

<https://doi.org/10.1093/ej/ueaa111>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

The Geography of Unconventional Innovation

Enrico Berkes

Ruben Gaetani*

17th June 2019

Abstract

Using a newly assembled dataset of U.S. patents, we show that overall innovation activity is less concentrated in high-density urban areas than commonly believed, but inventions based on atypical combinations of knowledge are indeed more prevalent in high-density cities. To interpret this relation, we propose that informal interactions in densely populated areas help knowledge flows between distant fields, but are less relevant for flows between close fields. 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

*Enrico Berkes: Ohio State University, berkes.8@osu.edu; Ruben Gaetani (corresponding author): University of Toronto, ruben.gaetani@rotman.utoronto.ca. We thank for their comments our discussant Matthew B. Ross, as well as Treb Allen, David Berger, Lawrence Christiano, Matthias Doepke, Martin Eichenbaum, Richard Florida, 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, 2018 European Meeting of the Urban Economics Association in Dusseldorf, 2018 Meeting of the Urban Economics Association in New York, Northwestern, Ohio State University, Toulouse School of Economics, Collegio Carlo Alberto, University of Toronto (Economics), Rotman School of Management, Chicago Fed, Atlanta Fed, Federal Reserve Board, 2014 Conference of Swiss Economists Abroad. Fernando Crupi provided excellent research assistance. This research was funded in part by the Ewing Marion Kauffman Foundation. The contents of the paper are solely our responsibility.

1 Introduction

The idea that informal interactions are central to innovation and knowledge diffusion has become a cornerstone of recent theories of economic growth (Lucas, 1988). This idea implies that economic geography, by determining the extent of those interactions, should play a first-order role in the creation and diffusion of knowledge. A sizable literature has built on this intuition to emphasize the role of cities and agglomeration in driving technological progress and growth (Glaeser et al., 1992; Black and Henderson, 1999; Glaeser, 1999).

In this paper, we empirically examine the link between density and innovation using narrowly geo-referenced information on patenting activity 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 finding reconciles the intuition that density fosters creativity with the observation that the locus of innovation in the U.S. is far from being limited to dense urban areas. 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 places of different density and degree of diversification. This rationale is grounded in the process of knowledge creation and resolves the tension between returns to local specialization (Marshall, 1890) and returns to diversity (Jacobs, 1969) without relying on agents whose productivity is ex-ante heterogeneous. While existing spatial theories of innovation and knowledge diffusion have focused on explaining heterogeneity in size (Davis and Dingel, 2018) or diversification (Duranton and Puga, 2001), our model can account simultaneously for both dimensions, as well as their empirical relation, opening up novel insights for policy analysis. We show 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 empirical analysis is based on the universe of patents granted by the USPTO in the years 2002-2014 georeferenced at the County Sub-Division (CSD, henceforth) level. At this narrow level of disaggregation, the relationship between patenting per capita and density of population appears non-monotonic, peaking around the density of San Jose and declining for higher levels of density. This finding seems to contrast with the common intuition, confirmed by numerous studies (e.g., Bettencourt et al., 2007; Abel et al., 2011) that knowledge-intensive activities tend to concentrate in the most densely populated metropolitan areas.¹

¹This fact is line with Carlino et al. (2007), who find that the density of employment in the urbanized core of a metropolitan area and patenting intensity have an inverted-U shaped relationship that peaks at intermediate levels of density.

However, a closer look to the data reveals a more nuanced connection between density and innovation outcomes. First, innovation produced in densely populated areas is more likely to be built upon unconventional combinations of prior knowledge. To show this fact, we propose a notion of technological distance, based on the observed network of patent citations, that proxies for the intensity of idea flows between fields. We develop an algorithm in the spirit of Uzzi et al. (2013) to evaluate the atypicality of the 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 assigns an index of conventionality (*c-score*) for each citation pair: combinations are conventional if their empirical frequency is large compared to their random frequency. The *c-score* ranks inventions along a dimension that is economically meaningful: unconventional patents are significantly more likely to be highly cited compared to conventional ones, and significantly less likely to be produced by large, publicly traded firms. We find that unconventional innovations disproportionately originate 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.

Second, dense cities host a more diversified pool of learning opportunities. Computing the technological distance between any pair of patents produced in each CSD, we find that pairwise combinations of inventions in high-density CSDs are more technologically distant than combinations in low-density ones. Therefore, inventors in dense cities are more likely to be exposed to ideas from distant backgrounds. This higher degree of local diversification can translate into a higher degree of unconventionality provided that the local pool of innovation is a predictor of the technologies combined into new inventions. To verify this, we adopt a difference-in-difference strategy and look at the patenting activity of pre-existing firms upon arrival in town of companies in different technological fields. We find that such arrival significantly biases the citation behavior of pre-existing entities toward the field of the incoming 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 interpretation of how technological change interacts with economic geography. Overall, suburban areas play a prominent role in the innovation process. Big innovative companies such as IBM or Motorola tend to perform their research in large office parks located outside the main city centers. One possible interpretation is that these companies can organize knowledge flows efficiently within the firm, and do not need to rely on casual 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, aca-

demic departments, or research labs) may not internalize them efficiently. As a result, innovations originating in high-density areas will display more uncommon combinations of prior knowledge. This 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 study its implications for place-based policy analysis.

In our setting, innovators are specialized in a given scientific field. They choose where to locate, balancing congestion costs and innovation opportunities. New product lines are created by combining an unconventional idea, which assembles *diversified* knowledge from multiple fields, with a conventional idea, that embodies *specialized* knowledge from a single field. 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, developing unconventional ideas demands interactions with inventors from different fields, which require search through informal channels. This friction amplifies the benefits from agglomeration in the form of *inter-field* spillovers, and implies that, in equilibrium, diversified cities are more densely populated than specialized ones.

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

This unexplored link opens up novel possibilities for welfare improving place-based transfers. Market forces produce inefficiencies in the balance between the rate of invention and urban congestion, and in the balance between the supply of conventional and unconventional ideas. We study optimal policy in this setting, and characterize conditions under which a planner would use place-based policies to increase urbanization and boost unconventional innovation. We also show that welfare gains from the optimal set of transfers are significantly larger when the planner has the ability to affect the urban structure by creating new cities and reconverting the nature of existing ones, compared to a planner who can only intervene by relocating agents within the current urban structure.

This paper contributes to the empirical and theoretical literature on the role of localized

knowledge spillovers for innovation and growth. The importance of localization and geography for the spreading of knowledge, which dates back to Marshall (1890),² has been the subject of extensive research in recent years since Lucas (1988) and Krugman (1991) seminal papers on economic development and geography. A sizable literature has provided empirical estimates of the size and properties of local knowledge and productivity externalities. Jaffe et al. (1993) find that patent citations display a significant bias towards patents that were produced in the same state and metropolitan area. Audretsch and Feldman (1996) find that, in state-level regressions, the geographical concentration of innovation in an industry is positively related to its knowledge intensity after controlling for the concentration of production.³ Greenstone et al. (2010) estimate significant agglomeration spillovers on TFP by comparing winning and losing counties bidding to attract large plants. Kerr and Kominers (2014) propose a theory of cluster formation based on firm's location and interaction choices and confront its predictions using data on patent citations by technology class, finding that the geographical properties of innovation clusters are controlled by the spatial range of knowledge transmission which is specific to each technology class. We contribute to this literature by providing evidence of a specific class of local knowledge externality, which operates via the *composition* of the knowledge base upon which new ideas are built.

A rich body of literature has investigated the specific role of large cities in the process of economic growth, providing evidence and foundations for Jacobs's (1969) view of density and diversity as central drivers of innovation. A number of studies have focused on the role of specialization and diversity in cities in driving innovation and economic outcomes (Glaeser et al., 1992; Florida and Gates, 2001; Feldman and Audretsch, 1999; Delgado et al., 2014). Duranton and Puga (2001) emphasize the prevalence of young firms in diversified cities, developing a model in which firms locate in diversified places to find their optimal production process, and later relocate to specialized towns to benefit from lower production costs. Lin (2011) uses revisions to occupation classifications to show that "new work" is more prevalent in cities with high density of college graduates and industry variety. Our paper contributes to this literature by suggesting, documenting and studying the implications of a novel channel through which urban agglomeration contributes to economic growth, that is, by increasing the availability and the effectiveness of knowledge exchanges across remote technological fields, while acknowledging that a significant share of patenting occurs

²In 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."

³In a related paper, Rosenthal and Strange (2001) find that industry-level measures of knowledge spillovers have a significant effect on industry agglomeration only at a very narrow geographical level (specifically, ZIP codes).

in low-density, specialized locations. Our interpretation is consistent with Agrawal et al. (2010), who use patent citations to shed light on the reason why we observe the prevalence of large firms locating in “company towns”. The authors find that those companies tend to cite disproportionately previous own inventions, suggesting large benefits for them to internalize within-firm knowledge flows through formal channels, and limited returns to gain access to a diversified learning environment. 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.⁴ A comprehensive review of the existing literature on the geography of innovation can be found in Carlino and Kerr (2015).

This paper also contributes to the theoretical literature on spatial equilibria in a knowledge-based economy. Glaeser (1999) proposes one of the first models of knowledge flows in a spatial setting. The coexistence of diversified and specialized cities in an innovation economy was first analyzed by Duranton and Puga (2001). In our model, density and diversification emerge jointly and endogenously as a consequence of the role of density in promoting knowledge exchange across fields. Davis and Dingel (2018) develop a model in which productivity in cities is fostered by informal interactions among people with heterogeneous abilities. In their setting, the heterogeneity in city size is determined by the comparative advantage of high-skilled individuals in an environment with high learning opportunities. Our setting rationalizes heterogeneity in city size through the complementarity of different forms of innovation, while maintaining homogeneity in agents’ productivity.

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

⁴Packalen 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. On the contrary, 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.

2 Empirical Analysis

The analysis is performed using 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 B.1 reports the total number of patents by filing year. There are several advantages of focusing on this recent sample. First, the recent digitization of the patent archive has made it easier for authors and reviewers to refer to earlier patents. Second, by focusing on a short period, we minimize long-run changes in the propensity to patent and the technological composition of the sample. Third, we can reliably link the location reported in the patents to a wide range of socio-economic and demographic characteristics from the Census and the American Community Survey.

Every patent is assigned to one of the International Patent Classification (IPC) categories.⁵ For each grant, we gather information on the identity and location of the original assignee, of the inventors, as well as on the full list of cited patents (up to a maximum of 1,500 citations per patent). Since our focus is understanding how firm location decisions contribute to explain the observed geography of innovation, in the benchmark analysis, we geolocate each patent according to the location of the original assignee.⁶ To address the well-known concern of the assignee location representing the headquarters of the company instead of the research facility, we restrict attention to the sub-sample of grants in which the location of the assignee and that of at least one of the inventors are in the same state. We also run several robustness checks in which we consider the universe of patents located at the CSD of residence of the first inventor. The results are confirmed in all the specifications. We only consider patents that reference at least two citations. This procedure leads to a final sample of 634,470 patents filed over an 11-year period.

The analysis is conducted at a County Sub-Division (CSD) level. The CSD is 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.⁷ A CSD 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, we interpolate

⁵Since each grant is associated with several IPC classes but only one main USPTO class, we map each USPTO class to a single IPC class based on the associations that recur more often.

⁶Most of the literature on the subject, since Jaffe et al. (1993) uses the location of the inventor. Both alternatives raise a number of issues. For example, when a patent lists multiple inventors whose locations are too far apart to suggest any interaction through spatial proximity, the location of the institution can represent a more accurate indication of the geographical origin of the invention. Many companies issue patents under several locations, corresponding to different establishments or research facilities.

⁷The socio-economic and demographic indicators at a CSD level are available at <https://nhgis.org>.

the values of the demographic variables between 2000 and 2010 assuming a constant growth rate throughout the years.

2.1 A sizable share of innovation originates from low-density areas

The literature on the geography of innovation has long emphasized the importance of density of population in determining innovation outcomes and documented the concentration of innovation activities in densely populated regions. Most studies have focused on large geographical units, such as states, commuting zones (CZs) or metropolitan areas (Carlino et al., 2007; Bettencourt et al., 2007). A visual inspection of the geography of patenting in the U.S. confirms this intuition. The map in Figure B.1 shows the distribution of continuously innovative CSDs in the United States, defined as locations that filed at least one grant per year between 2000 and 2010. There is a clear tendency for innovation activity to concentrate around main urban areas, highlighting the pattern that one would expect. For example, the East-Coast, the Chicagoland, the Texas Triangle and the Bay Area, among others, are all highly innovative regions.

However, the picture becomes more subtle when we narrow down the unit of observation to the CSD-level. The close-up maps in Figure B.2 show the distribution of continuously innovative CSDs in the four most densely populated U.S. metropolitan areas. Two less obvious observations emerge. First, a substantial part of the 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 urban agglomerations, a relevant share of innovation takes place in their suburban portion (e.g., Schaumburg, IL and Mountain View, CA).

When measured at the CSD-level, patenting intensity does not appear to be monotonically increasing in density. The left and right panel of Figure 2.1 display bin-scatter plots⁸ of the empirical relationship between the logarithm of density of population (measured as residents per square kilometer) and patenting intensity (measured as patents per capita) for the balanced panels of the least and most dense CSDs, respectively, each making up 50% of the U.S. population in 2005.⁹ We winsorize the measure of innovation at the 1% level. Table B.2 reports the summary statistics of the main variables. We weight observations by total population, and include year fixed effects to control for aggregate trends in patenting and density.

⁸To obtain bin-scatter plots, we divide the variable on the x -axis into equally-weighted bins and take the mean of the y -variable across the observations falling in each bin. Chetty et al. (2013) show that this methodology graphically captures the correlation between two variables. See <http://michaelstepner.com/binscatter/> for a discussion.

⁹For graphical convenience, we discard observations with density below 5 people per square km.

