NORTHWESTERN UNIVERSITY

The Geography, Determinants, and Effects of Innovation

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Enrico Berkes

EVANSTON, ILLINOIS

June 2018

Abstract

The Geography, Determinants, and Effects of Innovation

Enrico Berkes

Endogenous growth theory has long recognized innovation as one of the key drivers of growth. Understanding what factors encourage or discourage innovative activities and how, in turn, these affect our communities is therefore crucial to inspire policies that promote inclusive growth. This dissertation tries to broaden our comprehension of the innovative process and its consequences. The first chapter shows that knowledge intensive activities cause an increase in income segregation within U.S. cities and proposes a framework that can be used to study how to mitigate this effect. In the second chapter, we explore how population density is related to the kind of innovation produced in a certain area. More densely populated places tend to promote the creation of unconventional ideas. Finally, the third chapter describes a newly developed data set of geographically referenced historical patents that will allow researchers to get a long run perspective and better understanding of the innovation process as a whole.

Acknowledgements

Completion of this doctoral dissertation was only possible thanks to the support of many people. I would like to express my sincere gratitude to all of them. First, I must acknowledge my Master's thesis advisor, Ugo Panizza, and my colleagues at the IMF, Andrew Berg, Rafael Portillo, and Felipe Zanna. They showed me how fascinating the research process is and without them I would have never pursued an academic career.

At Northwestern, I found a vibrant community of researchers that have inspired and motivated me during these years. I am particularly grateful to the members of my committee, Treb Allen, Matthias Doepke, Ben Jones, and Lorenz Kueng, for always having an open door and guiding me through this adventure.

I also have been surrounded by people armed with immeasurable patience and wisdom. Their words and encouragement have been an irreplaceable source of renewable energy when I needed it the most. A special thank goes to my parents, my siblings, Catherine, and the GS.

Finally, I am extremely indebted to Ruben Gaetani, without whom, this dissertation would not exist. He has been a great co-author always trying to push our limits.

This research was funded in part by the Ewing Marion Kauffman Foundation. The contents of this publication are solely my responsibility.

Preface

The first chapter of this dissertation, which is joint work with Ruben Gaetani, analyzes the effect of the rise of knowledge-based activities on spatial inequality within U.S. cities, exploiting the network of patent citations to instrument for local trends in innovation. We find that innovation intensity is responsible for 20% of the overall increase in urban segregation between 1990 and 2010. This effect is mainly driven by the clustering of employment and residence of workers in knowledge-based occupations. We develop and estimate a spatial equilibrium model to quantify the contribution of productivity and residential externalities in explaining the observed patterns. Endogenous amenities account for two thirds of the overall effect. We illustrate the relevance of the model for policy analysis by studying the impact of four proposed projects for Amazon's HQ2 on the structure of Chicago.

In the second chapter, we use a newly assembled dataset of U.S. patents to show that innovation activities are far from being limited to densely populated urban areas, 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 technologically 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.

In the last chapter, I describe a newly assembled data set of historical patents. Patents are commonly used as the main source of data for empirical studies related to innovation and technological change. The large amount of information about the underlying innovative process contained in each patent has certainly contributed to their popularity. Nevertheless, due to the lack of reliable data, historical analysis has focused on relatively small time frames or on specific dimensions of patents data. The goal of this paper is to fill this gap. I build and release a comprehensive time series of the universe of U.S. patents. The data set contains all the variables commonly used in the literature and, importantly, geolocates every inventor and assignee reported in each grant over the period 1836-2016.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Income Segregation and Rise of the Knowledge Economy

with Ruben Gaetani

1.1. Introduction

The knowledge economy is a set of economic activities relying on non-manual and non-routine technical skills, scientific knowledge, and intellectual creativity. Over the past 40 years, these activities have become the main engine of economic prosperity in advanced countries. Since 1975, the share of value added generated by knowledge-intensive sectors in the United States has increased by almost 15 percentage points, and the number of patents per capita issued by the United States Patent and Trademark Office (USPTO) has doubled (Figure A.1). The same trend is observed when considering several other measures of knowledge intensity, including educational attainment, number of scientific publications, ratio of intangibles to assets, and share of workers employed in R&D activities and creative sectors. The suggested explanations for this structural shift include globalization, automation of routine jobs, and the steady increase in the burden of knowledge that requires an ever-increasing number of R&D workers to sustain a constant productivity growth (Jones, 2009).

This trend is believed to be associated with major social and cultural changes. Individuals with different education levels, abilities, and social connections have been differentially exposed to the opportunities offered by this new economic landscape and, as a result, have experienced diverging economic fortunes. Moretti (2012) argues that the geographical dimension is the most striking aspect of this divergence. The rise of the creative class (Florida, 2002) has allowed and induced waves of gentrification and re-urbanization of metropolitan cores, as well as the development of specialized innovation clusters in suburban areas. The reorganization of production and consumption activities within cities, driven by supply factors (e.g., thick labor markets and knowledge spillovers) as well as demand factors (e.g., preferences for local amenities), appears to be correlated with the emergence of intellectually creative jobs in many fast-growing local economies (Florida and Mellander, 2015).

One of the most evident signs of this reorganization of the urban structure is the sharp increase in income segregation in U.S. cities. Our preferred measure of income segregation, the cross Census tracts (CTs) within commuting zone (CZ) Gini index, increased by 3 Gini points over the period 1990-2010, which corresponds to 70% of the increase in overall inequality over the same period of time (Table B.1). However, the extent to which the rise in income segregation in U.S. metropolitan areas reflects a causal effect of the expansion in knowledge-intensive activities remains an open question. Theoretically, there are several reasons to believe that such effect exists. First, innovation and other creative jobs crucially depend on knowledge transmission, which has been shown to be strongly localized (e.g., Jaffe et al., 1993). An increase in the returns to new ideas makes clustering in space with individuals who offer high learning opportunities more convenient for creative people. Second, workers in the knowledge economy tend to be disproportionately sensitive to urban and social dimension, such as quality of schooling and social relationships, which are often strictly local in nature.

Uncovering the fundamental causes of the increase in urban segregation is of great importance, as segregation has been shown to have a first-order impact on several policy relevant outcomes, including schooling (Baum-Snow and Lutz, 2011), health (Acevedo-Garcia et al., 2003; Alexander and Currie, 2017), and inter-generational mobility (Chetty and Hendren, 2016). However, inferring the direct impact of an expansion in creative jobs is problematic because of potential reverse causation and the presence of unobservable factors affecting, at the same time, the explanatory and dependent variables. Examples of these factors include financial or housing shocks, that affect, at the same time, the urban environment and the ability of a geographical area to develop innovation-based activities.

In this study, we address this challenge by adopting an instrumental variable approach, that exploits exogenous variation in knowledge intensity across U.S. cities. Our analysis suggests that innovation intensity is responsible for 20% of the aggregate trend in income segregation. The analysis further reveals that the effect we measure can be explained only in part by diverging income paths of initially segregated neighborhoods. A major part of the effect is, in fact, explained by an increase in the geographical sorting of households along the income dimension.

To measure (and instrument for) the knowledge intensity of the local economy, we use a newly assembled dataset of geo-referenced USPTO patents in the years

1975–2014. By comparing citation patterns in the early period (1975–1994) with the ones in the late period (1995–2014), we document the existence of a stable network of knowledge diffusion across geographical areas and technological classes. This persistence suggests that knowledge links established in the past are broadly orthogonal to changes in the economic environment. Using the network in combination with actual patenting in the period 1995–2004, we build a credible instrument for current innovative activities at the local level. We run an extensive set of validation exercises to address the remaining endogeneity concerns.

Our two-stage least squares (2SLS) results imply that a one standard deviation increase in patenting between 1990 and 2010 leads to an increase in the measured income segregation of 1.19 Gini points, equal to 39% of the overall increase in segregation over the considered period. Educational and occupational segregation, which is the extent to which residents of different educational backgrounds and occupations sort themselves in the city, also surges. The estimated effect is stronger for high-learning sectors (including IT and electronics) and even negative for low-learning ones, such as textiles. The IV analysis reveals that the bias in the OLS estimates is negative. This bias suggests that unobserved shocks affecting, at the same time, segregation and innovation tend to operate on the two variables in opposite directions, overall. Financial shocks that generate widespread housing and neighborhood dismantlement are possible examples.

These results can be explained as the outcome of two (related but) inherently different phenomena. On the one hand, an increase in inequality in a metropolitan area that is perfectly segregated induces a one-to-one increase in measured segregation (we will refer to this case as the *inequality effect*). On the other hand, the measured segregation increases even in the absence of any change in inequality when people move closer to other people with a similar level of income (we will refer to this second case as the *sorting effect*). The analysis strongly supports the sorting effect as the primary cause of the increase in urban segregation resulting from the expansion of innovation activities, with the inequality effect only explaining a limited portion of it.

In the second part of the paper, we explore two possible mechanisms. We argue that innovation shocks increase the returns from local learning externalities and generate incentives for firms to cluster in space to benefit from them. As a result, high-education, high-salary workers move close to these areas to reduce commuting costs, thereby affecting residential segregation. We provide evidence that employment in knowledge-intensive occupations becomes more geographically concentrated in cities experiencing larger innovation shocks. We also propose that the endogenous response of residential amenities plays an important role in amplifying this effect. Consistent with this interpretation, we find that the impact is significantly stronger in cities whose variation in residential amenities is not anchored to persistent or natural amenities. The magnitudes of the estimated effects suggest that localized knowledge spillovers and residential amenities play an important role in linking innovative activities to income segregation.

To quantitatively disentangle the relative importance of these two forces in determining the trends in segregation that we observe in the data, we build a general equilibrium model of the city structure in the spirit of Ahlfeldt et al. (2015) – ARSW hereafter – that embeds endogenous amenities and productivities. We extend the model in ARSW by introducing heterogeneity in workers' occupations: workers in creative occupations enjoy local learning externalities that are directly affected by a city-wide knowledge shock, whereas workers in non-creative occupations have stagnant productivity that is unaffected by the surrounding economic activity. Both types of workers perceive local residential externalities that are determined by the density and background of their neighbors.

To estimate the strength of local externalities, we rely on the exogenous cross-city variation in knowledge intensity inferred in the empirical analysis. To this end, we impose that residual factors affecting the spatial distribution of economic activity do not vary systematically with the predicted patenting growth. In particular, our identifying assumption is that the within-city average of the change in the exogenous components of productivity and residential amenities is independent of the value of the knowledge shock. The structural estimation reveals the existence of steep, localized residential externalities for agents in creative sectors. This finding confirms that the endogenous response of residential amenities in neighborhoods where knowledge workers concentrate is disproportionately valued by knowledge workers themselves, and it operates as a powerful amplification channel in driving the increase in segregation. This asymmetry accelerates the effect of an initial shock to geographical sorting in the city. The model suggests that about two thirds of the overall impact on urban segregation can be explained through the endogenous response of localized, occupation-specific residential amenities.

We illustrate the relevance of the model for policy analysis by running four counterfactual exercises that analyze the impact of four Chicago-based bids for Amazon's new headquarters. Our simulations suggest that although some high-knowledge workers relocate to the high-amenity neighborhoods by the lake in all scenarios, the location of the campus has a sizable effect on the local development of the neighborhoods around, as well as on the overall increase in income segregation. The impact on segregation would be the smallest when the campus is located in the southern part of the city, as it would attract high-salary workers where low-income neighborhoods currently prevail.

Related Literature. This study contributes to the literature on the causes of income segregation in cities in advanced countries in general, and the United States, in particular. Jargowsky (1996) documents a steady increase in economic segregation in U.S. metropolitan areas since 1970, and confronts this trend with the slow decline in racial segregation. In more recent research, Reardon and Bischoff (2016) document that the trend in residential segregation that started in the 1980s continued, to a lesser extent, until very recently. They also show that residential segregation in cities is correlated with the increase in income inequality. Income inequality at the city level has been intensively analyzed by Baum-Snow and Pavan (2013), and Baum-Snow, Freeman and Pavan (2016), who document a positive relationship between city size and an increase in the dispersion of earnings; they interpret this relation as evidence of a skill-biased change in agglomeration economies. Diamond (2016) studies the geographical sorting of college graduates *across* U.S. cities between 1980 and 2010, whereas the current study focuses on the determinants of income and occupational sorting *within* cities.

Income segregation has been widely studied, particularly in relation to the role that neighborhood effects play in social and economic outcomes, such as education, health, and inter-generational mobility. Education and segregation have a strong two-way link, especially in countries (like the United States) where public spending in schooling is very localized. For example, Baum-Snow and Lutz (2011) analyze the response of white families in schooling enrollment (that took the form of migration to the suburbs and private school enrollment) following the racial desegregation of U.S. metropolitan areas in the 1960s and 1970s. Chetty and Hendren (2016) use tax records in a quasi-experimental setting to measure the strength of neighborhood effects on children and their ability to explain differences in inter-generational mobility across areas.

This study examines the distributional effects of innovation, but focuses specifically on the process of knowledge creation. A similar approach is adopted by Aghion et al. (2015), who use cross-state variation and find that changes in innovation intensity can explain the rise in top income inequality in the United States. Florida and Mellander (2015) conduct a comprehensive study of urban segregation in U.S. metro areas and link this increase to the emergence of the creative class and the expansion of jobs in the high-technology industry. In the present study, we provide causal evidence that supports their interpretation.

On the theory side, we augment the model developed by Ahlfeldt et al. (2015) by allowing for agents of different backgrounds (specifically workers in creative and non-creative occupations). While their strategy uses cross-neighborhood exogenous variation in the concentration of economic activity given by Berlin's division and reunification, our structural estimation relies on exogenous *cross-city* variation in the intensity of knowledge spillovers for the innovative sector.

The rest of the paper is organized as follows. Section 1.2 introduces the data and the measures of inequality, segregation, and knowledge intensity. Section 1.3 describes the empirical strategy and results. Section 1.4 introduces the model setting, discusses the structural estimation, and presents the quantitative results. Section 1.5 concludes.

1.2. Data and Measurement

We combine data on innovation, captured by patenting activity, with social and economic indicators from the Census and the American Community Survey (ACS). For the purposes of our empirical analysis, we interpret Commuting Zones (CZs) as cities and Census Tracts (CTs) as neighborhoods (and use the terms interchangeably throughout the text). CZs are defined with respect to actual commuting flows in the U.S. and, contrary to MSAs, constitute a complete partition of the country.[1]

[1]We use the definition of 2000 Commuting Zones provided by Data.gov.

Given that our objective is to assess how innovation shocks affect residential and employment concentration within a local labor market, CZs are the natural unit of geographical aggregation for our analysis.

We now proceed to describe the data sources and main variables in more details.

1.2.1. Patents data

Our preferred measure of knowledge intensity is patenting within a local labor market. Patent data are collected from the United States Patents and Trademark Office (USPTO). The USPTO has digitized the full text of all the patents issued from 1976 onwards, and made the files available for download. We download and parse all the files up to March 2015 and construct a new dataset that includes, for each grant, information on filing and issuing year, technological class,[2] forward and backward citations as well as residence (city and state) of its inventors. Grants are then assigned to a CZ based on the location of their first inventor. From the publicly available documents, we identify a total of 5,030,264 patents out of which 2,634,606 are located in the United States.

1.2.2. Segregation, Inequality and other economic outcomes

Our preferred measures of inequality and segregation in cities are based on the Gini index which has the advantage of being widely used, and therefore offers a natural

[2]Although each patent is associated to multiple classes, the USPTO assigns a single main class to each grant. This main class is available only in the US classification system, although in our analysis we use the international patent classification. Since each grant is associated with several IPC classes but only one main USPTO class, we build a many-to-one function that maps every USPTO class to a single IPC class based on the associations that recur more often.

reference point for our empirical analysis. Mathematically, the Gini index is defined as twice the area between the Lorenz curve and the 45-degree line. More precisely, letting $\{i\}_{i=1}^{N_{cz}}$ be the set of basic units (e.g., individuals or households) in a CZ ordered from the poorest to the richest, the Gini index of city cz is defined as:

$$Ineq_{cz} = 100 \times \left[1 - 2 \times \sum_{i=1}^{N_{cz}} \sum_{i'=1}^{i} \frac{x_{i'}}{x_{cz}}\right] \tag{1.1}$$

where x_i is the income of the basic unit i, whereas x_{cz} total city income. Equivalently, we can construct a measure of income segregation in city cz, defined as inequality of income across neighborhoods, where each unit in neighborhood ct is assigned the average income of the neighborhood itself. In particular, letting $\{ct\}_{ct=1}^{M_{cz}}$ be the set of neighborhoods in a CZ, ordered from the poorest to the richest, we define segregation in city cz as:

$$Segr_{cz} = 100 \times \left[1 - 2 \times \sum_{ct=1}^{M_{cz}} \left(\frac{N_{ct}}{N_{cz}} \sum_{ct'=1}^{ct} \frac{x_{ct'}}{x_{cz}}\right)\right] \tag{1.2}$$

where x_{ct} is total neighborhood income and $\frac{N_{ct}}{N_{cz}}$ is the population share of neighborhood ct in city cz. In other words, $Segr_{cz}$ measures the variation of income within a CZ, once the variation within neighborhoods has been removed.[3] In the extreme case in which average income of each neighborhood is the same, our measure takes

[3]In the implementation of (1.2) we use a piecewise linear, instead of a step function, to approximate the Lorenz curve. This guarantees that $Segr_{cz}$ is always between zero and one. The empirical results are robust to using the Theil index, that has the advantage of being decomposable into between and within components of income dispersion, but it has the disadvantage that its upper bound is determined by the size of total population. This makes it difficult to use this index to analyze the evolution of inequality over time.

value zero. On the other extreme, when households are perfectly sorted across neighborhoods, $Segr_{cz}$ is equal to $Ineq_{cz}$.

Information on income is provided at the CT level by the National Historical Geographic Information System (NHGIS).[4] The NHGIS assembles data from the Census and the American Community Survey (ACS) and aggregates them at various geographical levels. Data at the CT level divide households into 15 income bins.[5] To measure inequality and segregation, we need the income distribution (or its approximation). The problem arises from the fact that the top bin is unbounded, with an average that potentially varies substantially across CTs. The literature has approached this issue in different ways, each with its own advantages and limitations. Appendix C.1.1.1 discusses them and provides a detailed description of the procedure we use to approximate the income distribution.[6]

From the NHGIS, we also extract data at the CT level on population, education and rents. These are used either as controls or in ancillary analyses throughout the text. The structural estimation of the model requires data on the distribution of residence and employment by occupation in each CT, average earnings by occupation at the CZ level, and measures of bilateral commuting times and commuting flows across

[4] https://www.nhgis.org/.

[5] The lower bounds of each income bracket are 0$, 10,000$, 15,000$, 20,000$, 25,000$, 30,000$, 35,000$, 40,000$, 45,000$, 50,000$, 60,000$, 75,000$, 100,000$, 125,000$, and 150,000$.

[6] To validate our procedure further, we compute segregation in (1.2) using income per capita in each CT provided by the NHGIS, that does not require to make assumptions on the distribution of the top bin. The correlation between the two variables is 90% in 1990 and 91% in 2010 (see Figure A.9).

CTs. The distribution of residence by occupation is obtained by matching information from the NHGIS and the Integrated Public Use Microdata Series (IPUMS).[7] The distribution of employment by occupation is gathered from the National Establishment Time Series (NETS). The NETS provides data on employment, geographical location and industry for the universe of establishments over the period 1990-2015.[8] Contrary to the County Business Pattern, this dataset has the advantage of also including jobs in the public sector. Industry is then mapped into occupations by using the crosswalks provided by the BLS. Average earnings by occupation in each CZ are compiled from the IPUMS.

Bilateral commuting times across CTs are taken from the Open Source Routing Machine (OSRM).[9] This routing engine allows us to compute travel time by car for each pair of coordinates. We collect data on commuting times for each pair of neighborhoods within each city for a total of 16.2 million pairs.[10] Finally, bilateral commuting flows are collected at the Census Block level from the Longitudinal Employer-Household Dynamics (LEHD) dataset.[11]. Data at a block level are then

[7]https://www.ipums.org/

[8]In particular, the dataset includes about 10 million observations in 1990 and about 30 million observations in 2010. The vast majority of the establishments can be univocally assigned to a CT. The establishments for which we can only identify the ZIP code are proportionally distributed to the corresponding CTs based on their area. We discard the establishment for which the geographical information is only available at a State level. More details in Appendix C.1.1.2.

[9]http://project-osrm.org/

[10]The OSRM can be run locally and has therefore the advantage of not being subject to query limits. However, real-time data on traffic are not available, as it is the case for more popular services such as Google Maps. The commuting times collected this way are therefore to be interpreted as lower bounds.

[11]https://lehd.ces.census.gov/

aggregated to obtain commuting flows at our preferred level of geographical aggregation (CTs).

Appendix C.1.1.2 provides summary statistics and further details on the construction the main variables.

1.2.3. Data Timeline

In this paper, we study the long-run impact of local innovation activities on income segregation and inequality within U.S. cities. For most of the analysis we look at changes in local labor market outcomes over a 20-year period, specifically, between 1990 and 2010. The structure of the data, schematized in Figure A.2, is especially suitable for this purpose.

Socio-economic outcomes at the CT level are available every ten years, whereas patent data cover a 40-year period that can be conveniently divided into two 20-year samples. The early sample (1976-1995) is used to infer knowledge links across geographical and technological areas in the U.S. and to measure innovation for the 1990 observation. The late sample (1996-2014) is itself divided into two time periods. The first decade (1995-2004) is used in conjunction with the knowledge links previously estimated to calculate the local shocks to innovation used as an instrument. The second decade (2005-2014) is used to measure innovation for the 2010 observation. To avoid our results to be driven by transitory shocks to innovation, we compute the patenting activity for each data point (1990 and 2010) as ten-year averages (1985-1994 and 2005-2014, respectively).

1.3. Empirical Analysis

The main question of this paper is whether CZs that experience an expansion in innovation and knowledge activities also experience an increase in income segregation, defined as variation of income across neighborhoods within the city. We first identify a causal nexus between those phenomena and empirically investigate its features. We then use a quantitative model to infer the relative importance of economic forces behind our findings as well as some prevailing features of production and consumption in a knowledge economy.

The empirical model studies the relationship between income segregation at the city level and the size of local patenting activity:

$$Y_{cz,t} = \alpha_t + \beta_{cz} + \gamma \, \log(1 + Patents_{cz,t}) + \delta \, X_{cz,t} + \epsilon_{cz,t} \tag{1.3}$$

where Y_{cz} is segregation, X_{cz} a set of controls for city cz, and $t \in \{1990, 2010\}$. Our instrument for patenting allows us to generate exogenous variation for the late sample ($t = 2010$), while taking patents in the early sample ($t = 1990$) at their observed level. This requires us to estimate the model in differences:

$$\Delta \, Y_{cz} = \tilde{\alpha} + \gamma \, \Delta \, \log(1 + Patents_{cz}) + \delta \, \Delta \, X_{cz} + \tilde{\epsilon}_{cz} \tag{1.4}$$

and instrument for $\Delta \, \log(1 + Patents_{cz})$ in the 2SLS analysis. Since we include the logarithm of population in the set of controls, the results would be identical if patents per capita are used instead. To avoid having to drop observations with zero patents either in 1990 or 2010, we adopt the convention of taking the logarithm of one plus total patents.[12] For robustness, we also estimate (1.4) including the set of controls at their 1990 level.

1.3.1. Correlations and OLS

Figure A.3 shows the unconditional correlation between the change in income segregation and the growth rate of total patents between 1990 and 2010. The Figure (like most of the regressions throughout the text) is weighted by total number of households in the first period (1990). The R^2 of the weighted regression is 0.10 and the coefficient is statistically and economically significant. A one standard deviation increase in patenting growth is associated with an increase of 31% of one standard deviation in segregation in the cross-section of CZs.

In Table B.2, we include a set of control variables that might naturally confound this correlation. First, since the number of CTs changes substantially between 1990 and 2010, there might be the risk that a dimensionality bias in the construction of our segregation measures leads us to mis-measure the increase in segregation in cities where the number of CTs has grown more. To account for this possibility, in column (2) we control for the growth in the number of CTs within the city.[13] In

[12]Since all the regressions are weighted by total population in 1990 and zeros are concentrated in scarcely populated areas, this strategy yields virtually identical results as alternative strategies used in the literature (e.g. including dummies for zeros, taking growth rates through midpoint method). Also note that, since we consider 10 year averages for patenting activity, only 25 commuting zones have a patenting activity which is equal to 0 either in 1990 or in 2010. The total population of these is about 208,000 people in 1990 (or 0.08% of the U.S. population).

[13]It is possible that controlling for the growth rate in the number of CTs is not enough to account for the potential dimensionality bias in the construction of our segregation measures. To address this concern, we run a set of simulations in which we reassign CTs to CZs under the constraints

columns (3)-(4), we include the growth rate of population and income, respectively. Local industry composition at the beginning of the sample could be a major confounding factor if aggregate shocks at the industry level (notably, trade shocks) had an impact both on a location's expansion in knowledge-intensive activities and on other variables affecting the urban environment. Hence, in column (5) we control for trade shocks using the measure of exposure to import from China developed by Autor et al. (2013).[14] Finally, the role of the public sector in providing at the same time local services for residents and financial support to innovation activities may generate a significant bias. In column (6), we control for the growth rate of local public spending, provided by the Census at the County level.[15] Although some of the controls attenuate the size, the coefficient for patent growth remains positive, statistically significant and economically large.[16]

Table B.13 reports the results for the OLS regressions when the controls are included in levels at their 1990 value, instead of growth rates. Results are virtually unchanged. As shown in Appendix B.16, we uncover a similar pattern when we consider segregation along an educational or occupational dimension. To measure

that (1) each CZ is assigned the same number of CTs as the original dataset, and (2) each CZ has approximately the same population as the original dataset. This random assignment experiment reveals that the pure dimensionality bias is zero for all practical purposes.

[14]This measure is constructed at the CZ level as: $\Delta IPW_{uit} = \sum_j \frac{L_{ijt}}{L_{ujt}} \frac{\Delta M_{ucjt}}{L_{it}}$, where L_{it} is 1990 employment in CZ i and ΔM_{ucjt} is the change in US import from China in industry j, between 1990 and 2007. Since the authors use 1990 CZs (instead of 2000 CZs), we construct a crosswalk between the two partitions based on the intersection with the highest population.

[15]These data are available for download at http://www2.census.gov/pub/outgoing/govs/special60/.

[16]Data for the last two controls is not available for all the commuting zones in our sample, so that the number of observations is lower than 703. Data are mainly missing in low populated areas. We exclude the last two controls in our benchmark specification, and in tables where full controls are included but not reported. Results change to a negligible extent when these two variables are included.

educational segregation, we use a modified version of the Gini index, where individuals are assigned 1 unit of "income" if they have a college degree and 0 otherwise. As for occupational segregation, we use the classification of individuals into creative and non-creative occupations, as outlined in Appendix C.1.1, which constitutes the basis for our structural model in Section 1.4. In this case, residents are assigned 1 unit of "income" if they are employed in a creative occupation, and 0 otherwise. Both measures display a positive and significant correlation with patenting growth.

1.3.2. Instrumenting for patenting activity

The evidence discussed up to this point must be interpreted with caution. To claim the existence and identify the strength of a causal relationship, we need to identify variation in patenting that is orthogonal to unobserved factors that might affect at the same time the expansion of a knowledge-based economy and urban segregation. The range of such possible factors is large and the direction of the bias is ex-ante ambiguous. Examples of unobserved factors include short-run phenomena such as housing shocks and financial shocks, or long-run trends such as technological obsolescence of local industries, that have a direct impact on the urban context, as well as potentially affecting patenting and other innovative activities. One might also be worried about inverse causality, with income segregation being the cause, rather than the consequence, of the emergence of the knowledge economy in U.S. local labor markets.

In this section, we propose an instrument for innovation activities at the local level that can be used to tackle this identification challenge. The strategy we propose is general and can be applied to other contexts in which channels of knowledge diffusion are observable and measurable. We use the observed network of patent citations to infer the existence of persistent diffusion links across technological classes and geographical areas. Observing a patent that cites another invention reveals the existence of an underlying link between the technological classes and the geographical areas of the two grants. The more citations we observe from and to the same class-CZ pair, the stronger the underlying link. In the reminder of this section, we provide details on the mathematics and intuition behind the instrument. Section 1.3.4 discusses conditions and evidence for its validity.

1.3.2.1. Construction of the instrument. The idea behind the instrument is that local patenting is determined, at least partly, by ideas that are generated elsewhere in the economy, and that transmit to local innovative activities through channels of knowledge transmission that are pre-determined, stable over time, and inferable from the network of patent citations. In order to be used to draw conclusions on the causal effect of innovation on segregation and inequality, this instrument must (1) have predictive power on actual patenting in 2005-2014 and (2) identify variation in patenting that is uncorrelated (conditional on controls) with unobservable factors that can affect at the same time innovation and the dependent variable. We extensively discuss the first point in the next sub-section, where we show that the network of diffusion inferred in the early sample is in fact persistent and can be used to predict innovation in the late sample. As for the second point, our identification assumptions can be summarized in two main points: (1) Innovation shocks that occur in other geographical areas do not have a direct impact on local outcomes (relative to the aggregate impact), other than the effect that operates through knowledge diffusion, and (2) there are not unobservable factors that affect at the same time the ability to form knowledge links with specific areas in the past and local segregation and inequality outcomes 20 years later. Section 1.3.4 discusses the conditions for and the evidence in support of the validity of the instrument.

