^G\left[\frac{\mathcal{V}}{\mathcal{W}}\right]^{\frac{\alpha(\phi+1)}{1-\alpha(\phi+2)}} 2^{-(\phi+1)}\mathcal{V} - \frac{\mathcal{W}}{\alpha q^{\frac{1}{\alpha}}}\left(C_1^G\right)^{\frac{1-\alpha}{\alpha}}\left[\frac{\mathcal{V}}{\mathcal{W}}\right]^{\frac{1-\alpha}{1-\alpha(\phi+2)}} =$$

$$= d\,(1+\phi)\,C_1^C\left[\frac{1}{\mathcal{W}}\right]^{\frac{\alpha\phi}{1-\alpha(\phi+1)}} - \frac{\mathcal{W}}{\alpha q^{\frac{1}{\alpha}}}\left(C_1^C\right)^{\frac{1-\alpha}{\alpha}}\left[\frac{1}{\mathcal{W}}\right]^{\frac{1-\alpha}{1-\alpha(\phi+1)}} = \mathcal{U}^C$$

where C_1^G and C_1^C are constant that only depend on the structural parameters (defined as in the proof of Proposition 3.3). Rearranging this equation yields:

$$\mathcal{W} = C^W \mathcal{V}^{\frac{1-\alpha(\phi+1)}{\alpha}} \tag{A.1}$$

where, again, C^W only depends on structural parameters. The equilibrium conditions can be restated as:

$$\mathcal{I}^G N^G + \mathcal{I}^C N^C = N \qquad \mathcal{I}^G \left(\frac{N^G}{q}\right)^{\frac{1}{\alpha}} + \mathcal{I}^C \left(\frac{N^C}{q}\right)^{\frac{1}{\alpha}} + L^F = L$$

$$\zeta = 2^{-(\phi+1)} z d \mathcal{I}^G \left(N^G\right)^{(\phi+2)} \qquad \psi = d\mathcal{I}^C \left(N^C\right)^{(\phi+1)}$$

in addition to (A.1), (3.16), (3.17) and the two equations in (3.20). This defines a system of 9 equations in 9 unknowns ($\mathcal{I}^C, \mathcal{I}^G, L^F, \zeta, \psi, \mathcal{W}, \mathcal{V}, N^C$ and N^G) that can be solved analytically up to a system of two equations in two unknowns ($\mathcal{I}^G$ and $\mathcal{V}$):

$$C^W \mathcal{V}^{\frac{1-\alpha(\phi+1)}{\alpha}} = \frac{(1+\lambda_1)\left[\rho + 2^{-(\phi+1)} z d \mathcal{I}^G \left(\frac{C_2^G}{\mathcal{V}}\right)^{(\phi+2)}\right] + d\left(\frac{N\mathcal{V}}{C_2^C} - \frac{\mathcal{I}^G C_2^G}{C_2^C}\right)\left(\frac{C_2^C}{\mathcal{V}}\right)^{(\phi+1)}(1-b)\lambda_1}{b\lambda_1\left[L - \left(\frac{N\mathcal{V}}{C_2^C} - \frac{\mathcal{I}^G C_2^G}{C_2^C}\right)\left(\frac{C_2^C}{\mathcal{V}q}\right)^{\frac{1}{\alpha}} - \mathcal{I}^G\left(\frac{C_2^G}{\mathcal{V}q}\right)^{\frac{1}{\alpha}}\right]} \tag{A.2}$$

$$\begin{aligned}\mathcal{V} \;=\; & \frac{\left\{(1+\lambda_0)\lambda_1 d\left(\frac{N\mathcal{V}}{C_2^C} - \frac{\mathcal{I}^G C_2^G}{C_2^C}\right)\left(\frac{C_2^C}{\mathcal{V}}\right)^{(\phi+1)}(1-b) + \lambda_0(1+\lambda_1)\left[\rho + 2^{-(\phi+1)} z d\mathcal{I}^G\left(\frac{C_2^G}{\mathcal{V}}\right)^{(\phi+2)}\right]\right\}}{\left[\rho + 2^{-(\phi+1)} z d\mathcal{I}^G\left(\frac{C_2^G}{\mathcal{V}}\right)^{(\phi+2)}\right] b 2^{-(\phi+1)} d\mathcal{I}^G\left(\frac{C_2^G}{\mathcal{V}}\right)^{(\phi+2)}\lambda_1(1+\lambda_1)} \times \\ & \times \left[(1+\lambda_1)\, 2^{-(\phi+1)} z d\mathcal{I}^G\left(\frac{C_2^G}{\mathcal{V}}\right)^{(\phi+2)} + \lambda_1 d\left(\frac{N\mathcal{V}}{C_2^C} - \frac{\mathcal{I}^G C_2^G}{C_2^C}\right)\left(\frac{C_2^C}{\mathcal{V}}\right)^{(\phi+1)}\right].\end{aligned} \tag{A.3}$$