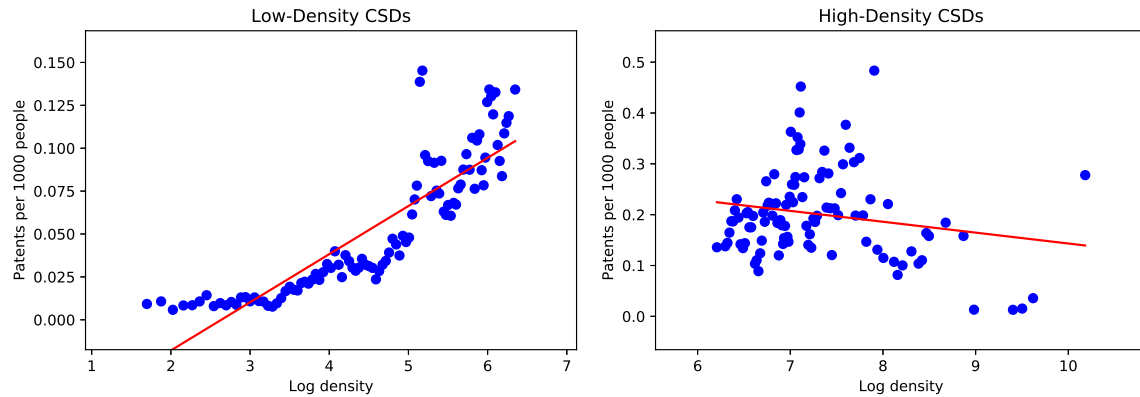


Figure 2.1: Bin-scatter plot of (log) population density and patents per capita in the least dense CSDs (left) and densest CSDs (right) hosting 50% of the U.S. population. The plot is weighted by total population and controls for year fixed effects. The measure of innovation is winsorized at the 1% level. For graphical convenience, we discard observations with density below 5 people per square km.

The relationship is strongly positive in the left portion of the distribution, but becomes weak and, if anything, decreasing in the right portion, with the underlying coefficient implying that doubling density is associated with 0.02 less patents per 1,000 residents, although the estimated coefficient is not statistically different from zero. Patenting intensity peaks at around 1,300 residents per square km (roughly the density of CSDs such as San Jose-Palo Alto, Austin, and Raleigh) and declines for CSDs with higher density.

There are two major measurement concerns related to the interpretation of Figure 2.1. First, 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. Second, the choice of density of population could bias the empirical correlation if units with a high density of skill-rich employment tend to have a low density overall (as would be the case for places like Mountain View, CA, and Armonk, NY).

To address these concerns, we analyze two extreme cases. In the first case, we assume that all the relevant interactions only occur at the workplace. In this case, since we make use of the address of the assignee in our benchmark analysis, 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. The top panel of Figure B.3 reproduces the results of Figure 2.1 by using density of workers computed from the National Establishment Time Series, that contains information on the close-to-universe of establishments in the U.S., including industry indicators. In the middle panel of Figure B.3, we use density of employment in high-tech industries, which controls for the skill composition of the local labor force and provides a more accurate measurement of the interactions that are relevant

for innovation.¹⁰ Although the relationship between density and patenting intensity is now tilted upwards compared to Figure 2.1, the qualitative patterns are preserved, with patenting intensity peaking at intermediate levels of density and flattening at the right tail of the density distribution.

In the second case, we assume that all the relevant interactions only occur at the inventor's residence and its surroundings. This time, learning opportunities would be correctly measured by population density, but the geographical origin of patents would be better approximated by the location of the inventor. In the bottom panel of Figure B.3, we replicate the analysis by geo-locating all the patents at the CSD of residence of the first inventor. The patterns appear even more pronounced than in Figure 2.1, with a significant negative relationship emerging among high-density CSDs.

Finally, in Figure B.4, we show the bin-scatter plot of patenting intensity against log-density of population on the entire sample. The red line represents a non-parametric regression estimated using the actual underlying data. Figure B.4 (a) displays the benchmark case, in which the peak in patenting intensity clearly emerges among CSDs with log-density corresponding to approximately 1,300 people per square km.¹¹ In Figure B.4 (b)-(d), we perform the same analysis using density of (high-tech) employment or geo-locating patents at the CSD of residence of the first inventor. The same qualitative patterns emerge in all those cases.

2.2 Dense locations produce more unconventional innovation

Once we account for a narrow level of aggregation to tell apart highly urbanized centers of larger metropolitan areas from their suburban parts, 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. Agglomeration positively correlates with the rate of invention in low-density places, but not in high-density ones. A possible explanation is that density catalyzes the flow of knowledge across fields that are not fully connected through established networks, whereas formal organizations are able to efficiently internalize knowledge flows within their own field without relying on density-driven informal interactions. As a result, higher density eventually does not translate into more intense patenting, but rather into a shift in the *type* of innovation produced.

¹⁰To identify "high-tech" industries, we use the classification of high-tech NAICS codes provided by the NETS.

¹¹This peak is remarkably close to the tipping point estimated by Carlino et al. (2007), who find that patent intensity is maximized at an employment density of about 849 jobs per square km.

In this subsection, we show that the innovation produced in high-density areas tends to be constructed on a more diversified set of prior knowledge. To assess this fact, we build a measure of atypicality of the knowledge base of each invention. 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 observation that a given pair of technologies is rarely referenced together denotes the lack of knowledge transmission between the two.

2.2.1 Measurement

To 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 patents. The basic idea is to compare the frequency of a bundle of classes in the observed network of citations with the frequency one obtains by assigning citations at random in a replicated network. In this process, the structure of the network is kept constant. In other words, references in the replicated network are randomly reshuffled under the constraint that the total number of citations from each class \mathcal{A} to each other class \mathcal{B} is the same in the two networks.¹² The conventionality-score (or *c-score*) of the pair $(\mathcal{A}, \mathcal{B})$ is then defined as the ratio between the observed frequency and the random frequency:

$$c(\mathcal{A}, \mathcal{B}) = \frac{f_{obs}(\mathcal{A}, \mathcal{B})}{f_{rand}(\mathcal{A}, \mathcal{B})} \times 100.$$

The interpretation of the *c-score* is straightforward: a high value of c implies that we observe classes \mathcal{A} and \mathcal{B} cited together relatively more often in the data than what we would expect if citations were assigned randomly. We refer to such a citation pair as “conventional” and infer that knowledge 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 at random. In this case, the combination is defined as “unconventional”. The details of the

¹²This is a departure from UMSJ that only keep constant the total number of citations from *and* into each class. We do this so that our measure does not depend on the size of a given class relative to the whole sample. While aligning with the basic intuition in UMSJ, we differ from their implementation along two additional dimensions. First, we do not consider the time dimension explicitly 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 that the law of large numbers applies, which allows us to have an analytical expression for the random frequency. This delivers an exact formula that can be computed without simulating the random networks.

algorithm are provided in Appendix A.¹³

Figure B.6 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 c-score of citation pairs of the form $(\mathcal{A}, \mathcal{A})$. We use a chromatic scale in which brighter pixels denote more unconventional pairs. The figure highlights two patterns that validate our measure. First, combinations on the diagonal tend to be more conventional than other citation pairs. This is exactly what we would expect to observe: once a patent cites a certain class, it is likely to cite it again, since that class is likely to play a role in the patent development. Second, around the diagonal we observe some “clusters” of conventionality. This happens because of the hierarchical nature of the International Patent Classification system that assigns assigns 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 a certain technological group is more likely to appear together with another citation in the same group.¹⁴

We assign to each patent an entire distribution of c-scores, one for each pairwise combination of references (hence, a grant with N references is assigned $\binom{N}{2}$, possibly identical, scores).¹⁵ The citation pairs in the left tail of such distribution correspond to the most unconventional combinations of technology classes that contributed to the development of the patent. To measure the degree of conventionality of such combinations, or “tail-conventionality”, we follow UMSJ and use the 10th percentile of the distribution.¹⁶ Figure B.5 plots the cumulative distribution function of tail-conventionality in our final sample.

We now show that having an unconventional tail is a powerful predictor of technological impact. In the spirit of UMSJ, we define a hit patent as an invention that, in terms of citations received, is in the top 5% among grants issued in the same year and belonging to the same class. We then estimate a logit model of the form:

$$\text{logit}(\text{Hit}_{ict}) = \alpha + \mu_c + \mu_t + \beta \times \text{TailConv}_{ict} + \sum_r \delta_r \text{Cit}_{ir} \quad (2.1)$$

where Hit_{ict} is a dummy that takes value 1 if grant i is a hit patent, TailConv_{ict} takes value 1 if the tail-conventionality is above the median of class c in year t , and μ_c and μ_t are class

¹³An alternative measure used in the literature to capture the frequency of co-citation between patent classes is the propinquity score used by Akcigit et al. (2016).

¹⁴However, the c-score identifies technological proximity also between classes that belong to different IPC clusters. The following are some notable 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.

¹⁵We winsorize the c-score measure at the 1% level.

¹⁶Our results are robust to using the minimum.

	Probability of being a hit patent	
	(1)	(2)
Conventional Tail	0.0489 [0.0480, 0.0498]	0.0483 [0.0472, 0.0494]
Unconventional Tail	0.0524 [0.0517, 0.0531]	0.0542 [0.0533, 0.0552]
Year and time f.e.	yes	yes
Reference count indicators	yes	yes
Reference count	≥ 2	≥ 10
N. Obs	634,467	364,499
Pseudo- R^2	0.0531	0.0539

Table 2.1: Implied probability of becoming a hit patent in the logit regression (2.1). The dependent variable is defined as being equal to 1 if the patent is in the top 5% in the forward citation distribution within its issue year - technology class group. The independent variable is defined as being equal to 1 if the patent has tail conventionality larger than 50% of the patents in its issue year - technology class group. Column (1) includes all the patents in the sample. Column (2) only considers patents with a backward citation count of at least 10 references. 95% confidence intervals in parentheses.

and time fixed effects. To correct for the non-linear relationship between number of citations given and number of citation pairs, which might mechanically bias the likelihood of having a conventional tail, we include a set of 10 indicator variables, Cit_{ir} , one for each decile of the number of citations given, that control flexibly for the reference count in each patent.

Column (1) of Table 2.1 reports the implied probabilities of becoming a hit patent conditional on having a conventional or unconventional tail. Patents with an unconventional tail are 0.33 percentage points more likely to become hit patents, relative to a baseline probability of 5%, when (2.1) is estimated using the entire sample. Column (2) performs the same exercise by restricting the sample to patents with at least ten referenced grants, which insures a rich distribution of citation pairs. The marginal effect increases to 0.61 percentage points. The strong correlation between tail-unconventionality and technological impact shows that the c-score ranks patents along an economically meaningful dimension.

2.2.2 Finding

We now explore the hypothesis that density plays a decisive role in 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 networks. This translates into new ideas being obtained by assembling a more hybridized set of prior knowledge.

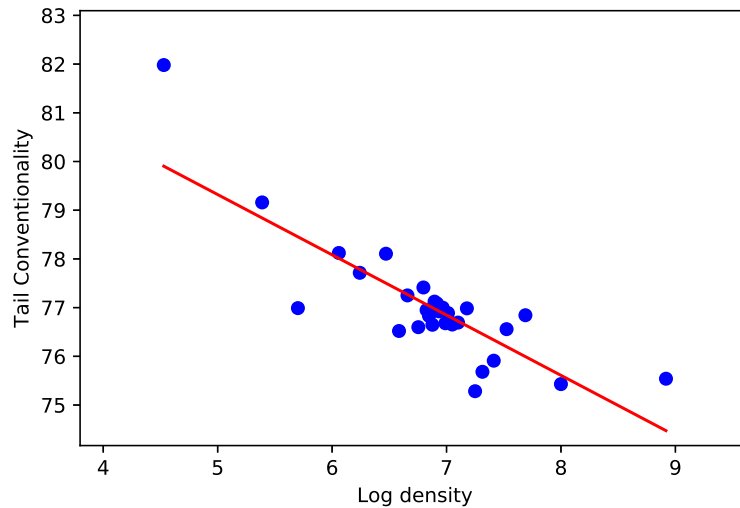


Figure 2.2: The dependent variable is defined as the tail-conventionality of the median patent in the CSD-year observation. The bin-scatter plot is weighted by the total number of patents filed in the CSD/Year observation and controls for CZ and Year fixed effects.

Table 2.2 and Figure 2.2 show several CSD-level correlations between (log) density of population or (knowledge-intensive) employment and the tail-conventionality of the median patent filed in a given CSD and year. In all the specifications, increasing density of population has a negative and significant impact on the tail-conventionality of the median patent. In the baseline specification (Column 2), an increase in density of population equal to the weighted residual inter-quartile range decreases tail-conventionality by 22.5% of its weighted residual inter-quartile range.

To control for possible confounding factors, we add to the specification various CSD-level socio-economic indicators, including (log) median income, the percentage of people with a college degree, and inequality (measured by the Gini index). The results are reported in Table B.3. The effect of density on tail-conventionality remains negative and statistically significant. Interestingly, the coefficient on median income is always positive and statistically significant. This is probably driven by specialized high-income company towns. The share of college graduates and the degree of inequality both have a negative effect, but the coefficients are not significant.

Table B.4 in Appendix reports the marginal effects of (log) density on the probability that the patent has an unconventional tail, obtained from a patent-level logit regression. Consistently with the CSD-level results, the coefficient is positive and significant. An advantage of the patent-level regression is that it allows us to control for whether the patent is produced by a publicly traded firm. The results show that traded firms tend to produce more conventional innovation, which is consistent with the interpretation of unconventional innovation

	Median Tail-conventionality					
	(1)	(2)	(3)	(4)	(5)	(6)
Log population density	-1.13*** (0.22)	-1.24*** (0.33)				
Log employment density			-1.14*** (0.21)	-1.28*** (0.30)		
Log knowledge employment density					-0.79*** (0.14)	-1.01*** (0.18)
CZ and year f.e.	no	yes	no	yes	no	yes
Weighted	Pat	Pat	Pat	Pat	Pat	Pat
N. Obs	33,677	33,677	33,669	33,669	33,669	33,669
R ²	0.015	0.112	0.019	0.117	0.017	0.114

Table 2.2: The dependent variable is defined as 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. C-scores are winsorized (1%) at the patent level. ***p<0.01; **p<0.05; *p<0.1.

as creative destruction events. We believe that this is an interesting fact per se and would deserve further research.

Table B.5 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). In the bin-scatter plots in Figure B.7 we show that the results are fully consistent when using the alternative measures of density and rules for patent geo-location discussed in Section 2.1.

Taken together, Figures 2.1 and 2.2 show that density of population and innovation are indeed tightly related. However, density seems to be more powerful in affecting the type, rather than the rate, of local innovation activities. 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 next two subsections, 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 document can be explained as a result of the local interactions available in densely populated areas.

2.3 Dense locations are more technologically diversified

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

areas are specialized in technologically close fields.

2.3.1 Measurement

In addition to assessing the degree of unconventionality of a single patent, the idea of the c-score can also be employed to evaluate the technological diversification of a given subset of grants: a group of patents is highly diversified if two items drawn at random from the group are likely to give rise to a low c-score, i.e. they 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 of time.