Formally, we proceed in two steps. In the first step, we use the observed citation patterns to isolate knowledge links across space, time and the technology spectrum. For each patent of class μ issued in CZ r at time $t-\Delta$, we first calculate the share of citations that it receives from patents produced in other commuting zones at time t. We then sum up over the time period that goes from 1985 to 1994 and, to account for size effects in the citations distribution, we divide by the total number of patents of class μ issued in CZ r at time $t-\Delta$. Mathematically, we calculate a coefficient of diffusion as:

$$d^{75-94}_{r,s,\mu,\nu,\Delta} = \begin{cases} \dfrac{\displaystyle\sum_{t=1985}^{1994} \sum_{p \in (\mathcal{S},\mathcal{N},\mathcal{T})} s_{p \rightarrow (r,\mu,t-\Delta)}}{\displaystyle\sum_{t=1985}^{1994} \sum_{q} \mathbb{1}_{\{q \in (r,\mu,t-\Delta)\}}} & r \neq s \\ 0 & r = s \end{cases} \qquad \text{for } \Delta \in \{1,\ldots,10\} \tag{1.5}$$

where $s_{p \to (r,\mu,t-\Delta)}$ is the share of citations that a patent $p \in (\mathcal{S}, \mathcal{N}, \mathcal{T})$ (i.e., of class ν produced in CZ s at time t) gives to patents of class μ produced in CZ r at time $t - \Delta$ for all the s's different from r. To reduce endogeneity concerns, we set the coefficient to zero for links that start and end in the same CZ. The coefficient can be interpreted as "how much" of a new patent in (ν, s), the *destination* class-CZ pair, is "induced" by a previous patent in ($\mu\, r$), the *origin* class-CZ pair, Δ years after filing. The idea is that existing patents are perfectly substitutable building blocks for future innovation. Note that since we use $\Delta \in \{1, \ldots, 10\}$, we need to use the entire early sample (1975-1994) to compute the coefficients of diffusion. Note that this approach implicitly assumes an input-output model for the production of ideas. In particular, $D^{75-94}_{r,s,\cdot,\cdot,\Delta}$ is equivalent to an input-output matrix specific to each pair of cities, (r, s),[17] and time lag, Δ. Each entry, $d^{75-94}_{r,s,\mu,\nu,\Delta}$, of this matrix determines how many patents of class μ produced in CZ r are necessary to produce an extra patent of class ν in CZ s after Δ years. The main departure from a classic input-output model of production is that in our case ideas are non-rival, non-excludable inputs. As a result, the sum of all the inputs that appear in the production of new patents can be larger than the overall amount of available inputs. [18]

[17]Note that the network is not symetric in cities, so that $D^{75-94}_{r,s,\cdot,\cdot,\Delta} \neq D^{75-94}_{s,r,\cdot,\cdot,\Delta}$

[18]To fix ideas, consider a world with two CZs (San Francisco and Detroit) that only produce two types of patents (Vehicles and Computers) and that only exists between 1975 and 1978. Assume that one patent of class Vehicles is filed in Detroit in 1975 and that San Francisco in 1976 produces 100 patents of class Computers that only cite the one patent filed in Detroit the year before. In this case, our measure of knowledge diffusion between the pairs (Detroit, Vehicles) and (San Francisco, Computers) at lag 1 would be:

$$d_{DT,SF,VH,CPU,1} = 100.$$

In the second step, the coefficients of diffusion constructed using the 1975-1994 sample are used to predict patenting in each class-CZ pair for the 2005-2014 period. More precisely, to estimate the patenting activity in the destination CZ s in 2005, we apply the adjacency matrix of the network with lag 1 to the actual patenting activity of all the other (origin) CZs in 2004 and then add up the results. In a similar way, we then apply the adjacency matrix with lag 2 to the actual patenting activity that occurred in 2003, and so on until lag 10. To obtain the predicted patenting activity, we sum the numbers we obtained at all lags. Mathematically,

$$\hat{pat}_{s,2005} = c_{2005} \sum_{\Delta=1}^{10} \sum_{r \in \mathcal{S}} \sum_{\nu \in \mathcal{N}} \left(\boldsymbol{D}^{75-94}_{r,s,\cdot,\nu,\Delta} \right)^T \boldsymbol{pat}_{r,\cdot,2005-\Delta}$$

where $\boldsymbol{D}^{75-94}_{r,s,\cdot,\nu,\Delta}$ is a column of the adjacency matrix that contains the coefficients of diffusion from CZ r to CZ s and class ν. Each row in the vector represents a technological class in the origin CZ. The vector $\boldsymbol{pat}_{r,\cdot,2005-\Delta}$ contains the actual number of patents for each class filed in CZ r in year $2005-\Delta$. The term c_{2005} is a rescaling term that makes sure that the total number of patents we estimate nationwide is the same as the one we observe in the data. The prediction of the patenting activity in the subsequent years follows the same strategy, with the only exception that when predicting total patents for 2006, the network with lag 1 is

Now, further assume that in 1978 Detroit files another patent of class Vehicles that cites 30 of the patents produced in San Francisco 2 years before. In this case, we would have,

$$d_{SF,DT,CPU,VH,2} = \frac{1}{30}$$

The intuition is that, from what we observe in the citations network, one single patent of class Vehicles in Detroit produces enough ideas to "generate" 100 patents of class Computers in San Francisco. On the contrary, we need 30 patents of class Computers in San Francisco to produce a single patent of class Vehicles in Detroit.

applied to predicted patents in 2005, instead of the actual ones (and similarly for all the years between 2006 and 2014).[19] We do this to avoid endogeneity concerns that might arise when using contemporaneous patenting activity. Table B.12 graphically outlines the exact structure used to build the instrument. Predicted patents in the second sub-period (2005-2014) are then averaged to obtain the instrument for the $t = 2010$ observation.

Note that the network we build is a directed one. If a class-CZ pair is linked to another pair, the opposite is not necessarily true. This contrasts with more common IV approaches used in the past in similar settings. For example, the Bartik instrument relies on the mere geographical distribution of innovative activities in the pre-sample period, and implicitly assumes that the coefficient of diffusion of ideas from any origin class-CZ pair is given by the national share of patents of the same class in the destination region. For our purposes, this approach carries some undesirable properties, most notably the inability to separate innovation shocks from industry or technology-specific nationwide trends that ultimately affect innovation, but also have an impact on the dependent variable. As we extensively discuss in Section 1.3.2.2, our approach significantly dampens this concern. First, we exploit the richness of the citation data to isolate directed technological linkages, including *across classes* links, and use it to diffuse lagged innovation output (1995-2004), rather than contemporaneous one (2005-2014). Second, our approach is robust to

[19]The role of c_{2005} is now evident. We add it to our estimation to avoid that the predicted number of patents in the later years is smaller just because predicted number of patents is used alongside actual patenting activity.

setting to zero the coefficient of diffusion not only for the citations coming from the same region but also for those coming from the same technological class, reducing the concern that predicted patenting growth simply reflects correlated industry trends. Third, we can directly control for those nationwide trends by including a Bartik-like variable directly into our set of controls.

1.3.2.2. First-stage results. One of the conditions for the instrument to be valid is that the network of knowledge inferred from the citations patterns is determined in the past but stable over time. This condition can be directly tested by comparing the network in the early sample with its counterpart in the late sample. This is done in three steps. First, we build the network of citations and compute the coefficients of diffusion separately for the two samples (1975-1994 and 1995-2014). For each $\Delta \in \{1, \ldots, 10\}$, we take the difference of the two adjacency matrices and calculate its Frobenius norm as follow:

$$real_{\Delta} = \left\| \boldsymbol{D}_{\Delta}^{75-94} - \boldsymbol{D}_{\Delta}^{95-14} \right\|_2 = \sqrt{\sum_{r,s,\mu,\nu} \left(\boldsymbol{D}_{\Delta}^{75-94} - \boldsymbol{D}_{\Delta}^{95-14} \right)^2}.$$

Second, for each year between 1975 and 2014, we reshuffle all the patents filed in that year under the constraint that after the reshuffling each commuting zone is assigned the same amount of patents as in the real dataset.[20] We repeat the same

[20] We also run the exercise under the constraint that each commuting zone is assigned the same number of patents it started with for each technological class. The results are virtually the same.

exercise performed in the first step for this new sample of patents and calculate,

$$reshuf_{\Delta} = \left\| \widetilde{\boldsymbol{D}}_{\Delta}^{75-94} - \widetilde{\boldsymbol{D}}_{\Delta}^{95-14} \right\|_2 = \sqrt{\sum_{r,s,\mu,\nu} \left(\widetilde{\boldsymbol{D}}_{\Delta}^{75-94} - \widetilde{\boldsymbol{D}}_{\Delta}^{95-14} \right)^2}$$

where $\widetilde{\boldsymbol{D}}_{\Delta}^{75-94}$ and $\widetilde{\boldsymbol{D}}_{\Delta}^{95-14}$ are the citation networks built using the reshuffled patents.

Finally, we calculate the percentage difference between $reshuf_{\Delta}$ and $real_{\Delta}$ for each Δ. This number tells us how far the two real networks are compared to two networks that, while maintaining the same structure and properties of the original ones, are uninformative of each other. A positive value indicates that the two networks built using the actual data are more similar than the two reshuffled networks.[21] Figure A.4 plots the difference (in percentage) for all the values of Δ together with the 95% confidence interval we obtained by repeating this procedure 50 times. The difference of the reshuffled networks is around 26% higher than the one obtained with the actual networks for the first lag and it gradually declines until it is indistinguishable from zero at lags 9 and 10. The decline implies that the more years pass after a new idea is generated the less citation patterns are distinguishable from links that are generated at random. This result is quite intuitive. With time a new technology becomes more and more public knowledge and is adopted or embedded in patents produced in areas that do not have any direct link with the city where the technology was originally produced.

Consistently with the results in the left-panel of Figure A.4, the right-panel shows a scatter plot of the first stage relationship between predicted and actual growth rate

[21] Note that this difference is only interpretable in relative terms.

of patenting. We plot the residuals of a regression of patent growth on the full set of controls. It is visually clear that the two variables are strongly but not perfectly correlated. The residual R^2 is 0.24, while the coefficient of the regression is 0.57. The Cragg-Donald Wald F statistics in the regression with full set of controls is 223.4, which rules out weak instrument concerns.

Figure A.14 in Appendix visually compares actual and predicted patent growth at the CZ level on a map of the United States, and can be useful to gain intuition on the validity of the instrument. Areas that are anecdotally associated with a large expansion of innovation and other knowledge-intensive activities (notably, Austin TX and Durham-Raleigh NC) are properly captured by the instrument.

1.3.3. IV Results

Our identification strategy captures local changes in patenting that are due to knowledge created in *other* geographical areas, linked to the original CZ through the channels of knowledge diffusion computed in (1.5). These channels are pre-determined with respect to new ideas themselves. When new knowledge becomes available in a city, innovation-intensive activities expand. In this section, we explore the effects of such an expansion on income segregation.

Table B.3 shows the 2SLS estimates of the relationship between innovation and segregation, as depicted in (1.4). All regressions are weighted by total number of households in 1990. The coefficient on patent growth is positive and statistically significant. Columns (2)-(6) introduce the set of controls considered for the OLS estimates. The coefficient on income growth reveals that segregation has increased more in areas with better economic performance. Columns (1)-(4) of Table B.15 in Appendix report the results when controls are included at their 1990 values. The coefficient on early sample population reveals that segregation has increased more in larger cities (consistently with the findings in Baum-Snow and Pavan 2013). Contrary to the OLS regressions, where the full set of controls had a significant dampening effect on the size of the coefficient, the 2SLS estimates are not significantly affected by the introduction of the controls.

A 10% increase in patenting between 1990 and 2010 is estimated to increase income segregation by $0.17 - 0.23$ Gini points, depending on the specification. Since the (population weighted) average growth rate of patents is 16% and the average increase in segregation 2.94, the effect is economically large. The effect is particularly significant in accounting for the cross-sectional variation in changes in segregation. Taking the specification with the basic set of controls in growth rates as a reference point, a one residual standard deviation increase in patenting growth increases segregation by 56% of a residual standard deviation in segregation change.

The 2SLS estimates are more than twice as large as the ones in the OLS regressions. This suggests that unobservable factors affecting at the same time innovation and segregation tend to operate on the two variables in opposite directions. This is hardly surprising. For example, financial shocks that generate widespread turmoil on the urban structure are likely to increase segregation while having a dampening effect on the local potential to develop a knowledge-based economy.

Table B.16 in Appendix shows that a similar effect is observed for segregation defined in terms of educational achievement and occupation type (as defined in Section 1.3.1), instead of income level. Patent intensity appears to have a strong positive impact on both measures. However, occupational segregation appears to be more tightly connected with income segregation than its educational counterpart: a regression of the change in occupational segregation on the change in income segregation yields an R^2 of 12%, whereas the corresponding figure for educational segregation is only 1.5%.

1.3.4. Instrument validation: Exclusion restriction

The instrument used in the IV analysis is a composite one, as it combines a pre-established network of knowledge links and a collection of innovation shocks that are then diffused through it. Hence, it requires two main identifying assumptions. First, the network of patent citations should not be capturing long-run trends in innovation and segregation. Second, shocks that affect innovation in the origin commuting zones should not be correlated with other shocks that affect innovation and segregation in the destination commuting zone other than through the channel identified by our instrument.

To address the first point, we run a number of falsification tests to verify to what extent the growth rate of patenting predicted by our instrument reflects long-term trends in innovation and segregation. We start by regressing predicted patenting growth (1990-2010) on past changes in segregation (1980-1990). Figure A.11 and columns (1)-(2) of Table B.17 show the correlation between our instrument and the pre-sample trend in segregation. This correlation is practically equal to 0.[22] Then, we check whether the instrument is correlated with previous trends in innovation, and to what extent this could affect our second stage results. Figure A.13 shows the correlation between the residuals of regressions of predicted patenting growth and past trends in patenting growth (1980-1990) on the basic set of controls. Although the coefficient of the two variables have a slightly positive correlation (the coefficient of the regression is 0.13 and is statistically significant), the R^2 of the regression is just 0.03, reflecting a very weak correlation. Column (2) in Table B.14 shows the 2SLS regression with the basic set of controls once the past trend in innovation is explicitly controlled for. The coefficient on patenting growth remains positive and significant, and is slightly larger in magnitude. This suggests that the correlation of the instrument with past trends in innovation is weak at best and is unlikely to confound our estimated effects.

As for the second point, the main concern is that geographical areas that are linked in the knowledge network have similar characteristics, such as a similar industry structure, geographical proximity, common regulation, or exposure to other shocks that make it hard to disentangle the genuine effect of knowledge shocks from the effect of other factors that have an impact on innovation in the origin CZ and

[22]The years we selected to calculate past changes in segregation are dictated by data availability from the Census. Note that, in the 1980 Census, CTs were not covering the entirety of the United States, but only the most densely populated areas. For this reason, not all the CZs are available for our analysis. This is unlikely to affect our results significantly, since all our regressions are weighted by the number of households. However, to make the two exercises readily comparable we re-run our benchmark regressions only using the CZs available in 1980. Columns (3)-(4) of Table B.17 report the results, which remain mainly unchanged.

segregation in the destination CZ. To control for the effect of nationwide industry or technology-specific shocks, we include a Bartik-like variable in the set of controls. Namely, for each CZ r we define a vector $S_r^{1990} = \left\{s_{1,r}^{1990}, ..., s_{N,r}^{1990}\right\}$, where $s_{\mu,r}^{1990}$ denotes the share of patents in the early sample that belong to technological class μ and was produced in CZ r. Then, for each class-CZ pair (μ, r), we compute the growth rate $g_{\mu,-r}$ of the number of grants in that technological class, counting only patents produced outside r, between 1990 and 2010. We then compute the Bartik-like variable in r as:

$$\hat{g}_r = \sum_{\mu \in \mathcal{N}} s_{\mu,r}^{1990} \cdot g_{\mu,-r}.$$

This prediction replicates the idea behind a Bartik shock, with the distribution of patents across technological classes used in place of the distribution of employment across industries. Column (3) in Table B.14 shows the 2SLS regression once the Bartik shock is included in the set of controls. The coefficient on patenting growth is robustly positive and larger in magnitude. Again, the inclusion of variables that control directly for industry performance (via their correlation with the distribution of innovation across classes) increases the size of the coefficient, confirming that unobservable shocks tend to operate on income segregation and innovation output in opposite directions.

To provide further evidence that our instrument is not capturing correlated industry trends across technologically linked CZs, column (4) of Table B.14 replicates the main 2SLS, with a version of our instrument in (1.5), in which the coefficient of diffusion is set to zero not only when the origin and destination CZs coincide, but also when the origin and destination technological classes are the same.[23] This version of the instrument displays a weaker correlation with observed patenting growth (the R^2 of the first stage regression drops from 0.23 to 0.16) but the coefficient of the IV regression is robustly positive and, again, larger in magnitude compared to our benchmark regression.

Lastly, we address the concern of changes in legislation and other geographically correlated unobservable factor by introducing state fixed effects in the 2SLS estimation of (1.4).[24] In this case, we are evaluating changes in segregation resulting from an expansion in innovation activities only through within-state variation. The results are reported in column (5) of Table B.14. The estimated coefficient is smaller, but the share of explained within-state variation is still sizable. One residual standard deviation in patenting growth explains 42.6% of a residual standard deviation the change in segregation. Column (6) reports the results when all the controls introduced in this section are included in the IV regression. Also in this case the results are robust.

1.3.5. Which technologies are driving the effect?

Our analysis can be disaggregated to investigate what types of technology are mainly responsible for the estimated effect. This decomposition is possible because our instrument delivers a separate predicted value for patenting in each technology class. It is a widespread belief that segregation has increased more in areas that are intensive

[23]In other words, we set $d^{75-94}_{r,s,\mu,\nu,\Delta} = 0$ whenever either $r = s$ or $\mu = \nu$.
[24]This implies that time fixed effects in (1.3) are state specific.

in high-tech industries. The following quote is taken from Florida (2015): *"Economic segregation tends to be more intensive in high-tech, knowledge-based metros. It is positively correlated with high-tech industry [...]".* By disaggregating the analysis at a technology class level, we can test whether this observation can be interpreted as causal.

The International Patent Classification (IPC) classifies patents into 8 main technological areas (each one divided into several technology sub-classes). We aggregate patents from each technology sub-class into their respective main technological area (which are labelled by letters from A to H). We then run a set of 8 separate 2SLS regressions, analogous to the ones shown in Section 1.3.3, with the exception that patenting growth is measured (and instrumented for) only within a given technological area.

Results are shown in Table B.4. The positive effect of patenting on segregation seems to be entirely driven by 4 out of 8 technological areas: class A (Human Necessities), which include Medicine and Pharmaceuticals among the others; class C (Chemistry); class G (Physics) which include all IT and Computer sectors; and class H (Electricity) which includes all major electronics products. Class D (Textiles and Paper), which is arguably the least knowledge intensive one in the IPC, has a negative and significant coefficient.

These results are obtained with the full set of controls, including income growth, so they are unlikely to capture exclusively differences in economic outcomes brought about by different types of jobs. However, the reason *why* knowledge intensive sectors (like Medicine, Chemistry and Information Technology) have a disproportionate effect on urban segregation, while less knowledge-intensive ones (like Textiles) have a negative effect is not obvious. Two explanations are the most likely candidates. On the one hand, learning-intensive sectors benefit more from learning spillovers and the proximity that such spillovers require. This produces higher incentives to cluster in space for people in areas where returns from learning are higher. On the other hand, people employed in those sectors might be disproportionally sensible to residential amenities, giving them a higher incentive to cluster in space. The spatial equilibrium model in Section 1.4 will be used to disentangle the contribution of the two candidate explanations to the observed effect.

1.3.6. Segregation and Inequality: Is it sorting?

Results up to this point show that an expansion of innovation activities has a positive impact on measured segregation, that is, on the variation of income across neighborhoods, within cities. Disregarding migration, there are two main phenomena that can induce this.

On the one hand, starting from a city with positive segregation (i.e. a condition in which the distribution of income is not the same in every neighborhood), a *divergence* in household income (e.g. a spread in the income distribution of the city) leads to an increase in measured segregation, even in the absence of any reallocation of residents across neighborhoods in the city. We refer to this phenomenon as *inequality effect.* On the other hand, measured segregation can increase even if within-city inequality stays the same, if residents choose to relocate across neighborhoods and sort themselves along the income dimension. We refer to this case as *sorting effect*.

The two phenomena can be used to think about the link between segregation and inequality in an intuitive way. The inequality effect allows us to connect changes in inequality with changes in segregation in the case where initial segregation is complete (e.g. where each household is the only resident in its neighborhood). In this case, it is clear that the following identity holds:

$$\Delta\, Ineq_{cz} = \Delta\, Segr_{cz}.$$

Since in reality initial segregation is never complete, in the absence of relocation an increase in inequality will in general induce a smaller change in segregation:

$$\Delta\, Ineq_{cz} \geq \Delta\, Segr_{cz}.$$

Hence, changes in inequality can always be interpreted as upper-bounds in terms of the effects on measured segregation.

As for the sorting effect, segregation can increase, as a result of the relocation of high (low) income households towards initially high (low) income neighborhoods, even if $\Delta\, Ineq_{cz} = 0$. In what follow, we discipline how much of the observed effect can be due to inequality and how much to sorting effects.

In Table B.5, we provide a comparison of the impact of patenting on segregation and inequality within-city. Specifically, we estimate (1.4) using alternatively $Segr_{cz}$

and $Ineq_{cz}$ as dependent variables. Innovation does have a positive impact on inequality. However, since the effect on segregation is larger than the one on within-CZ inequality, the two regressions taken together imply that the sorting effect is contributing significantly to the change in segregation.

In Column (3), we estimate (1.4) using $\Delta Segr_{cz}$ as dependent variable, and including $\Delta Ineq_{cz}$ as a control. The coefficient on $\Delta Ineq_{cz}$ is 0.93, suggesting an almost complete transmission of inequality to segregation. Moreover, the coefficient that measures the effect of patenting growth on $\Delta Segr_{cz}$ drops accordingly by roughly one third, but remains positive and significant. This implies that roughly two thirds of the impact of innovation shocks on segregation can be explained as a sorting effect, whereas the remaining third as an inequality effect.

The impact of an innovation shock on within-neighborhood inequality is ex-ante ambiguous, since the inequality and sorting effects operate on opposite directions. On the one hand, the positive impact on $\Delta\, Ineq_{cz}$ implies that, if people were not allowed to relocate, we would observe a positive effect on within-CT inequality, as well.[25] On the other hand, the sorting effect works to counteract the impact of within-city on within-CT inequality. The last two columns of Table B.5 report the parameter estimates using average within-CT inequality as left-hand side variable. Patenting growth has a small negative coefficient, that becomes statistically indistinguishable from zero when we add the baseline controls to the regression. This suggests that the

[25]In the extreme case in which the income distribution for each CT is identical to the one in the city, an increase in inequality at a city level would translate into a one-to-one increase of average within-CT inequality. On the other hand, if people were perfectly sorted along the income dimension, an increase in city-level inequality would have no impact on within-CT inequality.

sorting effect completely offsets the increase in the dispersion of income within-CT that stems from the inequality effect.

1.3.7. Exploration of the mechanism

In the previous subsections, we showed the existence of a strong, causal relationship between the expansion of local knowledge-based activities and income segregation in U.S. cities. We further showed that this effect is also visible along an educational and occupational dimension and is mostly driven by technological fields with high technological content and learning intensity such as Physics and Chemistry. This result suggests that high returns from learning spillovers can increase incentives for companies whose output has a high knowledge content to cluster in space to take advantage of highly localized learning opportunities, inducing a positive link between innovation intensity and concentration of knowledge-intensive firms. In addition, high-education, high-salary workers might optimally relocate in the surrounding areas to minimize their commuting costs. The endogenous response of residential externalities (e.g. local services that are valued more by workers in the knowledge economy, such as schools and organic grocery stores) can play an important role in amplifying this effect.

The structural model presented in Section 1.4 formalizes this mechanism. The goal of this subsection is to provide suggestive reduced-form evidence in its support. First, we show that innovation shocks promote the geographical concentration of knowledge workers towards neighborhoods with high learning opportunities.

Second, we show that the impact of innovation shocks is stronger in cities whose neighborhoods are less anchored to natural (or persistent) amenities, highlighting the potential role of endogenous residential externalities in driving the process.

1.3.7.1. Clustering of employment. One possible mechanism behind the results described in Section 1.3.3 is the change in the concentration of employment of knowledge-intensive occupations that is induced by a knowledge shock. The fact that knowledge spillovers are strongly localized has been confirmed by multiple studies, starting from Jaffe et al. (1993). When useful knowledge becomes available and innovation opportunities emerge, incentives to cluster in space to benefit from them are positively affected. This in turn has a direct effect on residential segregation, provided that work location affects residential choices, (for example, if people are averse to spending time commuting).

To confront this intuition, we first verify that in cities with high innovation shocks, knowledge intensive employment moves towards neighborhoods with strong learning opportunities. Our measure of knowledge spillovers at the neighborhood level is adapted from Ahlfeldt et al. (2015), and is based on the structural model outlined in Section 1.4.3. The index captures the concentration of knowledge workers surrounding a given neighborhood.[26] Specifically, for each CT j in city cz, knowledge externalities in 1990 are computed as:

$$\Lambda_j^{kk} = \sum_{l \in \mathcal{S}_{cz}} e^{-\delta_k \tau_{jl}} \frac{W_l^k}{K_l},$$

[26]See Appendix C.1.1 for details on the classification of occupations and the construction of the distribution of residents by occupation at the neighborhood level.

where $\mathcal{S}_{cz}$ is the set of neighborhoods in cz, τ_{jl} is the commuting time (in minutes) between CTs j and l, W_l^k is the number of knowledge workers employed in l in 1990 and K_l is the area of l. The parameter δ_k controls the rate of decay of knowledge externalities and is estimated in Section 1.4.6.

Our conjecture is that, in cities that receive strong knowledge shocks, knowledge occupations will cluster into neighborhoods with high externalities. Letting $s_{j,cz}^k$ be the percentage of knowledge workers in CT $j \in \mathcal{S}_{cz}$, and letting $rank_{j,cz}$ be the percentile of j in the distribution of Λ^{kk} within cz in 1990, we estimate via 2SLS the following equation:

$$\Delta s_{j,cz}^k = \alpha_{cz} + \beta\, rank_{j,cz} + \gamma\, rank_{j,cz} \times \Delta \log(1 + Patents_{cz}) + \epsilon_{j,cz}. \tag{1.6}$$

We cluster standard errors at the CZ-level and weight each CT by the total number of workers in 1990. A positive sign for the coefficient of the interaction, γ, suggests that neighborhoods with high learning opportunities in 1990, in cities where the knowledge shock has been stronger, have experienced a more pronounced shift towards knowledge-intensive occupations. The OLS and IV estimates of 1.6 are displayed in Table B.6. The interaction term has a positive and significant coefficient, that is meaningful in magnitude. Combining the estimates of β and γ, we can see that in cities at the 95th percentile of the distribution of innovation shocks, CTs at the top of the distribution of Λ^{kk} in 1990 experienced a shift in the composition of employment towards knowledge occupations about 3.52 percentage points higher than CTs at the bottom of the distribution of Λ^{kk} in 1990. The corresponding figure, in cities at the 5th percentile of the distribution of innovation shocks, is significantly smaller (1.40).

1.3.7.2. The role of residential amenities. As mentioned in the previous subsection, the fact that high-knowledge firms cluster in space might directly influence residential choices of workers through commuting costs considerations. This process could be amplified by the existence of endogenous residential spillovers that are disproportionately valuable to high-education, high-salary workers. For example, a high concentration of creative workers might attract amenities such as elite schools or fitness centers, to which other types of workers might be less sensible.

To check whether residential amenities play a role in promoting the increase in segregation observed in the data, we exploit the index of natural amenities assembled by Lee and Lin (2017). The authors build an index based on the distance to natural amenities (e.g., ocean coast) or the presence of steady features (e.g., fountains) for each Census Tract contained in a Metropolitan Statistical Area (MSA). In their paper, they show that MSAs where the index variance is higher are also MSAs whose spatial income distribution has remained more persistent over time. Our idea is that cities that incorporate residential amenities whose valuation is unlikely to be altered by the surrounding distribution of residents, should also be cities where the residential spillover channel is weaker. In other words, the presence of extremely valuable amenities that are exogenous relative to the geography of the city should have a dampening effect to the residential agglomeration forces documented in the previous sections, since the endogenous spillovers would play a more marginal role.

We first assign every CT contained in the Lee and Lin's (2017) dataset to a CZ and, following their methodology, we calculate the standard deviation of the amenities index for each city.[27] We then introduce this term and its interaction with patenting growth to our baseline regression model. A negative coefficient for the interaction term indicates that cities whose variation in residential amenities is more anchored to natural or persistent features, experience a less pronounced change in income segregation following an innovation shock. Columns 2 and 4 of Table B.7 report the OLS and IV results of such a regression. As expected, the parameter associated with the interaction term is negative and statistically significant at a 10% level. The magnitude of the coefficient is economically large. The point estimate implies that cities ranked at the 95th percentile in their degree of persistent residential amenities display a marginal effect of knowledge shocks on income segregation equal to 0.84 Gini points, less than a quarter of the marginal impact in a city at the 5th percentile of the distribution (2.96). This suggests that residential amenities play indeed an important role in amplifying the effect of innovation shocks on income segregation.