It is easy to see that equations (A.2) and (A.3) define two implicit functions from $\mathcal{V}$ to $\mathcal{I}^G$ that we can rename $\gamma(\cdot)$ and $\chi(\cdot)$. In particular, for fixed $\mathcal{V}$, equation (A.2) can be rewritten as linear equation in $\mathcal{I}^G$, which admits one and only one solution, while for fixed $\mathcal{V}$, the right-hand side of (A.3) is strictly decreasing in $\mathcal{I}^G$.

In order to guarantee a solution to (A.2), we need that when $\mathcal{I}^G$ is maximum, i.e. $\mathcal{I}^G = \frac{N\mathcal{V}}{C_2^G}$ and $\mathcal{I}^C = 0$, RHS>LHS. This defines a level $\bar{\mathcal{V}}$ such that any equilibrium must have $\mathcal{V} \leq \bar{\mathcal{V}}$. Also, we need that when $\mathcal{I}^G = 0$, LHS>RHS. This defines a level $\underline{\mathcal{V}}$ such that any equilibrium must have $\mathcal{V} \geq \underline{\mathcal{V}}$.[37] Finally, we need to impose that in (A.3) when $\mathcal{I}^C = 0$, LHS>RHS (when $\mathcal{I}^G = 0$ the RHS is equal to infinity, so the other inequality is always true).

[37] It can be shown that $\bar{\mathcal{V}}$ and $\underline{\mathcal{V}}$ are uniquely identified as functions of the structural parameters.

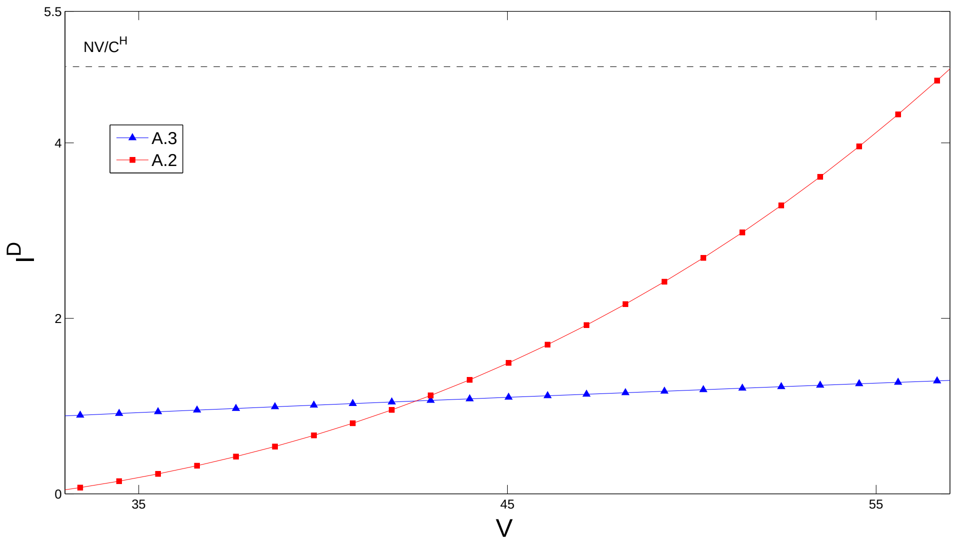


Figure A.15: Intersection of (A.2) and (A.3) under condition (A.4), at benchmark parameter values.

This defines another minimum level of $\mathcal{V}$ as:

$$\mathcal{V}^* = \frac{\lambda_0 (1 + \lambda_1)}{b\lambda_1}.$$

In order to guarantee the existence of the equilibrium, we need to impose conditions on the parameters such that:

$$\mathcal{V}^* < \underline{\mathcal{V}} < \bar{\mathcal{V}}. \tag{A.4}$$

Under condition (A.4), the equilibrium can be found as the intersection of (A.2) and (A.3) as shown in Figure A.15.