Specifically, we consider all the pairwise combinations of patents filed in each CSD-year pair. Each of these combinations is assigned the c-score corresponding to the pair of patent classes to which the two grants belong. For example, a CSD that has produced N patents in a given year will be assigned $\binom{N}{2}$ c-scores.¹⁷ We then compute the median c-score of those combinations. This procedure delivers an index of concentration for the County Sub-Division CSD in year t defined as:

$$\text{Concentration}(CSD_t) \equiv \text{median}(\{c(CLASS_i, CLASS_j) \mid (i, j) \in CSD_t\}). \quad (2.2)$$

2.3.2 Finding

The bin-scatter plot in Figure 2.3 shows the correlation between density of population and the concentration 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 concentration index yields a coefficient of -3.36 , which implies that an increase in density of population equal to the weighted residual inter-quartile range increases diversification by 37% of its weighted residual inter-quartile range.

A well-established empirical regularity is the correlation between city size and degree of industrial diversification (Duranton and Puga, 2000). To verify that the relationship in Figure 2.3 does not simply reflect this well-known regularity, in Table 2.3 we report the results of the underlying linear regression, directly controlling for the HHI index of the local employment shares in 3-digit NAICS industries, computed using the National Establishment Time Series. The coefficient of density on concentration of innovation drops slightly in absolute value from -3.36 to -3.15 but remains strongly significant.

Since the measure in (2.2) computes the median of a set whose cardinality grows at a binomial rate with respect to the number of local patents, a possible concern is that CSDs

¹⁷To 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.

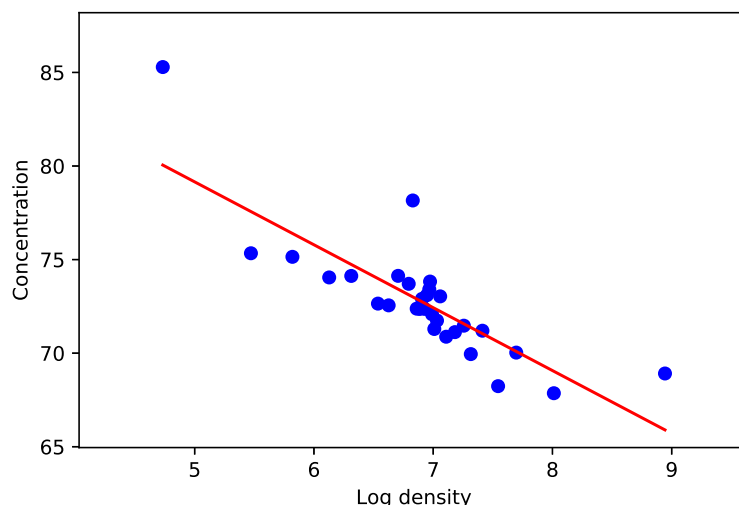


Figure 2.3: Concentration of innovation output and log density of population. The bin-scatter plot is weighted by the total number of patents filed in the CSD/Year observation and controls for CZ and Year fixed effects.

with a higher number of patents (as it is typically the case with dense cities) mechanically have a low index of concentration. To address this possibility, we conduct a placebo experiment in which we generate 30 datasets identical to the original one in terms of total number of patents produced by each CSD-year pair, but reshuffling the geographical allocation of individual patents at random. We then run 30 regressions of the simulated indexes of concentration on log-density. The resulting coefficients are plotted in Figure B.8. Although the distribution of coefficients for the simulated datasets still has a slightly negative average, showing that the index in (2.2) has indeed a dimensionality bias, the estimated coefficients range between -0.14 and 0.03 , with a mean of -0.05 , two orders of magnitude smaller than the estimated coefficient on the original sample (-3.36). This fact illustrates that the correlation in Figure 2.3 cannot be explained by this bias.

2.4 The local pool of ideas predicts local inventions

A key implication of Figure 2.3 is that, if local interactions are important in determining the knowledge embedded into new inventions, people in densely populated regions will have a more diversified pool of possible ideas to draw from and will, as a result, have a higher chance of producing unconventional innovation. 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 is a necessary condition for generating unconventional patents. In this section, we look at the citation behavior of geographically close patenting firms to provide evidence

	Concentration	
	(1)	(2)
Log population density	-3.36*** (0.63)	-3.15*** (0.58)
HHI index (3-digit NAICS)		0.27* (0.14)
CZ and year f.e.	yes	yes
Weighted	Pat	Pat
N. Obs	20,487	20,487
R ²	0.264	0.267

Table 2.3: The dependent variable is defined as the index of concentration in (2.2). 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. The number of observations reflects that (2.2) is only defined for CSD-year observations with at least two patents. ***p<0.01; **p<0.05; *p<0.1.

of this type of cross-field knowledge spillovers.

Inferring the existence of these externalities from the citation behavior of local firms raises the obvious challenge that it is hard to disentangle knowledge spillovers from endogenous locational choices. Places that produce (or are expected to produce) significant knowledge flows between two fields are naturally attractive to firms belonging to those fields. For example, a company that aims to produce high-tech wearable goods, might find it optimal to locate in a town hosting strong CPU and apparel 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. Each incoming firm is assigned to the technology class corresponding to the most recurring class among its patents. Then, for each class-CSD-year observation, we construct an arrival shock as:

$$A_{cdt} = \frac{\sum_{\tau=2001}^t R_{cd\tau}}{P_{d,2000}} \quad (2.3)$$

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.

Percentage of citations to class \mathcal{A} from class $\neq \mathcal{A}$				
	(1)	(2)	(3)	(4)
Arrival of new firm of class \mathcal{A}	0.44*** (0.15)	0.35*** (0.15)	0.45*** (0.19)	0.31*** (0.18)
Class-CSD f.e.	yes	yes	yes	yes
Class-Year f.e.	no	yes	no	yes
Shock arrival year	2001	2001	2002	2002
Average \bar{S}	0.43	0.43	0.43	0.43
N. Obs	569,196	569,196	569,196	569,196
within R^2	0.001	0.005	0.001	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 2002. Standard errors clustered at the CSD/class level are reported in parenthesis. ***p<0.01; **p<0.05; *p<0.1.

The regression model that we estimate is the following:

$$S_{cdt} = \delta_{ct} + \delta_{dc} + \beta A_{cdt} + \epsilon_{cdt} \quad (2.4)$$

where δ_{ct} and δ_{dc} are class-year and CSD-class fixed effects, respectively, that control for aggregate trends in the importance of a given class, and for the time-invariant relevance of a given class in the local innovation output. The dependent variable S_{cdt} is the percentage of citations that class c receives in patents filed by firms that are *pre-existing* in place d and belong to a class different than c .¹⁸ Its unconditional average is 0.43%.¹⁹ 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 (2.4) between year t and year $t + r$ for class c in places d_1 and d_2 :

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

¹⁸For example, how frequently patents that belong to any class different from *CPU* reference items in *CPU*.

¹⁹Given that we have 107 classes, if citations were distributed at random, every class should receive a share of citations from other classes equal to $\frac{1}{106} = 0.94\%$ on average. The fact that the unconditional average is about half that number is a sign of same-class bias in patents citations, since on average half of the citations go to items in the same class of the citing patent itself.

If $\beta > 0$, it means that pre-existing firms producing, say, laptops in a town that has received a high inflow of apparel firms (compared to its size) 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} close in size to its unconditional mean. We also report results where we construct the shocks only considering incoming firms that arrive in or after 2002 (columns 3 and 4). The results are robust.

2.5 Discussion

We provided evidence that a significant share of innovation activity concentrates in low-density CSDs and, as a result, the relationship between density and patenting is non-monotonic. Above a certain threshold higher density does not translate into a higher rate of patenting. We show that it is possible to reconcile this finding with the common wisdom that cities play a key role in fostering innovation. In particular, we show that denser places produce innovation with a higher degree of unconventionality, i.e. innovation that is built upon a more uncommon combination of existing knowledge. We propose that the observed geographical pattern stems from the fact that density is crucial in facilitating learning across distant fields, where ideas are more efficiently transmitted through informal channels. However, this requires dense cities to attract a diversified innovation pool, at the cost of weakening intra-field externalities, which may result in a lower rate of invention. Finally, we provide evidence that the local technological mix predicts the composition of the knowledge background upon which new inventions are built, suggesting that local learning externalities across fields are indeed an important determinant of innovation outcomes.

In the next section, we develop a model of innovation in a spatial economy that accounts for these empirical facts, and generates novel implications for place-based subsidies and innovation policy.

3 Model

In this section, we explore the interaction between economic geography and composition of innovation in a general equilibrium 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 for the co-existence of heterogeneous cities (both in terms of size and degree of diversification) without assuming agents whose ability is ex-ante heterogeneous, differentiated products, or intrinsic productivity differences across different locations. From a normative perspective, the model highlights previously unexplored dimensions for the use of local policies and shows that a system of place-based subsidies can have sizable effects on welfare by affecting the intensity and direction of innovation.

3.1 Production and consumption

A representative household has access to a homogeneous final good that aggregates a set \mathcal{X} of available, perfectly substitutable, varieties:

$$C = \int_{\mathcal{X}} c_i \, di.$$

The economy is closed and there is no investment. Total consumption of the final good is equal to total output. The representative household receives and consumes a lump-sum transfer from the other agents in the economy (innovators, unskilled workers, absentee landlords, city developers and absentee managers).

Active varieties are produced by firms whose production facilities are located outside urban centers, in a congestion-free area where rent is zero. Firm producing variety i decides how much unskilled labor l_i to hire in order to maximize:

$$\max_{l_i} \pi_i \equiv l_i^\beta - w l_i \quad (3.1)$$

where w is the wage of unskilled workers and $\beta \in (0, 1)$. Labor demand for active varieties is equal to:

$$l = \left(\frac{\beta}{w} \right)^{\frac{1}{1-\beta}} \quad (3.2)$$

so that total labor demand in production is equal to $L_F = x l$, where x denotes the measure of active product lines. Firm's profits are equal to:

$$\pi = \gamma w^{-\frac{\beta}{1-\beta}} \quad (3.3)$$

where $\gamma = \left(\beta^{\frac{\beta}{1-\beta}} - \beta^{\frac{1}{1-\beta}} \right)$.

In equilibrium, labor demand is constant across firms, which implies that total output

(and total consumption) is equal to:

$$C = x^{1-\beta} L_F^\beta, \quad (3.4)$$

from which it emerges that consumption depends positively on the mass of product lines, x , and the mass of workers employed in production, L_F .

3.2 Economic Geography

Innovation takes place in a system of cities whose mass and size is endogenous.

3.2.1 Agents, Cities and Housing

The economy is populated by a measure L of unskilled workers and a measure $N = 1$ of skilled innovators. Each innovator is specialized in a given technological field $i \in \{\mathcal{A}, \mathcal{B}\}$.²⁰ For simplicity, we focus on the symmetric case in which the mass of innovators is equal for the two fields, so that $N_{\mathcal{A}} = N_{\mathcal{B}} = 1/2$.

Skilled and unskilled agents are fully mobile. Skilled innovators live in cities with positive rent. Unskilled agents live either in rural areas (close to production facilities), or in the outskirts of the city in which they are employed as construction workers, and do not pay rent. There is a large mass of potential settlements. Each settlement has an area equal to 1. That 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,²¹ who have the ability to tax and provide subsidies to the local economy. Developers have three options for how to utilize their own site:

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.

²⁰The model easily generalizes to the case of multiple technological fields.

²¹As in Becker and Henderson (2000).

In order to attract innovators, developers commit to provide type-specific subsidies, τ_A and τ_B , to the research activity of local inventors. The subsidies are financed by taxing the absentee landlords' profits. 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.²² Housing services, H^k , are provided by competitive landlords, who face a local housing production function:

$$N^k = \left(L^k\right)^\alpha \quad (3.5)$$

where the parameter $\alpha \in (0, 1)$ controls the strength of the congestion force. The rent paid by residents of city k is equal to the marginal cost of producing housing services:

$$R^k = \frac{w}{\alpha} \left(H^k\right)^{\frac{1-\alpha}{\alpha}}. \quad (3.6)$$

Each skilled individual inelastically demands one unit of housing, which implies that, in equilibrium, $N^k = H^k$. The entire landlord's profit is taxed by the local developer, whose revenue is equal to $N^k R^k - wL^k$. To clarify, city developers are large agents at the local level but are small from the point of view of the aggregate economy: 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. The innovation process takes place in two steps:

1. Agents of type $i \in \{A, B\}$ living in a city with N_i^k innovators receive ideas with probability $(N_i^k)^\phi$, where $\phi > 0$ controls the extent of the learning externalities. Namely, individuals receive *intra-field spillovers* by agents of the same type that live in the same location. Being surrounded by a high number of "peers" increases the rate of arrival of ideas.²³
2. Upon receiving an idea, the agent must either execute it conventionally, through the local formal network, or search for an innovator *of the other type* to execute the idea unconventionally:

²²We denote skilled population of type A and B by N_A^k and N_B^k , respectively.

²³This source of agglomeration externality is akin to the cost-reduction externality considered by Duranton and Puga (2001) in that it only affects agents of the same industry.

- (a) The first option (execute it conventionally) is only available to agents living in *company towns*: in this case, the agent makes her conventional idea available.
- (b) The second option (execute it unconventionally) is only available to agents living in *generic towns*: the innovator of class i starts a search process in which she finds an innovator of the opposite type with frequency N_{-i}^k . If the search is unsuccessful, the idea is lost. If the search process is successful, the innovator makes her unconventional idea available.

This process generates an aggregate supply of conventional ideas, denoted by ψ , and an aggregate supply of unconventional ideas, denoted by ζ . The expected monetary value of these ideas depends on their adoption in the production of intermediate varieties. In order to preserve the analytical tractability of the model, we model adoption in a reduced form way. We assume that, in order for a product line to become successfully active (we refer to this case as a “successful innovation”), it must receive ideas from *both* an unconventional *and* a conventional innovator (of either type \mathcal{A} or \mathcal{B}).²⁴ Conventional and unconventional ideas are matched via undirected search according to the following aggregate matching function:

$$x = \zeta^\mu \psi^{1-\mu}, \quad (3.7)$$

where $\mu \in (0, 1)$. The resulting mass of matched ideas, x , is also equal to the total mass of active product lines.