[27]Note that since the MSAs do not cover the whole U.S. territory, for this exercise we are able to use data from 337 cities only.

1.3.8. Taking Stock

The empirical analysis shows a robust and economically meaningful causal relationship between the expansion of innovation activities and the increase in income segregation in U.S. cities between 1990 and 2010. This effect is stronger for learning-intensive fields (Medicine, Chemistry, IT, Electronics) and weaker (or negative) for less knowledge-intensive fields (Textiles). Less than 50% of this effect can be explained by an increase of income inequality, suggesting that knowledge intensity generates incentives for people to sort in space along income, occupational and educational dimensions. As a potential mechanism, we provide evidence suggesting that (1) innovation shocks induce an increase in the geographical concentration of employment of knowledge-intensive occupations, which can affect income segregation if the location of employment is linked to residential choices, and (2) the endogenous response of residential amenities can work as an important amplification channel.

In the next section, we propose a structural model of the internal structure of cities that formalizes and quantifies such mechanism. We augment the model developed in ARSW to allow for a creative, knowledge-intensive sector and a residual non-creative sector. The model features occupation-specific productivity and residential externalities, generating a variety of motives for job clustering and residential sorting. The exogenous innovation shocks derived in the empirical analysis allow us to structurally estimate the parameters controlling the strength of such externalities. The model is successful in replicating the key empirical relationships, and can be used to investigate the factors that drive them.

1.4. Model

We consider an economy comprising a finite set of cities $\mathcal{C}$. In what follows, we present the model for an arbitrary city $c \in \mathcal{C}$, and suppress the city index for notational convenience. Our setting expands ARSW by allowing for multiple cities and worker types. We refer to the original paper and its Appendix for some of the derivations and details.

1.4.1. Demand

A city $c \in \mathcal{C}$ comprises a finite set of neighborhoods (CTs) $\mathcal{S}$. Agents differ intrinsically by their background, or sector in which they operate. There is a creative sector k and a residual sector n, to which each worker inelastically supplies one unit of labor. The utility function of worker o of type $x \in \{k, n\}$, living in neighborhood i and working in j is given by

$$U^x_{ijo} = \frac{z_{ijo}}{d_{ij}} B^x_i \left(\frac{c^x_{ijo}}{\beta}\right)^{\beta} \left(\frac{h^x_{ijo}}{1-\beta}\right)^{1-\beta} \tag{1.7}$$

where c_{ijo} is a tradable consumption good (the numeraire), h_{ijo} is consumption of housing of price q_i, B^x_i represents residential amenities, and z_{ijo} is a Frechet-distributed random variable with shape parameter $\varepsilon > 1$. The term $d_{ij} = e^{\kappa \tau_{ij}}$ represents iceberg commuting costs, with τ_{ij} denoting commuting times (in minutes) from i to j, and $\kappa > 0$ a parameter controlling the sensitivity to commuting. Every worker maximizes her utility subject to

$$c_{ijo}^{x} + q_i h_{ijo}^{x} \leq w_j^{x},$$

where w_j^x is the wage that workers of type x receive when working in CT j. Utility maximization yields

$$h_{ijo}^{x} = (1-\beta)\frac{w_j^x}{q_i}, \qquad c_{ijo}^{x} = \beta w_j^x.$$

Using the two optimality conditions, we can write the indirect utility function as

$$u_{ijo}^{x} = B_i^x \frac{z_{ijo}}{d_{ij}} w_j^x \left(q_i\right)^{\beta-1}. \tag{1.8}$$

Upon moving to the city, each agent receives a collection of Frechet-distributed independent draws, one for each $(i,\, j)$ pair of residence and workplace neighborhoods, and chooses the pair that delivers the highest utility. Using the indirect utility function in (1.8) and the properties of the Frechet distribution, we can calculate the share of workers o of type x choosing to live in CT i and work in CT j:

$$\pi_{ij}^{x} = \frac{\left(B_i^x w_j^x\right)^{\varepsilon} \left(d_{ij} q_i^{1-\beta}\right)^{-\varepsilon}}{\sum\limits_{l,m \in \mathcal{S}\times\mathcal{S}} \left(B_l^x w_m^x\right)^{\varepsilon} \left(d_{lm} q_l^{1-\beta}\right)} \equiv \frac{\Phi_{ij}^{x}}{\Phi^{x}}. \tag{1.9}$$

Summing over the work locations, we get the share of people of type x who live in neighborhood i:

$$\pi_{Ri}^x = \sum_{j \in \mathcal{S}} \pi_{ij}^x = \frac{\sum_{j \in \mathcal{S}} \Phi_{ij}^x}{\Phi^x}. \tag{1.10}$$

Similarly, the share of workers of type x who work in j can be expressed as

$$\pi_{Wj}^x = \sum_{i \in \mathcal{S}} \pi_{ij}^x = \frac{\sum_{i \in \mathcal{S}} \Phi_{ij}^x}{\Phi^x}. \tag{1.11}$$

The probability of commuting to j conditional on living in i is given by

$$\pi_{ij|i}^x = \frac{\left(w_j^x / d_{ij}\right)^\varepsilon}{\sum_{l \in \mathcal{S}} \left(w_l^x / d_{il}\right)^\varepsilon} \tag{1.12}$$

Therefore, the measure of people of type x who work in j, denoted by W_j^x, is given by

$$W_j^x = \sum_{l \in \mathcal{S}} \frac{\left(w_j^x / d_{lj}\right)^\varepsilon}{\sum_{m \in \mathcal{S}} \left(w_m^x / d_{lm}\right)^\varepsilon} \pi_{lj}^x R^x, \tag{1.13}$$

where R^x is the amount of residents of type x living in the city.[28]

[28]Note that $R^x = \sum_j W_j^x = \sum_i R_i^x$, where R_i^x is the mass of residents of type x in CT i.

Using the conditional probability derived in (1.12), we can calculate the expected wage of type x conditional on living in neighborhood i:

$$\mathbb{E}\left[w^x \mid i\right] = \sum_{k \in \mathcal{S}} \frac{\left(w_k^x / d_{ik}\right)^{\varepsilon}}{\sum_{l \in \mathcal{S}} \left(w_l^x / d_{il}\right)^{\varepsilon}} w_k^x,$$

which is the average wage received by workers of type x in CT k weighted by the probability of working there, conditional on living in i.

The distribution of utilities of type x for each type in the city is given by

$$G^x\left(u\right) = e^{-\Phi^x u^{-\varepsilon}}.$$

To see this, note that the probability that the utility of an agent of type x chosen at random in the city is higher than u, $1 - G^x\left(u\right)$, is equal to the probability that her utility is bigger or equal to u for at least one residence-workplace pair, or equivalently to 1 minus the probability that her utility is smaller than u for all the residence-workplace combinations:

$$1 - G^x\left(u\right) = 1 - \prod_{l,m \in \mathcal{S} \times \mathcal{S}} G_{lm}^x\left(u\right),$$

where $G_{ij}^x\left(u\right) = e^{-\Phi_{ij}^x u^{-\varepsilon}}$ is the utility distribution of workers of type x living in i and working in j. From here it is easy to see that

$$\mathbb{E}\left[u^x\right] = \Gamma\left(1 - \frac{1}{\varepsilon}\right)\left(\Phi^x\right)^{1/\varepsilon}$$

where $\Gamma(\cdot)$ is the Gamma function. The expected utility when deciding to move into the city must be equal to the reservation utility $\bar{U}^x$, that is constant across cities.

1.4.2. Production

Each neighborhood j hosts a representative, perfectly competitive firm of each sector $x \in \{k, n\}$. The firm hires sector-specific labor and rents office space, and aggregates them into a homogeneous final good according to a Cobb-Douglas production function:

$$y_j^x = A_j^x \left(H_j^x\right)^{1-\alpha} \left(W_j^x\right)^{\alpha},$$

where y_j^x is output of firm x in CT j, A_j^x is its total factor productivity, and H_j^x is total office space rented by the representative firm.

Profit maximization gives

$$(1-\alpha) A_j^x \left(\frac{W_j^x}{H_j^x}\right)^{\alpha} = q_j, \qquad \alpha A_j^x \left(\frac{H_j^x}{W_j^x}\right)^{1-\alpha} = w_j^x. \tag{1.14}$$

Combining the FOCs with the zero profit condition yields

$$q_j^x = (1-\alpha) \left(\frac{\alpha}{w_j^x}\right)^{\alpha/(1-\alpha)} \left(A_j^x\right)^{1/(1-\alpha)}. \tag{1.15}$$

1.4.3. Residential and productivity externalities

The terms B_i^x and A_j^x summarize the location's residential and productivity characteristics. We assume them to be geometric functions of the concentration of economic activity around the relevant location. Elasticities are occupation-specific, so that the intensity of the externalities depend on the type of resident or worker who is generating and benefiting from them.

We define density of residents of type $x_2 \in \{k, n\}$ around residents of type $x_1 \in \{k, n\}$ in neighborhood i as

$$\Omega_i^{x_1 x_2} = \sum_{l \in \mathcal{S}} e^{-\rho_{x_1} \tau_{il}} \frac{R_l^{x_2}}{K_l}, \tag{1.16}$$

where ρ_{x_1} is the rate of decay of residential externalities perceived by residents of type x_1, and K_l is the area in CT l.[29] Then, residential amenities for type x_1 in location i are

$$B_i^{x_1} = b_i^{x_1} \left(\Omega_i^{x_1 x_1}\right)^{\omega_{x_1 x_1}} \left(\Omega_i^{x_1 x_2}\right)^{\omega_{x_1 x_2}}, \tag{1.17}$$

where $\omega_{x_1 x_1}$ ($\omega_{x_1 x_2}$) represents the elasticitiy of residential externalities from residents of type x_2 (x_1) to residents of type x_1, and $b_i^{x_1}$ is an exogenous term that captures the component of residential amenities that is not affected by the surrounding economic activity.

Similarly, we define the density of employment of type x around workers of type k (creative occupations) in neighborhood j as

$$\Lambda_j^{kx} = \sum_{l \in \mathcal{S}} e^{-\delta_k \tau_{jl}} \frac{W_l^x}{K_l}, \tag{1.18}$$

[29]This functional form is consistent with the intuition given by Lucas and Rossi-Hansberg (2003) on how knowledge spillovers are generated.

where δ_k is the rate of decay of productivity externalities perceived by workers of type x. Then, the productivitiy term for type k in location j is

$$A_j^k = a_j^k \left(\Lambda_j^{kk}\right)^{\lambda_{kk}} \left(\Lambda_j^{kn}\right)^{\lambda_{kn}}, \tag{1.19}$$

where λ_{kk} (λ_{kn}) represents the elasticity of productivity externalities from workers of type k (n) to workers of type k, and a_j^k is an exogenous term that captures the component of productivity that is not affected by the surrounding economic activity. In the structural estimation of Section 1.4.6, we allow λ_{kk} (the intensity of local learning among workers in the creative sector) to depend on the city-specific knowledge shocks that were estimated in the empirical analysis.

We maintain the assumption that the productivity terms for the non-creative occupations, A_j^n, are stagnant, and are not affected by local externalities, so that $A_j^n = a_j^n$ for all neighborhoods. This assumption is consistent with Davis and Dingel (2016), in which only workers who select themselves in knowledge intensive occupations benefit from the concentration of learning opportunities in large cities. As discussed in Section 1.4.6, our quantitative results support this interpretation.

1.4.4. Equilibrium

We now have all the elements to define an equilibrium of the model.

Definition 1.4.1. Given quantities $\left\{L_i,\, K_i,\, \{a_i^x, b_i^x\}_{x\in\{k,n\}}\right\}_{i\in\mathcal{S}} \in (0,\, \infty)$ and $\{\tau_{ij}\}_{i,j\in\mathcal{S}\times\mathcal{S}} \in (0,\, \infty)$ and reservation utilities $\left\{\bar{U}^k,\, \bar{U}^n\right\}$, an **equilibrium** is a set of quantity and prices $\left\{ \left\{ \pi^x_{Ri},\, \pi^x_{Wi}, R^x_i, W^x_i, w^x_i, A^x_i, B^x_i \right\}_{x\in\{k,n\}},\, q_i \right\}_{i\in\mathcal{S}}$, so that, for each type $x \in \{k,\, n\}$:

- Expected utility of moving into the city equals the reservation utility

$$\Gamma\left(1-\frac{1}{\varepsilon}\right)\left[\sum_{l\in\mathcal{S}}\sum_{m\in\mathcal{S}}\left(d_{lm}\left(q_l\right)^{1-\beta}\right)^{-\varepsilon}\left(B^x_l w^x_m\right)^{\varepsilon}\right]^{1/\varepsilon} = \bar{U}^x \tag{1.20}$$

- The share of population living in i is given by (1.10)
- The share of population working in j is given by (1.11)
- Land markets clear for each $i \in \mathcal{S}$:

$$\sum_{x\in\{k,n\}}\left(\frac{(1-\alpha)\,A^x_i}{q_i}\right)^{1/\alpha} W^x_i + (1-\beta)\sum_{x\in\{k,n\}}\left[\sum_{l\in\mathcal{S}}\frac{\left(w^x_l/d_{il}\right)^{\varepsilon}}{\sum_{m\in\mathcal{S}}\left(w^x_m/d_{im}\right)^{\varepsilon}} w_l\right]\frac{R^x_i}{q_i} = L_i \tag{1.21}$$

- Productivity and residential externalities are determined by (1.17) and (1.19), respectively
- Factor prices satisfy (1.14), so that firms make zero profits
- Labor markets clear:

$$R^x_i = \pi^x_{Ri}\sum_{l\in\mathcal{S}} R^x_l, \qquad W^x_j = \pi^x_{Wj}\sum_{l\in\mathcal{S}} W^x_l,$$

$$R^x \equiv \sum_{l\in\mathcal{S}} R^x_l = \sum_{l\in\mathcal{S}} W^x_l \equiv W^x.$$

The fact that residential amenities and productivities are subject to local externalities gives rise to the potential for multiple equilibria. As discussed by ARSW, the structure of the model allows to deal with this multiplicity directly by identifying a unique set of location characteristics that is compatible with the data, so that only the observed equilibrium is relevant for the estimation of the model's parameters.

1.4.5. Recovering wages and location characteristics from data

The structure of the model allows us to recover unobserved location characteristics starting from data on residents by sector, $\left\{R_i^k, R_i^n\right\}_{i\in\mathcal{S}}$, workers by sector, $\left\{W_j^k, W_j^n\right\}_{j\in\mathcal{S}}$, and rental price of floor space, $\{q_i\}_{i\in\mathcal{S}}$, bilateral commuting times, $\{\tau_{ij}\}_{i,j\in\mathcal{S}}$, and average wage by sector in the city, $\left\{\bar{w}_c^k, \bar{w}_c^n\right\}$, given knowledge of the parameters κ and ε_c. The equilibrium conditions can then be inverted to univocally identify wages by sector, $\left\{w_j^k, w_j^n\right\}_{j\in\mathcal{S}}$, residential amenities $\left\{B_i^k, B_i^n\right\}_{i\in\mathcal{S}}$, and productivities $\left\{A_j^k, A_j^n\right\}_{j\in\mathcal{S}}$.

We first discuss how we obtain an estimate for the city-specific parameter controlling the sensitivity to commuting, $\nu_c = \varepsilon_c\kappa$. We then discuss how to pin down local wages by sector. Finally, we show how to recover the values of residential amenities and productivities. The data sources used for this purpose are described in details in Appendix C.1.1.

Estimating sensitivity to commuting times. We allow the parameter that controls the sensitivity of the utility function to commuting times to vary by city. Taking logs of (1.9) yields a gravity equation for commuting flows from CT i to CT j:

$$\log\left(\pi_{ij}^x\right) = \alpha^x + \psi_i^x + \zeta_j^x + \nu_c\tau_{ij} + \eta_{ij}^x \tag{1.22}$$

where $\nu_c = \varepsilon_c \kappa$, and ψ_i^x and ζ_j^x are residence and workplace fixed effects, respectively. Since there are no comprehensive measures of commuting flows by occupation, we approximate a single gravity equation for commuting flows by estimating one equation of the same form for each city:

$$\log(\pi_{ij}) = \alpha + \psi_i + \zeta_j + \nu_c \tau_{ij} + \eta_{ij}. \tag{1.23}$$

We show in the Appendix (Figure A.16) that an alternative method for estimating ν_c, based on replicating the observed share of residents commuting for less than 60 minutes from their workplace, yields very consistent results.

We estimate (1.22) by OLS separately for each city using data on actual commuting flows from the Longitudinal Employer-Household Dynamics (LEHD) dataset. The distribution of estimates of ν_c is illustrated in Figure A.15. The median value is -0.041, which implies that one additional minute of commuting time decreases commuting probability by 4.1%.[30]

Recovering wages, residential amenities and productivities. For given values of κ and ε_c, wages by sector in each location are uniquely (up to a normalization) determined by the following system of $2 \times |\mathcal{S}|$ equations:

$$W_j^x = \sum_{i \in \mathcal{S}} \frac{\left(w_j^x\right)^{\varepsilon_c} / e^{\nu_c \tau_{ij}}}{\sum_{l \in \mathcal{S}} \left(w_l^x\right)^{\varepsilon_c} / e^{\nu_c \tau_{il}}} R_i^x, \tag{1.24}$$

[30] The results are consistent with the ones in ARSW, who estimate a value of -0.07 for the same parameter.

where $\{W_i^x,\, R_i^x\}_{i\in\mathcal{S}}$ are observed in the data, and where an appropriate normalization of the wages is chosen, so that the average wage in the city is equal to the observed counterpart in the data, $\bar{w}_c^x$. We choose units so that the geometric mean of the non-creative sector's wage in the CZ with the first index (Memphis) is equal to one.[31]

Given a value for α and knowing $\{q_i\}_{i\in\mathcal{S}}$ and $\left\{w_j^k,\, w_j^n\right\}_{j\in\mathcal{S}}$, productivities $\left\{A_j^k,\, A_j^n\right\}_{j\in\mathcal{S}}$ can be recovered from equation (1.15). Then, given values for $\{\delta_k,\, \delta_n\}$ and $\{\lambda_{x_1x_2}\}_{x_1,x_2\in\mathcal{S}}$ and observed areas $\{K_i\}_{i\in\mathcal{S}}$, the exogenous component of productivity $\left\{a_j^k,\, a_j^n\right\}_{j\in\mathcal{S}}$ can be obtained by combining (1.18) and (1.19).

Given values for ε_c and β, observed data for $\left\{R_i^k,\, R_i^n,\, q_i\right\}_{i\in\mathcal{S}}$, and the equilibrium wages $\left\{w_j^k,\, w_j^n\right\}_{j\in\mathcal{S}}$, combining (1.10) and (1.20) allows us to recover residential amenities $\left\{B_i^k,\, B_i^n\right\}_{i\in\mathcal{S}}$:

$$B_i^x = \left(\frac{R_i^x}{R^x}\right)^{\frac{1}{\varepsilon_c}} \left(\frac{\bar{U}^x}{\Gamma\left(1-\frac{1}{\varepsilon_c}\right)}\right) \frac{q_i^{1-\beta}}{\left(\tilde{w}_i^x\right)^{1/\varepsilon_c}} \qquad x\in\{k,n\}\,, \tag{1.25}$$

where $\tilde{w}_i^x = \sum_j \left(w_j^x/d_{ij}\right)^{\epsilon}$. We choose units so that the geometric mean of residential amenities for both types in the CZ with the first index (Memphis) is equal to one. This choice of units allows us to recover the unobserved value of the reservation utility $\bar{U}^x$ and to evaluate (1.25) for the remaining cities.[32]

[31] See Lemma S.7 in the Supplement to ARSW for a proof that the system of equations in (1.24) determine a unique (up to a normalization) vector of wages $\left\{w_j^x\right\}_{j\in\mathcal{S}}$.

[32] One additional normalization is required to define units in which floor space is denominated. We normalize the price of floor space, q_i, so that the geometric mean in Memphis is equal to one.

1.4.6. Structural estimation

We follow ARSW and set $\alpha = 0.8$, $\beta = 0.75$ and $\kappa = 0.01$ in our calibration, which implies $\varepsilon_c = \nu_c/0.01$.[33] In order to estimate the remaining parameters (the ones that control the strength of the agglomeration externalities) we exploit the differential change in the concentration of economic activity in cities between 1990 and 2010 that results from differential changes in knowledge intensity, as recovered in the empirical analysis. In particular, we rely on the orthogonality between the inferred innovation shocks and other factors that affect the geographical distribution of economic activity in the city. The model captures those residual factors as changes in the exogenous components of productivity and residential amenities, a_j^x and b_j^x. Our orthogonality condition imposes that changes in the average of the exogenous components within a city are independent from the innovation shock the same city receives.

To introduce innovation shocks, we assume that the elasticity of productivity externalities for the creative sector λ_{kk} is identical across cities in 1990 (λ_{kk}^{90}), but varies in 2010 depending on the city-specific value of the knowledge shock:

$$\lambda_{kk,c}^{10} = \lambda_{kk}^{90} + \theta_0 + \theta_1 \cdot bin_c \tag{1.26}$$

[33]Following Allen et al. (2017), we also estimated ε_c using a model generated instrument together with the fixed effects obtained from the gravity equation (1.22). Although the confidence interval includes values strictly greater than 1 for 96% of the commuting zones, the point estimate is smaller than 1 in 20% of the cases, including in some major cities such as Los Angeles and New York. For this reason, we use the ARSW estimates and set $\kappa = 0.01$ for our analysis. Interestingly, the weighted and unweighted mean of κ_c obtained through the procedure proposed by Allen et al. (2017) is very close to this value. Details of the procedure and results are provided in the Appendix.

where θ_0 and θ_1 are estimated jointly with the remaining parameters, and bin_c is the value of the knowledge shock for city c, as described below.

To make the orthogonality condition operational, we proceed in three steps. First, we compute for each city the predicted patenting growth, as outlined in Section 1.3.2.1:

$$\hat{g}_c = \log\ \hat{pat}_{c,05-14} - \log\ pat_{c,85-94}. \tag{1.27}$$

Second, we take the residuals of a regression of $\hat{g}_c$ on the set of basic controls (number of CTs, income and population growth). Third, we sort cities according to those residuals (in ascending order) into 10 bins, so that the sum of the population of all the cities in the bin is approximately equal for all the bins (and equal to $\frac{1}{10}$ of the total population). The resulting categorization determines the value of the knowledge shock (bin_c) introduced in (1.26). The orthogonality condition can then be expressed as

$$\begin{cases} \mathbb{E}_{c\in\mathcal{C}_{bin}}\left[\Delta_{10-90}\mathbb{E}_{i\in\mathcal{S}_c}\log\left(a_i^x\right)\right] = \mathbb{E}_{c\in\mathcal{C}}\left[\Delta_{10-90}\mathbb{E}_{i\in\mathcal{S}_c}\log\left(a_i^x\right)\right] \\ \mathbb{E}_{c\in\mathcal{C}_{bin}}\left[\Delta_{10-90}\mathbb{E}_{i\in\mathcal{S}_c}\log\left(b_i^x\right)\right] = \mathbb{E}_{c\in\mathcal{C}}\left[\Delta_{10-90}\mathbb{E}_{i\in\mathcal{S}_c}\log\left(b_i^x\right)\right] \end{cases} \tag{1.28}$$

for all $bin \in \{0, ..., 9\}$ and $x \in \{k, n\}$. In (1.28), all expectations are weighted by total population in the neighborhood. For a fixed set of parameters, and given observed data on residents, workers and price of housing (that also imply a unique vector of wages through (1.24)), residential and productivity fundamentals can be recovered by combining (1.25) with (1.17) and (1.15) with (1.19), respectively.

Condition (1.28) requires that cities with different knowledge shocks do not display systematic differences in the way residual fundamentals change between 1990 and 2010. Hence, the systematic difference in how the concentration of economic activity changes must be due to the combination of the the change in the production function of the creative sector induced by the knowledge shock, and the endogenous agglomeration forces in the model.

The system in (1.28) delivers 3×10 moment conditions for a set of 11 parameters to estimate:[34]

$$P \equiv \{\rho_n,\, \rho_k,\, \delta_k,\, \omega_{nn},\, \omega_{nk},\, \omega_{kn},\, \omega_{kk},\, \lambda_{kn},\, \lambda_{kk},\, \theta_0,\, \theta_1\}\,.$$

Our estimation routine sets the value of the parameters, P^*, in such a way as to minimize the sum of the squares of the moment conditions:

$$P^* = \text{argmin}_{P \in \mathcal{P}}\, m\left(P\right)\, \mathbb{W}\, m\left(P\right)'$$

where $\mathbb{W}$ is the optimal weighting matrix. Details on the estimation algorithm can be found in Appendix C.1.3.

The results of the estimation are displayed in the right panel of Table B.8. The rates of decay of residential externalities (ρ_n and ρ_k) are close to the corresponding estimates in ARSW ($0.55-0.90$), and suggest that residential externalities are slightly more localized for knowledge workers. The rate of decay of productivity spillovers for

[34]Note that since we assume $\lambda_{nn} = \lambda_{nk} = 0$, the moment conditions involving a_i^n do not identify any relevant parameter.

knowledge workers (δ_k) is lower than the estimate in ARSW ($0.35-0.92$) and points in the direction that learning externalities, albeit localized, have a larger geographical span than other types of productivity spillovers.[35] The estimated value of δ_k implies that a 10 minutes commuting time reduces the strength of the externality by roughly 42%.

Two additional considerations stand out. First, as suggested by the similar estimated values of ω_{nn} and ω_{nk}, residential externalities perceived by non-creative workers are closer across the two types than externalities that knowledge workers receive from neighbors of both types. These are very steep for knowledge workers (ω_{kk} is high), and significantly lower for residents of the opposite type (ω_{kn} is low). This dichotomy suggests that, following an initial shock to the distribution of employment, the amplification effect of local amenities on the distribution of residents can be large. Second, workers in the creative sector receive very steep productivity externalities from other knowledge workers, and less powerful externalities from non-creative workers.

1.4.7. Quantitative Exploration

In this section, we first explore to what extent the estimated model can account for the observed relationship between innovation and income segregation, and then perform counterfactual experiments to shed light on the underlying mechanism.

[35]Incidentally, when productivity spillovers for n-workers are included in the estimation, the routine delivers an explosive value of δ_n, roughly equal to 10, which implies that productivity spillovers for non-creative workers are extremely localized, possibly limited to the firm's boundaries.

We proceed as follows. For each city in the sample, we first compute the model equilibrium using data on residents and workers by type, and rental price of housing in 1990. We then recover the exogenous component of productivity and residential amenities, $\{a_i^x, b_i^x\}$, as described in Section 1.4.5. In running the counterfactuals, we keep the value of the location characteristics fixed at the inferred 1990 level, and change exclusively the value of λ_{kk} in order to reflect the corresponding knowledge shock, as in equation (1.26). The algorithm used to find the new equilibrium (adapted from ARSW) can be found in Appendix C.1.4. Note that the endogenous agglomeration forces can give rise to multiple equilibria. The recursion used in the following experiments looks for the equilibrium that is closer to the original one.

We present our results in bin scatter plots, so that each dot in the figure corresponds to the weighted average of the observations in the knowledge shock bin, as defined in Section 1.4.6. The dotted line represents the predicted values in the following weighted OLS regression:

$$\Delta Y_{cz}^{90-10} = \alpha + \gamma \cdot bin_{cz} + \epsilon_{cz},$$

where the left-hand-side variable varies according to the specification. Since the model does not target the average change in segregation, we shift the resulting values by a uniform factor, in such a way as to make the average for the first bin equal to zero, and explore the ability of the model to explain the differential change in segregation between cities with different knowledge shocks.

Figure A.5 shows the model performance in replicating the empirical relationship between the estimated knowledge shock (i.e. the predicted patenting growth, as in (1.27)) and the change in segregation between 1990 and 2010. The model replicates the empirical relationship closely: the slope of the regression line is 0.22 for the data, and 0.27 for the model. A weighted regression of the change in segregation in the data and in the model yields a coefficient of 0.13, which suggests a large correlation, even if the only perturbance in the model is the change in λ_{kk} prescribed by the bin.

The model is also successful in replicating the empirical relationship between knowledge shocks and change in occupational segregation (Table B.9). The model coefficient (0.60) is not significantly different from the empirical one (0.51). Table B.9 also clarifies that occupational segregation is one of the dimensions along which knowledge shocks translate into higher income segregation. As shown in the right columns of Table B.9, when controlling for the change in occupational segregation, the coefficients on income segregation drop by about a third in both the model and the data regressions. Since occupational segregation does not depend on changes in the level or the dispersion of income, this effect only translates into higher sorting, and does not appear in the inequality effect.

The model also captures the relationship between knowledge shocks and clustering of employment in knowledge intensive occupations. Table in B.18 Appendix replicates the results in Table B.6 using the bin value of the knowledge shock for the model counterfactuals (left column) and the data (right column). Neighborhoods with strong learning externalities in 1990 experience a more pronounced increase in the share of knowledge workers in high-bin cities rather than in low-bin cities. The coefficient of the interaction terms in the model regression is larger in magnitude than the empirical counterpart, but is consistent in sign and statistical significance. Notice that none of the quantities in Figure A.5 and Table B.9 and B.18 appear as a target in the structural estimation.