Eliminating the ex-ante distinction between generic and company towns

The model described in the main text assumes that city developers, upon establishing a town, must decide *ex-ante* whether to make it a generic or a company town, the first one allowing for inter-field interactions, while the second only allowing for intra-field interactions (e.g. interaction within a firm). We now argue that this assumption is useful but it is not restrictive in our case. This ex-ante distinction is necessary to eliminate a layer of choice from the side of the innovator: upon arrival of an idea, innovators in a company town can only sell it to an existing firm, while innovators in a generic town can only search for a partner from the other field. A less restrictive definition of the equilibrium eliminates this ex-ante distinction and leaves to the innovator the choice of looking for a partner or selling the idea to an existing firm. Innovators in town k would choose to look for a partner from

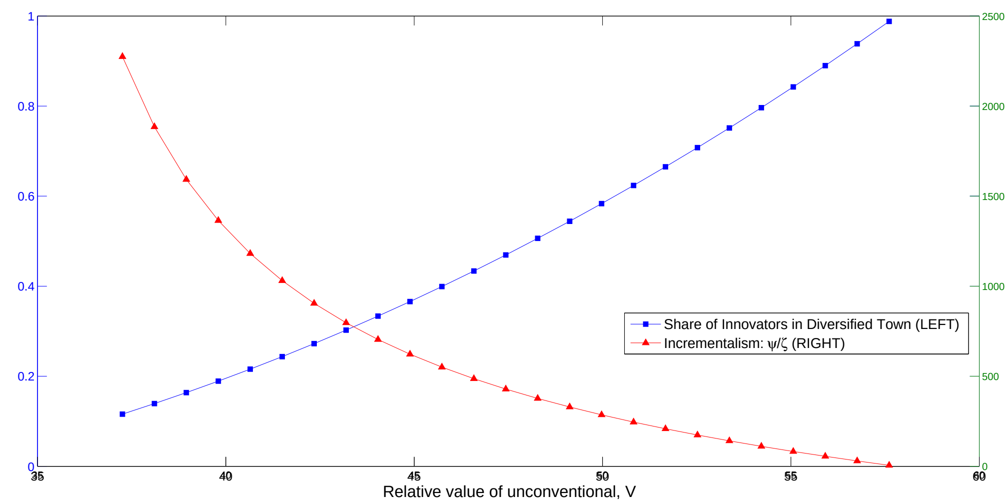


Figure A.16: Equilibrium share of innovators in diversified towns and degree of incrementalism as a function of the relative returns to unconventional ideas $\mathcal{V}$. The parameters are fixed at the baseline calibration.

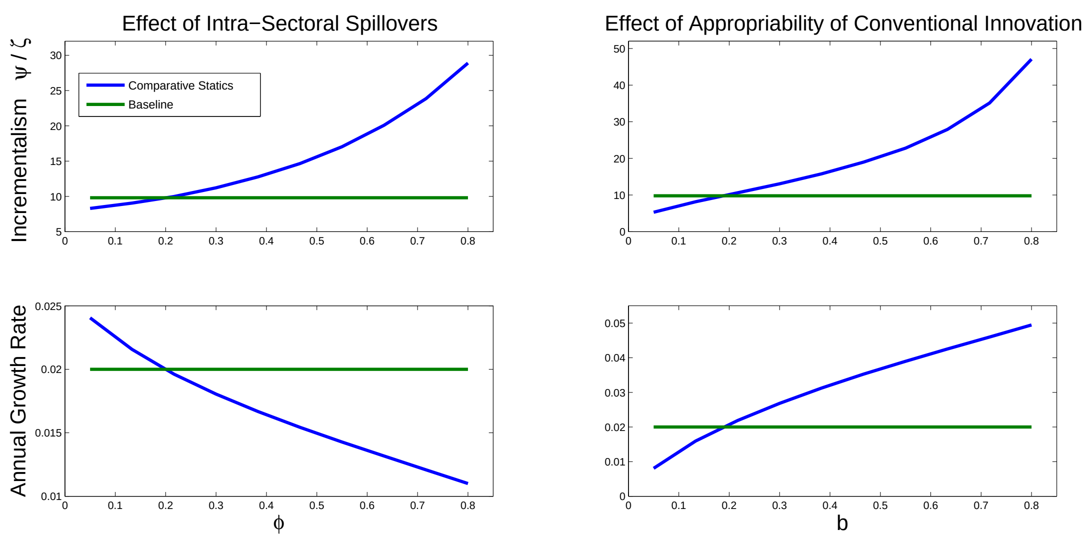


Figure A.17: Equilibrium: Comparative Statics. All the other parameters are kept at their baseline values.