The monetary value of the successful innovation is $a\pi$, where π is defined as in equation (3.3), and $a \in (0, 1)$ denotes the fraction of firm’s profits that is appropriated by the innovators. The remaining share, $1 - a$, is appropriated by absentee managers. The parameter a captures all the factors that contribute to the wedge between the social and the private returns to innovation, such as limited intellectual property protection and dynamic technological spillovers. The monetary value, $a\pi$, is split between the two matched innovators according to Nash bargaining, with the unconventional innovator receiving a fraction $b \in (0, 1)$ of the value, and the conventional innovator receiving the remaining share, $1 - b$. It is important to emphasize that this stage of the search process does not take place in cities, but rather on a decentralized, economy-wide marketplace in which geographical factors are irrelevant.

²⁴This ad-hoc assumption generates a complementarity between conventional and unconventional ideas in a way that preserves the analytical tractability of the model. In an extension available upon request, we micro-found this complementarity in a model of endogenous growth with firm heterogeneity, in which unconventional ideas are interpreted as creative destruction, and conventional ideas as incremental innovations (Aghion and Howitt, 1992; Klette and Kortum, 2004; Peters, 2016). The qualitative results are in line with the ones in the static model presented here, but the dynamic model does not allow for an analytical solution.

It follows from the matching function (3.7) that the rate of conversion of unconventional (conventional) ideas into successful innovations positively (negatively) depends on the ratio between the aggregate mass of conventional to unconventional ideas, $\kappa \equiv \frac{\psi}{\zeta}$. In particular, unconventional ideas become active product lines at rate $\kappa^{1-\mu}$, whereas conventional ideas are executed at rate $\kappa^{-\mu}$.

3.2.3 Utility and innovation rates

To save on notation, in what follows we conjecture that company towns will be *fully specialized* (i.e., they will only host innovators of one background). This conjecture will be proven formally in Proposition 3.2. Let \mathcal{K}^G denote the set of generic cities, and \mathcal{K}_i^C denote the set of i -specialized company towns. The utility of an inventor of type i living in city k can be written as:

$$U_i^k = \begin{cases} (1 + \tau_i^k) (N_i^k)^\phi N_{-i}^k b a \pi \kappa^{1-\mu} - R^k & k \in \mathcal{K}^G \\ (1 + \tau_i^k) (N_i^k)^\phi (1 - b) a \pi \kappa^{-\mu} - R^k & k \in \mathcal{K}_i^C \end{cases} \quad (3.8)$$

In (3.8), the utility of an innovator of type i in city $k \in \mathcal{K}^G$ is given by the frequency of idea generation, $(N_i^k)^\phi$, multiplied by the frequency of matching with an agent of type $-i$, N_{-i}^k , the frequency of finding a conventional idea on the economy-wide marketplace ($\kappa^{1-\mu}$), and the resulting share of profits, $b a \pi$, subsidized by the city developer at gross rate $(1 + \tau_i^k)$, minus the local rental price of housing, R^k . Analogously, the utility of an innovator of type i in city $k \in \mathcal{K}_i^C$ is given by the frequency of idea generation, $(N_i^k)^\phi$, multiplied by the frequency of finding an unconventional idea on the economy-wide marketplace ($\kappa^{-\mu}$), and the resulting share of profits, $(1 - b) a \pi$, subsidized by the city developer at gross rate $(1 + \tau_i^k)$, minus the local rental price of housing, R^k .

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

$$\zeta = \int_{\mathcal{K}^G} \left[(N_{\mathcal{A}}^k)^{\phi+1} N_{\mathcal{B}}^k + (N_{\mathcal{B}}^k)^{\phi+1} N_{\mathcal{A}}^k \right] dk \quad (3.9)$$

$$\psi = \int_{\mathcal{K}_{\mathcal{A}}^C} (N_{\mathcal{A}}^k)^{\phi+1} dk + \int_{\mathcal{K}_{\mathcal{B}}^C} (N_{\mathcal{B}}^k)^{\phi+1} dk. \quad (3.10)$$

In (3.9), the rate of unconventional innovation is given by the integral over all the generic locations of the probability of arrival of ideas for \mathcal{A} -type innovators, $(N_{\mathcal{A}}^k)^\phi$, multiplied by the mass of \mathcal{A} -type innovators in city k , $N_{\mathcal{A}}^k$, and multiplied by the frequency of successful search for a \mathcal{B} -type innovator, $N_{\mathcal{B}}^k$, plus the corresponding product for \mathcal{B} -type innovators. In (3.10), the aggregate rate of conventional innovation is given by the probability of arrival

of ideas for \mathcal{A} -type innovators in \mathcal{A} -specialized company towns, $(N_{\mathcal{A}}^k)^\phi$, multiplied by the mass of \mathcal{A} innovators in city k , $N_{\mathcal{A}}^k$, plus the same rate for \mathcal{B} -type company towns.

The following assumption, that will be maintained throughout, is necessary to guarantee that the solution to the developer's problem, as derived in the proof of Proposition 3.2, is well defined. The assumption ensures that agglomeration externalities are not sufficiently strong to perpetually dominate the congestion force.

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

3.3 Equilibrium

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

$$U_i^k = U_i^{k'} \quad \forall k, k' \in \mathcal{K}^G \cup \mathcal{K}_i^C \quad i \in \{\mathcal{A}, \mathcal{B}\}.$$

In what follows, we will focus on symmetric equilibria in which the contribution to aggregate growth of classes \mathcal{A} and \mathcal{B} is the same. This simply requires ex-ante utility to be equalized also across types:

$$U_{\mathcal{A}}^k = U_{\mathcal{B}}^{k'} \quad \forall k \in \mathcal{K}^G \cup \mathcal{K}_{\mathcal{A}}^C \quad k' \in \mathcal{K}^G \cup \mathcal{K}_{\mathcal{B}}^C.$$

A local developer's revenues are equal to the total profit made by the competitive landlord:

$$\text{Rev}^k = R^k N^k - w L^k = \frac{w (1 - \alpha)}{\alpha} (N^k)^{\frac{1}{\alpha}}.$$

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

$$\text{Exp}^k = \begin{cases} \left[\tau_{\mathcal{A}}^k (N_{\mathcal{A}}^k)^\phi N_{\mathcal{B}}^k + \tau_{\mathcal{B}}^k (N_{\mathcal{B}}^k)^\phi N_{\mathcal{A}}^k \right] b a \pi \kappa^{1-\mu} & k \in \mathcal{K}^G \\ \tau_i^k (N_i^k)^\phi (1 - b) a \pi \kappa^{-\mu} & i \in \{\mathcal{A}, \mathcal{B}\} \quad k \in \mathcal{K}_i^C \end{cases}.$$

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

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

To save on notation, in deriving the equilibrium, we work with the returns on unconventional ideas, \mathcal{V} , and the unskilled wage rate, \mathcal{W} , normalized by the expected returns on conventional ideas:

$$\mathcal{V} \equiv \frac{b a \pi \kappa^{1-\mu}}{(1 - b) a \pi \kappa^{-\mu}} = \frac{b}{1 - b} \kappa \quad (3.11)$$

$$\mathcal{W} \equiv \frac{w}{(1-b)^a \pi \kappa^{-\mu}} \quad (3.12)$$

where $\kappa \equiv \frac{\psi}{\zeta}$. The fact that the relative return on unconventional ideas \mathcal{V} depends linearly on κ highlights a complementarity that is at the root of equilibrium existence and uniqueness.

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

Definition 3.1. A symmetric equilibrium is a set of company towns and generic cities $\mathcal{K} = \{\mathcal{K}^C, \mathcal{K}^G\}$ a utility level U , aggregate innovation rates ζ , ψ and x , profit level π , wage rate w , subsidies $\{\tau_A^k, \tau_B^k\}_{k \in \mathcal{K}}$, local skilled populations $\{N_A^k, N_B^k\}_{k \in \mathcal{K}}$, local rents $\{R^k\}_{k \in \mathcal{K}}$, local unskilled labor $\{L^k\}_{k \in \mathcal{K}}$, firm's labor demand l and unskilled labor employed in production L_F such that:

1. City developers optimally choose $\tau_A^k, \tau_B^k, N_A^k$ and N_B^k and make zero profits
2. ζ, ψ and x are defined as in (3.9), (3.10) and (3.7), respectively
3. U is defined as in (3.8) and is equalized across types and active sites
4. Firm's labor demand l and profits π are defined by (3.2) and (3.3), respectively, and total labor in production is given by $L_F = xl$
5. L^k and R^k are defined as in (3.5) and (3.6)
6. Labor markets clear: $\int_{\mathcal{K}} N_A^k + N_B^k dk = N$ and $L_F = L - \int_{\mathcal{K}} L^k dk$.

3.4 Characterization

We start by solving the city developer's problem of determining the type, size and composition of her 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 (or leave the site deserted).

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

$$\begin{aligned} & \max_{N_A^k, \tau_A^k, N_B^k, \tau_B^k} \frac{\mathcal{W} (1 - \alpha)}{\alpha} \left(N_A^k + N_B^k \right)^{\frac{1}{\alpha}} - \tau_A^k \left(N_A^k \right)^{\phi+1} - \tau_B^k \left(N_B^k \right)^{\phi+1} \\ & \text{subject to :} \quad (1 + \tau_i^k) \left(N_i^k \right)^{\phi} - \frac{\mathcal{W}}{\alpha} \left(N_i^k + N_{-i}^k \right)^{\frac{1-\alpha}{\alpha}} \geq \mathcal{U} \quad i \in \{\mathcal{A}, \mathcal{B}\} \end{aligned}$$

In this problem, the maximand represents the developer's profits, while the constraints represent the level of utility the developer must guarantee to the inventors to convince them

to join the location. Notice that we have written the maximization problem normalizing all terms by the expected returns on conventional ideas, $(1 - b) a \pi \kappa^{-\mu}$, including normalizing inventor's utility, $\mathcal{U} = \frac{U}{(1-b)a\pi\kappa^{-\mu}}$. As a consequence, the returns on conventional ideas that enter the developer's cost and the inventor's utility are normalized to one.

The maximization of a city developer choosing to establish a *generic city* is

$$\max_{N_A^k, \tau_A^k, N_B^k, \tau_B^k} \frac{\mathcal{W} (1 - \alpha)}{\alpha} \left(N_A^k + N_B^k \right)^{\frac{1}{\alpha}} - \tau_A^k \left(N_A^k \right)^{\phi+1} N_B^k \mathcal{V} - \tau_B^k \left(N_B^k \right)^{\phi+1} N_A^k \mathcal{V}$$

subject to :

$$(1 + \tau_i^k) \left(N_i^k \right)^{\phi} N_{-i}^k \mathcal{V} - \frac{\mathcal{W}}{\alpha} \left(N_i^k + N_{-i}^k \right)^{\frac{1-\alpha}{\alpha}} \geq \mathcal{U} \quad i \in \{A, B\}$$

The following proposition characterizes the solution to the developer's problem and the equilibrium system of cities.

Proposition 3.2. *In a symmetric equilibrium, city developers in company towns (C) and generic towns (G) set the optimal subsidy to:*

$$\tau^C = \phi \quad \tau^G = 1 + \phi. \quad (3.13)$$

The optimal population in the two types of locations is:

$$\begin{aligned} N^C &= F^C \frac{1-b}{b} \kappa^{-1} \\ N^G &= F^G \frac{1-b}{b} \kappa^{-1} \end{aligned} \quad (3.14)$$

where F^C and F^G are constants that only depend on the primitives of the model. Company towns are perfectly specialized. Generic towns are perfectly diversified ($N_A^G = N_B^G = \frac{N^G}{2}$) and are more densely populated than company towns.

Proof. See Appendix. □

The city developer's optimal strategy is derived for given equilibrium relative prices \mathcal{V} and \mathcal{W} . In the Appendix, we show that, by substituting this optimal choice into the remaining equilibrium conditions and the definitions in (3.11) and (3.12), the system reduces to one equation in one unknown (the relative supply of conventional and unconventional innovation, κ), that admits one and only one solution, and can be solved analytically. Once the equilibrium value of κ has been determined, backing up the remaining variables becomes trivial. This leads to the following:

Proposition 3.3. *A symmetric equilibrium exists and is unique.*

Proof. See Appendix. □

3.5 Mechanism

Proposition 3.2 represents the model counterpart to Figures 2.2 and 2.3, that show the empirical correlation between density and the conventionality of patenting, and between density and the concentration of the knowledge pool, respectively. The intuition behind Proposition 3.2 is that agents receive 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, the elasticity is equal to ϕ , that is, 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 also increases the matching frequency for 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 developer's optimal strategy maximizes the value of local output per person, given the relative prices \mathcal{V} and \mathcal{W} ,²⁵ although at the aggregate level the equilibrium is in general constrained-inefficient. The equilibrium configuration does not maximize neither the rate of innovation, x , nor social welfare, C , that also depends on the mass of unskilled labor employed in the production of the final good. Two immediate sources of inefficiency are the fact that the equilibrium system of cities depends on the bargaining weight b , that does not enter social welfare, and the lack of full appropriability of the returns from innovation, captured by the parameter a .

The model unique symmetric equilibrium displays the coexistence of specialized company towns that produce conventional ideas, and diversified high-density cities that produce unconventional ideas. The coexistence of both types of cities is dictated by the complementarity between the two forms of innovation, which is transmitted to the equilibrium outcomes via the relative returns of unconventional to conventional ideas, \mathcal{V} . This relative value depends in turn on the relative supply of the two types of ideas, κ , and on the Nash bargaining weight of the unconventional innovator, b . The parameter b encapsulates all the residual forces that control the relative returns of unconventional to conventional ideas, such as the degree of competition and IP legislation.

For a given value of the relative supply, the total rate of invention is determined by the degree of agglomeration, characterized by the mass of active sites of each type, M^C and M^G , and their population density, N^C and N^G . These are in turn determined by the interplay between the agglomeration and congestion forces in the model. The social cost of conges-

²⁵This property was named Henry George Theorem by Stiglitz (1977).

tion is the increase in the demand of unskilled labor that is needed to produce housing, at the expense of the mass of unskilled labor employed in the production of the intermediate varieties (recall that, ultimately, only tradable goods enter consumption and utility). The concavity in the production of housing, controlled by the parameter α , implies that higher agglomeration leads to lower unskilled labor in production. Higher agglomeration can either result from a more concentrated geography (i.e., a lower mass of active sites, each displaying higher density), or, for a given mass of active sites M^C and M^G , from a higher share of skilled labor in the most densely populated sites (of type G).