Figure A.6 shows the baseline change in segregation in the model simulation (red line) and the change in segregation that results exclusively from the reallocation of workers across neighborhoods following the shocks, keeping the average income by occupation for each neighborhood and occupation fixed at its original 1990 level. This measure captures the portion of the sorting effect that realizes along the occupational dimension, and translates in units of income segregation the occupational sorting observed in Table B.9. The slope of the blue line (0.11, compared to 0.27 for the red line) can be interpreted as a lower bound for the contribution of the sorting effect to the overall response of segregation to knowledge shocks.

1.4.7.1. Endogenous vs Exogenous Residential Amenities. Finally, we use the model to isolate the role of learning externalities and evolving residential amenities in driving the response of income segregation to innovation shock. Disentangling the relative importance of those two candidate factors is of crucial importance for the design of policies aimed at attenuating the rise in segregation, from the improvement of the transit system to changes in the provision of local public goods. As discussed in Section 1.4.6, the estimated values for residential elasticities suggests that the endogenous amenities generated by the concentration of residents in the creative sector are valued disproportionately more by residents of the same type. Emblematic examples may include high-quality schools, walkable areas, fitness centers or organic grocery stores. Hence, an initial shock to the distribution of residence - generated, for example, by a reshuffling of the distribution of employment - can be significantly amplified by the endogenous response of residential amenities.

Figure A.7 shows the results of a counterfactual experiment, in which residential amenities are exogenously given (in other words, $B_i^x = b_i^x$). This is equivalent to assume that the elasticities of residential externalities ($\omega_{x_1 x_2}$) are equal to zero. The resulting relationship is significantly flatter than the benchmark, suggesting that the amplification mechanism can be quite large. The coefficient of the regression in the counterfactual is 0.09, whereas the coefficient in the benchmark model is equal to 0.27. Comparing the two coefficients, we conclude that two thirds (roughly 66%) of the overall estimated impact of knowledge shocks on income segregation can be attributed to the amplifying effect of localized, occupation-specific residential amenities.[36]

1.4.7.2. Chicago Bids for Amazon HQ2. At the beginning of September 2017, Amazon announced its intention of adding a second North American headquarter to the one already existing in Seattle. By the end of the month more than 50 cities across the United States and Canada, including Chicago, had publicly considered to

[36]Another implication is that the relationship between the change in segregation in the data and in the model is flatter than in the model with endogenous amenities. A regression of the change in segregation in the model with exogenous amenities and the data yields a coefficient of 0.03, compared to the 0.13 of the full model.

submit a bid, for a total of more than 100 projects.[37] In this Section, we illustrate how our model can be used for policy experiments by assessing the impact of each bid on Chicago's city structure. Six projects, each in a distinct location, were deposited by city developers. The project located furthest north plans to redevelop the area by the river that is now occupied by the former A. Finkl & Sons steel plant, which was demolished in 2011. The second proposal for Amazon's HQ2 would be located a couple of miles South East towards the Loop, and it would be composed of four new buildings overlooking the river in the property owned by Tribune Media at 700 W. Chicago Avenue. Three other projects were proposed just (South-)West of the Loop: One in the Old Main Post Office; another plans to redevelop Union Station in different stages; and the last one would be just be just South of the loop. Being the 3 projects in a radius of less than one mile from each other, for the purpose of our counterfactual analysis, we only consider the one in the Old Main Post Office which lies in the middle of the two. The last project is the only one located in the South Side and would be built over the Michael Reese Hospital which ceased activity in 2009. Figure A.19 in the Appendix shows the exact location of the four projects considered in this simulation.

For this analysis, we first estimate the equilibrium quantities in 2010, and we then shock the exogenous term of productivities. The shock is calibrated to attract about 50,000 high-knowledge workers in the considered neighborhood. This number

[37] "Amazon refuses Arizona's cactus as bidders for HQ2 climb to 118," *The Seattle Times*, September 19, 2017. Map updated October 19, 2017.

matches the number of workers Amazon expects to employ in its second headquarter. Figure A.8 shows the forecasted change in high-knowledge residents in the four scenarios. The first panel considers the scenario in which the headquarter is located on the premises of the former Michael Reese Hospital, the second panel the scenario in which it is located in the Old Main Post Office, the third panel the Tribune Media scenario, whereas the last panel the former A. Finkl & Sons steel plant. There are two main trends that is possible to identify by comparing the four counterfactuals. First, in all the considered scenarios high-skilled high-salary workers tend to move in the high-amenities areas by the lake and downtown. Second, despite this general trend, the location of the headquarter seems to matter a lot for local development. On the one extreme, in the Michael Reese Hospital's scenario, high-knowledge workers start moving in the South Side. The areas around the University of Chicago and along the coast seem to be the most attractive. On the other extreme, in the A. Finkl & Sons steel plant's project, the majority of the gains are concentrated in the richer North Side. This is also reflected in the estimated changes in segregation: 1 Gini point in the former case, and 1.3 Gini points in the latter. According to our simulations the city would experience the highest change in segregation (1.5 Gini points) if the Amazon campus was located on the Tribune Media's property. Figure A.20 in the Appendix shows the change in high-knowledge workers in the four cases.

It is important to point out two caveats. First, some of these projects also include an expansion of the public transportation system. This might reduce the overall segregation, although it should not affect the local development results. Second, our model does not include a notion of migration. All the 50,000 high-knowledge workers attracted by the new campus all come from the commuting zone of Chicago. Taking migration into consideration might make the segregation effect worse, since rents in high-demand neighborhoods would increase more than in our counterfactuals.

1.5. Conclusion

We have shown that the rise of an innovation based economy is causally linked to the surge in income segregation experienced by U.S. cities in the last decades. Our instrumental variable results imply that local innovation trends are responsible for 56% of the cross-sectional variation, and 20% of the overall change in measured segregation. We have further showed that the estimated effect is driven by innovation in learning-intensive sectors (including IT and Electronics), and can be only partially explained as a consequence of an increasing dispersion of income.

Our interpretation relies on the view that local knowledge shocks (e.g. the development of new scientific insights that are relevant for local innovation) increase the returns from localized learning externalities, providing incentives for companies in knowledge-intensive sectors to cluster geographically. This in turn affects residential segregation, as workers in creative occupations relocate to live closer to their place of employment. But therein lies a powerful amplification mechanism, as the endogenous response of residential amenities, valued disproportionately by the creative class, makes the overall change in residential segregation more pronounced. A quantitative model of the internal structure of cities, estimated using detailed neighborhood level data on residence and employment in U.S. cities, predicts that as much as 66% of the overall effect can explained as the result of the endogenous evolution of localized, occupation-specific residential amenities.

The rise of the knowledge economy is profoundly changing the way we live and interact. The increasing economic divide in areas experiencing rapid growth in their innovative sectors has often been cited as one of the main challenges that advanced economies will need to face in their near future, as it brings about social unrest and political instability. Understanding its causes is a crucial step in properly designing policies aimed at confronting it, and making sure this secular shift happens in an inclusive way. Those suggested policies include improvements in the public transit system, supply of affordable housing, and change in the way public goods, such as schooling, are provided. Our quantitative framework, which combines state-of-the-art techniques from urban economics with newly constructed datasets on patenting and on the geographical distribution of creative occupations in the universe of U.S. cities, is especially suitable to study the effects of those policies. This is left for future research.

CHAPTER 2

The Geography of Unconventional Innovation

with Ruben Gaetani

2.1. 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). If true, 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 origin 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 reconciles 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 2016) 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 full-text record of all the patents granted by the USPTO in the years 2002-2014, georeferenced at the County Sub-Division (CSD, henceforth) level. At this narrow level of geographical disaggregation, the concentration of innovation activities in high-density areas appears to be smaller than commonly thought. Over 77% of the patents in our sample originate from geographical units with density below 1,600 people per square kilometer.[1] The relationship between patenting per capita and density of population is non-monotonic,

[1]In 2005, the density of population of San Jose-Palo Alto was 1,547 residents per square km. The share of patents produced in CSDs with lower density is 63.4%. As a comparison, in 2005, density of population is 26,407 in Manhattan; 7,175 in Boston; 6,514 in San Francisco; 5,588 in Chicago.

peaking around the density of San Jose-Palo Alto and declining for higher levels of density.

However, our disaggregated data reveal 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 tend to 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 get at 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 arriving firm. To the best of our knowledge, this paper is the first to provide direct evidence of inter-sectoral localized knowledge spillovers operating through this channel.

The facts that we document suggest an alternative 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 main city centers. One possible interpretation is that these companies can organize knowledge flows efficiently within the organization, 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, academic 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 one out of a set of scientific fields. 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 one 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 additional 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 wedges 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.

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

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. 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.

Another body of literature has investigated the implications of knowledge externalities for the spatial concentration of innovation, and its interaction with the geographical distribution of economic activities more broadly. Audretsch and Feldman (1996) find that, in State-level regressions, knowledge intensity in an industry is positively related to its geographical concentration of innovation, after controlling for the concentration of production. 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 levels (specifically, ZIP codes). Carlino et al. (2007) document a positive relationship between employment density and patent intensity across MSAs. Agrawal et al. (2010) use patent citations to shed light on the reason why, despite the well known advantages from innovating in dense and diverse cities, 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 a limited scope for them to have access to broad and diversified learning opportunities. 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, Porter and Stern 2014). Our main finding is broadly consistent with Packalen and Bhattacharya (2015), who find that over the last century newer concepts have been implemented in inventions originating from high-density regions.[3] 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 knowlege-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 their model, young firms locate in diversified cities to experiment with different prototypes, while established firms move to specialized sites where cost advantages are stronger. Davis and Dingel (2016) develop a model in which productivity in cities is

[3] Packalen and Bhattacharya (2015) find that throughout the last century, patents produced in more densely populated urban areas have made more intense use of newer *concepts*, identified as new sequences of words. 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.

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 2.2 introduces the dataset and presents new empirical facts about the geographical organization of the innovative activity in the United States. Section 2.3 introduces the model, characterizes its solution, highlights the mechanism, and studies its implications. Section 2.4 analyzes optimal place-based policies under fixed and flexible urban structure. Section 2.5 concludes.

2.2. 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.23 reports the total number of patents by filing year. There are several advantages to focusing on this recent sample. First, the recent digitization of the patent archive has made it easier for authors and reviewers to look for earlier patents to reference. 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 patent with socio-economic and demographic characteristics from the Census and the American Community Survey.

Every patent is assigned to one of 107 International Patent Classification (IPC) categories.[4] For each grant, we gather information on the identity and location of the original assignee, the inventors, as well as on the full list of cited patents (up to a maximum of 1,500 citations per patent). Every patent is geo-located following a hierarchical rule: If the patent file reports the name of an institutional assignee (e.g. a company, a research lab or an academic institution), we assign the patent to the geographical coordinates of its location; if it does not report any assignee or its address is missing or located outside the United States, we attempt to geotag the grant according to the location of its first inventor, otherwise of its second inventor and so on until we are able to assign a location to each patent. Note that we choose to use the location of the assignee, whenever available, instead of the address of the inventor. Most of the literature on the subject, since Jaffe et al. (1993) uses the location of the inventor. Both alternatives raise a number of issues. For example, when a patent lists multiple inventors whose locations are too far apart to suggest any interaction through spatial proximity, the address of the institution can represent a more accurate indication of the geographical origin of the invention. Many companies issue patents under several addresses, corresponding to different establishments or research facilities. The main concern with our approach is that

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

the address of the assignee sometimes represents the headquarters of the company instead of the research facility.[5] To address this concern, we run robustness checks using two separate geotagging strategies: (1) the sub sample of patents assigned to the address of the firm, *only* when it is in the same state of the address of at least one of the inventors and (2) all the patents located at the address of the inventor. We mention these checks several times throughout the text. We only consider patents that reference at least two citations. The main analysis is performed on a final sample of 1,058,999 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.[6] The CSD is finer than a county. It typically coincides with city boundaries and, in a few cases (e.g. New York City) a city can be partitioned in multiple CSDs. Since demographic data at this level of disaggregation are only available every 10 years, we interpolate the values of the demographic variables between 2000 and 2010 assuming a constant growth rate throughout the years.

2.2.1. Low-density areas produce a substantial share of patents

The literature on the geography of innovation has long emphasized the importance of density of population in determining innovation outcomes, and documented the

[5]Aghion et al. (2015) report a 92% correlation between the two locations at a State-level.
[6]The socio-economic and demographic indicators at a CSD level are available at https://nhgis.org.

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; Abel et al. 2011). A visual inspection of the geography of patenting in the U.S. confirms this intuition. The map in Figure A.30 shows the distribution of continuously innovative CSDs, defined as locations that filed at least one grant per year between 2000 and 2010.[7] There is a clear tendency for innovative activity to concentrate around main urban areas, highlighting a pattern that one would expect. For example, the East-Coast, the Chicago Area, the Texas Triangle and the Bay Area, among others, are all highly innovative regions.

The left panel of Figure A.21 displays a bin-scatter plot of the CZ-level empirical relationship between the logarithm of density of population (measured as residents per square kilometer) and patenting intensity (measured as patents per capita)[8] in the balanced panel of 742 U.S. CZs between 2000 and 2010.[9] At this level of disaggregation, the common intuition that density is associated with higher innovation intensity is confirmed. More densely populated CZs have higher patenting per capita, and the relationship appears to be monotonically increasing even in the right portion of the density distribution. The coefficient of the underlying regression implies that

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

[8]We winsorize this variable at the 1% level. Table B.24 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.

[9]To obtain the bin-scatter plot, we divide the variable on the x-axis into bins (typically 50 or 100 bins) - each containing the same weighted number of observations - 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.

doubling CZ-level density is associated with 0.10 more patents per 1000 residents. The right panel of Figure A.21 shows the same correlation only for the subset of the densest CZs that make up 50% of the U.S. population in 2005. The correlation remains positive and significant, and the underlying coefficient implies that doubling density is associated with 0.08 additional patents per 1000 residents.

The picture changes substantially when we narrow down the unit of observation to the CSD-level. Figure A.31 shows the distribution of continuously innovative CSDs in close-up maps of the four most densely populated metropolitan areas. Two less obvious observations emerge. First, a substantial part of patenting activity occurs away from main urban centers, often in low-density areas that are geographically separated from major cities (notably, Armonk, NY and Schenectady, NY). Second, even within major urban agglomerations, a big share of the innovative action takes place in their suburban portion (e.g. Schaumburg, IL and Mountain View, CA). Overall, low density regions seem to play a key role in the innovation process. About 77% of the patents filed between 2000 and 2010 originate from CSDs with 2005 density below 1,600 residents per square km (slightly above San Jose-Palo Alto), and about 63% from CSDs with 2005 density below 1,500 (slightly below San Jose-Palo Alto).[10]

The left panel of Figure A.22 shows the CSD-level empirical relationship between the logarithm of density and patents per capita,[11] Patenting intensity is no longer monotonically increasing in density, peaking at around 1,300 residents per square km

[10]About 12.2% of all the grants in the dataset are assigned to the San Jose-Palo Alto CSD.
[11]Again, we winsorize this variable at the 1% level.

(roughly the density of CSDs such as San Jose-Palo Alto, Austin and Raleigh) and declining for CSDs with higher density. The right panel of Figure A.22 include only the densest CSDs that make up 50% of the U.S. population in 2005. The relationship is weak and, if anything, decreasing, with the underlying coefficient implying that doubling density in the right portion of the distribution is associated with 0.02 less patents per 1,000 residents, although the estimated coefficient is not statistically different from zero.

2.2.1.1. Robustness. There are three major measurement concerns related to the interpretation of Figure A.22. 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 using density of population can bias the empirical correlation if units with a high density of skill-rich employment tend to have low overall density (as would be the case for places like Mountain View, CA, and Armonk, NY). Third, the choice of locating the address of the firm whenever possible raises the concern that a firm files for the patent in a location that is different from the one of the research facility.

To address these concerns, we look at two extreme cases. In the first case, we assume that all the relevant interactions only occur at the workplace. To attenuate the possibility of incorrectly assigning patents at the firm's headquarters instead of the research facility, we consider only the subset of patents for which the assignee is in the same state of at least one of the inventors. In this case, 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 A.32 reproduces the results of Figure A.32 by using density of workers and innovation intensity for this subset of reliably geo-located patents. To measure density of employment, we use data from the National Establishment Time Series, that contains close-to-universe information on establishments in the U.S., including industry indicators.[12] In the middle panel of Figure A.32, we use density of knowledge-intensive employment, 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. Although the relationship between density and patenting intensity is now tilted upwards compared to Figure A.32, the qualitative patterns are preserved, with patenting intensity peaking for intermediate levels of density and declining for places in the right portion of the density distribution.

In the second case, we assume that all the relevant interactions only occur the inventor's residence and its surroundings. This time, learning opportunities would be correctly measured by population density, but the patents issued to institutional assignees would be wrongly geolocated. In the bottom panel of Figure A.32, we replicate the analysis by geo-locating all the patents at the address of the first inventor. The patterns appear even more pronounced than in Figure A.32, with a significant negative relationship emerging in the right panel.

[12]See the Appendix of Berkes and Gaetani (2018) for details on the geolocation of each establishment, on the crosswalk between industry and occupation, and on the definition of knowledge-intensive occupations.

2.2.2. Dense locations produce more unconventional innovation

Once we account for a narrow level of disaggregation 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 internalize knowledge flows efficiently 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.

In this subsection, we show that 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 fact that a given pair of technologies is rarely referenced together denotes the lack of frequent knowledge transmission between the two.

2.2.2.1. Measurement. We now describe how we measure the degree of interconnection between two technological classes. We adapt the methodology proposed by Uzzi et al. (2013, UMSJ henceforth) who study atypical citation patterns in the universe of academic papers. To the best of our knowledge, this paper is the first to apply a similar algorithm to patents. The basic idea is to compare the frequency of a bundle of classes in the observed network of citations with the frequency one would obtain by assigning citations at random in a replicated network. In this process, the structure of the network is kept constant. In other words, 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.[13] 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 pseudo-randomly. We refer

[13]This is a departure from UMSJ that only keep the total number of citations from *and* into each class constant. 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 in 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 such 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 replicas.

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 algorithm are provided in Appendix A.

Figure A.35 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 support the validity of the measure. First, combinations on the diagonal tend to be more conventional than other citation pairs. This is exactly what we would expect: once a patent cites a certain class, it is likely to cite it again, since that class plays some role in the patent development. Second, around the diagonal we observe some "clusters" of conventionality. This happens because the IPC classification system 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 that group is likely to appear with another citation in the same group.[14]

[14]However, the c-score identifies technological proximity also between classes that belong to different IPC clusters. The following are some significant examples: Food (belonging to the Human Necessities cluster) and Sugar (belonging to the Chemistry cluster) have a c-score of 1.17; Butchery (Human Necessity) and Weapons (Metallurgy) have a c-score of 1.14; Decorative Arts (Printing) and Photography (Instruments) have a c-score of 1.15; Knitting (Textiles) and Brushware (Human Necessity) have a c-score of 1.84.

We assign to each patent an entire distribution of c-scores, one for each pairwise combination of references (hence, a grant with N references is assigned $\binom{N}{2}$ possibly identical scores). Two statistics of the distribution are of particular interest. The 10th percentile (or "tail-conventionality") proxies for the most unconventional pair of classes listed by the patent.[15] The median c-score (or "core-conventionality") proxies for how tightly grounded the patent is in prior knowledge. Figure A.33 plots the cdf of the core and tail-conventionality in our final sample. Consistently with the findings in UMSJ, it shows that the median patent is highly conventional at the core (its core-conventionality is well above one).

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

$$\text{logit}\left(Hit_{ict}\right) = \alpha + \delta_c + \delta_t + \beta \times UTail_{ict} + \gamma \times Core_{ict} \tag{2.1}$$

where Hit_{ict} is a dummy that takes value 1 if grant i is a hit patent, $UTail_{ict}$ is a dummy that takes value 1 if the tail-conventionality is below the median of class c in year t, $Core_{ict}$ is a set of 4 indicators denoting the core-conventionality quartile (in class c and year t), δ_c and δ_t are class and time fixed-effects respectively.[16]

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

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

Figure A.23 shows the joint marginal effects of the two variables on the probability of becoming a hit patent. The conditional probability ranges from 3.7% of a patent with a conventional tail and an unconventional core to 6.2% of a patent with an unconventional tail and a somewhat conventional core. By construction, the unconditional probability is 5%. Having an unconventional tail increases this probability by about 1.7 percentage points. On the other hand, the core seems to have a smaller impact. If anything, having an unconventional core *decreases* the chances of being a hit patent. Our results are very similar to the ones obtained by UMSJ for academic papers: scientific research with the highest impact appears strongly rooted in existing knowledge and at the same time displays the intrusion of novel combinations. This surprising similarity suggests that the process of innovation, no matter if academic or applied, follows a somewhat universal pattern.[17]

The strong correlation between unconventionality and technological impact shows that the c-score is ranking patents along a meaningful dimension. Motivated by this result, in what follows we will use tail-conventionality as our reference measure.

2.2.2.2. Finding. Here we explore the hypothesis that density plays the decisive role of catalyzing knowledge diffusion across unrelated fields. If this intuition is correct, we should observe that patents from high-density regions display more unconventional references. By facilitating interactions, density allows people to gain

[17]The fact that high-impact research is novel and, at the same time, tightly grounded, is explained at least in part by UMSJ by the necessity to efficiently deliver an idea to an inertial audience. For example - as mentioned in their paper - Charles Darwin's *On The Origin of Species* does not address the groundbreaking idea of natural selection until the second part of the work, the first part being entirely dedicated to a much more uncontroversial subject: the selective breeding of cattle and dogs.

insights they cannot acquire through their formal network. This translates into new ideas being obtained by assembling a more hybridized set of prior knowledge.

Table B.19 and Figure A.24 show several CSD-level correlations between (log) density of population (or college educated workers) and the tail-conventionality of the median patent filed in a given CSD/Year observation. For the purpose of this section, we limit the sample to continuously innovative CSDs, which gives us a balanced panel of 1,645 locations over 11 years, for a total of 18,095 observation. 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, an increase in density of population equal to the weighted residual inter-quartile range decreases tail-conventionality by 27% of its weighted residual inter-quartile range.

To study this relationship more in depth, we add to the specification various CSD specific controls, including (log) median income, the percentage of people with a college degree, inequality (measured by the Gini index). The results are reported in Table B.20. The effect of density on tail-conventionality stays negative and statistically significant. 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 (Gini index) both have a negative effect, but the coefficients are not statistically significant.

Table B.21 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.

This patent-level regression allows us to control for whether the patent is produced by a publicly traded firm. The results show that traded firms produce conventional innovation, which is consistent with the interpretation of unconventional innovation as creative destruction events. This is an interesting fact per se and would deserve further research.

2.2.2.3. Robustness. Table B.25 in Appendix shows that these results are not driven by any of the four most densely populated urban centers (New York City, Boston, San Francisco and Chicago). The bin-scatter plots in Figure A.36 repeat the robustness checks mentioned in Section 2.2.1.1. The top panel shows the correlation between tail conventionality and density of employment for the subset of patents in which the assignee and at least one of the inventors coincide at the level of the state. The middle panel reproduces the same result using density of knowledge-intensive employment. The bottom panel uses all the patents located at the address of the first inventor. All these alternative specifications yield consistent results.

Figures A.22 and A.24 show that density of population and innovation are indeed tightly related. 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 together suggest that the geographical sorting that we document can be explained as a result of the local interactions available in densely populated areas.

2.2.3. Dense locations are more technologically diversified

In this subsection, we show that dense cities tend to be more diverse in their innovation output. In particular, we use the concept of the 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 a set of technologically close fields.

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

Specifically, we consider all the pairwise combinations of patents filed in each CSD/Year bin. Each of these combinations is assigned the 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.[18] We then compute the median c-score of those combinations. This procedure delivers an index

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

of concentration for County Sub-Division CSD in year t defined as:

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

2.2.3.2. Finding. The bin-scatter plot in Figure A.25 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 -2.7, which implies that an increase in density of population equal to the weighted residual inter-quartile range increases diversification by 30% of its weighted residual inter-quartile range.

2.2.3.3. Robustness. Since the measure in (2.2) computes the median of a set whose cardinality grows at a binomial rate with the number of local patents, a possible concern is that CSDs with a higher number of patents (as it is typically the case with dense cities) mechanically have a low index of concentration. To address this possibility, we conduct a placebo experiment in which we generate 50 datasets identical to the original one in terms of total number of patents produced by each CSD/Year bin, but reshuffling the geographical allocation of individual patents at random. We then run 50 regressions of log-density on the simulated indexes of concentration. The resulting coefficients are plotted in Figure A.34. 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 0.04 and -0.099, with a mean of -0.037, two orders of magnitude smaller than the estimated coefficient on the original sample (2.7). This illustrates that the correlation in A.25 is not explained by this bias.

2.2.4. The local pool of ideas predicts local inventions

The key implication of Figure A.25 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 will be a necessary condition for generating unconventional patents. In this subsection, we look at the citation behavior of geographically close patenting firms to provide evidence of this class 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 choice. Places that produce (or are expected to produce) significant knowledge flows between two fields can be endogenously populated by firms belonging to those fields. For example, a company that aims to produce high-tech 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 (we run a robustness exercise considering only firms entering from 2005 onwards). Each incoming firm is assigned to the technology class corresponding to the most recurring class among its patents. Then, for each class-CSD-year observation, we construct an arrival shock as:

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

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.

The specification of the regression is the following:

$$S_{cdt} = \delta_{ct} + \delta_{dc} + \beta A_{cdt} + \epsilon_{cdt} \tag{2.4}$$

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.[19] Its unconditional average is 0.43%.[20] 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:

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

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 B.22. The estimates of β are always positive and statistically significant, as well as economically meaningful: the arrival of a firm producing exactly as many patents as $P_{d,2000}$ results in an increase in S_{cdt} equal in size to its unconditional mean (column 3). We also report results where we construct the shocks only considering incoming

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

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

firms that arrive in or after 2005 (column 4). The results are robust and larger in magnitude.[21]

2.2.5. 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. However, 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 show 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 an important determinant of innovation outcomes.

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

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 policies.

2.3. 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 coexistence 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. In its normative implications, the model highlights previously unexplored room 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.

2.3.1. Production and consumption

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

$$C = \int_{\mathcal{X}} c_i \,\mathrm{d}i.$$

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:

$$\max_{l_i} \pi_i \equiv l_i^{\beta} - wl_i \tag{2.5}$$

where w is the wage of unskilled workers and $\beta \in (0,1)$. In order for product line i to become active, one conventional *and* one unconventional idea directed to i must be combined. We denote by x the measure of active product lines (or, equivalently, the innovation rate).

Labor demand for active varieties is equal to:

$$l = \left(\frac{\beta}{w}\right)^{\frac{1}{(1-\beta)}} \tag{2.6}$$

while total labor demand in the production sector is equal to $L_F = x\, l$. Firm's profits are equal to:

$$\pi = \gamma\, w^{-\frac{\beta}{1-\beta}} \tag{2.7}$$

where $\gamma = \left(\beta^{\frac{\beta}{1-\beta}} - \beta^{\frac{1}{1-\beta}}\right)$. A fraction $a \in (0,1)$ of firm's profits is appropriated by the innovators responsible for discovering the corresponding variety, while 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.

In equilibrium, labor demand is constant across firms, which implies that total output (and total consumption) is equal to:

$$C = x^{1-\beta}\, L_F^{\beta}, \tag{2.8}$$

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

For expositional simplicity, we assume that all the intermediate varieties in the economy are high-tech devices (e.g. smartphones) that are obtained by combining a software component ($\mathcal{S}$) with a design blueprint ($\mathcal{D}$). The model easily generalizes to the case of multiple components or multiple sectors.[22] In order for a variety to

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

become active, an unconventional idea (that mixes software *and* design) must be combined with a conventional idea (*either* in the software or in the design). Conventional and unconventional ideas are matched via undirected search: letting ζ denote the aggregate mass of unconventional, and ψ the aggregate mass of conventional ideas, the resulting mass of active varieties is determined by the following matching function:

$$x = \zeta^{\mu} \psi^{1-\mu} \tag{2.9}$$

where $\mu \in (0, 1)$.

2.3.2. Economic Geography

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

2.3.2.1. 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 born either as a programmer ($\mathcal{S}$) or a designer ($\mathcal{D}$). For simplicity, we focus on the symmetric case in which the mass of designers is equal to the mass of programmers, so that $N_{\mathcal{S}} = N_{\mathcal{D}} = 1/2$.

Skilled and unskilled agents are fully mobile. Skilled innovators live in cities with positive rent. Unskilled workers live either in rural areas (close to production facilities) or in the outskirts of cities, and do not pay rent. There is a large mass of potential settlements. Each settlement has an area equal to 1, which implies that we can think of local population and local density interchangeably. These sites are owned by absentee competitive landlords, and governed by city developers,[23] 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.

In order to attract innovators, developers commit to provide type-specific subsidies, $\tau_{\mathcal{S}}$ and $\tau_{\mathcal{D}}$, to the research activity of local inventors. The subsidies are financed by taxing the 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.[24] Each skilled individual inelastically demands one unit of housing. Housing services are provided by competitive landlords, who face a local housing production function:

$$N^k = \left(L^k\right)^\alpha \tag{2.10}$$

[23]As in Becker and Henderson (2000).
[24]We denote skilled population of type $\mathcal{S}$ and $\mathcal{D}$ by $N^k_{\mathcal{S}}$ and $N^k_{\mathcal{D}}$, respectively.

where the parameters $\alpha \in (0,1)$ 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:

$$R^k = \frac{w}{\alpha} \left(N^k\right)^{\frac{1-\alpha}{\alpha}}. \tag{2.11}$$

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.