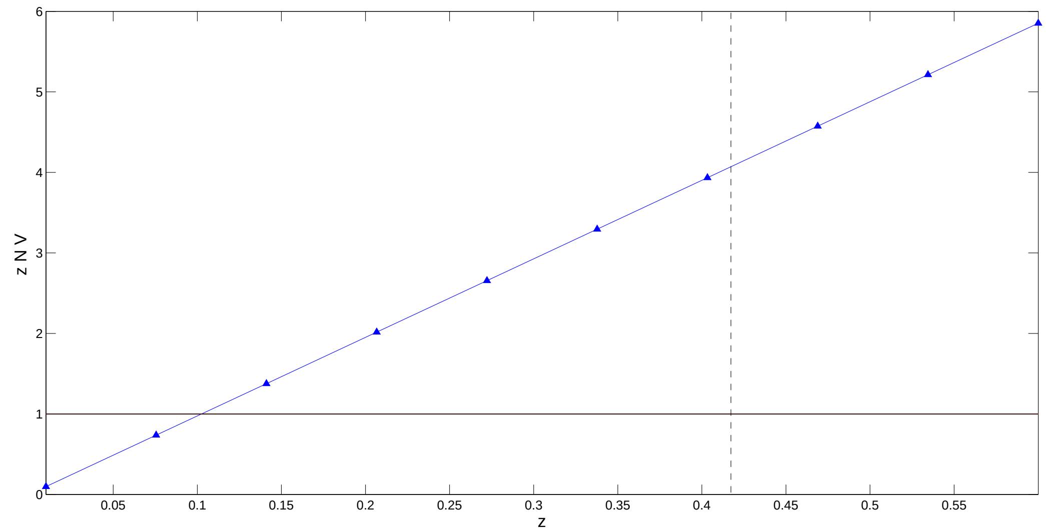


Figure A.18: Condition A.5 under baseline calibration for different values of z. The vertical dashed line corresponds to the calibrated value.

the opposite type if the following conditions are satisfied:

$$\begin{cases} zN_{\mathcal{D}}^{k}\mathcal{V} \geq 1 & \text{for type } \mathcal{S} \\ zN_{\mathcal{S}}^{k}\mathcal{V} \geq 1 & \text{for type } \mathcal{D} \end{cases} \tag{A.5}$$

where $zN_{\mathcal{D}}^{k}$ (or $zN_{\mathcal{S}}^{k}$) is the probability of being matched with a partner, and $\mathcal{V}$ is the return on finding the partner compared to selling the idea to a firm.

This alternative setting complicates the definition of the equilibrium as it requires the inclusion of a set of inequality constraints in the city developer's maximization problem. Nonetheless, here we show that the baseline equilibrium computed with the calibrated parameters in Table 3.1 satisfies Condition (A.5). Figure A.18 shows the left hand side of Condition (A.5) *for diversified towns* for several values of z, while keeping the other parameters at their baseline calibration. The vertical dashed line corresponds to the calibrated value of z. The Condition is satisfied for all values of z above $\bar{z} \simeq 0.1$, which is substantially below the calibrated value of 0.42. Finally, note that since company towns are perfectly specialized, Condition (A.5) is never satisfied: upon arrival of an idea, innovators in company towns will always prefer to sell the idea to the firm.