By substituting the equilibrium values of N^C and N^G into the expressions for ψ and ζ (3.10 and 3.9) it is easy to see that the relative mass of company towns to diversified cities is a decreasing function of the parameter b , which controls the relative returns of unconventional to conventional ideas:

$$\frac{M^C}{M^G} = F^\kappa \frac{1-b}{b},$$

where F^κ is a simple function of the other model primitives. The following proposition shows that also the relative supply of conventional to unconventional ideas is in fact a decreasing function of the the same parameter b .

Proposition 3.4. *The equilibrium relative supply of conventional to unconventional ideas, κ , is a decreasing function of the bargaining weight of the unconventional innovator, b .*

Proof. See Appendix. □

4 Welfare and Policy

We now turn to the study of the optimality of the equilibrium. City developers internalize knowledge externalities at the local level and the associated congestion costs, but they do not internalize non-local externalities such as the pecuniary effect on \mathcal{V} and \mathcal{W} and the imperfect appropriability of innovation, a . The existence of non-local externalities makes the equilibrium constrained-inefficient. In this section, we analyze the optimal local policy of a constrained planner who can tax and provide place-based subsidies to innovators.

Before turning to the study of optimal subsidies, we illustrate how the model delivers a simple additive decomposition of welfare. Since there is no investment in the model, consumption, production and welfare coincide. Using the definition of total consumption in (3.4), we can decompose welfare additively as:

$$\log(C) = \underbrace{(1-\beta)(1-\mu)\log(\psi)}_{\text{Conv. rate}} + \underbrace{(1-\beta)\mu\log(\zeta)}_{\text{Unconv. rate}} + \underbrace{\beta\log(L_F)}_{\text{Congestion}}. \quad (4.1)$$

The first term captures the contribution of the frequency of conventional innovation, ψ , on aggregate welfare. The second term identifies the contribution of the frequency of unconventional innovation, ζ . Finally, the third term captures the benefits from reducing congestion in cities and freeing up unskilled labor to be employed in the production of tradable goods.

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 generic/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 M^C and M^G is fixed. The planner can only reallocate workers across the pre-existing sites. This can be achieved through a simple system of lump-sum transfers $\{T_{\mathcal{A}}^k, T_{\mathcal{B}}^k\}_{k \in \mathcal{K}}$ 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 reduces to the choice of the share $\eta \in (0, 1)$ representing the fraction of innovators living in diversified cities:

$$\max_{\eta \in (0,1)} (1 - \beta) (1 - \mu) \log(\psi) + (1 - \beta) \mu \log(\zeta) + \beta \log(L_F) \quad (4.2)$$

$$\begin{aligned} \text{subject to :} \quad \psi &= M^C \left(\frac{(1-\eta)N}{|\mathcal{K}^C|} \right)^{\phi+1} \\ \zeta &= M^G \left(\frac{\eta N}{2|\mathcal{K}^G|} \right)^{\phi+2} \\ L_F &= L - M^G \left(\frac{\eta N}{M^G} \right)^{\frac{1}{\alpha}} - M^C \left(\frac{(1-\eta)N}{M^C} \right)^{\frac{1}{\alpha}} \end{aligned}$$

with M^C and M^G (the mass of company and generic towns, respectively) given.

Differentiating the problem in (4.2) with respect to η , and evaluating the first-order condition at the equilibrium, reveals that the planner chooses to incentivize agglomeration and unconventional innovation if and only if the following condition is satisfied:

$$- \frac{(1 - \beta) (\phi + 1) (1 - \mu)}{1 - \eta} + \frac{(1 - \beta) (\phi + 2) \mu}{\eta} + \beta \frac{\partial \log(L_F)}{\partial \eta} > 0 \quad (4.3)$$

Since Proposition 3.2 implies that $N^G > N^C$, and due to the concavity in the housing production function (3.5), the third term of condition (4.3) is negative. Hence, the planner can decide to pay additional congestion costs to increase the share of skilled labor in di-

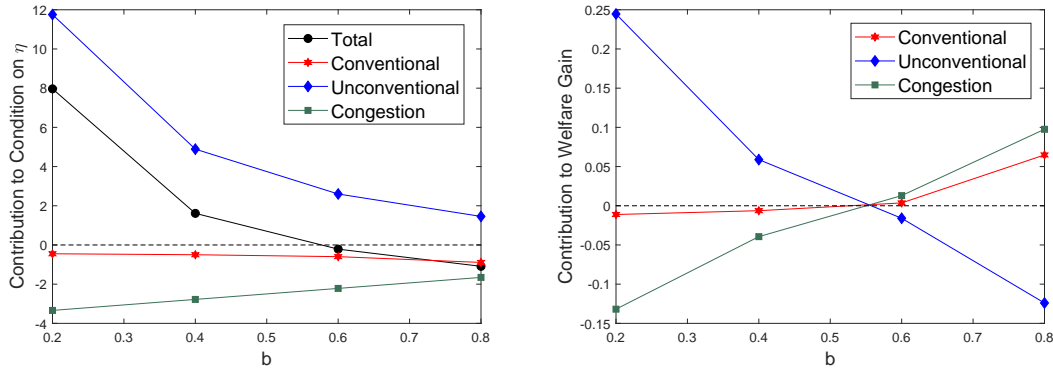


Figure 4.1: Left panel: Contribution of each term to condition 4.3. Right panel: Contribution of rate of conventional ideas, rate of unconventional ideas and congestion costs to overall welfare gain from optimal policy under fixed urban structure.

versified cities and the supply of unconventional ideas. The planner will choose to do so if the benefit from increasing the supply of unconventional innovation, $\frac{(1-\beta)(\phi+2)\mu}{\eta}$, is large enough to outweigh the cost from the loss of conventional innovation, $-\frac{(1-\beta)(\phi+1)(1-\mu)}{1-\eta}$, and the increase in congestion, $\beta \frac{\partial \log(L_F)}{\partial \eta}$.

The bargaining weight b plays a central role in determining departures from optimality in the case of a fixed urban structure. To see this, note that in equilibrium, the share of skilled agents in company towns is proportional to:

$$1 - \eta \propto \frac{\frac{1-b}{b}}{F^I F^G + F^G \frac{1-b}{b}},$$

from which it is immediate to see that it is a decreasing function of b .

The black line in the left panel of Figure 4.1 displays the value of condition (4.3) for a simple parametrization of the model²⁶ and for b spanning between 0 and 1. For sufficiently low values of b , condition (4.3) is positive, which implies that the decentralized equilibrium supplies too little unconventional ideas. The planner chooses to increase the share of agents in diversified cities to increase the supply of unconventional ideas (blue line), reduce the supply of conventional ideas (red line) and increase congestion costs (green line). The opposite policy is implemented if b is sufficiently high to make condition (4.3) negative, that is, the direction of the market forces is such that the equilibrium supply of unconventional ideas is above the socially optimal level.

The right panel of Figure 4.1 displays the contribution of each component of (4.1) to the

²⁶We set $a = 0.5$, $\beta = 0.3$, $\mu = 0.5$, $\phi = 0.2$ and $\alpha = 0.4$. Note that the values of ϕ and α satisfy Assumption 1. The qualitative pattern displayed in Figure 4.1 holds irrespectively from the value of these parameters.

improvement of aggregate welfare for the same range of values of b . When market forces favor inefficiently low supply of unconventional ideas (b low), welfare gains emerge from increasing agglomeration and congestion costs, and reducing the supply of conventional innovation. This kind of efficiency gain, which originates from a tradeoff between congestion and aggregate innovation, is observationally equivalent to the gain that would emerge in a standard model of geography with homogeneous innovation: if social returns to innovation are not fully captured by innovators and other local agents, the market outcome tends to provide too little agglomeration. On the other hand, when market forces push for an inefficiently high supply of unconventional ideas, welfare gains emerge from a *decrease* in congestion costs and an *increase* in the supply of conventional ideas. Contrary to the first case, here efficiency gains originate from a *reduction* in agglomeration. Innovation activities relocate from high-density to low-density areas, freeing up unskilled labor for production, and simultaneously increasing the supply of conventional ideas.

4.2 Flexible urban structure

We now analyze optimal policy when the urban structure is not fixed. We consider a planner whose policy tool consists of class-location specific transfers that multiplicatively subsidize innovation outcomes, financed through a lump-sum tax. In this case, the system of cities is not predetermined: new cities can be created, specialized towns can be converted into diversified cities (or vice versa), and existing cities can be shut down. For a given policy choice, the zero profit condition of city developers will work to determine the size and mass of active sites. This gives the planner some flexibility to affect the urban structure.

There are several possible interpretations of a setting in which policy is not constrained by a fixed urban structure, including a social planner who faces a sufficiently mobile skilled labor force, adopts a sufficiently long-run perspective in its policy implementation, or faces a geography in which the system of cities is not anchored to the presence of natural or historical amenities.

The planner chooses net transfer rates $\{T_A^k, T_B^k\}_{k \in \mathcal{K}}$ between -1 and $+\infty$ and pays to successful innovators the corresponding transfer rate times the effective value of the innovation. Assuming symmetry in the planner's solution, the optimal system of transfers reduces to a pair of net transfer rates $\{T^G, T^C\}$ for diversified and specialized cities, respectively. Given this policy choice, the resulting equilibrium can be found as in Proposition 3.2, with the inventor's income now augmented with the multiplicative transfer. The resulting

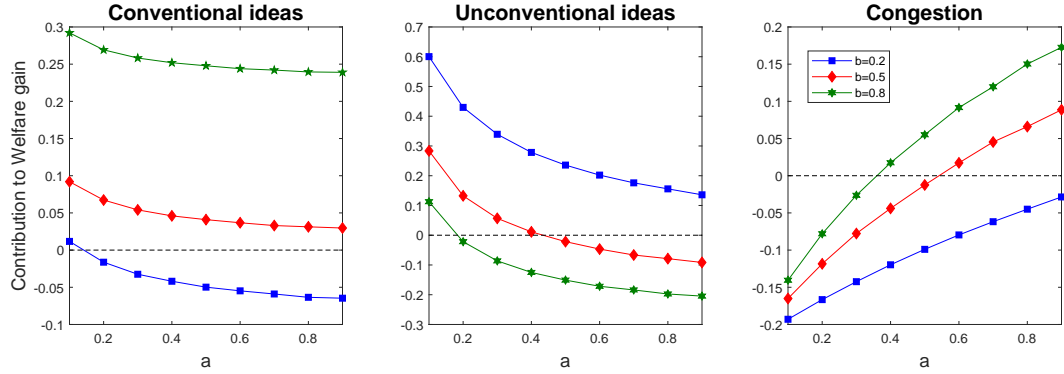


Figure 4.2: Contribution of each component in 4.1 to overall welfare gain from optimal policy under flexible urban structure.

geography has the following solution:

$$\begin{cases} N^C = \frac{1+T^C}{1+T^G} F^C \frac{1-b}{b} \kappa^{-1} \\ N^G = \frac{1+T^C}{1+T^G} F^G \frac{1-b}{b} \kappa^{-1} \end{cases} \quad (4.4)$$

Figure 4.2 plots the contribution of the three components in (4.1) to the welfare gains resulting from the optimal system of transfers, for a range of values of the parameter a (which controls the appropriability of the returns to innovation) and for values of the bargaining weight of the unconventional innovator ranging from $b = 0.2$ to $b = 0.8$). Figure 4.2 reveals two key patterns.

First, as the appropriability of the returns to innovation increases, the contribution of conventional and unconventional ideas to overall welfare gains (left and central panel) decreases, and the contribution of congestion (right panel) increases. This pattern is not specific to a setting with heterogeneous innovation, as a similar result would emerge in an analogous model that only allows for one type of ideas. What is peculiar to this setting is that the contribution of congestion to overall welfare gains becomes *positive* for sufficiently high (but still strictly lower than one) values of a (i.e., even with incomplete appropriability), provided that the bargaining weight of the unconventional innovator b is sufficiently high (red and green lines, right panel). The intuition is that when the bargaining weight is sufficiently high, the decentralized equilibrium supplies an inefficiently high amount of unconventional ideas. Reducing the supply of unconventional ideas requires relocating innovators towards low-density company towns, which reduces congestion.

Second, for sufficiently low values of the appropriability parameter, the optimal policy can increase welfare via a contemporaneous increase in *both* conventional and unconventional ideas, at the cost of a corresponding increase in congestion. This outcome is achieved

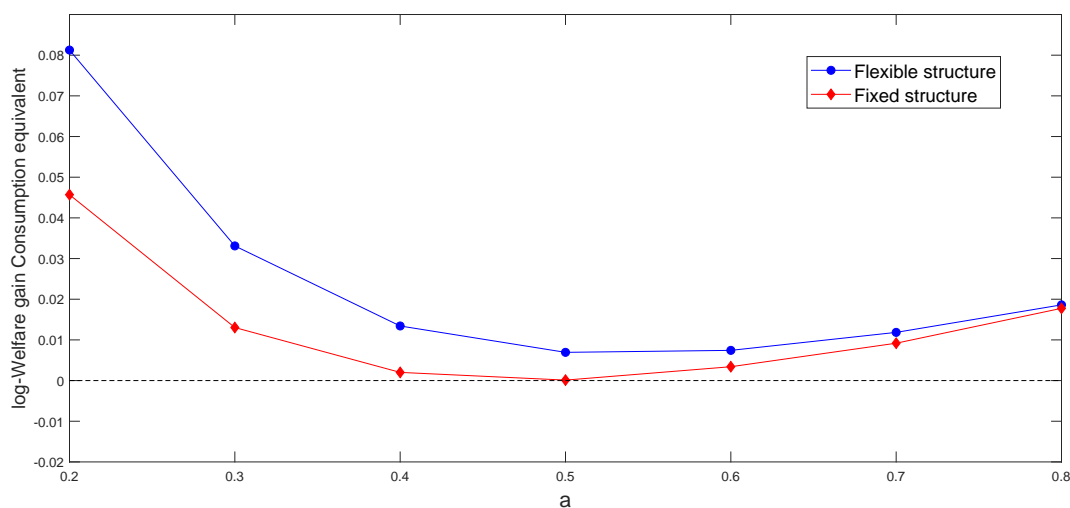


Figure 4.3: Welfare gain (% consumption equivalent) from optimal policy with fixed structure (red line) and flexible structure (blue line).

by shrinking the mass of active sites, and increasing the density of population in both specialized and diversified cities. An implication of this fact is that when the margin of adjusting the urban structure is available, the tradeoff between the two types of ideas disappears, and composition and rate of innovation can both be improved at the same time.