2.3.2.2. 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 three steps:

(1) Agents of type $\mathcal{S}$ living in a city with $N^k_{\mathcal{S}}$ innovators receives an idea with probability $\left(N^k_{\mathcal{S}}\right)^{\phi}$, where $\phi > 0$ controls the extent of the learning externalities. Similarly, inventors of type $\mathcal{D}$ living in a city with $N^k_{\mathcal{D}}$ peers receive idea with probability $\left(N^k_{\mathcal{D}}\right)^{\phi}$. 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.[25]

[25]This 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.

(2) Upon receipt of 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:

  (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 programmer (or designer) starts a search process in which he finds an innovator of the opposite type with frequency $N^k_{\mathcal{D}}$ (or $N^k_{\mathcal{S}}$). If search is unsuccessful, the idea is lost. If search is successful, the innovator makes her unconventional idea available.

(3) 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 $\mathcal{S}$ or $\mathcal{D}$). Letting ζ denote the total mass of unconventional ideas in the economy, and ψ the total mass of conventional ideas, a total mass of $x = \zeta^{\mu}\psi^{1-\mu}$ is formed. The resulting monetary value of the successful innovation is $a\pi$, where π is defined as in equation (2.7). The monetary value is split between the two innovators according to Nash bargaining, with the unconventional innovator receiving a fraction $b \in (0, 1)$ of the profits, and the conventional innovator receiving the remaining share $1-b$. Note 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.

Please note that the search process at point 2(b) is separated from the search process at point 3. The former allows an innovator in a generic town to be matched with an agent from the opposite background, and execute her idea unconventionally. The latter allows an innovator (of any background) endowed with a conventional idea to be matched with another innovator (of any background) endowed with an unconventional idea. Importantly, the process at point 3 occurs in an economy-wide marketplace that is unaffected by geography. If this matching process is successful, a product line becomes active and the two matched innovators split the resulting surplus ($a\pi$) according to the bargaining weight b.

It follows from the discussion that the probability that an unconventional (conventional) idea is turned into a successful innovation positively (negatively) depends on the ratio between the aggregate mass of conventional to unconventional ideas, $\kappa \equiv \frac{\psi}{\zeta}$. In particular, the probability that an unconventional idea becomes an active product line is equal to $\kappa^{1-\mu}$, whereas the probability that a conventional idea is executed is equal to $\kappa^{-\mu}$.

2.3.2.3. Utility and innovation rates. To save on notation, in what follows we conjecture that company towns will be *fully specialized* (i.e. they will host innovators of only one background). This conjecture will be proven formally in Proposition 2.3.1. Let $\mathcal{K}^G$ denote the set of generic cities, and $\mathcal{K}^C_{\mathcal{S}}$ (or $\mathcal{K}^C_{\mathcal{D}}$) denote the set of $\mathcal{S}$-specialized (or $\mathcal{D}$-specialized) company towns. The utility of an inventor of type

$\mathcal{S}$ living in city k can be written as:

$$U_{\mathcal{S}}^{k} = \begin{cases} \left(1+\tau_{\mathcal{S}}^{k}\right) \left(N_{\mathcal{S}}^{k}\right)^{\phi} N_{\mathcal{D}}^{k}\, b\, a\, \pi\, \kappa^{1-\mu} - R^{k} & k \in \mathcal{K}^{G} \\ \left(1+\tau_{\mathcal{S}}^{k}\right) \left(N_{\mathcal{S}}^{k}\right)^{\phi} (1-b)\; a\, \pi\, \kappa^{-\mu} - R^{k} & k \in \mathcal{K}_{\mathcal{S}}^{C} \end{cases} \tag{2.12}$$

with the one for type $\mathcal{D}$ being analogous, but with inverted indexes. In (2.12), the utility of an innovator of type $\mathcal{S}$ in city $k \in \mathcal{K}^{G}$ is given by the frequency of idea generation, $\left(N_{\mathcal{S}}^{k}\right)^{\phi}$, multiplied by the frequency of matching with a $\mathcal{D}$-type agent, $N_{\mathcal{D}}^{k}$, the frequency of finding a conventional idea on the economy-wide marketplace ($\kappa^{1-\mu}$), and the resulting share of profits, $ba\pi$, subsidized by the city developer at gross rate $\left(1+\tau_{\mathcal{S}}^{k}\right)$, minus the local rental price of housing, R^{k}. Analogously, the utility of an innovator of type $\mathcal{S}$ in city $k \in \mathcal{K}^{C}$ is given by the frequency of idea generation, $\left(N_{\mathcal{S}}^{k}\right)^{\phi}$, 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 $\left(1+\tau_{\mathcal{S}}^{k}\right)$, 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:

$$\psi = \int_{\mathcal{K}_{\mathcal{S}}^{C}} \left(N_{\mathcal{S}}^{k}\right)^{\phi+1} \mathrm{d}k + \int_{\mathcal{K}_{\mathcal{D}}^{C}} \left(N_{\mathcal{D}}^{k}\right)^{\phi+1} \mathrm{d}k. \tag{2.13}$$

$$\zeta = \int_{\mathcal{K}^{G}} \left[\left(N_{\mathcal{S}}^{k}\right)^{\phi+1} N_{\mathcal{D}}^{k} + \left(N_{\mathcal{D}}^{k}\right)^{\phi+1} N_{\mathcal{S}}^{k}\right] \mathrm{d}k \tag{2.14}$$

In (2.14), the rate of unconventional innovation is given by the integral over all the generic locations of the probability of arrival of ideas for $\mathcal{S}$-type innovators, $\left(N_{\mathcal{S}}^{k}\right)^{\phi}$, multiplied by the mass of $\mathcal{S}$-type innovators in city k, $N_{\mathcal{S}}^{k}$, and multiplied by the frequency of successful search for a $\mathcal{D}$-type innovator, $N_{\mathcal{D}}^{k}$, plus the corresponding product for $\mathcal{D}$-type innovators. In (2.13), the aggregate rate of conventional innovation is given by the probability of arrival of ideas for $\mathcal{S}$-type innovators in $\mathcal{S}$-specialized company towns, plus the same rate for $\mathcal{D}$-type company towns.

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

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

2.3.3. Equilibrium

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

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

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

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

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

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

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

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

$$\mathrm{Exp}^k = \begin{cases} \left[\tau_{\mathcal{S}}^k\,\left(N_{\mathcal{S}}^k\right)^{\phi}\,N_{\mathcal{D}}^k + \tau_{\mathcal{D}}^k\,\left(N_{\mathcal{D}}^k\right)^{\phi}\,N_{\mathcal{S}}^k\right]\,b\,a\,\pi\,\kappa^{1-\mu} & k \in \mathcal{K}^G \\ \tau_{\mathcal{S}}^k\,\left(N_{\mathcal{S}}^k\right)^{\phi}\,(1-b)\;a\,\pi\,\kappa^{-\mu} & k \in \mathcal{K}_{\mathcal{S}}^C \\ \tau_{\mathcal{D}}^k\,\left(N_{\mathcal{D}}^k\right)^{\phi}\,(1-b)\;a\,\pi\,\kappa^{-\mu} & k \in \mathcal{K}_{\mathcal{D}}^C \end{cases}.$$

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

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

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

$$\mathcal{V} \equiv \frac{b\,a\,\pi\,\kappa^{1-\mu}}{(1-b)\;a\,\pi\,\kappa^{-\mu}} = \frac{b}{1-b}\,\kappa \tag{2.15}$$

$$\mathcal{W} \equiv \frac{w}{(1-b)\;a\,\pi\,\kappa^{-\mu}} \tag{2.16}$$

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

(1) City developers optimally choose $\tau_{\mathcal{S}}^k$, $\tau_{\mathcal{D}}^k$, $N_{\mathcal{S}}^k$ and $N_{\mathcal{D}}^k$ and make zero profits

(2) ζ, ψ and x are defined as in (2.14), (2.13) and (2.9), respectively

(3) U is defined as in (2.12) and is equalized across types and active sites

(4) Firm's labor demand l and profits π are defined by (2.6) and (2.7), respectively, and total labor in production is given by $L_F = xl$

(5) L^k and R^k are defined as in (2.10) and (2.11)

(6) Labor markets clear: $\int_{\mathcal{K}} N_{\mathcal{S}}^k + N_{\mathcal{D}}^k \,\mathrm{d}k = N$ and $L_F = L - \int_{\mathcal{K}} L^k \,\mathrm{d}k$.

2.3.4. Characterization

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

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

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

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

In this problem, the maximand represents the developer's profits, while the 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 normalizing all terms by the expected returns on conventional ideas, $(1-b)\;a\,\pi\,\kappa^{-\mu}$.[26] 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 *diversified city* is

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

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

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

[26] This includes normalizing inventor's utility $\mathcal{U}=\frac{U}{(1-b)a\pi\kappa^{-\mu}}$.

Proposition 2.3.1. *In a symmetric equilibrium, city developers in company towns (C) and generic towns (G) set the optimal subsidy to:*

$$\tau^C = \phi \qquad \tau^G = 1 + \phi. \tag{2.17}$$

The optimal population in the two types of locations is:

$$\begin{aligned} N^C &= F^C \tfrac{1-b}{b}\, \kappa^{-1} \\ N^G &= F^G \tfrac{1-b}{b}\, \kappa^{-1} \end{aligned} \tag{2.18}$$

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 ($N^G_{\mathcal{S}} = N^G_{\mathcal{D}} = \frac{N^G}{2}$) 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 (2.15) and (2.16), 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 2.3.2. *A symmetric equilibrium exists and is unique.*

PROOF. The proof is constructive. See Appendix. □

2.3.5. Mechanism

Proposition 2.3.1 represents the model counterpart to Figures A.24 and A.25, that show the empirical correlation between density and conventionality of patenting, and concentration of the knowledge pool, respectively. The intuition behind Proposition 2.3.1 is that agents perceive an additional benefit from agglomerating in diversified cities compared to specialized clusters, and this induces them to trade off additional congestion costs and lower intra-field spillovers for the opportunity of having a higher exposure to inter-field interactions. To see this, compare the elasticity of the local externalities in a specialized company town with the elasticity in a diversified city. In the former case, it is 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 $\mathcal{W}$,[27] 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

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

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, $\left|\mathcal{K}^C\right|$ and $\left|\mathcal{K}^G\right|$, and their population density, N^C and N^G. These are in turn determined by the interplay between the agglomeration and the congestion forces in the model. The social cost of congestion 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 $\left|\mathcal{K}^{\mathcal{C}}\right|$ and $\left|\mathcal{K}^{G}\right|$, 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 (2.13 and 2.14) it is easy to see that the relative mass of company towns to diversified cities is a decreasing function of the parameter b:

$$\frac{\left|\mathcal{K}^{C}\right|}{\left|\mathcal{K}^{G}\right|} = F^{\mathcal{K}} \frac{1-b}{b},$$

where $F^{\mathcal{K}}$ 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 2.3.3. *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. □

2.4. Welfare and Policy

We now turn to studying 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 (2.8), we can decompose welfare additively as:

$$\log\left(C\right) = \underbrace{\left(1-\beta\right)\left(1-\mu\right)\log\left(\psi\right)}_{\text{Conv. rate}} + \underbrace{\left(1-\beta\right)\mu\log\left(\zeta\right)}_{\text{Unconv. rate}} + \underbrace{\beta\log\left(L_F\right)}_{\text{Congestion}}. \tag{2.19}$$

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.

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

The planner's problem reduces to the choice of the share $\eta \in (0,1)$ representing the fraction of innovators living in diversified cities:

$$\max_{\eta\in(0,1)} \; (1-\beta)(1-\mu)\log(\psi) + (1-\beta)\mu\log(\zeta) + \beta\log(L_F) \tag{2.20}$$

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

with $\left|\mathcal{K}^C\right|$ and $\left|\mathcal{K}^G\right|$ (the mass of company and generic towns, respectively) given.

Differentiating the problem in (2.20) with respect to η, and evaluating the first-order condition at the equilibrium, it is easy to see that the planner chooses to incentivize agglomeration and unconventional innovation if and only if the following condition is satisfied:

$$-\frac{(1-\beta)(\phi+1)(1-\mu)}{1-\eta} + \frac{(1-\beta)(\phi+2)\mu}{\eta} + \beta\frac{\partial\log(L_F)}{\partial\eta} > 0 \tag{2.21}$$

Since Proposition 2.3.1 implies that $N^G > N^C$, and due to the concavity in the housing production function, the third term of condition (2.21) is negative. Hence, the planner can decide to pay additional congestion costs to increase the share of skilled labor in diversified 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 A.26 displays the value of condition (2.21) for a simple parametrization of the model[28] and for b spanning between 0 and 1. For sufficiently low values of b, 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 (red line), reduce the supply of conventional ideas (blue line) and increase congestion costs (green line). The opposite policy is implemented if b is sufficiently high, 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 A.26 displays the contribution of the components in (2.19) to the improvement of aggregate welfare for the same range of values of b.

[28]We 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 A.26 holds irrespectively from the value of these parameters.

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 and innovation, where social returns to innovation are not fully captured by innovators and other local agents, and, as a result, the market outcome provides 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, this kind of efficiency gain originates 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.

2.4.2. 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 lamp-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 in affecting 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 $\left\{T_{\mathcal{S}}^k, T_{\mathcal{D}}^k\right\}_{k\in\mathcal{K}}$ 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 $\left\{T^G, T^C\right\}$ for diversified and specialized cities, respectively. Given this policy choice, the resulting equilibrium can be found as in Proposition 2.3.1, with the inventor's income now augmented with the multiplicative transfer. For a given choice of transfers, the resulting geography has the following solution:

$$\begin{cases} N^C = & \frac{1+T^C}{1+T^G}\, F^C\, \frac{1-b}{b}\, \kappa^{-1} \\ N^G = & \frac{1+T^C}{1+T^G}\, F^G\, \frac{1-b}{b}\, \kappa^{-1} \end{cases} \tag{2.22}$$

Figure A.27 plots the contribution of the three components in (2.19) 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 two extreme values of the bargaining weight of the unconventional innovator ($b = 0.2$ and $b = 0.8$). Figure A.27 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 pattern would emerge in an analogous model that only allows for one type of ideas. What is peculiar to our 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 (namely, even with incomplete appropriability), provided that the bargaining weight of the unconventional innovator b is sufficiently high (red line, 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 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 A.28 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 two extreme values of the bargaining weight ($b = 0.2$, left panel, and $b = 0.8$, right panel). The welfare gains are significantly larger under a flexible structure. Under the baseline parametrization ($a = 0.5$), the welfare gain from the optimal policy under a fixed urban structure are equal to 5.7% of consumption, against 8.69% under a flexible structure when $b = 0.2$. The difference is even larger when the bargaining weight of the unconventional innovator increases to $b = 0.8$, with the gain under a flexible structure increasing to 15.2%, against 5.86% achievable under a fixed structure.

2.5. Conclusion

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

We integrate these findings into a model of heterogeneous innovation and spatial sorting. In our setting, the choice between producing conventional and unconventional ideas depends on their relative price and, crucially, on the local degree of density and diversification. In equilibrium, low-density specialized cities coexist with high-density diversified ones. This asymmetry is dictated by the complementarity of unconventional and conventional ideas in the innovation process and does not depend on the existence of agents with ex-ante heterogeneous productivity. The composition of innovation determines the balance between rate of innovation and congestion costs, which in equilibrium is suboptimal. Our analysis reveals that whether the planner has some flexibility in adjusting the urban structure makes a big difference in determining the welfare benefits from place-based policies. This supports the widespread idea that a fully mobile skilled labor force can be 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.

CHAPTER 3

Comprehensive Universe of U.S. Patents (CUSP): Data and Facts

3.1. Introduction

Patents have been the main source of data for empirical studies on innovation and technological change. Despite being an imperfect proxy for technological input and output,[1] the fact that patent data are easily accessible, offer a wide range of information about the invention content and the underlying innovation process, and are available for a large number of developed countries has contributed to their popularity in the literature. With some notable exceptions (e.g., Nicholas, 2010), until recently, research papers on the topic have mostly focused on the past 50 years. Similarly, historical analysis has concentrated on relatively small time frames (e.g., Moser, 2005) or on specific dimensions of patents data. The likely underlying reason is the lack of a reliable source of data for historical patents. In fact, the U.S. Patent and Trademark Office (USPTO) provides detailed data for all the patents issued from 1976 on, and studies on innovative activities prior this year often required the collection of data by hand.

[1]For example, Moser (2005) uses data from two World's Fairs at the end of the 19th century and shows that inventors from countries without patent laws focused on sectors that relied more on secrecy than patenting. This suggests that, at least in that period, patenting activity was skewed towards a certain set of industries.

More recently, thanks to the availability of increasingly reliable Optical Character Recognition (OCR) software, cheap computational power, and the publication of high-quality scan of historical patents by the USPTO various scholars started working on historical patent data. Notable examples are Akcigit et al. (2017), Sarada et al. (2017), Packalen and Bhattacharya (2015), and Petralia et al. (2016). The first two match patents data to the recently released decennial Census data and therefore mainly focus on the decades between 1880 and 1940. Packalen and Bhattacharya (2015) study the importance of physical proximity for innovation throughout history. To do so, they extract the name of the city, or county, from the text of each patent and study how the tendency of using new ideas in inventions changes with population density. Finally, Petralia et al. (2016) digitalize the images provided by the USPTO and extract information about the county of residence of inventors and assignees. The parsing of the text is supplemented with some machine learning techniques, such as neural networks, that are used to measure the plausibility of the collected data, as well as to infer the values of missing observations.

Despite the important contribution of these papers, different data sets contain different sets of variables and cover different time frames. Moreover, not all the data collected in these projects are readily available to other researchers. The aim of this paper is to fill this gap. Three years ago, I started working on a newly assembled data set of historical patent. The idea was to collect all the variables that are commonly used in the innovation literature using a consistent methodology and data sources and share the result with the rest of the community. Traditional sources, such as the USPTO and Google Patents, are complemented by newly digitalized patent documents and an extensive use of fuzzy matching is employed to extract information about the patent itself (e.g., technology classes, filing year, and backward citations), as well as about inventors and assignees. Each inventor and assignee is geolocated at the town level, the most disaggregated geographical level that is possible to identify from the patent text. The outcome is what I called the Comprehensive Universe of U.S. Patents (CUSP). It spans almost two centuries of patent data (1936-2015) and contains the richest set of variables available so far. Various sanity checks show a high degree of accuracy.

The first part of the paper describes in details the data sources and the techniques used to extract the data. I also compare CUSP with HistPat (Petralia et al., 2016), one of the most promising data sets of historical patents readily available on the Harvard Dataverse. The analysis shows a broader coverage of CUSP and a similar level of accuracy in terms of geolocation of the patents, the dimension that is most stressed in HistPat. In the second part, I report some stylized facts. Some of these are new and might point to interesting directions for future research. Some others confirm well-known patterns already discussed in the literature (e.g., the upward trend in the average number of inventors per patent is already described by Wuchty et al., 2007). Nevertheless, this new data set offers for the first time a long-term perspective and allows us not only to observe trends but also to pin down when the trends started in history.

The rest of this paper is structured as follows. Section 3.2 describes in details the data contained in the CUSP and how they were assembled. Section 3.3 briefly compares the CUSP with some other historical patent data sets. Section 3.4 provides some stylized facts that might be source of inspiration for future research. Section 3.5 concludes.

3.2. Data

The data set collects a comprehensive set of variables for the entire universe of patents issued by the USPTO between 1836 and 2015. To do this, I use five distinct data sources:

(1) Patent text and information reported on the USPTO website;[2]
(2) State-, or in one case city-, level databases. Such databases are usually maintained either by universities or public libraries and contain all the inventions a (not always) comprehensive list of the patents whose inventor was resident in that state (or city). In many cases, these databases only cover historical patents. I was able to identify seven local inventors databases:
    (a) Cincinnati Inventors Database
    (b) Iowa Inventors Database
    (c) Nevada Inventors Database
    (d) Oklahoma Inventors Database
    (e) South Carolina Inventors
    (f) The Portal to Texas History

[2] http://patft.uspto.gov/netahtml/PTO/search-adv.htm

   (g) Wyoming Inventors Database;

(3) High quality patent images digitalized with an OCR software;

(4) Google Patents;

(5) Patents issued after 1920 digitalized by Google and made available on the USPTO website. This was the first attempt made by Google to OCR historical patents and the result is generally of poor quality. Nevertheless, these data are used as a last resort in case it is not possible to extract the needed information from the previous sources.

Using multiple sources reduces the probability that the information I am looking for is not available in any of them, and allows me to select the most reliable one. Given the peculiarities of each database in terms of the degree of accuracy and data availability, the database of choice is based on the year in which the patent was issued and the piece of information I am trying to collect. First, since the USPTO makes readily available all the information for the patents issued after 1976, their website is my preferred source for all the patents issued after that year. Additionally, from there it is also possible to collect information about the technology classes for all the grants going back to 1836. Second, I use the local inventor databases to extract information about inventors and assignee and their town of residence for all the patents listed there. These data have a limited coverage in terms of space and time, but the information contained in the local databases is reliable and easy to extract. Finally, I parse the patent text obtained from the three digitalization processes (mine, Google Patents, and USPTO) for all the remaining variables and patents. It was not possible to extract all the pieces of information for the universe of patents, but all the patents are listed in each table.

The rest of this section describes in more details the variables available in CUSP and the strategy employed to extract them.

3.2.1. Issue and Filing Years

A patent's issue year is readily available from the USPTO website for all the patents ever granted. The same is not true for the year in which the patent was filed. This piece of information is often missing for historical patents. However, filing years are arguably a better indicator of when the invention was completed than issuing years.[3] When not available from digital sources, it is possible to retrieve the date in which the patent was filed directly from the patent text starting from patent number 137,279 and issued on April 1, 1873. The filing date appears in the patent header preceded by "application filed on". Figure A.38 shows the header of this patent. The parsing process follows two increasingly less stringent steps. First, I look for sequences of exactly four numbers preceded by the words "application", "filed", "tiled", "fiied", or "fledi"[4] and followed by a month or its abbreviation (e.g., january or jan).[5] Second, if this procedure is not successful, I look for sequences of exactly four numbers that are on the same line as the keywords listed above.

[3]Figure A.46 in Section 4 shows that the up until the 19th century the issue and filing years were very close, with an issuance time of less than one year for the average patent. However, the two kept diverging until the late 40s when on average a patent had to wait 4 years before being issued.
[4]"tiled", "fiied" and "fledi" are common mistakes made by the OCR software when reading "filed".
[5]Note that before this process, I substitute all the occurrences of "l9" ("el" followed by a nine) and "—9" with "19". Similarly, for "l8" and "—8".

Since the likelihood of error is different for each of the two steps, each observation in the data set is assigned a flag that will help researchers to understand how confident we should be with the value reported. The flag is set equal to 1 if the filing year comes from the USPTO (or Google Patents) website; 2 if the filing year was obtained through the first round of parsing; 3 if it was obtained by searching for sequences of four numbers appearing on the same line as the keywords "application" and "filed" (and its variations). At the end of this process, the filing year of 8,178,429 patents (or 93.2%) was obtained from an official source, 446,184 (or 5.1%) from the first round of parsing and 88,270 (or 1.01%) from the second round.[6]

Finally, issue and filing years are checked for consistency. If the first two digits are a 9 and a 1, respectively, I swap them;[7] if the issue year is outside the time frame of the dataset (1790-2015), then I replace it with a missing value; if the filing year is outside the time frame of the dataset, is larger than the issue year, or the difference between issue and filing years is bigger than 30, then I set it to missing value. In the end, issue and filing years are available for a total of 8,712,883 patents (or 99.3%).

3.2.2. Technological Classes

Technological classes are assigned to each patent by patent reviewers. The USPTO regularly updates class definitions and corrects the classification of patents backwards. Each patent is associated with multiple technology classes according to three

[6]Note that the percentages here is calculated using 8,772,775 as denominator, that is the total number of patents minus the number of patents for which the filing year is unknown since it is not reported anywhere in the text (i.e., patents whose number is smaller than 137,279).

[7]Swapping the first two digits is a relatively common typo in patent documents.

different classification systems: the U.S. Patent Classification (USPC), the Cooperative Patent Classification (CPC) and International Patent Classification (IPC) schemes. Technology classes are indicative of the pieces of knowledge embedded in the invention patented. For example, a patent that describes an image processing method for TVs might be classified under the US patent categories 382 (Image Analysis) and 348 (Television). Documents classified in the USPC system are also assigned one (and only one) principal class. The principal class captures the scope of the invention as a whole or the main inventive concept using the claims as a guide (USPTO, 2012). The CPC and the IPC do not specify any main technological category. If a patent cannot be classified within the current classification system, or its principal class is unclear at that point in time, then it is assigned a 1/1 as main class. Since the USPTO reviews and updates its classification system every couple of months, the total number of 1/1 patents is quite small (15,819 patents or 0.2% of my sample).

In my data set, I collect technology classes for all the three classification schemes directly from the USPTO website.[8] Classes were collected in June 2016 and therefore represent the classes assigned at that point in time. The data set contains two tables that report the technology classes according to the USPC. The first contains the classes as they were made available by the USPTO. In the second, I assign a main class to the patents whose principal class is 1/1 based on the frequency of its

[8]See for example http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PALL&S1=0137279.PN.&OS=PN/137279&RS=PN/137279 for patent number 137,279. The principal class according to the USPC is in bold.

secondary classes (at a three-digit level). When two or more secondary classes appear with the same frequency, the one with the smallest number is selected. For example, if a patent was assigned classes 1/1, 324/12, 324/121, 345/67, 345/87, then 1/1 is substituted with 324.[9]

3.2.3. Backward and Forward Citations

Backward and forward citations have been extensively used in the empirical literature to understand knowledge flows across firms and inventors, as well as as a measure of patent quality, the idea being that the more a patent is cited the more the invention it describes is valuable.[10] That is why a patent data set would not be complete without this piece of information. For each patent in the dataset, I collect the patent number of all the U.S. patents referenced in the grant. Once the backward citation matrix is populated, it is possible to obtain the list of forward citations, that is the patent numbers of all the patents that cite a certain invention, simply by "inverting" that matrix. Starting in 1947, all the patents issued by the USPTO include a section that lists all the references cited.[11] Before that year, prior art upon which the invention was built was reported on the file history which is not publicly available. Nevertheless, some patents were directly referenced in the patent text and it is therefore possible to get a sense, albeit noisy, of knowledge flows across technology fields and regions.

---

[9] Note that even after this procedure some patents are assigned class 1/1. This is due to the fact that they do not report any additional class making the frequentistic procedure impossible.
[10] Alcácer et al. (2009) discuss the merits and demerits of using citations in empirical work.
[11] The first patent to include a "References Cited" section is Patent Nr. 2,415,068.

The data collection strategy for backward citations crucially depends on when the patent was issued. For patents issued after 1976, citations are directly collected from the computerized patent information available on the USPTO's website. For patents issued between 1947 and 1975 I parse the text and extract the contents of the section titled "References Cited" that lists the references to other patents or, in some cases, scientific articles. From this section, I collect the number of all the U.S. patents cited. Finally, for the patents issued before 1947, I look for references to other patents directly in the patent text. In particular, I look for sentences that contain the keywords "patent" or "patents" followed by "no", "number", "numb", "num", "nos", or "numbers" and get the patent number referenced afterwards. Figure A.39 shows an extract of patent no. 46,101 in which the inventor describes how his patent differs from a previously issued patent and states his claims. This strategy finds a total of 182,044 patents cited by patents issued before 1947. I apply the same strategy for all the post-1947 patents that do not include a "References Cited" section.

The two tables that contain backward and forward citations are structured with a long form. The first column contains the number of the citing patent, whereas the second column the number of the patent cited. Each line correspond to a single citation. A patent that cites multiple grants will appear on multiple lines. The table containing backward citations has two additional columns. The first is a binary variable that takes value 1 if the citations was added by the examiner, as reported by Google Patents. The second column contains a flag that take value 1 if the citations are collected from a digital source (i.e., either Google Patents or the USPTO website);

value 10 and 11 if the citations are obtained from the "References Cited" section of patents OCR'ed by the USPTO and by myself, respectively; value 5 if the citations come from the main text of the patent as in Figure A.39.[12]

3.2.4. Inventors Name and Location

The collection of the inventors names and locations is the most challenging and sensitive task. For this reason, particular attention was devoted to this phase of the data collection. Fuzzy matching techniques are employed to overcome some of the problems that occur due to the fact that the performance of OCR programs heavily relies on the original image quality and sometimes the digitalized text displays various typos. As for backward citations, use the information available on the USPTO website to extract the name of the inventors and their residence for all the patents issued after 1976.[13] The maximum number of inventors in a single patent in the sample is 76. This is the number of inventors of grant number 7,581,231, a software patent filed by Microsoft.