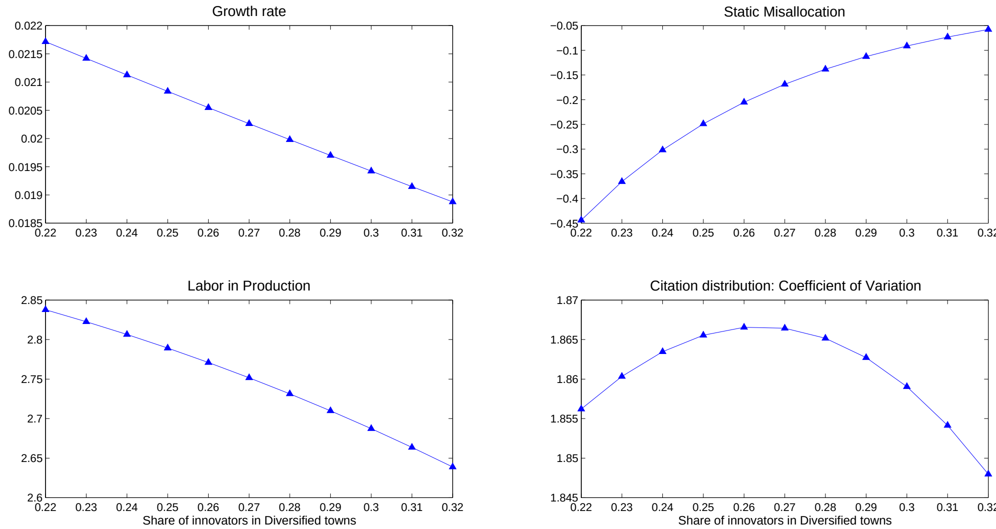


Figure A.19: The figure shows a set of equilibrium outcomes when the urban structure $\mathcal{K} = \{\mathcal{K}^G, \mathcal{K}^C\}$ is fixed at the benchmark equilibrium, but the share of innovators living in generic towns ($\mathcal{K}^G$) is let vary. The growth rate is defined in (3.10), Static Misallocation is defined in (3.6), Labor in Production is defined in (3.18) and the Coefficient of Variation for the citation distribution is computed according to the distribution in (A.6) and (A.7).

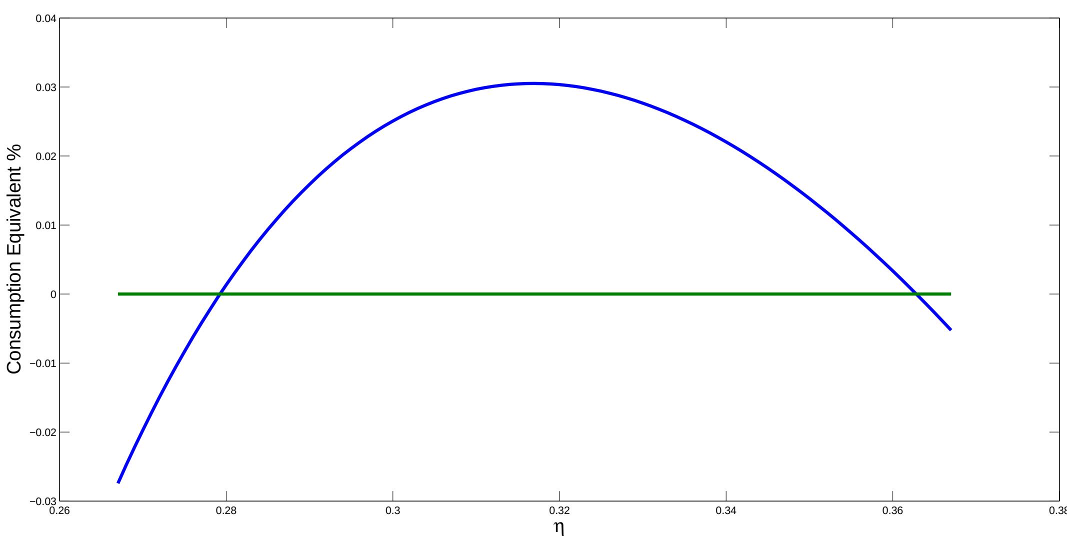


Figure A.20: Welfare gains in annualized consumpiton equivalent units as a function of the share of innovators living in diversified towns. Parameters fixed at their baseline values.

Expected citations

We assume a simple citation structure for patents in the model that is used to pin down the relative frequency of conventional and unconventional innovation. As we show below, higher dispersion in the citations received is associated with more incrementalism, while a more compressed distribution is associated with higher creative destruction.

The citation structure works as follows. Every patent is born with one citation. Upon receiving a conventional improvement, every patent in the product line (since the last unconventional patent) receives $p\,C$ where $p > 0$ is a parameter that we calibrate to match the average number of citations in $T = 12$ years, and C is the number of citations received by the patent prior to the shock. Hence, after the incremental improvement is received, each patent in the product line will have $(1+p)\,C$ citations. In this section, we derive an expression for the measure of citations at a generic time T of patents produced at time 0. We denote this function $\tilde{M}(C,T)$. Since $\tilde{M}(1,0) = 1$ and every mass point in $\tilde{M}$ can be expressed as $(1+p)^{\Delta-1}$ for some integer $\Delta \geq 1$, we can think of $\tilde{M}$ without loss of generality as $M(\Delta,T) \equiv \tilde{M}\left((1+p)^{\Delta-1}, T\right)$. We will distinguish between patents in product lines that have been innovated upon by unconventional inventions and will no longer receive citations (Inactive) and patents that have still the ability of receiving citations (Active). The corresponding measures will be denoted by M_I and M_A respectively.[38] The low of motion of the measure of Inactive patents can be described by the following differential equation:

$$\dot{M}_I(\Delta,T) = \zeta\, M_A(\Delta,T)$$

for any integer $\Delta \geq 1$, with associated initial condition $M_I(\Delta,0) = 0$. The low of motion of the cdf of Active patents can be described by the following recursive system:

$$\begin{cases} \dot{M}_A(\Delta,T) = -M_A(\Delta,T)\,[\zeta+\psi] & \Delta = 1 \\ \dot{M}_A(\Delta,T) = \psi\, M_A(\Delta-1,T) - M_A(\Delta,T)\,[\zeta+\psi] & \Delta \geq 2 \end{cases}$$

The associated initial conditions are $M_A(1,0) = 1$ and $M_A(\Delta,0) = 0$ for $\Delta \geq 2$.

The solution of the system described above has a simple solution:

$$M_A(\Delta,T) = \frac{1}{(\Delta-1)!}\,(\psi\,T)^{\Delta-1}\, e^{-(\zeta+\psi)T} \tag{A.6}$$

[38] Clearly, $M(\Delta,T) = M_A(\Delta,T) + M_I(\Delta,T)$.

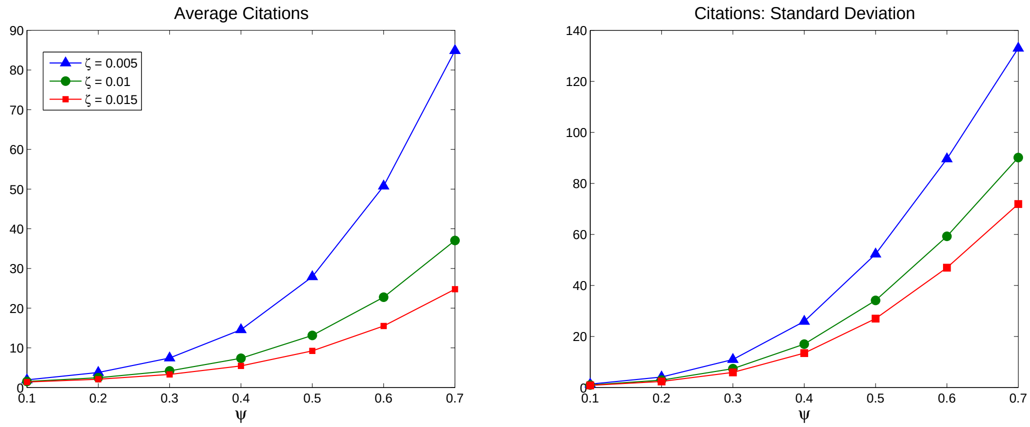


Figure A.21: Average citations and dispertion under different values of ψ and ζ at $T = 12$.

Indeed, for a fixed T, summing over Δ's yields the measure of patents that are still Active:

$$\sum_{\Delta=1}^{\infty} \frac{1}{(\Delta-1)!} (\psi T)^{\Delta-1} e^{-(\zeta+\psi)T} = e^{-\zeta T}.$$

Similarly for Inactive patents:

$$M_I(\Delta, T) = \frac{\zeta \psi^{\Delta-1}}{(\zeta+\psi)^{\Delta}} \left[1 - e^{-T(\zeta+\psi)} \sum_{i=0}^{\Delta-1} \frac{1}{i!} ((\zeta+\psi) T)^i \right].$$

Using the fact that

$$\sum_{i=0}^{\Delta-1} \frac{1}{i!} ((\zeta+\psi) T)^i = e^{T(\zeta+\psi)} \frac{\Gamma(\Delta, (\zeta+\psi) T)}{(\Delta-1)!}$$

yields the close-form solution for the mass of inactive patents with $(1+p)^{\Delta-1}$ citations:

$$M_I(\Delta, T) = \frac{\zeta \psi^{\Delta-1}}{(\zeta+\psi)^{\Delta}} \left[1 - \frac{\Gamma(\Delta, (\zeta+\psi) T)}{(\Delta-1)!}\right] \tag{A.7}$$

where $\Gamma(\cdot, \cdot)$ is the incomplete Gamma function. Note that, by construction:

$$\sum_{\Delta=1}^{\infty} M_I(\Delta, T) + M_A(\Delta, T) = 1.$$