The welfare benefits from having access to this additional margin of adjustment are potentially large. Figure 4.3 compares the welfare gains from optimal policy in the cases of fixed and flexible urban structures for a range of values of a and for an intermediate value of the bargaining weight ($b = 0.5$). The welfare gains are consistently larger under a flexible structure, and the difference is more pronounced for low values of the appropriability parameter. Under the baseline parametrization ($a = 0.5$), the welfare gain from the optimal policy under a fixed urban structure are equal to 0.11% of consumption, against 0.69% under a flexible structure.

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 them a measure of creativity that is novel to this literature. Our empirical

analysis reveals that the combination of ideas embedded into inventions is determined by the local technology mix. This fact 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 the innovation 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 a model of heterogeneous innovation and spatial sorting. In equilibrium, low-density specialized cities, producing conventional ideas, coexist with high-density diversified ones, producing unconventional ideas. 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 productivity. The composition of innovation determines the balance between rate of innovation and congestion costs, which in equilibrium is suboptimal. Our analysis shows that a planner with the flexibility of adjusting the urban structure can achieve welfare benefits from place-based policies that are significantly larger than the ones achievable by a planner that lacks such flexibility. This supports the widespread idea that a fully mobile skilled labor force is an important accelerator for growth in advanced economies. The archetypal geographical mobility of the U.S. labor force was crucial in the development of some of the most innovative areas on the planet (e.g. Silicon Valley, Research Triangle, etc...) and can help explain why over the last decades the United States outperformed Europe in terms of technological leadership and creativity. Future research will be devoted to exploring this nexus.

References

- [1] Abel, Jaison R., Ishita Dey, and Todd M. Gabe. 2012. "Productivity and the density of human capital." *Journal of Regional Science*, 52(4): 562-586.
- [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] Agrawal, Ajay, Iain Cockburn, and Carlos Rosell. 2010. "Not invented here? Innovation in company towns." *Journal of Urban Economics*, 67(1): 78-89.
- [4] Aghion, Philippe, Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323-351
- [5] Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell and David Hémous. 2015. "Innovation and Top Income Inequality." NBER Working Paper No. 21247.
- [6] Akcigit, Ufuk, Murat Alp Celik, and Jeremy Greenwood. 2016. "Buy, keep, or sell: Economic growth and the market for ideas." *Econometrica*, 84(3): 943-984.
- [7] Akcigit, Ufuk, and William R. Kerr. 2010. "Growth Through Heterogeneous Innovations." NBER Working Paper No. 16443.
- [8] 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.
- [9] Audretsch, David B., and Paula E. Stephan. 1996. "Company-Scientist Locational Links: The Case of Biotechnology." *American Economic Review*, 86(3): 641-52.
- [10] 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.
- [11] Berkes, Enrico and Ruben Gaetani. 2018. "Income Segregation and Rise of the Knowledge Economy." Mimeo.
- [12] Berliant, Marcus, and Masahisa Fujita. 2011. "The Dynamics of Knowledge Diversity and Economic Growth." *Southern Economic Journal*, 77(4): 856-84.
- [13] Bettencourt, Luis MA, Jose Lobo, and Deborah Strumsky. 2007. "Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size." *Research policy*, 36(1): 107-120.

- [14] Black, Duncan and Vernon Henderson. 1999. "A Theory of Urban Growth." *Journal of Political Economics*, 107(2): 252-85.
- [15] Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt. 2007. "Urban density and the rate of invention." *Journal of Urban Economics*, 61(3): 389-419.
- [16] Carlino, Gerald, and William R. Kerr. 2015. "Agglomeration and innovation." In *Handbook of regional and urban economics*, 5: 349-404.
- [17] 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.
- [18] 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
- [19] Davis, Donald R., and Jonathan I. Dingel. 2014. "The Comparative Advantage of Cities." NBER Working Paper No. 20602.
- [20] Davis, Donald R., and Jonathan I. Dingel. 2018. "A Spatial Knowledge Economy." Mimeo.
- [21] Delgado, Mercedes, Michael Porter and Scott Stern. 2014. "Clusters, convergence, and economic performance." *Research Policy*, 43(10): 1785-99
- [22] Duranton, Gilles, and Diego Puga. 2000. "Diversity and Specialisation in Cities: Why, Where and When Does it Matter? ." *Urban Studies*, 37(3): 533-555.
- [23] Duranton, Gilles, and Diego Puga. 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products." *American Economic Review*, 91(5): 1454-1477.
- [24] 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.
- [25] Florida, Richard, and Gary Gates. 2001. "Technology and Tolerance: The Importance of Diversity to High-Technology Growth." in *The Brookings Institution Survey Series*, June.
- [26] Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114 (3): 739-767.

- [27] Gaubert, Cecile. 2018. "Firm sorting and agglomeration." NBER Working Paper No. w24478.
- [28] Glaeser, Edward L., Hedi Kallal, Jose' Scheinkman and Andrei Schleifer. 1992. "Growth in Cities." *Journal of Political Economics*, 100(6): 1126-52.
- [29] Glaeser, Edward L. 1999. "Learning in Cities." *Journal of Urban Economics*, 46(2): 254-77.
- [30] 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.
- [31] Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings." *Journal of Political Economy*, 118(3) : 536-598.
- [32] Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28(4): 1661-1707.
- [33] 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.
- [34] Hanley, Douglas. 2015. "Innovation, Technological Interdependence, and Economic Growth." Mimeo.
- [35] Imbs, Jean, and Romain Wacziarg. 2003. "Stages of Diversification ." *American Economic Review*, 93(1): 63-86.
- [36] Inoue, Hiroyasu, Kentaro Nakajima and Yukiko U. Saito. 2015. "Localization of Collaborations in Knowledge Creation." Mimeo.
- [37] Jacobs, Jane. 1969. *The Economy of Cities*. New York: Vintage.
- [38] 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.
- [39] Jovanovic, Boyan. 2015. "Learning and Recombination." Mimeo.
- [40] Katz, Bruce and Julie Wagner. 2014. "The Rise of Innovation Districts: A New Geography of Innovation in America." in *Metropolitan Policy Program at Brookings*, May

- [41] Kerr, William R., and Scott Duke Kominers. 2015. "Agglomerative forces and cluster shapes." *Review of Economics and Statistics*, 97(4): 877-899.
- [42] Klette, Tor J., and Samuel Kortum. 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy*, 112(5): 986-1018.
- [43] Kogan, Leonid, Dimitris Papanikolaou, Amit Seru and Noah Stoffman., 2012, "Technological Innovation, Resource Allocation, and Growth." NBER Working Paper No. 17769.
- [44] Kogan, Leonid, Dimitris Papanikolaou, and Noah Stoffman. 2013. "Technological Innovation: Winners and Losers." NBER Working Paper No. 18671.
- [45] Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy*, 99(3): 483-99.
- [46] Lucas, Robert Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1): 3-42.
- [47] Lucas, Robert Jr. and Benjamin Moll. 2014. "Knowledge Growth and the Allocation of Time." *Journal of Political Economy*, 112(1): 1-51.
- [48] Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2017. IPUMS National Historical Geographic Information System: Version 12.0 [Database]. Minneapolis: University of Minnesota. <http://doi.org/10.18128/D050.V12.0>
- [49] Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan and Co., Ltd.
- [50] Matray, Adrien. 2014. "The Local Innovation Spillovers of Listed Firms." Mimeo.
- [51] Moretti, Enrico. 2004. "Workers' Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*, 94(3): 656-90.
- [52] Moretti, Enrico. 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt.
- [53] Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. "German Jewish Émigrés and US Invention." *American Economic Review*, 104(10): 3222-55.
- [54] Packalen, Mikko, and Jay Bhattacharya. 2015. "Cities and Ideas." NBER Working Paper No. 20921.
- [55] Peters, Michael. 2013. "Heterogeneous Mark-Ups, Growth and Endogenous Misallocation." Mimeo.

- [56] Porter, Micheal E. 1998. *The Competitive Advantage of Nations*. New York: Free Press.
- [57] Rosenthal, Stuart S., and William C. Strange. 2001. "The determinants of agglomeration." *Journal of urban economics*, 50(2): 191-229.
- [58] Rossi-Hansberg, Esteban and Mark Wright. 2007. "Urban Structure and Growth" *Review of Economic Studies*, 74(2): 597-624.
- [59] Saxenian, Annalee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and route 128*. Cambridge, MA: Harvard University Press.
- [60] Stiglitz, Joseph. 1977. "The Theory of Local Public Goods." in *The Economics Of Public Services*. New York, NY: Macmillan
- [61] Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Benjamin Jones. 2013. "Atypical Combinations and Scientific Impact." *Science*, 342(6157): 468-472.
- [62] Weitzman, Martin. 1998. "Recombinant Growth." *The Quarterly Journal of Economics*, 113(2): 331-60.

Appendix

A C-Score: Details and Example

The c-score of the class pair $(\mathcal{A}, \mathcal{B})$ is calculated according to the following algorithm:²⁷

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 inverse of the number of possible pair combinations in a certain patent. Mathematically,

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

where N is the total number of patents in the dataset, C_n is the total number of citations in patent n , c_m and c_l are the m -th and l -th citation of patent n , respectively. It is easy to see that $f_{obs}(\mathcal{A}, \mathcal{B})$ 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, but the citations were distributed randomly subject to such constraint. Formally,

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

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 given by patent g , 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 \mathcal{A} 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 \mathcal{B} 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

²⁷To compute the conventionality score, we also use foreign patents. The algorithm is adapted from Uzzi et al. (2013).

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 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:

$$c(\mathcal{A}, \mathcal{B}) = \frac{f_{obs}(\mathcal{A}, \mathcal{B})}{f_{rand}(\mathcal{A}, \mathcal{B})} \times 100.$$

when the c-score is smaller than 100, the pair $(\mathcal{A}, \mathcal{B})$ is observed in the data less often than what one would expect by taking the same paper in a pseudo-random fashion. We consider this a sign of novelty. On the contrary, when the c-score is bigger than 100, the pair is observed more frequently than the pseudo-random distribution. We consider this a sign of commonality.

4. Each of the $\binom{C_n}{2}$ 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 a fictitious example of how a distribution of c-scores is assigned to a patent. 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:

$$\underbrace{(CPU, CPU)}_{c=140} \times 3 \quad \underbrace{(MON, MON)}_{c=125} \times 1 \quad \underbrace{(CPU, MON)}_{c=110} \times 6 \quad \underbrace{(CPU, SH)}_{c=90} \times 3 \quad \underbrace{(SH, MON)}_{c=75} \times 2$$

This generates a distribution of c-score for this specific patent from which we can extract the tail conventionality as its 10th percentile (in this case, $c = 75$).

B 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 B.1: This table reports the number of patents issued from January 2002 to August 2014 and re-arranged by filing year. All patents (including foreign grants) are counted.

Variable	Level of obs.	Mean	Min	Max	Winsor	Weight	# Obs.
Tail Conventionality	Patent	85	40	164	1%	No	634,470
Median Tail Conventionality	CSD	77	40	164	No	Pat	33,677
Patents per 1,000	CSD	0.196	0	208	No	Pop	390,854
Patents per capita (winsor)	CSD	0.126	0	0.866	1%	Pop	390,854
Density of population ($/km^2$)	CSD	1,217	0.001	26,821	No	Pop	390,854

Table B.2: Summary statistics for the main variables used in the analysis.

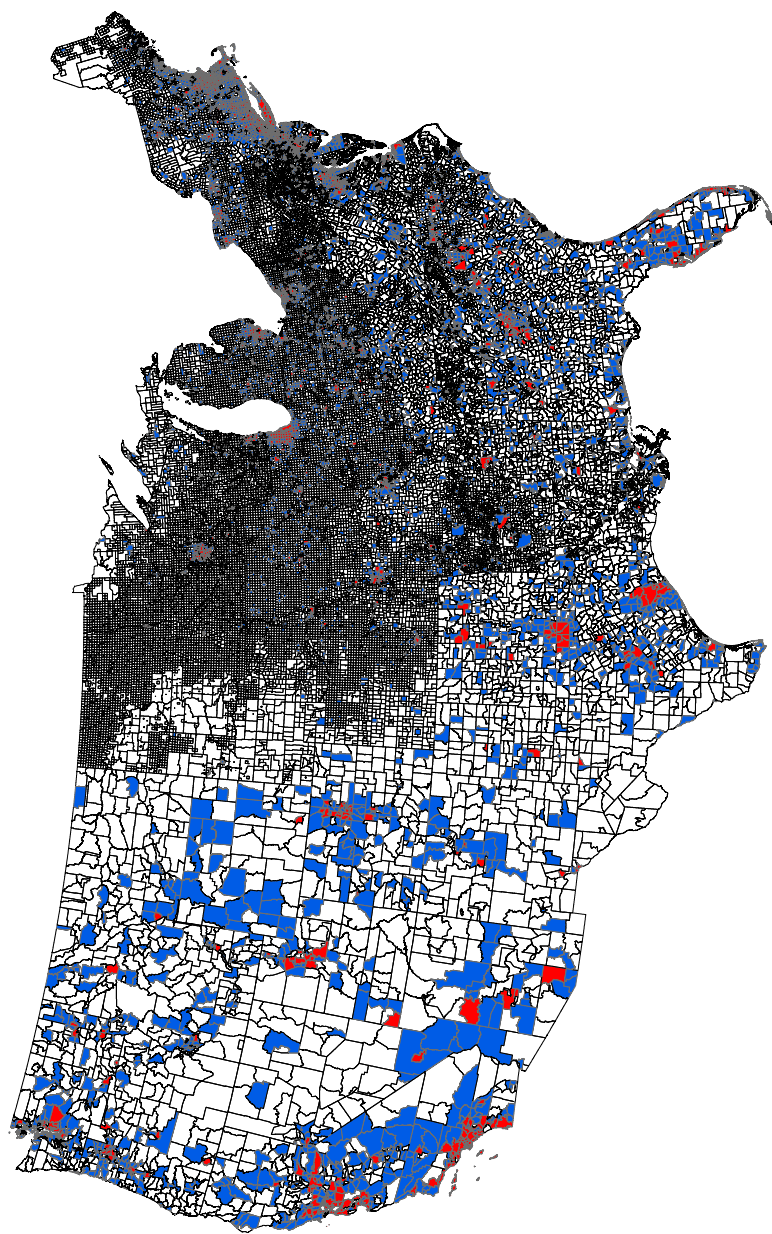
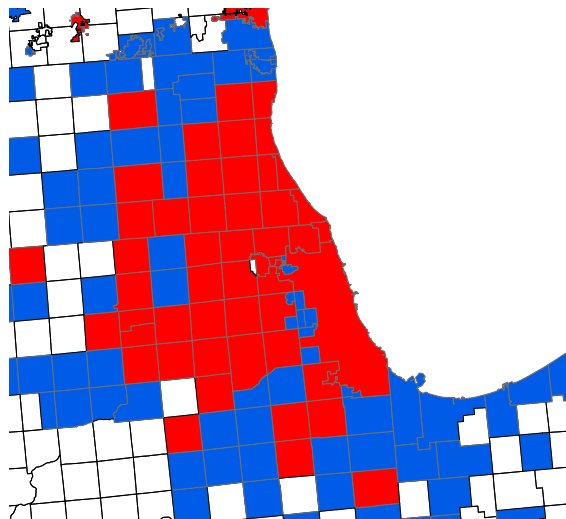
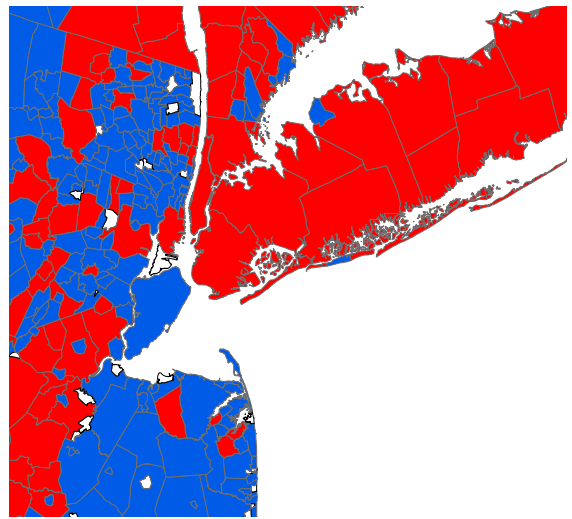