For patents whose patent number is smaller than 1,583,767, I use a three step approach to collect the relevant information. First, I parse the end of the patent and identify the inventors' signatures (in print). Figure A.40 shows the very end of patent number 580, a bee hive. The name of the inventor is reported in capital letters together with the name of two witnesses. The fact that signatures are printed

[12]This step only uses the patents OCR'ed by myself since the quality of the text is generally superior.
[13]Note that the USPTO reports the details of some patents even before 1976. In this case, all the details are collected from there.

in capital letters and with a larger font minimizes the amount of typos during the digitalization process. Second, I parse the patent header (Figure A.38) looking for the residence of the inventors identified at the end of the grant. The patent header is characterized by the keywords "United States Patent Office" or "assignor" (and some variations that take into account frequent typos), whereas the inventors' location is extracted by looking for keywords like "of" or checking weather the name of a state is contained in the header string. Third, if the code is unable to extract the information from the header (for example because none of the keywords listed above is present), then I parse the beginning of the patent text (Figure A.41). All the patents prior to 1,583,767 start with a formula similar to the one in the Figure: "*To all whom it may concern:* Be it known that I, <inventor name>, residing at <name of city>, in the county of <name of county> and State of <name of state>". By searching for this pattern, it is possible to extract the city and state of residence of a certain inventor. This technique is however left as last resort, since parsing the patents text is prone to more typos than the header (which is written in a bigger capitalized font), and the formula changes from time to time, making the pattern matching task more difficult.

The strategy used to extract inventors names and locations for the patents whose number is between 1,583,767 (included) and 1,920,165 (excluded) is similar to the one above, with the exception that the third step had to be dropped, since the patent text does not contain information about the inventors anymore. This is the same strategy employed to extract information for all the patents issued after patent 1,920,165. However, for these the parser needed to be modified to take into account the new structure of the header (e.g., the keywords used to identify the header are different). Note that for the majority of patents issued starting from the end of the 19th century, the name of the inventors is readily available from Google Patents. In that case, the name of the inventors is taken from there and the steps described above are used for the sole purpose of getting information about their residence.

Possible typos in the location names are then fixed by using a frequentistic approach. First, I count how often a city/state pair appears in the data set. Second, I iterate over all the inventors in the data set and compare the reported location with those in the previously built dictionary. If the dictionary contains a city in the same state with a Levenshtein distance of 2 or less that appears more frequently in the original data set, then I assign that city to that inventor. Similarly, if the data set contains a city/state pair with a Levenshtein distance of 1 or less for both the city and state (e.g., Chicago, IL and Chicag, HI) and a higher frequency then I assign to that inventor the more recurring pair.

Finally, when no location is reported for an inventor, I check in the previous and if there is an inventor whose location is not missing with the same name and who filed a patent one year before or after. If that is the case I assign the location of the latter to the former.

3.2.5. Assignees Name and Location

Extracting the location and name of assignees from the patent documents is a more straightforward task compared to extracting information about inventors, but also one that is prone to more mistakes. In fact, there is no redundancy in the documents: details about the assignees appear one and only one place: the header of the patent. Using the same procedure developed for inventors, I identify the header of each patent and check for the presence or absence of the string "assign". If this sequence of characters is not contained in the header, then I conclude that the patent has no assignees, otherwise I parse the rest of the line searching for the name and location of the assignees. Unfortunately, their location is not always available: sometimes only the assignee name is reported, while other times the assignee name is followed by "a corporation of <name of state>" without any further detail. From a careful review of a number of patents, it seems that the state reported there represents where the firm is registered, and does not necessarily indicate the location of the branch where the inventor works.[14] For this reason, when either the location is missing or the assignee name is followed by "a corporation of ¡name of state¿" without any reference to the city where the assignee is actually located, I assign the company to the same location of the first inventors, when they are all reported to live in the same location.[15] This approach biases the distance between inventors and assignees towards zero. Some of the facts reported in Section 4 should therefore be interpreted as a lower bound.

Similarly to what I did for the inventors, I fix possible typos using a frequentistic approach and missing values looking at assignees with the same name one year before and after the filing year of the patent.

[14]For example patent number 1,898,054 is assigned to the National Lead Company of New York, N.Y., a corporation of New Jersey.

[15]In the next iteration of the data set, I will flag these instances to allow researchers to drop them for robustness checks.

3.3. Validation and Comparison

An in-depth comparison with other data sets of historical patents is beyond the scope of this paper. The interested reader is referred to Andrews (2017) who presents a newly assembled data set of geo-referenced historical patents and, in doing so, he compares it with other four existing data sets.[16] However, to give credibility and motivate the data collection effort, it might be useful to contrast CUPS with HistPat, a very renowned publicly available data set of historical patents described in Petralia et al. (2016) and available on the Harvard's Dataverse. Table B.26 schematically shows the variables available in the two data sets.[17] Figure A.42 compares the number of patents contained in the two data sets and the actual number of patents reported by the USPTO by issue year. The dashed yellow line shows the official number of patents issued by the USPTO in each year. The total number of patents in CUSP almost perfectly match this series.[18] Since HistPat seems to only include patents for which all the inventors and assignees are located in the U.S., the red line shows the number of patent that satisfy this requirement in CUSP, whereas the green line represents the number of patents in HistPat. The difference between the two series is always relatively small except for the period between the two World Wars when HistPat systematically covers less patents. Although CUSP contains a

[16]Unfortunately, it was not possible to gain access to the data (or their aggregate statistics) underlying the work of Akcigit et al. (2017) and Packalen and Bhattacharya (2015). He therefore excludes them from the analysis in the present version of the paper.
[17]The rows and first column of the table are taken from Andrews (2017) and reported here upon the generous agreement of the author.
[18]Some minor differences are due to the fact that some patents were withdrawn after being issued. Those patents are discarded from my data set.

larger amount of variables and patents, Petralia et al. (2016) put a lot of intellectual and computational effort in identifying the county of residence of each inventor and assignee listed in the patent. The real test is therefore to compare the two data sets along a geographical dimension.

When multiple inventors and assignees are reported on a patent, HistPat assigns to that grant multiple locations without giving any information about whose residence is the one reported. For comparison purposes, I therefore extract all the counties assigned to the assignees and inventors of a patent and compare them with the ones listed in HistPat. If CUSP reports all the counties reported by HistPat for a given patent,[19] I count that as a success, otherwise the patent is categorized as a non-match. The resulting matching rate is about 80%. The exercise includes all the patents issued in the period 1836-1976 and available in both data sets. Figure A.43 reports the share of non-matched patents by issue year. The share remains quite stable around 20% over the whole period with a peak of about 35% in 1919-1920. Analyzing by hand a random sample of the patents not matched shows mixed results. Sometimes CUSP contains the right location of the assignee but the wrong location of the inventor (or viceversa), whereas HistPat contains the wrong location of the assignee but the right county of the inventor (or viceversa); sometimes CUSP is off track and other times HistPat is off track.[20] From time to time the mismatch is due to the fact that while HistPat reports the county stated in the text, CUSP reports

[19]Note that in some cases CUSP actually contains more entries than HistPat.

[20]Note that sometimes patents contain contradictory information. For example patent number 9 states in the header that the inventor is from New York, which is what is reported in CUSP, whereas in the text the inventor is from Springfield, MA, which is what is reported in HistPat.

the county that contained the town in 2000. For example, when patent number 48 was granted Portsmouth, VA was part of Norfolk County, whereas nowadays is an independent city. I include in the data set a list of the patent numbers that, according to the procedure described above, do not match in the two data sets. The list will provide guidance on where to concentrate my efforts for the next iteration of the data set.

3.4. Stylized Facts

3.4.1. Numbers

Fact 1.1: Patenting activity in the U.S. has steadily increased over time; the growth started accelerating in the 80s. The number of patents filed at the USPTO has experienced an important acceleration starting in the 80s. This trend seems to be mainly driven by two factors. First, the number of U.S. patents that had been decreasing since the 60s shows a dramatic reversal of the trend in that decade. The change might be due to the growing importance of software patents. Second, the number of foreign patents also accelerated in those years, although the upward trend started already in the 50s. Figure A.44 plots the total number of patents issued by the USPTO according to their filing year and country of residence of their inventors. The blue line represents the total number of patents by filing year, whereas the red and green lines show the patents whose inventors are foreign or a U.S. residents, respectively.[21] The graph highlights two additional interesting facts. First, the share

[21] Note that the graph excludes what in Fact 1.2 I define international collaborations, so the green and red lines do not necessarily sum up to the blue line. However, as it is shown in Fact 1.2 the

of foreign patents before the 60s was almost negligible. Second, in 2010 the number of patents whose inventors are foreign residents passed the number of patents filed by inventors whose residence is in the United States.

Fact 1.2: The share of patents resulting from international collaborations started increasing in the 50s. An international collaboration is defined as a patent for which at least one inventor is a U.S. resident and at least one other has her residence outside the United States. The number of international collaborations has importantly increased over the years with a steady growth that started in the 80s. Despite this, international collaborations still remain a small fraction of the total number patents filed at the USPTO. Figure A.45 shows this pattern graphically. In 2010, less than 5% of the patents filed were the result of international collaborations.

Fact 1.3: The time needed to issue a patent was negligible in the 19th century; it was on average 2-3 years in the 20th century. Since information about the year in which the grant was filed is often absent in data sets of historical patents data, it is common practice in the literature to proxy the filing year with the year in which the patent was granted. Authors often argue that in the past the time necessary to examine a patent was shorter due to the smaller amount of applications and their relative simplicity (see for example, Akcigit et al., 2017). Figure A.46 tests this hypothesis. The average issuance time for patents filed before 1900 was indeed below one year, but it was already almost 2 years by 1915 and more than 2.5 years in the 1920s. The average issuance time experienced an important increase

share of foreign collaborations is small. The main patterns of the graph do not change if I used the country of residence of the first inventor to classify patents into U.S. and foreign patents, instead.

during WWII reaching 4 years in 1947 and gradually went back to about 2 years in the period between the 1970s and the 2000s, when it started rising again to reach another peak in 2005 when the average patent had to wait about 4 years before being issued. The decrease at the end of the sample might be due to the decrease in the number of applications received by the USPTO during and after the Great Recession, or might be simply due to truncation problems (patents filed in 2010 with an issuance time larger than 5 years do not appear in the sample).

3.4.2. Inventors

Fact 2.1: The share of single-authored patents was around 80% up until 1920 when it started declining. Single-authored patents have become increasingly rare in the past century. Figure A.47 provides evidence for this fact from two different angles. Panel a shows the share of patents filed by a single inventor, whereas panel b the average team size by filing year. The share of patents filed by a single inventor has steadily decreased over time since the end of the 1920s. In the 19th century between 70% and 80% of the inventions patented were single-authored. By 2010, this share had decreased below 20%. Similarly, average team size remains surprisingly stable, around 1.2 inventors per patent, up until the late 40s when it starts a rapid increase. In 2010, the number of inventors for the average patents is about 2.7, more than double compared to 60 years before. Wuchty et al. (2007) document this pattern for the period (1975-2000). Thanks to the larger time frame covered by CUSP, it is possible to put this finding into a historical context and pin down the moment when the shift happens. Interestingly, de Solla Price (1963) documents that the cost of research as a share of GDP did not increase before WWII when it started and exponential growth. It would be interesting to understand what factors have driven the decline of single-authored patents which started in the 1920s and accelerated in the late 1940s, and if technology fields contributed differential to this trend. This is left for future research.

Fact 2.2: Average and maximum distance among the inventors of the median patent started an upward trend in the 50s; minimum distance increased at first and then plateaued. An important idea in the innovation literature is that the decline in communication costs have made the collaboration with people living in other cities or countries less costly and hence proximity less important. Consistently with this intuition, Packalen and Bhattacharya (2015) find that inventors in more dense cities were adopting ideas faster throughout the 20th century, but the advantage of living in a large city has disappeared more recently. However, this insight appears to be in contrast with other observations, such as the existence of large innovation hubs, or of seminars and conferences that allow scholars to personally discuss with their peers.

A possible explanation to these two seemingly contradictory facts is that proximity still matters at the very beginning of a project and for certain specific tasks. For example, informal exchanges of ideas might play a crucial role in first stages of a project and proximity might be important to, say, analyze and brainstorm about the outcomes of lab experiments. If this was the case, I would expect to observe an increase in the geographical dispersion of teams of inventors over time. The inventors who need to work on tasks that require proximity should be clustered in space, but could potentially be geographically disconnected from the other members of the team. Figure A.48 tries to shed some light on this by plotting the minimum, mean, and maximum distance among the inventors of the median patent. More precisely, I calculate the minimum, average, and maximum distance among the inventors of each U.S. patent filed in a given year by two or more inventors.[22] The left panel of Figure A.48 reports the median of these distribution. The graph shows a clear increase in the three series between 1950 and 1970, when they started diverging. After 1970, the minimum distance of the median patent stabilized around 10 kilometers, whereas mean and maximum distances kept their growth and reached 30 and 40 kilometers, respectively, in 2010. The right panel shows the share of patents for which at least one inventor is reported to live at least 100 kilometers away from any other inventor in the patent. This series shows two breaks. One between 1930 and 1940 that brought the share of these patents from about 7% to about 22%, and one in the 1970s when a still ongoing upward trend started. In 2010, about 33% of the filed patents had at least one inventor more than 100 kilometers apart. Figure A.49 shows the share of patents for which at least two inventors live in the same city.[23] This share has also experienced an important decline between 1930 and 1950, but it then stabilized just below 40%.In future research, it could be interesting to study whether

[22]Note that these three statistics coincide when there are only two inventors and that solo patents are discarded for this analysis.

[23]Note that this is not necessarily the specular image of Figure A.48, panel b, since a patent with three inventors, two living in Chicago and one in Columbus, would contribute to both graphs.

some technology fields have contributed differentially to this trends and also if they are confirmed when keeping the team size constant. Since team size has also been increasing over the same period of time, a null model in which inventors are added in a pseudo-random way could be consistent with the pattern described in Figures A.48 and A.49.

3.4.3. Assignees

Fact 3.1: The share of patents with an assignee has steadily increased over time. Figure A.50 shows the share of patents whose rights were assigned, in full or in part, to a third-party. A third-party could be an individual or a company that commissioned or sponsored the development of the invention described in the grant. The share of patents without assignee has been shrinking over time and has been less than 20% for the past 40 years. The increasingly capital intensive nature of R&D activities or a trend towards market concentration might be at the root of this trend.

Fact 3.2: Average and maximum distance between inventors and assignees of the median patent started an upward trend in the 50s; minimum distance increased at first and then plateaued. Similarly to what we did for inventors, it is possible to analyze the distance between inventors and their assigness. A priori it is not obvious what to expect. On the one hand, the advantage in terms of resources of big firms and a tendency towards concentration (see for example Grullon, 2017) should reduce the average distance between inventors and assignees. On the other hand, more outsourcing and the decline in communication and transportation prices should work as a centrifugal force. Figure A.51 suggests that centrifugal forces dominate centripetal ones. The left reports the evolution of the minimum, average, and maximum distance between the inventors their assignees for the median patent in the sample.[24] The right panel shows the share of patents for which the maximum distance between the inventors and their assignee is at least 100 kilometers. The graphs show a clear tendency towards decentralization, although the minimum distance has remained constant since the 1980s. Similarly to what was argued for Fact 2.1, it might the case that R&D operations are directed by researchers working for the assignee and some specific tasks are outsourced to other labs.

3.4.4. Citations

Fact 4.1: The average number of backward citations per patent has steadily increased over time. The average number of patents cited by each patent has been steadily increasing over time. Figure A.52 shows this trend over time. The left panel shows the series for the years between 1836 and 1940, whereas the right panel for the years after 1940. The data are split into two figures to take into account the introduction of a mandatory section containing the list of references cited in 1947 that has importantly increased the number of citations successfully extracted from the data. As it is possible to see in the left panel of Figure A.52, the average number of citations prior to the mandatory disclosure of the references is order of magnitudes

[24]Note that these three statistics coincide for solo authored patents with an assignee.

smaller, but not negligible. Nevertheless, as the right panel of Figure A.52 highlights, the amount of references obtained in this way is probably only a small fraction of the full list of prior art considered when examining the patent.[25]

Citations might have steadily increased over time for two main reasons. First, digitalization have made it easier for inventors and reviewers to find inventions related to the one described in the patent. This would explain the acceleration in the average number of citations after 1980. Second, the number of inventions upon which newer inventions are built on has also increased over time. Inventions have become increasingly complex and if in the past a new idea relied on basic knowledge, nowadays it builds on a large number of previous discoveries (e.g., Jones, 2009). Such an increase in complexity would translate in an increase in the number of references.

Fact 4.2: The share of patents without forward citations was around 90% until 1910; it has then declined to 10% and remained mostly stable. The share of patents without forward citations has dramatically decreased between 1910 and 1940, but was stable in the years before and after this period. Before 1910 about 90% of patents did not receive any citation since they were filed, whereas after 1940 this share was around 10%. Figure A.53 reports this pattern over time. The low share before the 20s might be related to the introduction of the mandatory reference sector in 1947. More interesting is the extremely low share of patents without forward citations in the second half of the 20th century. Three facts might explain this trend. First, higher patenting costs might have contributed to attract

[25]Future research should investigate the informativeness of pre-1947 citations.

more meaningful patents. Second, the second part of the 20th century witnessed a significant increase in the number of foreign patents filed in the United States. Because of the costs involved in the patenting process, it is usually believed that grants filed to multiple patent offices are particularly valuable. Finally, there might have been an increase in the amount of self-citations. As suggested by Jones et al. (2007), an increase in the number of inventors per patent is likely to increase the number of self-citations. It could be interesting to explore these explanations in future research.

Fact 4.3: The average distance of citations received by the median patent in the first 10 years after filing was 0 up to the 40s; it has been increasing ever since. Figure A.54 analyzes the average distance of the citations received by the median patent in the 10 years after being filed. More precisely, I calculate the average distance between first inventors for each patent filed in each year. The figure reports the median of this distribution. With the exception at the beginning of the sample which is manly due to the small number of citations in the 19th century, the series shows a clear upward trend that started in the mid-40s and is still ongoing. This trend seems to support the idea that the decreasing cost of communication facilitate the diffusion of knowledge across space.

3.4.5. Classes

Fact 5.1: In the past 200 years only 9 classes made it to the top 1% terms of citations received per decade. Given the long time span provided by these data, we can ask what technologies were the most valuable in each decade. To do so, I exploit a standard measure of patent relevance used in the literature, namely the number of citations received by each grant. More precisely, I rank all the patents filed in each decade by the number of citations received and I select those in the top percentile. I define the most frequent principal class among the patents selected as the leading technology for that decade.[26] Table B.27 reports the results of this procedure. The table highlights two interesting facts. First, despite its simplicity, this methodology is able to capture the well-known technological waves in the United States over the past two centuries. The industrial revolution at the beginning of the twentieth century, the rise of medical science after the second world war with the development of vaccines and antibiotics, and finally the digital revolution in the second part of the 90s. Second, the length of the technological waves seems to have increased over time. Although this might be due to the nature of the data that are more noisy at the beginning of the sample, this fact might be explained by two other observations. On the one hand, it might be that since innovation becomes more complex in every field over time, it is more rare to have a breakthrough that moves the center of gravity towards another technology. On the other hand, it might be that the more recent waves enjoy more ideas to build upon and it takes longer to exhaust their creative momentum.

[26]Using the top 5%, instead of the top percentile, leads to similar results. Berkes et al. (2018) explore more refined definitions of leading technology exploiting the network structure of patent citations.

3.5. Conclusion

Since Hall et al. (2001), patents have been the preferred measure of innovation in the literature. The more than 3000 citations received by that paper alone in less than 20 years testify the high-demand for high-quality data on the topic. Because of the new opportunities offered by newly released or collected historical data, such as the historical decennial Census of Population, researchers have started moving their attention to pre-1976 data. In the past few years, the efforts to digitalize and extract meaningful information from historical patents have multiplied. The lack of a single data set that offers all the variables of interest collected with a consistent methodology and the fact that these data are sometimes not share with the rest of the community might constitute an important barrier for researchers who do not have access to them. This paper fills this gap and describes a freely available newly assembled data set of historical patents containing all the variables usually employed in the literature. I anticipate that some issues might surface at the beginning when using them for actual research, the same way I found and fixed some problems while writing Section 3.4. Based on the feedback I will receive, I expect to make the data set more reliable over time and potentially include additional variables. The comparison with HistPat performed in Section 3.3 validates the data at least from a coverage and geographical points of view. Finally, some of the stylized facts presented in Section 3.4 show that the data are able to replicate some already well-known trends in the literature and gives a novel historical perspective to them. Some other trends described there are new and could spur ideas for future research.

Bibliography

[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] Acemoglu, Daron, Ufuk Akcigit and William Kerr. 2016. “Innovation Network.” NBER Working Paper No. 22783.

[4] Acevedo-Garcia, Dolores, Kimberly A. Lochner, Theresa L. Osypuk, and S. V. Subramanian. 2003. “Future directions in segregation and health research: A multilevel approach.” *American Journal of Public Health,* 93(2): 215–221.

[5] Aghion, Philippe, Peter Howitt. 1992. “A Model of Growth Through Creative Destruction.” *Econometrica*,60(2): 323–351.

[6] Agrawal, Ajay, Iain Cockburn, and Carlos Rosell. 2010. “Not invented here? Innovation in company towns.” *Journal of Urban Economics*, 67(1): 78–89.

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

[8] Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. “The Economics of Density: Evidence from the Berlin Wall.” *Econometrica,* 83(6): 2127–89.

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

[10] Akcigit, Ufuk, John Grigsby, and Tom Nicholas. 2017. "Immigration and the Rise of American Ingenuity." *American Economic Review: Papers and Proceedings*, 107(5): 327–331.

[11] Alcácer, Juan, Michelle Gittelman, and Bhaven Sampat. 2009. "Applicant and examiner citations in U.S. patents: An overview and analysis." *Research Policy*, 38(2): 415–427.

[12] Alexander, Diane, and Janet Currie. 2017. "Is It Who You Are or Where You Live? Residential Segregation and Racial Gaps in Childhood Asthma." NBER Working Paper No. 23622

[13] Allen, Treb, Costas Arkolakis, and Yuta Takahashi. 2017. "Universal Gravity." NBER Working Paper No. 20787.

[14] Andrews, Michael. 2017. "Comparing Historical Patent Datasets." https://sites.google.com/site/michaeljeffreyandrews/research.

[15] 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.

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

[17] Autor, David, David Dorn and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review,* 103(6): 2121–68.

[18] Autor, David, Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics*, 90(2): 300–323.

[19] Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009. "Skills in the City." *Journal of Urban Economics*, 65(2): 136–153.

[20] Baum-Snow, Nathaniel, and Byron F. Lutz. 2013. "Inequality and City Size." *The Review of Economics and Statistics*, 95(5): 1535–48.

[21] Baum-Snow, Nathaniel, and Ronni Pavan. 2011. "School Desegregation, School Choice, and Changes in Residential Location Patterns by Race." *American Economic Review*, 101(7): 3019–46.

[22] Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588–638.

[23] 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.

[24] Berkes, Enrico and Ruben Gaetani. 2018. "Income Segregation and Rise of the Knowledge Economy." Mimeo.

[25] Berkes, Enrico, Ricardo Dahis, and Martí Mestieri. 2018. "Technology and City Cycles." Mimeo.

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

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

[28] Carlino, Gerald, and William R. Kerr. 2015. "Agglomeration and innovation." In *Handbook of regional and urban economics*, 5: 349–404.

[29] 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.

[30] Chetty, Raj, and Nathaniel Hendren. 2016. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." NBER Working Paper No. 23001.

[31] 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): 2543–94.

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

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

[34] Davis, Don R., and Jonathan I. Dingel. 2016. “A Spatial Knowledge Economy.” NBER Working Paper No. 18188.

[35] Davis, Don R., Jonathan I. Dingel, Joan Monras and Eduardo Morales. 2016. ”How Segregated is Urban Consumption?” Working Paper.

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

[37] De Solla Price, Derek. 1963. Little Science, Big Science. New York: Columbia University Press.

[38] Diamond, Rebecca. 2016. “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000.” *American Economic Review,* 106(3): 479–524.

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

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

[41] Eaton, Jonathan, and Samuel Kortum. 2002. “Technology, Geography, and Trade.” *Econometrica*, 70(5): 1741–1779.

[42] 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.

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

[44] Florida, Richard, and Charlotta Mellander. 2015. “Segregated City. The Geography of Economic Segregation in America’s Metro” Martin Prosperity Institute, February.

[45] Florida, Richard. 2017. "The New Urban Crisis." New York: Basic Books.

[46] Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114 (3): 739–767.

[47] Gaubert, Cecile. 2018. "Firm sorting and agglomeration." NBER Working Paper No. w24478.

[48] Glaeser, Edward L., Hedi Kallal, José Scheinkman and Andrei Schleifer. 1992. "Growth in Cities." *Journal of Political Economics*, 100(6): 1126–52.

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

[50] 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.

[51] Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature*, 28(4): 1661–1707.

[52] Grullon, Gustavo, Yelena Larkin, and Roni Michaely. 2017. "Are U.S. Industries Becoming More Concentrated?" https://ssrn.com/abstract=2612047.

[53] Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. "Endogenous Gentrification and Housing Price Dynamics." *Journal of Public Economics*, 100: 45–60.

[54] 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.

[55] Hanley, Douglas. 2015. "Innovation, Technological Interdependence, and Economic Growth." Mimeo.

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

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

[58] 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.

[59] Jargowsky, Paul A. 1996. "Take the Money and Run: Economic Segregation in U.S. Metropolitan Areas." *American Sociological Review*, 61(6): 984–98.

[60] Jones, Benjamin F. 2009. "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *The Review of Economic Studies*, 76(1): 283–317.

[61] Jovanovic, Boyan. 2015. "Learning and Recombination." Mimeo.

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

[63] Kerr, William R., and Scott Duke Kominers. 2015. "Agglomerative forces and cluster shapes." *Review of Economics and Statistics,* 97(4): 877–899.

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

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

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

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

[68] Lee, Sanghoon, and Jeffrey Lin. 2017. "Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income." FRB of Philadelphia Working Paper No. 17-3.

[69] Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *The American Economic Review*, 96(3): 461–98.

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

[71] Lucas, R. E., and E. Rossi-Hansberg. 2003. "On the Internal Structure of Cities." *Econometrica*, 70(4): 1445–1476.

[72] Lucas, Robert Jr., and Benjamin Moll. 2014. "Knowledge Growth and the Allocation of Time." *Journal of Political Economy*, 112(1): 1–51.

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

[74] Matray, Adrien. 2014. "The Local Innovation Spillovers of Listed Firms." Mimeo.

[75] Moretti, Enrico. 2004. "Workers' Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*, 94(3): 656–90.

[76] Moretti, Enrico. 2012. *The New Geography of Jobs*. New York: Houghton Mifflin Harcourt.

[77] Moser, Petra. 2005. "How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World's Fairs." *American Economic Review*, 95(4): 1214–1236.

[78] Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. "German Jewish émigrés and US Invention." *American Economic Review*, 104(10): 3222–55.

[79] Nicholas, Tom. 2010. "The Role of Independent Invention in U.S. Technological Development, 1880-1930." *Journal of Economic History*, 70(1): 57–82.

[80] Packalen, Mikko, and Jay Bhattacharya. 2015. "Cities and Ideas." NBER Working Paper No. 20921.

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

[82] Petralia, Sergio, Pierre-Alexandre Balland, and David L. Rigby. 2016. "Unveiling the geography of historical patents in the United States from 1836 to 1975." *Scientific Data*, 3.

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

[84] Reardon, Sean F., and Kendra Bishoff. 2016. "The Continuing Increase in Income Segregation, 2007-2012." Cepa Report, February.

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

[86] Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens. 2010. "Housing Externalities." *Journal of Political Economy*, 118(3): 485–535.

[87] Sarada, Michael Andrews, and Nicolas L. Ziebarth. 2017. "The demographics of inventors in the historical United States." Unpublished.

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

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

[90] Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Benjamin Jones. 2013. "Atypical Combinations and Scientific Impact." *Science*, 342(6157): 468–472.

[91] Von Hippel, Paul T., Samuel V. Scarpino, and Igor Holas. 2014. "Methods for Estimating Inequality from Binned Incomes: The Simplest Method is Competitive with the Most Complex." Working Paper.

[92] Weitzman, Martin. 1998. "Recombinant Growth." *The Quarterly Journal of Economics*, 113(2): 331–60.

[93] White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica,* 48(4): 817–838.

[94] Wuchty, Stefan, Benjamin F. Jones and Brian Uzzi. 2007. "The Increasing Dominance of Teams in the Production of Knowledge." *Science*, 316(5827): 1036–1039.

APPENDIX A

Figures

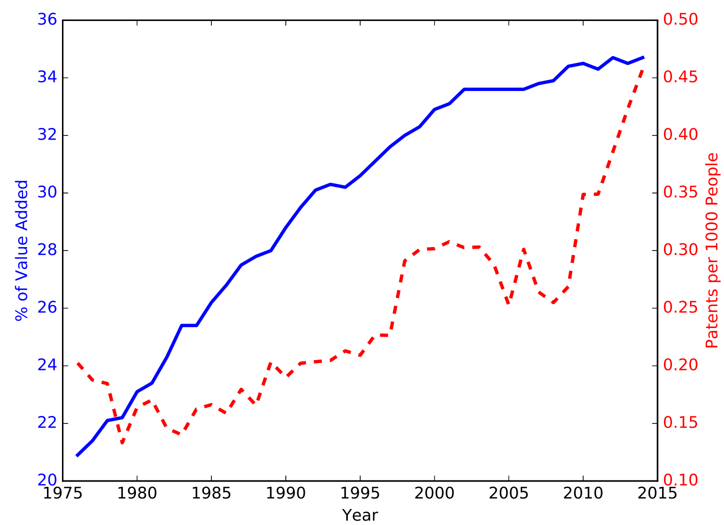


Figure A.1. The blue line is the contribution to U.S. GDP (value added) of computer and electronic products, electrical equipment, appliances and components, information, finance and insurance, professional and business services, educational services, health care and social assistance, arts, entertainment and recreation (data from the BEA). The dashed red line is the number of patents per 1,000 people issued to U.S. inventors by the USPTO.