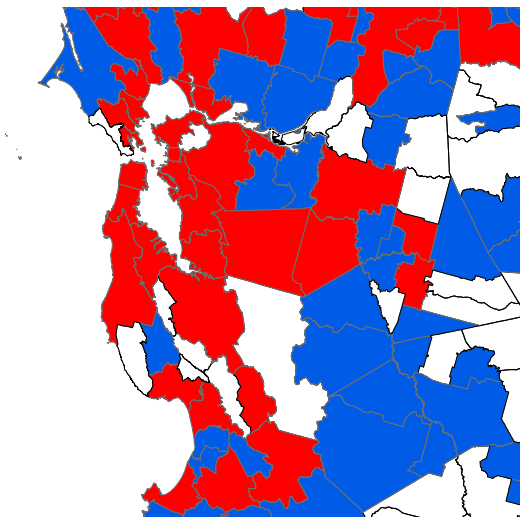
Figure B.1: County sub-divisions in the United States. A given CSD is colored in red if it produced at least one patent per year between 2000 and 2010 (“continuously innovative”), in blue if it produced innovation only occasionally. No patents have been filed in the CSDs colored in white. Note: The map is in vectorial format, so it is possible to zoom in without any loss of resolution.



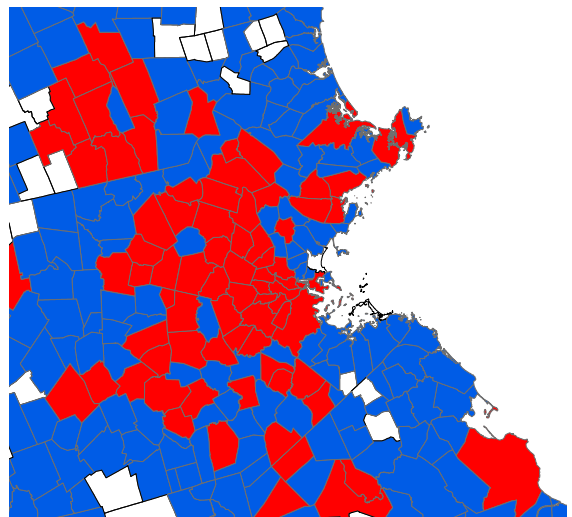
(a) Chicago



(b) New York

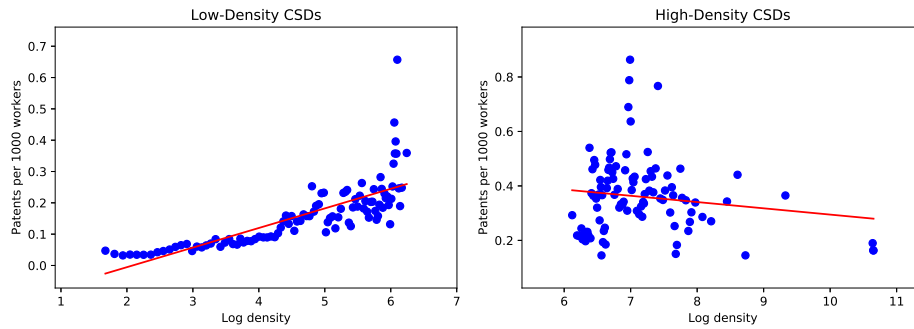


(c) San Francisco

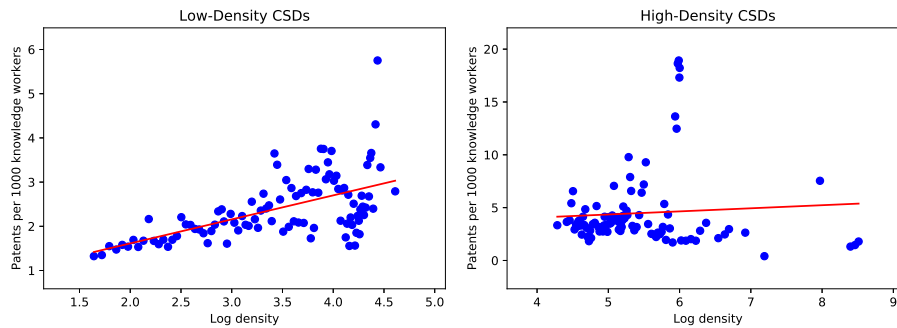


(d) Boston

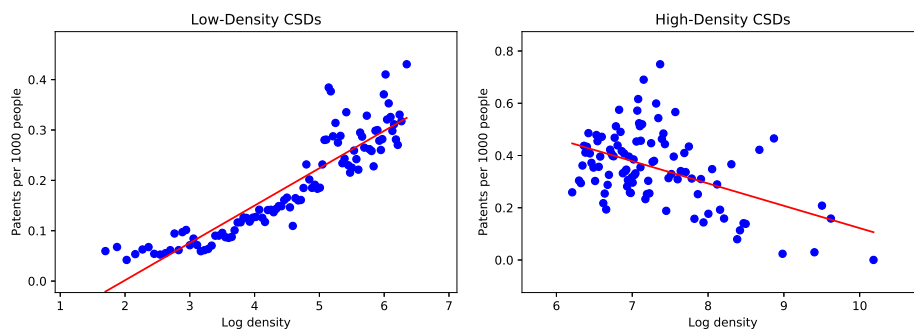
Figure B.2: The figures show maps of county sub-divisions in four of the main metropolitan areas. A given CSD is colored in red if it produced at least one patent per year between 2000 and 2010 (“continuously innovative”), in blue if it produced innovation only occasionally. No patents have been filed in the CSDs colore in white.



(a) Patents per 1000 workers and log-density of employment, least dense CSDs (left) and densest CSDs (right) hosting 50% of the U.S. employment. Only patents for which the state of the assignee coincides with at the state of at least one of the inventors are included.

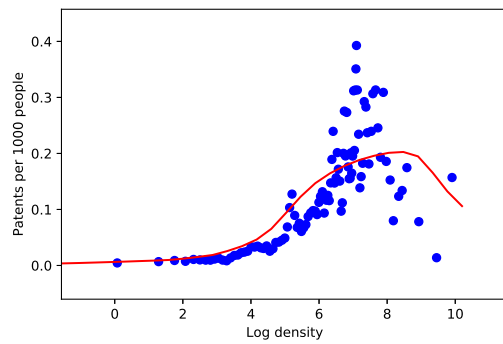


(b) Patents per 1000 high-tech workers and log-density of high-tech workers, least dense CSDs (left) and densest CSDs (right) hosting 50% of the U.S. knowledge employment. Only patents for which the state of the assignee coincides with at the state of at least one of the inventors are included.

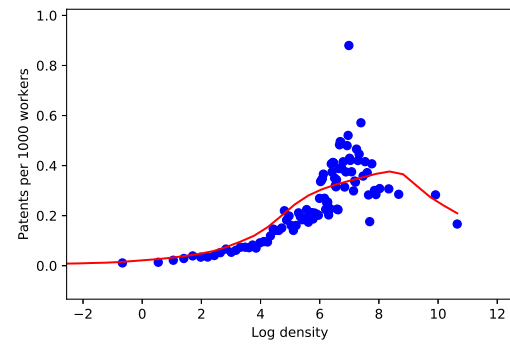


(c) Patents per 1000 residents and log-density of population, least dense CSDs (left) and densest CSDs (right) hosting 50% of the U.S. population. All patents are geo-located at the CSD of residence of the first inventor.

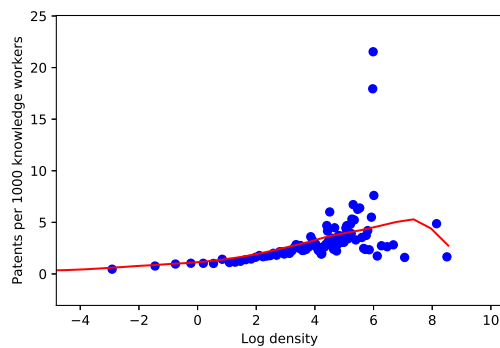
Figure B.3: All the bin-scatter plots are weighted by total population or (knowledge) employment, and control for year fixed effects. The measure of innovation is winsorized at 1% level.



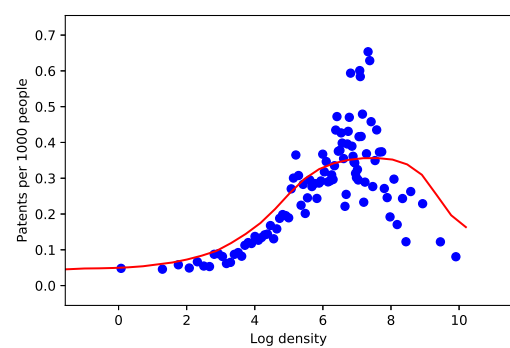
(a) Benchmark



(b) Density of employment



(c) Density of high-tech employment



(d) Inventor's residence

Figure B.4: Patents per 1000 people or (high-tech) workers and log-density of population or (high-tech) employment. All the bin-scatter plots are weighted by total population or (knowledge) employment, and control for year fixed effects. The measure of innovation is winsorized at the 1% level. The red line represents a local polynomial regression with kernel bandwidth of 1.

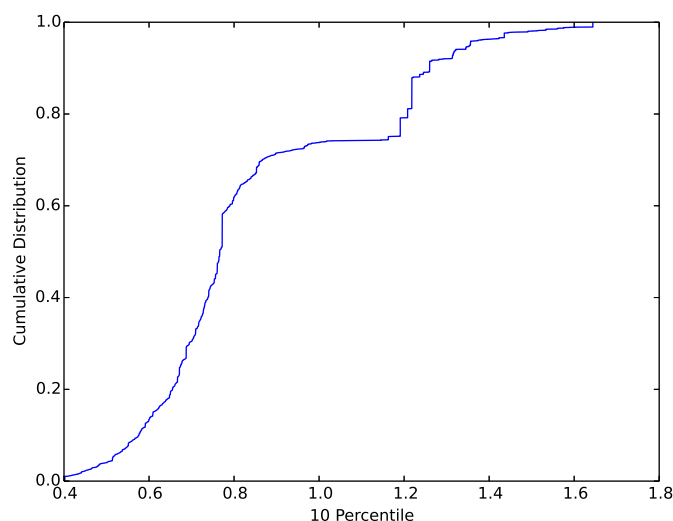


Figure B.5: Cumulative distribution functions of tail conventionality in the universe of patents.

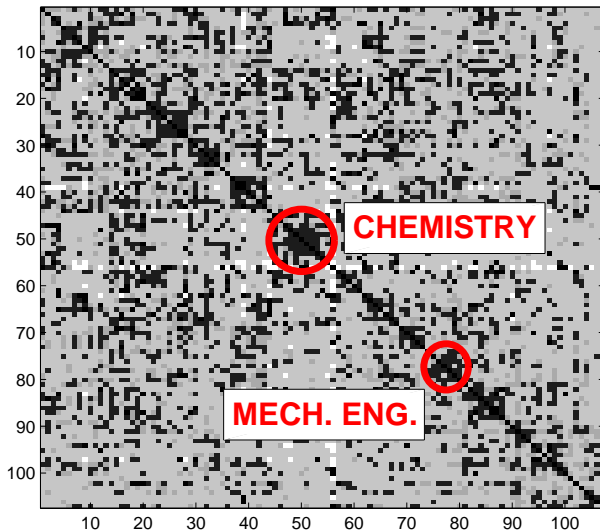


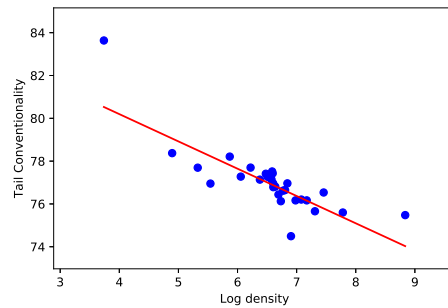
Figure B.6: Every pixel in the matrix indicates a patent class pair. The darker the pixel the higher the c-score assigned to that class pair, the lighter the lower the c-score. Diagonal elements of the matrix show a clear red tendency compared to the rest of the matrix. The “class-clusters” of Chemistry and Mechanical Engineering, among the others, are clearly visible around the diagonal.