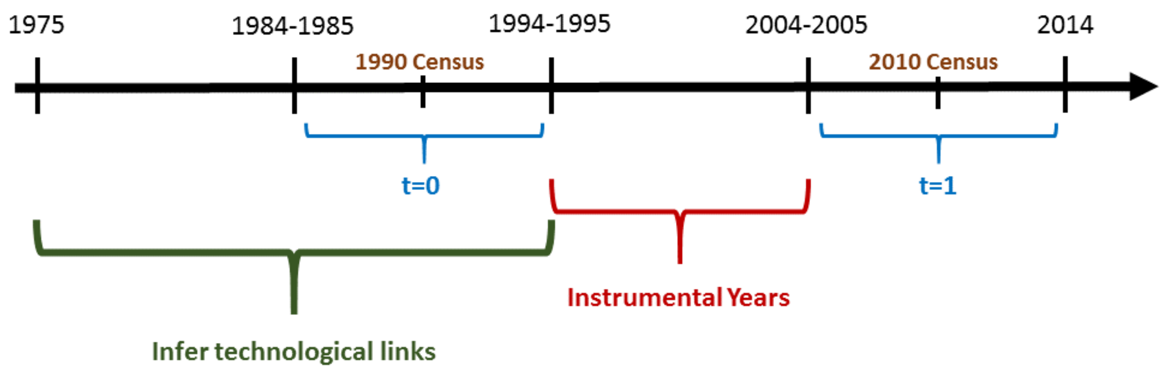


Figure A.2. The $t = 0$ observation corresponds to 1985-1994 data for patenting, and the 1990 Census for economic and demographic variables. The $t = 1$ observation corresponds to 2005-2014 data for patenting, and the 2008-2012 ACS for economic and demographic variables.

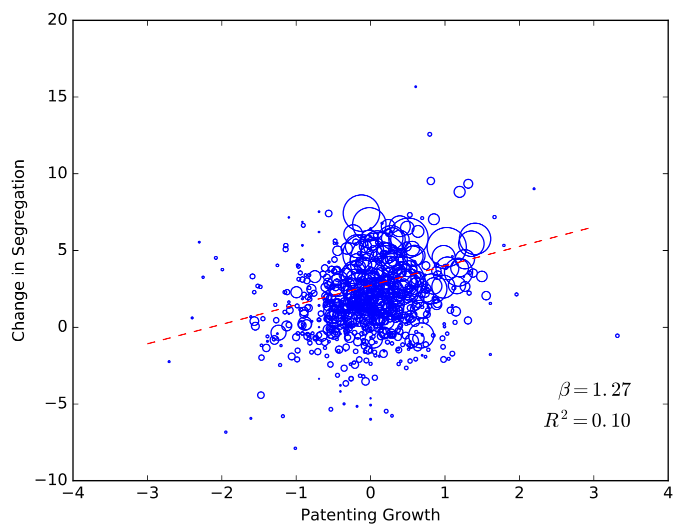


Figure A.3. Unconditional correlation between growth in patenting and change in income segregation between 1990 and 2010, weighted by total number of households in 1990.

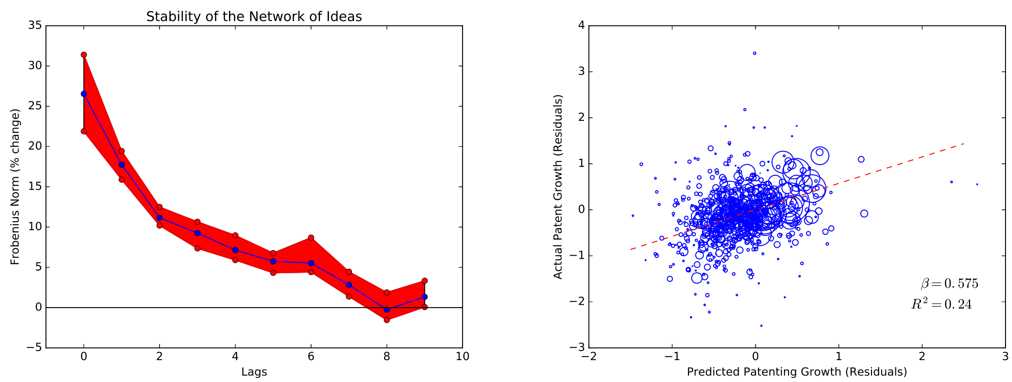


Figure A.4. Left-panel: Comparison between the Frobenius norm of the difference between the real diffusion matrices in the early and in the late samples, and the Frobenius norm of the difference between the reshuffled diffusion matrices in the early and in the late samples. Right-panel: Scatter plot of the residuals of actual and instrumented patent growth, after partialling out the standard controls (number of CTs, household growth and income growth). The scatter plot is weighted by total households in 1990.

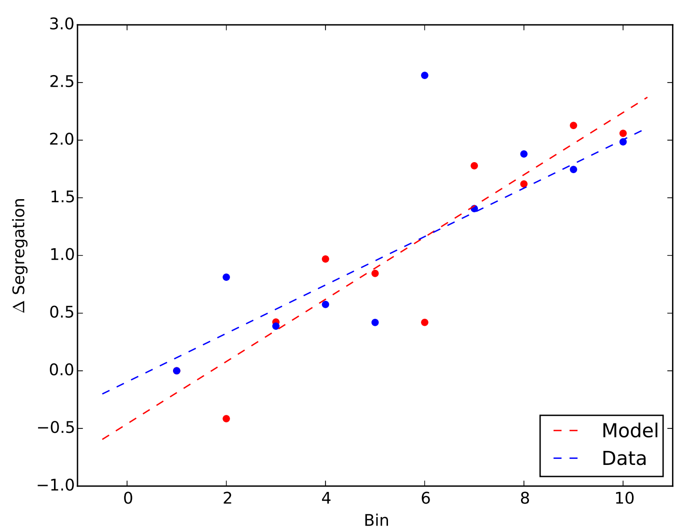


Figure A.5. Knowledge shock (bin) and change in segregation, 1990-2010: Data and Model.

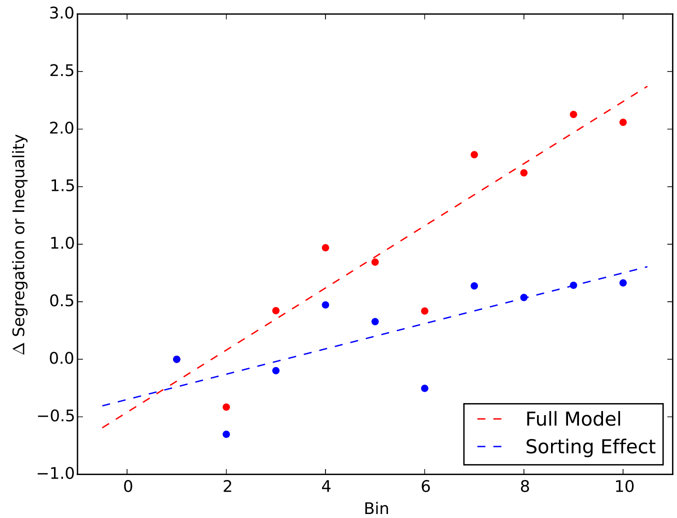


Figure A.6. Knowledge shock (bin) and change in segregation, 1990-2010: Full effect (red line) and Sorting effect (blue line) computed as segregation with 1990 distribution of average wages by CT/occupation and 2010 (model based) distribution of residents by CT/occupation.

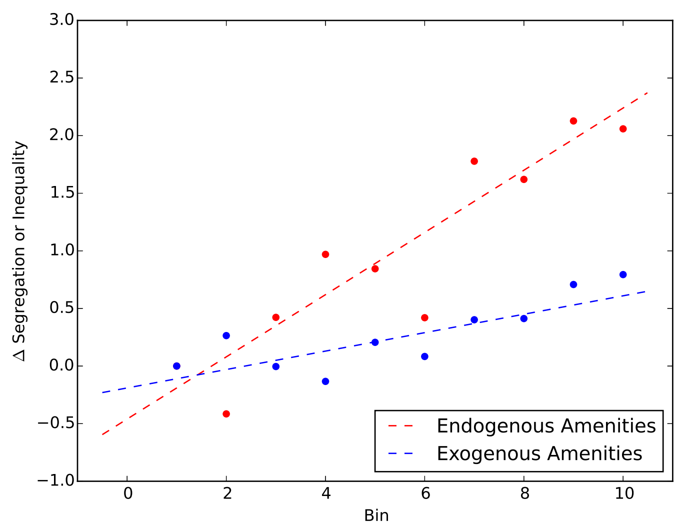


Figure A.7. Knowledge shock and change in segregation 1990-2010: Exogenous vs Endogenous residential amenities.

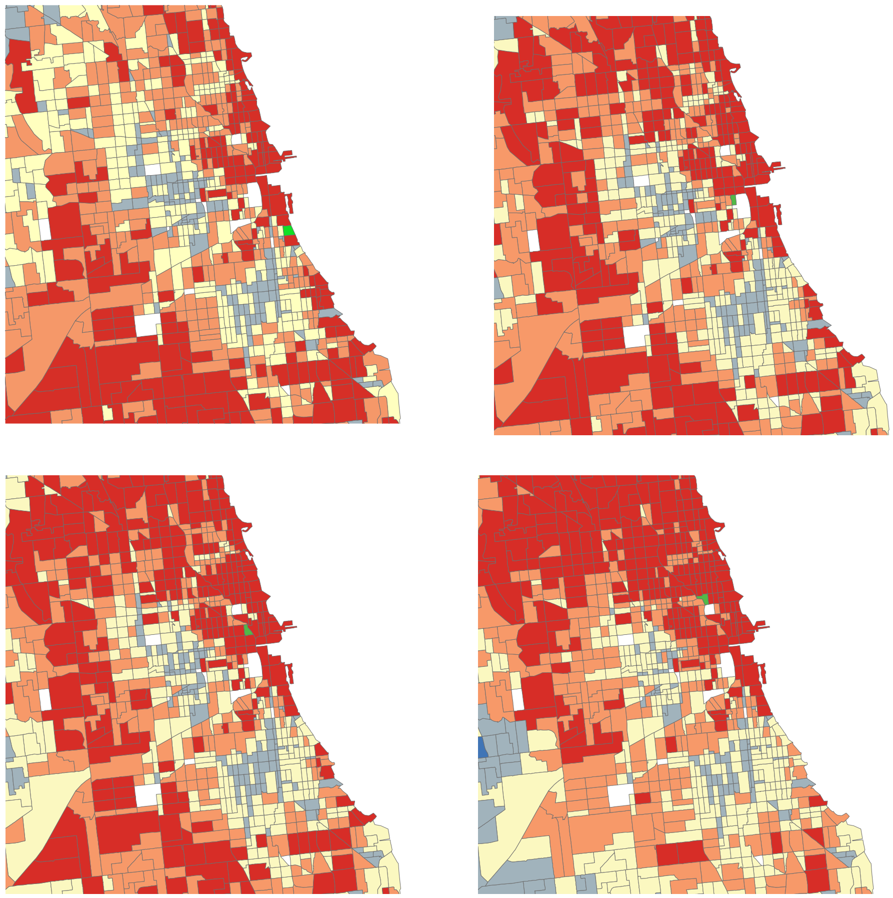


Figure A.8. Change in high-knowledge residents in each census tract of Chicago as a result of Amazon's new headquarter locating in a specific neighborhood (colored in green). Panel (a) considers the case in which Amazon's HQ2 is located on the old Michael Reese Hospital premises; panel (b) when it is located in the Old Main Post Office; panel (c) in the Tribune Media River Front property; panel (d) in the old A. Finkl & Sons steel plant. For each counterfactual, the distribution of the change is divided in 5 quantiles. The census tracts colored in bright red correspond to the top quantile, the ones in bright blue to the bottom quantile.

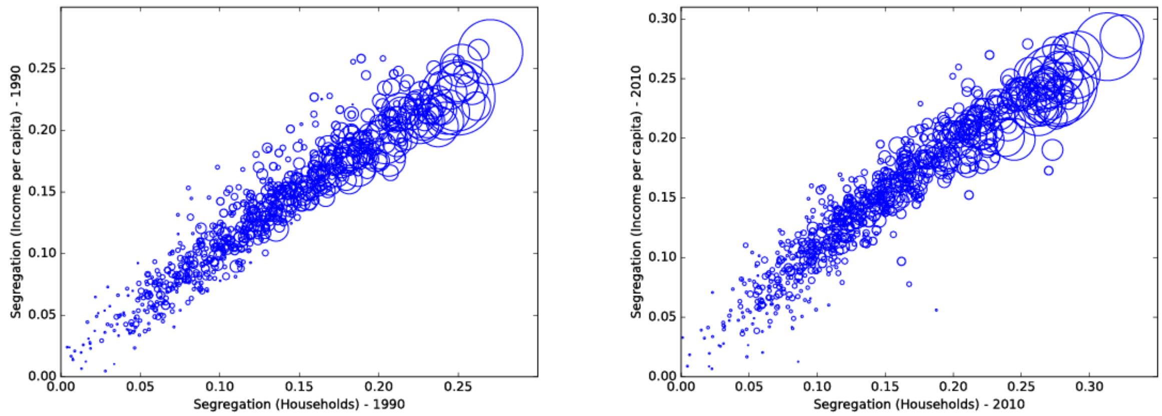


Figure A.9. Correlation between $Segr_{cz}$ computed using the approximated distribution of household income (horizontal axis) and the actual distribution of average personal income across CTs (vertical axis) in 1990 (left panel) and 2010 (right panel).

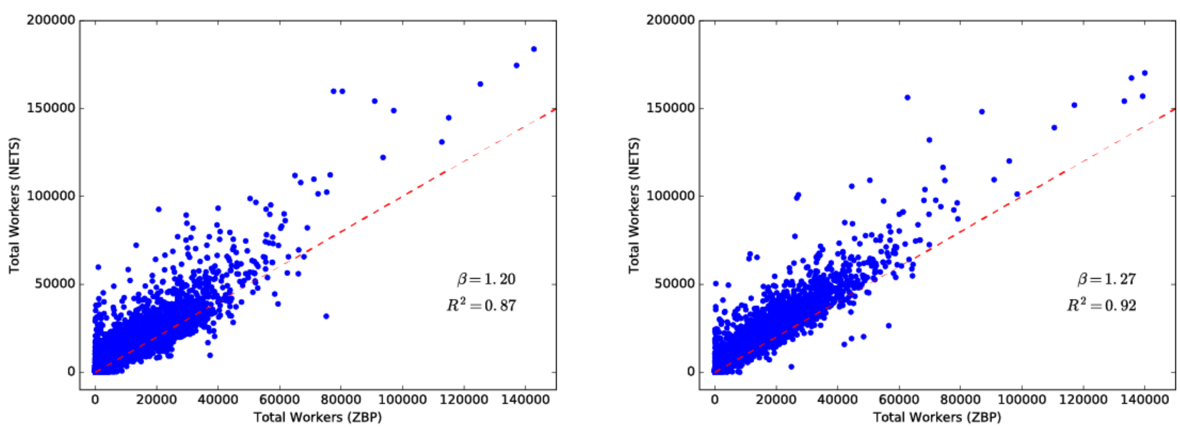


Figure A.10. Correlation between workers as reported in the ZBP (x-axis) and in the NETS (y-axis) in 1994 (left) and 2010 (right). Each point represents a ZIP code. The dashed red line is the 45-degree line.

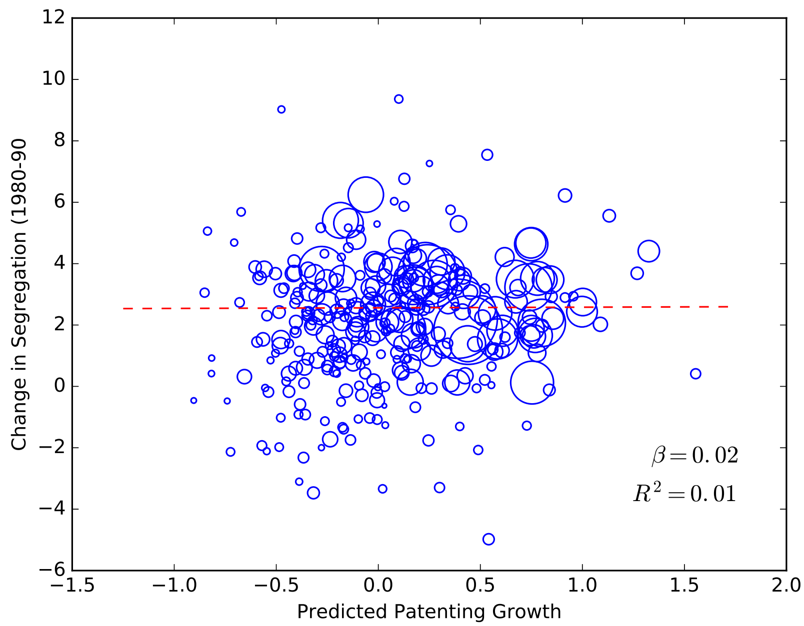


Figure A.11. Scatter plot of predicted patenting growth (instrument) and pre-trend in segregation (1980-1990). The scatter plot is weighted by total households in 1990.

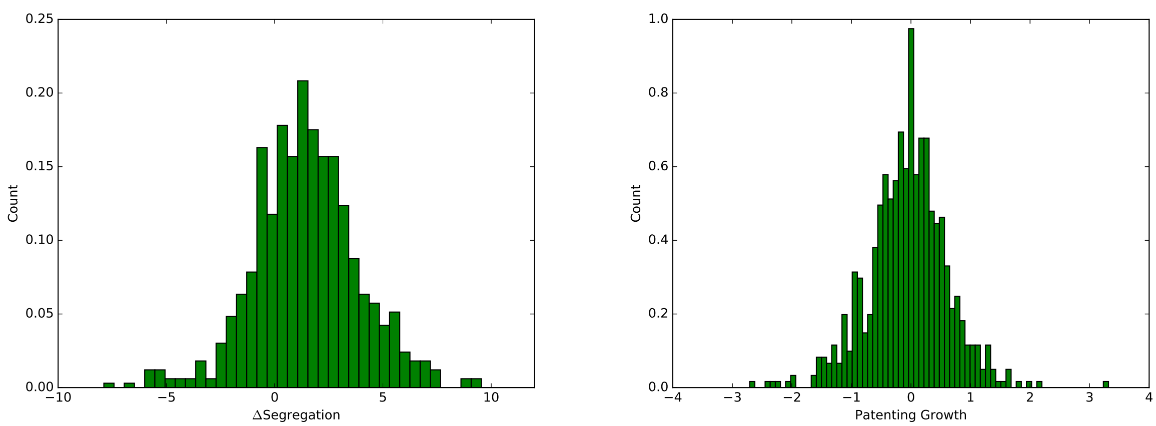


Figure A.12. Distribution of changes in measured segregation and patenting growth in the cross section of U.S. CZs, 1990-2010.

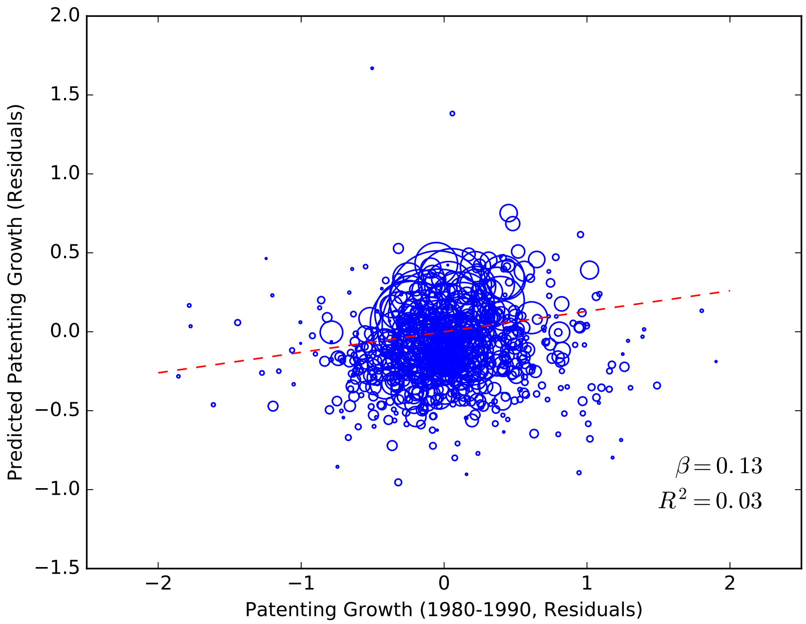


Figure A.13. Scatter plot of predicted patenting growth (instrument) and pre-trend in patenting (1980-1990). The scatter plot is weighted by total households in 1990.

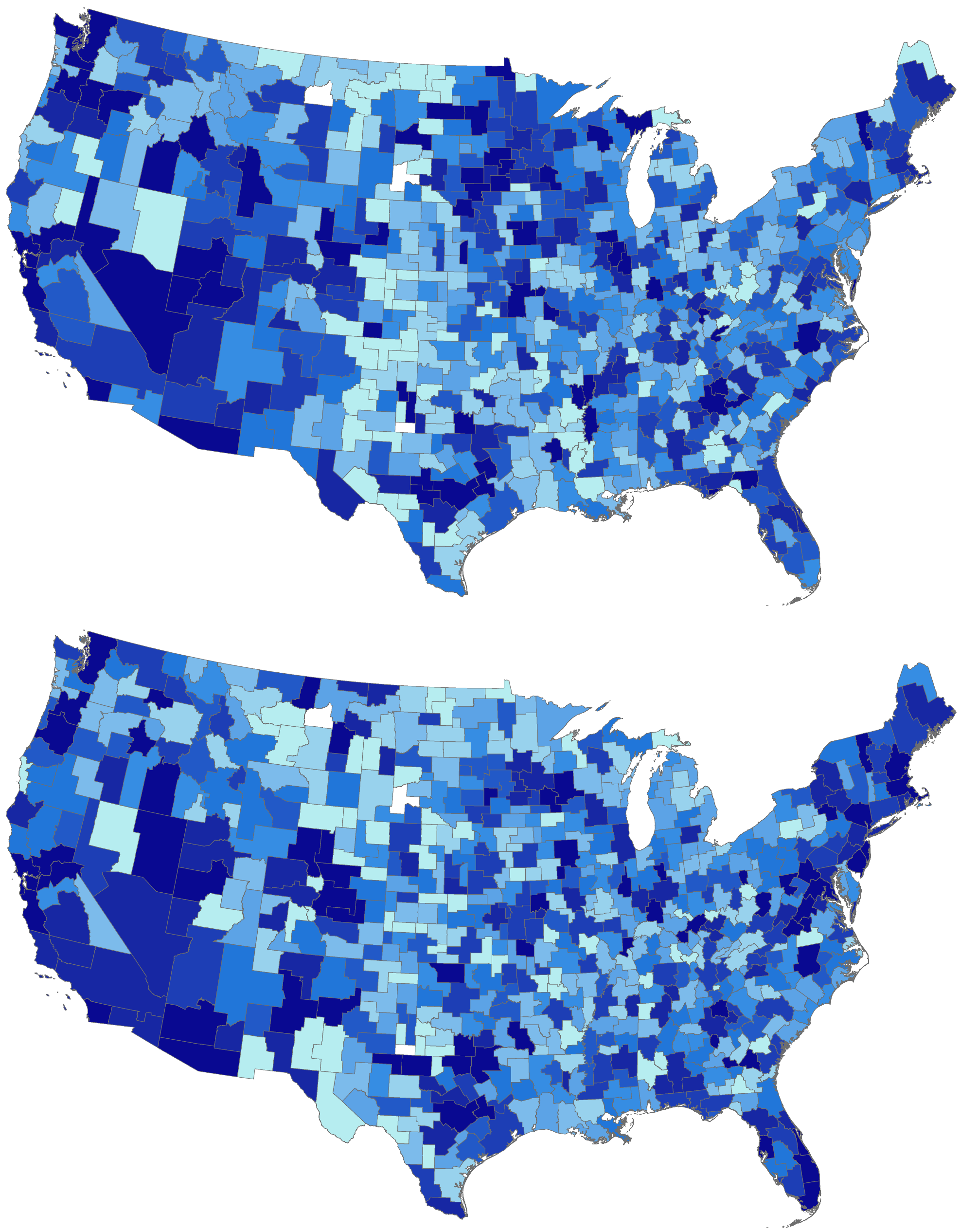


Figure A.14. Predicted (top map) and actual (bottom map) growth rate of patents in U.S. commuting zones, 1990-2010.

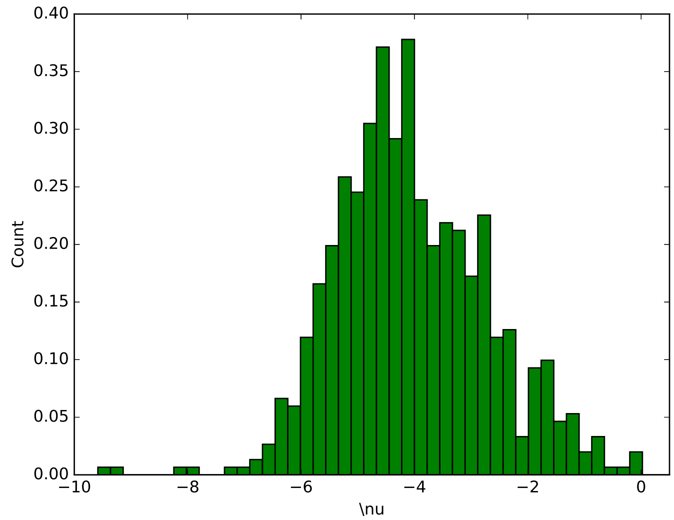


Figure A.15. Distribution of estimated values of ν_c in U.S. commuting zones

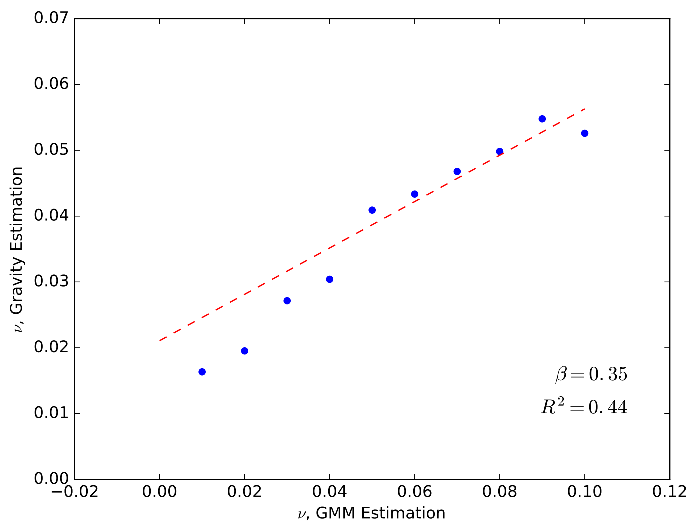


Figure A.16. The bin scatter plot compares the value of ν_c estimated using (1.22) with the value of ν_c that minimizes the difference between the share of people in city c commuting for less than 60 minutes and the model generated counterpart.

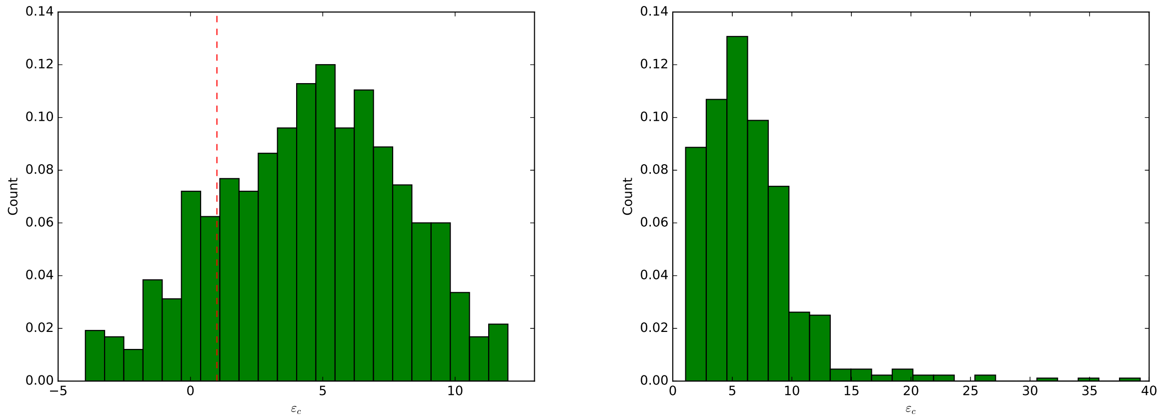


Figure A.17. The two histograms show the distribution of ε_c estimated through a 2SLS procedure that uses a model-generated instrument. The left histogram reports the entire distribution after dropping the top and bottom 5% of the values. The right panel reports the distribution for $\varepsilon_c > 1$ only.