	Median Tail-conventionality			
	(1)	(2)	(3)	(4)
Log population density	-1.24*** (0.33)	-1.28*** (0.32)	-1.03*** (0.34)	-0.77** (0.36)
Share College Graduates		-1.93 (1.46)	-4.45** (1.88)	-3.86* (2.04)
Log median income			1.80** (0.83)	1.34 (0.97)
Gini				-0.11 (0.07)
CZ/year f.e.	yes	yes	yes	yes
Weighted	Pat	Pat	Pat	Pat
N. Obs	33,677	33,674	33,674	32,771
R ²	0.11	0.11	0.11	0.11

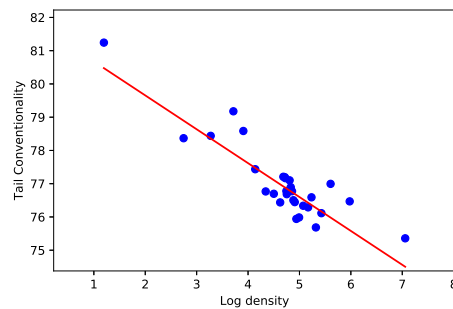
Table B.3: The dependent variable is defined as the tail-conventionality of the median patent in the CSD-year observation. All regressions are weighted by the total number of patents filed in the CSD/Year observation. Standard errors in all the regressions are clustered at the CSD level. U-scores are winsorized (1%) at the patent level. ***p<0.01; **p<0.05; *p<0.1.

	Dep. Variable: Conventional Tail		
	(1)	(2)	(3)
Log population density	0.0090*** (0.0034)	0.0083** (0.0033)	0.0101*** (0.0037)
Publicly Traded		-0.01712** (0.0080)	-0.0149* (0.0083)
Log total patents			-0.0031 (0.0022)
CZ/year/class f.e.	yes	yes	yes
N. Obs	634,261	419,240	419,240

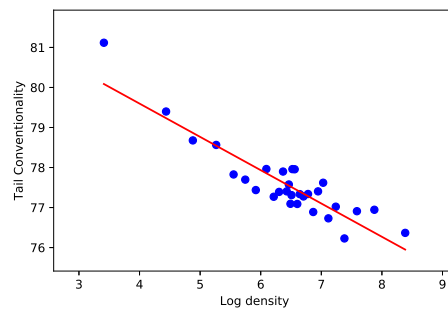
Table B.4: Marginal effects of a patent-level logit regression. Dependent variable is a dummy that takes value 1 if the Tail Conventionality of the patent is below the median of its year-class bin. Standard errors in all the regressions are clustered at the CSD level. ***p<0.01; **p<0.05; *p<0.1.



(a) Median tail-conventionality and log-density of employment in continuously innovative CSDs. Only patents for which the state of the assignee coincides with at the state of at least one of the inventors are included.



(b) Median tail-conventionality and log-density of high-tech workers in continuously innovative CSDs. Only patents for which the state of the assignee coincides with at the state of at least one of the inventors are included.



(c) Median tail-conventionality and log-density of population in continuously innovative CSDs. All patents are geo-located at the CSD of residence of the first inventor.

Figure B.7: All the bin-scatter plots are weighted by total patents, and control for year and CZ fixed effects.

	Median Tail Conventionality	
Log population density	-1.23*** (0.32)	-1.54*** (0.36)
Chicago		2.02 (1.28)
Boston		0.92 (1.68)
Manhattan		5.56*** (1.67)
San Francisco		3.13*** (0.97)
CZ/year f.e.	yes	yes
N. Obs	33,677	33,677
R^2	0.11	0.11

Table B.5: 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. U-scores are winsorized (1%) at the patent level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

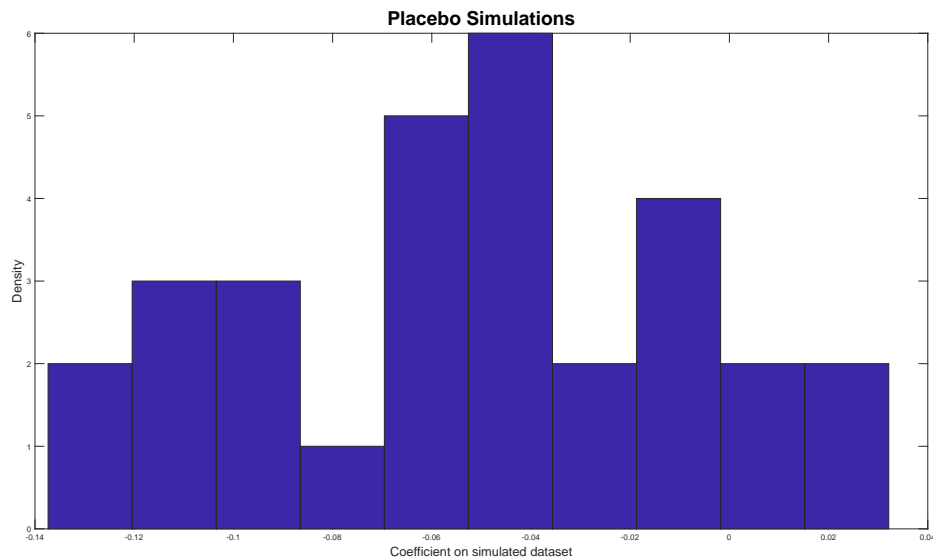


Figure B.8: Placebo experiment: Estimated coefficients from 30 regressions of log-density on concentration index on simulated patent networks.

C Proofs and Derivations

C.1 Proof of Proposition 3.2

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. Letting θ_i^C denote the Lagrange multiplier on the developer's participation constraint in a town specialized in i , the first order conditions of her problem can be expressed as:

$$\theta_i^C = N_i$$

$$\tau_i^C = \phi.$$

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

$$N_i^C = \left[\frac{\phi\alpha}{1-\alpha} \right]^{\frac{\alpha}{1-\alpha(\phi+1)}} \left[\frac{1}{\mathcal{W}} \right]^{\frac{\alpha}{1-\alpha(\phi+1)}}. \quad (\text{C.1})$$

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

$$\theta_{\mathcal{A}}^G = N_{\mathcal{A}}^G \quad \theta_{\mathcal{B}}^G = N_{\mathcal{B}}^G$$

$$\tau_{\mathcal{A}}^G = \phi + \left(\frac{N_{\mathcal{B}}^G}{N_{\mathcal{A}}^G} \right)^{\phi} \quad \tau_{\mathcal{B}}^G = \phi + \left(\frac{N_{\mathcal{A}}^G}{N_{\mathcal{B}}^G} \right)^{\phi}$$

while symmetry implies that $\mathcal{U}_{\mathcal{A}}^G = \mathcal{U}_{\mathcal{B}}^G$, which gives:

$$\left(\frac{N_{\mathcal{A}}^G}{N_{\mathcal{B}}^G} \right)^{1-\phi} = \frac{1+\tau_{\mathcal{A}}}{1+\tau_{\mathcal{B}}}.$$

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

$$\tau_{\mathcal{A}}^G = \phi + 1 \quad \tau_{\mathcal{B}}^G = \phi + 1.$$

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

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

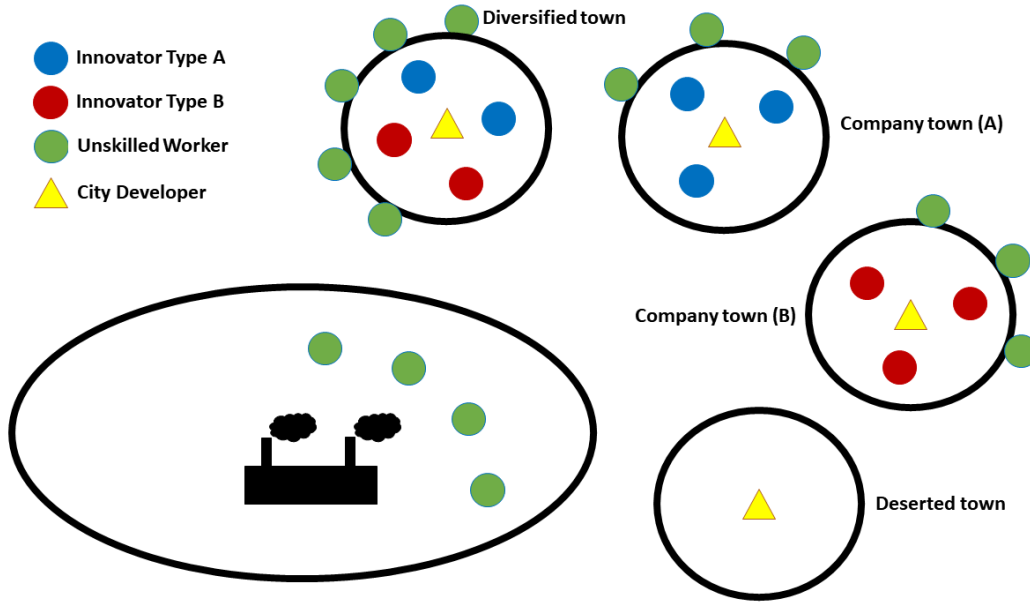


Figure C.1: Spatial economy: Illustration. Innovators from background \mathcal{A} and \mathcal{B} sort themselves into the downtown areas of cities. Unskilled labor lives in the outskirts of cities and in the rural areas. Production takes place in rural areas between cities.

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:

$$\mathcal{W} = \left[\frac{2^{-(\phi+1)} (2 + \phi) (C^G)^{\phi+1} - \frac{1}{\alpha} (C^G)^{\frac{1-\alpha}{\alpha}}}{(1 + \phi) (C^C)^\phi - \frac{1}{\alpha} (C^C)^{\frac{1-\alpha}{\alpha}}} \right]^{\frac{[1-\alpha(\phi+1)][1-\alpha(\phi+2)]}{\alpha(1-\alpha)}} \mathcal{V}^{\frac{1-\alpha(\phi+1)}{\alpha}} \quad (\text{C.3})$$

where

$$C^G \equiv \left[\frac{2^{-(\phi+1)} (1 + \phi) \alpha}{1 - \alpha} \right]^{\frac{\alpha}{1-\alpha(\phi+2)}} \quad C^C = \left[\frac{\phi \alpha}{1 - \alpha} \right]^{\frac{\alpha}{1-\alpha(\phi+1)}}.$$

Plugging (C.3) into (C.1) and (C.2), and using the fact that $\mathcal{V} = \frac{b}{1-b} \kappa$, yields (3.14), where F^G and F^C are constants that only depend on the primitives of the model. In particular, define:

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

Then, the expressions for F^C and F^G can be written as:

$$F^G = C^G (C^W)^{-\frac{\alpha}{1-\alpha(\phi+2)}} \quad F^C = C^C (C^W)^{-\frac{\alpha}{1-\alpha(\phi+1)}}.$$

Finally, we need to show that, in equilibrium, generic cities are more densely populated than company towns. This is true if and only if:

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

Writing down the expression explicitly, reveals that this is always the case as long as $\phi > 0$.

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 $N^k = N_A^k$ or $N^k = N_B^k$. \square

C.2 Proof of Proposition 3.3

In equilibrium, the rate of conventional (ψ) and unconventional (ζ) innovation can be written, respectively, as:

$$\psi = M^C \left(N^C \right)^{\phi+1} \quad \zeta = M^G \left(\frac{N^G}{2} \right)^{\phi+2}.$$

Using the equilibrium expressions for N^C and N^G , the ratio $\kappa = \frac{\psi}{\zeta}$ can be written as:

$$\kappa = \frac{\psi}{\zeta} = \frac{M^C (F^C)^{(\phi+1)} \left(\frac{1-b}{b} \right)^{(\phi+1)} \kappa^{-(\phi+1)}}{M^G 2^{-(\phi+2)} (F^G)^{(\phi+2)} \left(\frac{1-b}{b} \right)^{(\phi+2)} \kappa^{-(\phi+2)}}.$$

Solving this expression to eliminate κ from both sides, we can derive the equilibrium relative mass of generic and company towns:

$$\frac{M^G}{M^C} = \frac{b}{1-b} C^K, \tag{C.4}$$

where $C^K = \frac{(F^C)^{\phi+1}}{2^{-(\phi+2)} (F^G)^{\phi+2}}$.

The labor market clearing condition for skilled labor is:

$$M^G N^G + M^C N^C = 1.$$

Using (3.14) and (C.4) to substitute for M^G , N^G and N^C , we obtain:

$$M^C = \left[C^K (F^G) + \frac{1-b}{b} \right] \kappa.$$

The total amount of unskilled labor used in the production of housing in generic towns is equal to $M^G (N^G)^{\frac{1}{\alpha}}$, and the total amount of unskilled labor used in the production of housing in company towns is equal to $M^C (N^C)^{\frac{1}{\alpha}}$. The total amount of unskilled labor used in the production of intermediate variaties is $L_F = x \left(\frac{\beta}{w}\right)^{\frac{1}{1-\beta}}$, with $x = \zeta^\mu \psi^{1-\mu}$. The labor market clearing condition for unskilled labor is:

$$M^G (N^G)^{\frac{1}{\alpha}} + M^C (N^C)^{\frac{1}{\alpha}} + x \left(\frac{\beta}{w}\right)^{\frac{1}{1-\beta}} = L. \quad (\text{C.5})$$

We have showed that all the terms in (C.5), with the exception of w , can be written as function of the relative supply of conventional to unconventional ideas, κ . To obtain an expression for w , combine (3.12) with (C.3):

$$w = \left[C^W \left(\frac{b}{1-b}\right)^{\frac{1-\alpha(\phi+1)}{\alpha}} (1-b) a \gamma \kappa^{\frac{1-\alpha(\phi+1)-\alpha\mu}{\alpha}} \right]^{(1-\beta)},$$

which again illustrates that w can be written as a function of κ only. We can then write the left-hand-side of (C.5) as a function of κ only, and, in particular, it is easy to show that $\kappa^{-\frac{1-\alpha}{\alpha}}$ can be factored out from the expression, yielding:

$$\kappa^{-\frac{1-\alpha}{\alpha}} F = L,$$

where F is the sum of the constant terms in the addends of the left-hand-side of (C.5). This leads to the unique solution for κ :

$$\kappa = \left(\frac{F}{L}\right)^{\frac{\alpha}{1-\alpha}}.$$

Once the value of κ is obtained, recovering the equilibrium value of the remaining variables is trivial. \square

C.3 Proof of Proposition 3.4

Once we factor out the term $\kappa^{-\frac{1-\alpha}{\alpha}}$, the labor market clearing condition for unskilled labor can be rewritten as:

$$L \kappa^{\frac{1-\alpha}{\alpha}} = B_1 \left(\frac{1-b}{b}\right)^{\frac{1-\alpha}{\alpha}} + B_2 \left(\frac{1-b}{b}\right)^{\frac{1}{\alpha}} + B_3 \left(\frac{1-b}{b}\right)^{(\phi+1)},$$

where B_1 , B_2 and B_3 only depend on other parameters. From this expression, it is immediate that the relative supply of conventional to unconventional innovation, κ , is a decreasing function of the bargaining weight of the unconventional innovator, b . \square