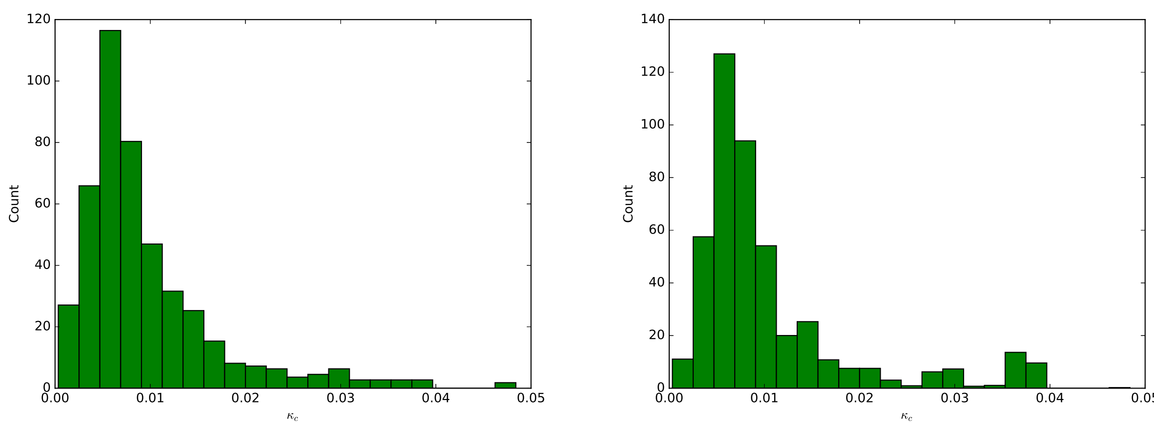


Figure A.18. The two histograms report the distribution of κ_c for those commuting zones with $\varepsilon_c > 1$. The left and right histograms show the unweighted and weighted distribution of this variable, respectively.

Figure A.19. Map of Chicago divided by census tract. The areas highlighted in black are the ones that were proposed as suitable places to host the Amazon's HQ2.

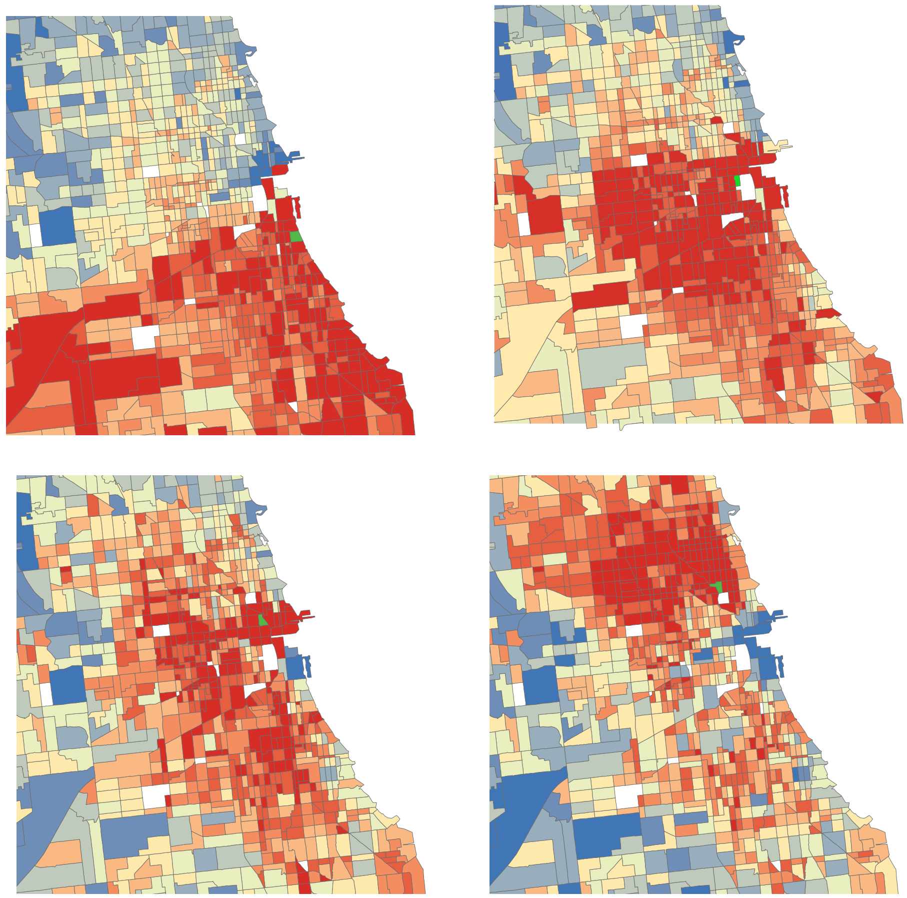


Figure A.20. Change in high-knowledge workers in each census tract of Chicago as a result of Amazon's new headquarter locating in a specific neighborhood (colored in green). Panel (a) considers the case in which Amazon's HQ2 is located on the old Michael Reese Hospital premises; panel (b) when it is located in the Old Main Post Office; panel (c) in the Tribune Media River Front property; panel (d) in the old A. Finkl & Sons steel plant. For each counterfactual, the distribution of the change is divided in 5 quantiles. The census tracts colored in bright red correspond to the top quantile, the ones in bright blue to the bottom quantile.

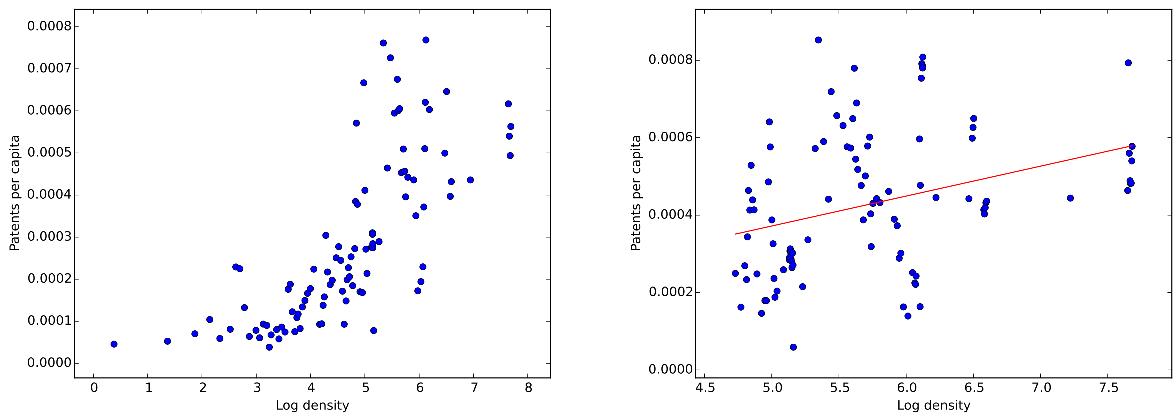


Figure A.21. Bin-scatter plot of patents per capita and (log) population density in all CZs (left) and densest CZs hosting 50% of the U.S. population. The plot is weighted by total population and controls for year fixed effects. The measure of innovation is winsorized at the 1% level.

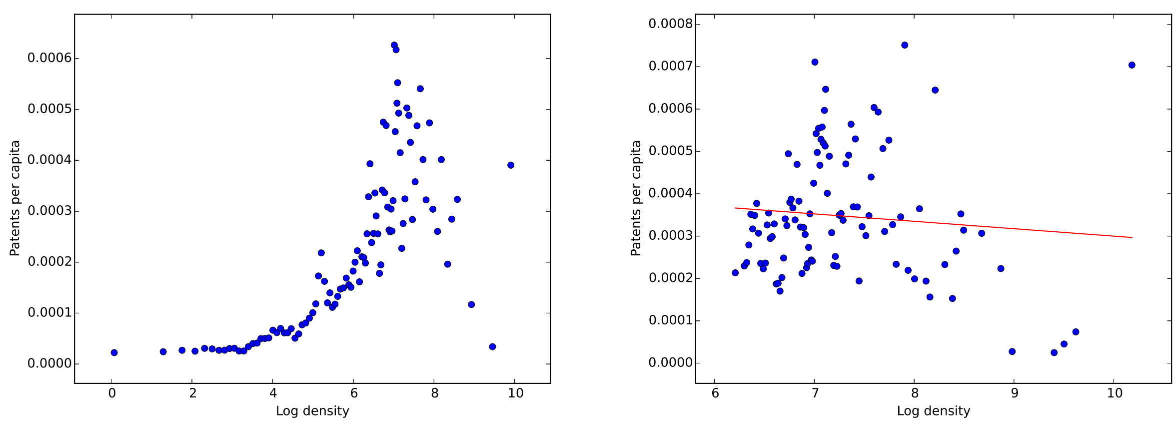


Figure A.22. Bin-scatter plot of patents per capita and (log) population density in all CSDs (left) and densest CSDs hosting 50% of the U.S. population. The plot is weighted by total population and controls for year fixed effects. The measure of innovation is winsorized at the 1% level.

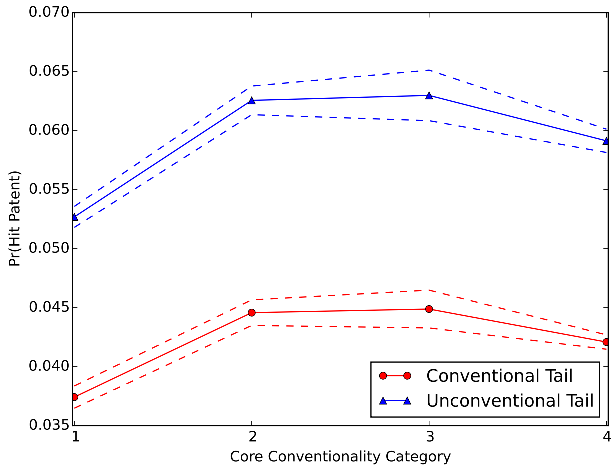


Figure A.23. Marginal effect of having a conventional tail and being in a certain core-conventionality category on the probability of being a hit patent.

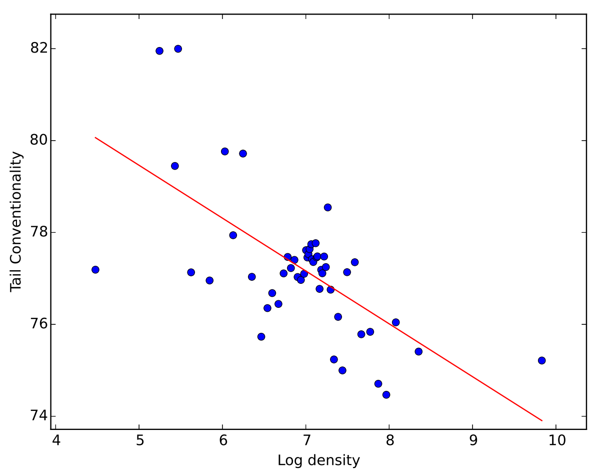


Figure A.24. The dependent variable is defined as the tail-conventionality of the median patent in the CSD-year observation. The bin-scatter plot is weighted by the total number of patents filed in the CSD/Year observation and controls for State and Year fixed effects.

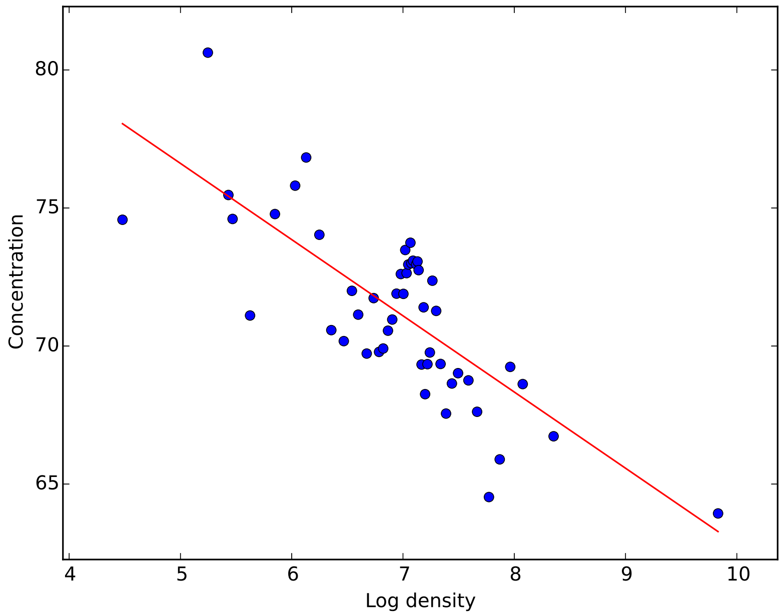


Figure A.25. Concentration of innovation output and log density of population. The bin-scatter plot is weighted by the total number of patents filed in the CSD/Year observation.

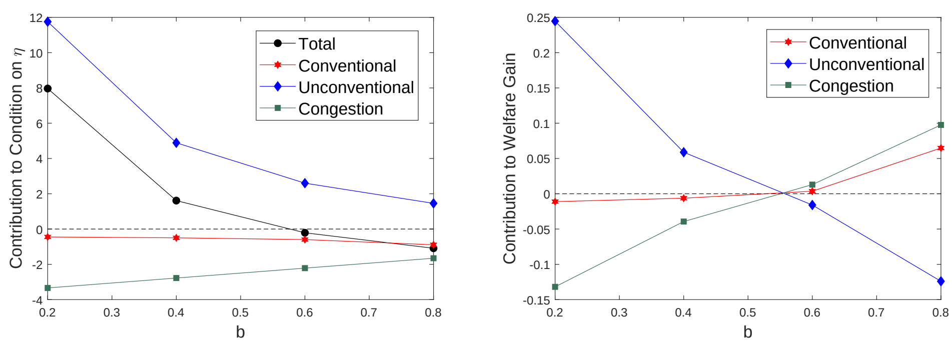


Figure A.26. Left panel: Contribution of each term to condition 2.21. Right panel: Contribution of rate of conventional ideas, rate of unconventional ideas and congestion costs to overall welfare gain from optimal policy under fixed urban structure.

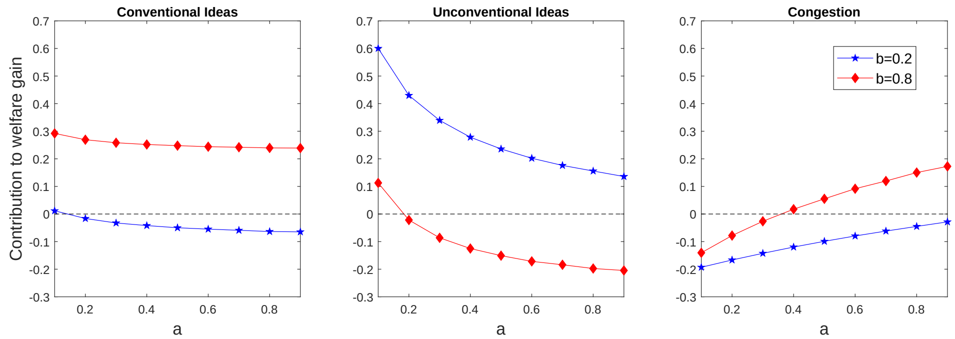


Figure A.27. Contribution of each component in 2.19 to overall welfare gain from optimal policy under flexible urban structure.

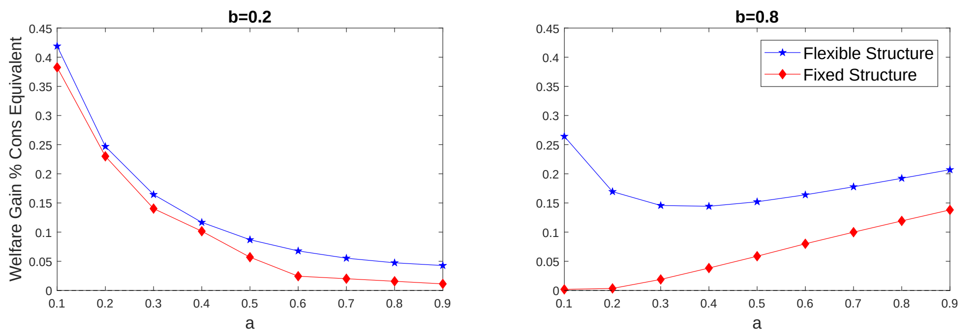


Figure A.28. Welfare gain (% consumption equivalent) from optimal policy with fixed structure (red line) and flexible structure (blue line).

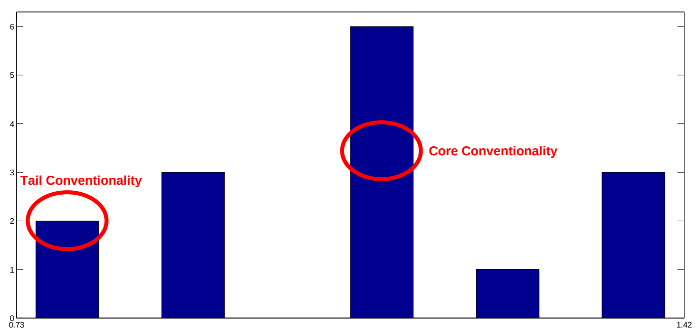


Figure A.29. Example of c-score distribution for a patent. Tail-conventionality corresponds to the 10th percentile of the distribution, core-conventionality corresponds to the median. Similarly, we define tail-unconventionality as one minus tail-conventionality.

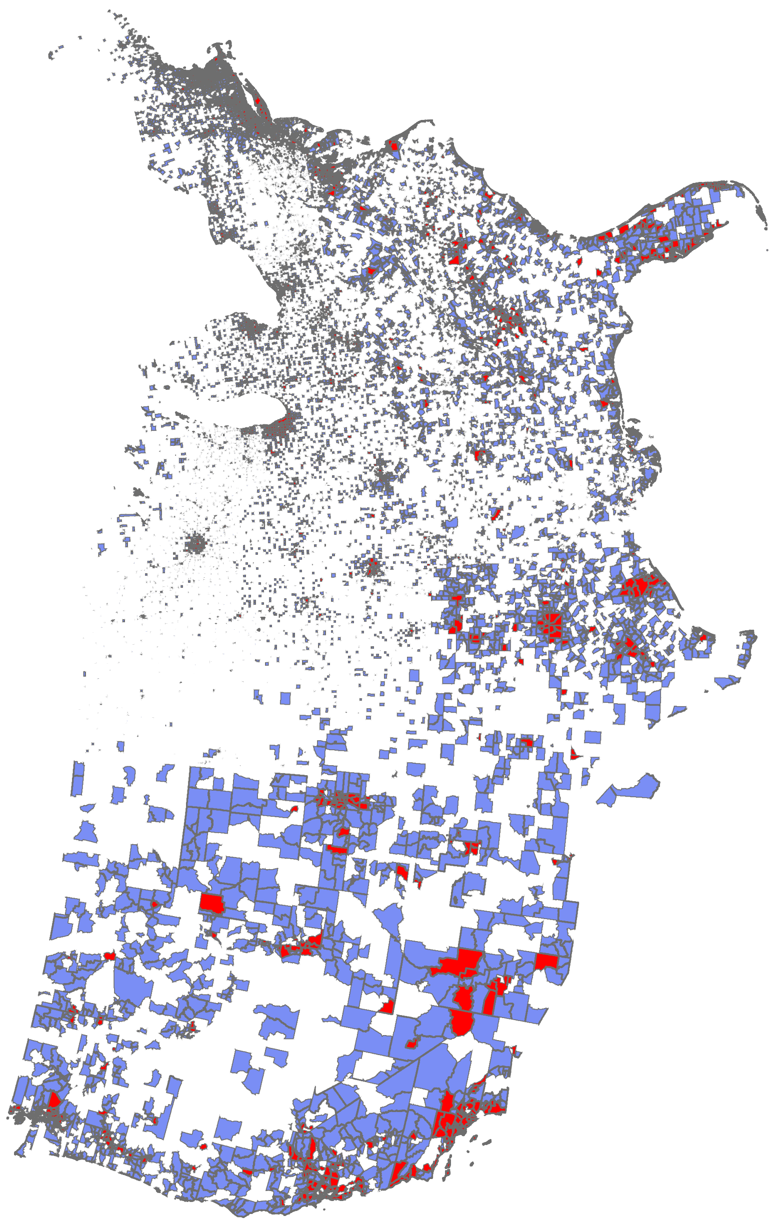
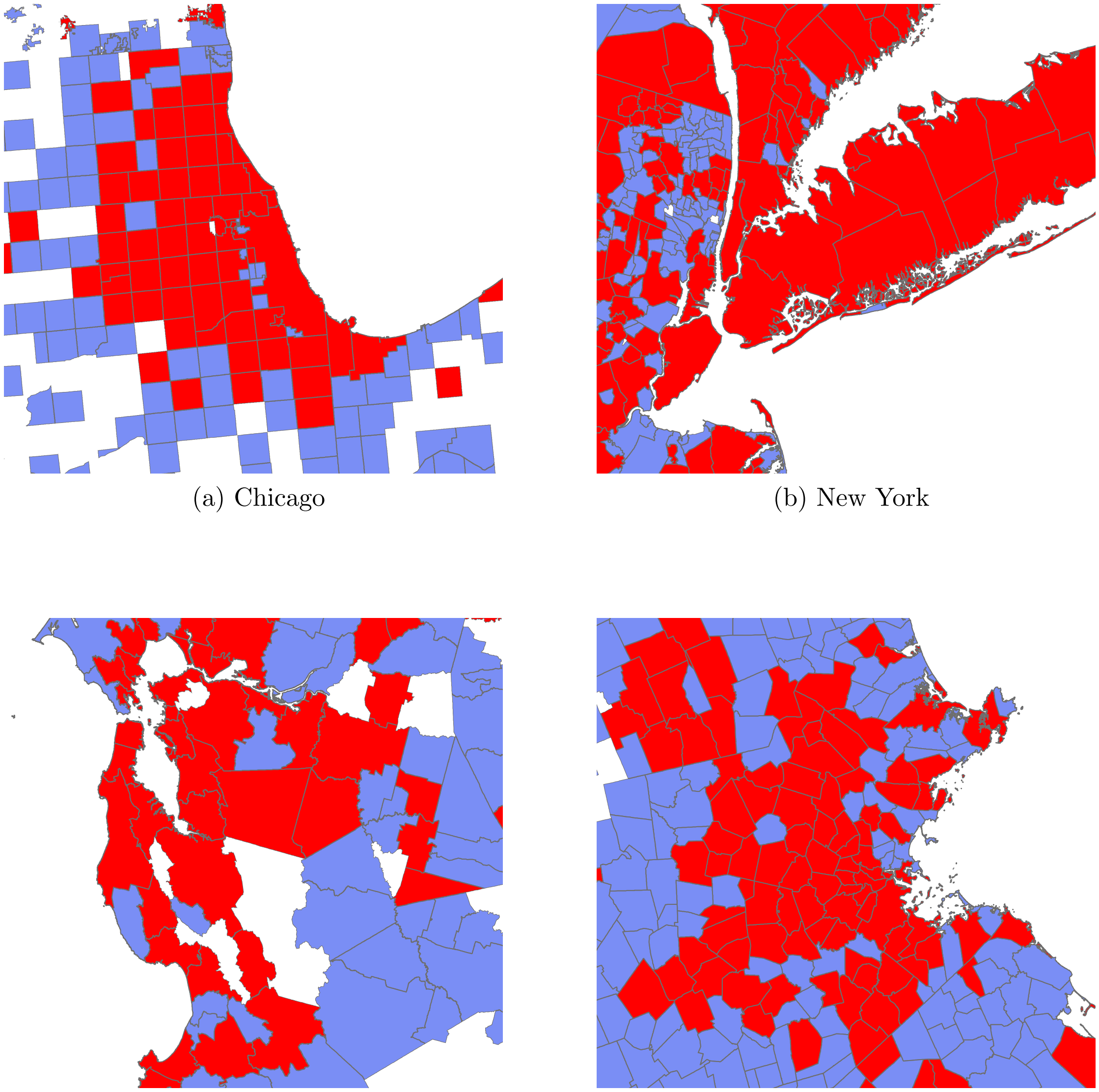


Figure A.30. The figure shows a map of county sub-divisions in the United States. A given CSD is colored in red if it produced at least one patent per year between 2000 and 2010 ("continuously innovative"), in blue if it produced innovation only occasionally. No patents have been filed in the CSDs that are missing in the map.



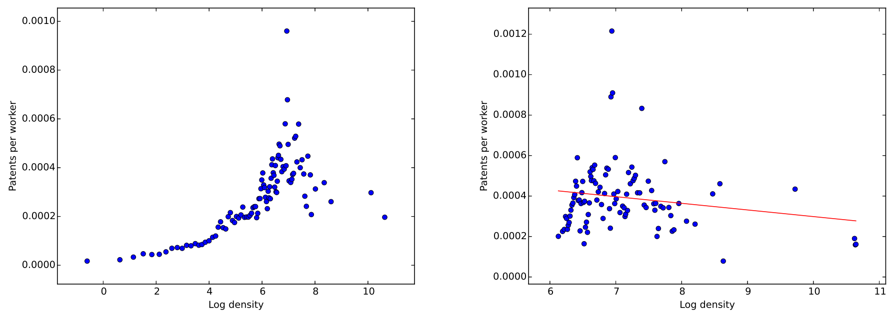
(a) Chicago

(b) New York

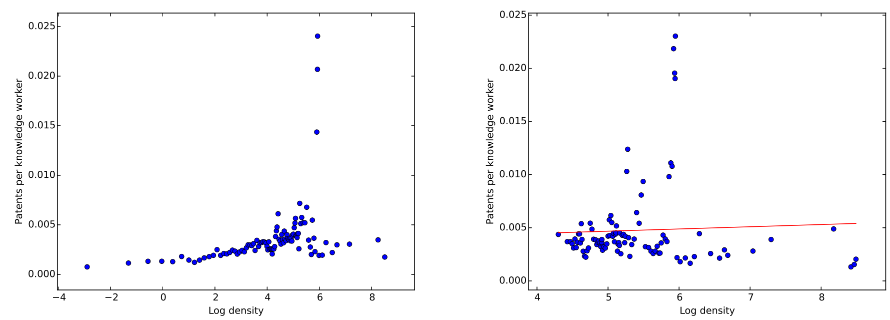
(c) San Francisco

(d) Boston

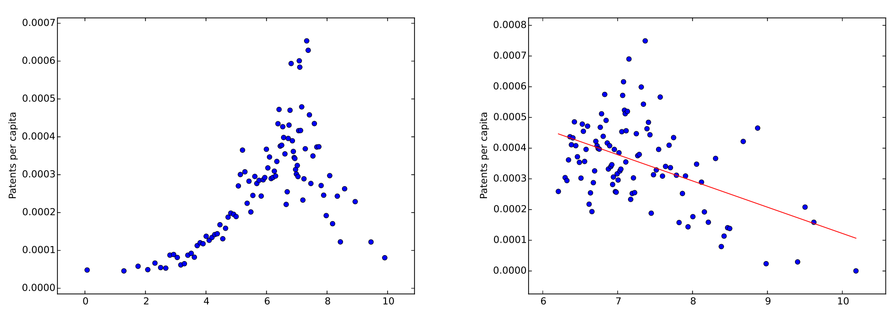
Figure A.31. The figure shows a map of county sub-divisions in four of the main metropolitan areas in the United States. A given CSD is colored in red if it produced at least one patent per year between 2000 and 2010 ("continuously innovative"), in blue if it produced innovation only occasionally. No patents have been filed in the CSDs that are missing in the map.



(a) Patents per worker and log-density of employment, all CSDs (left) and densest CSDs hosting 50% of U.S. employment (right). 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 knowledge worker and log-density of knowledge workers, all CSDs (left) and densest CSDs hosting 50% of U.S. knowledge employment (right). 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 capita and log-density of population, all CSDs (left) and densest CSDs hosting 50% of U.S. population (right). All patents are geo-located at the residence of the first inventor.

Figure A.32. 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.

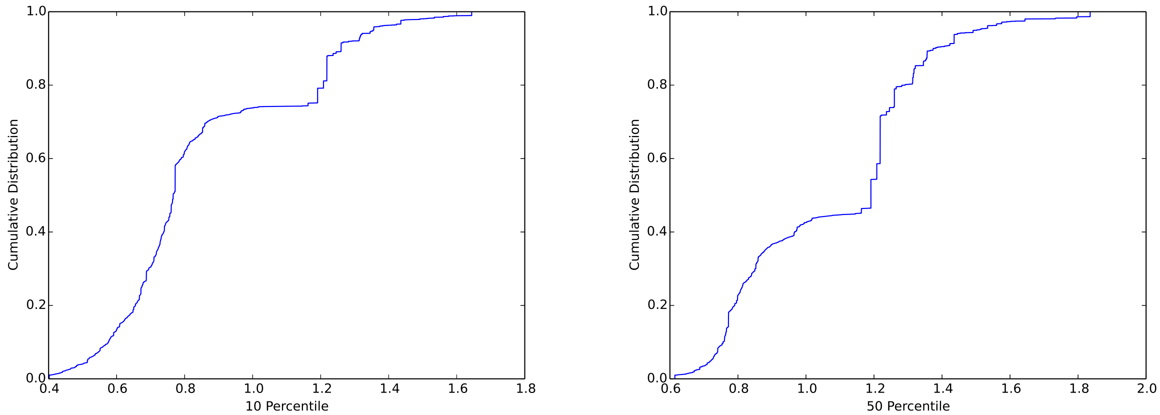


Figure A.33. Cumulative distribution functions of tail conventionality (left) and core conventionality (right) in the universe of patents.

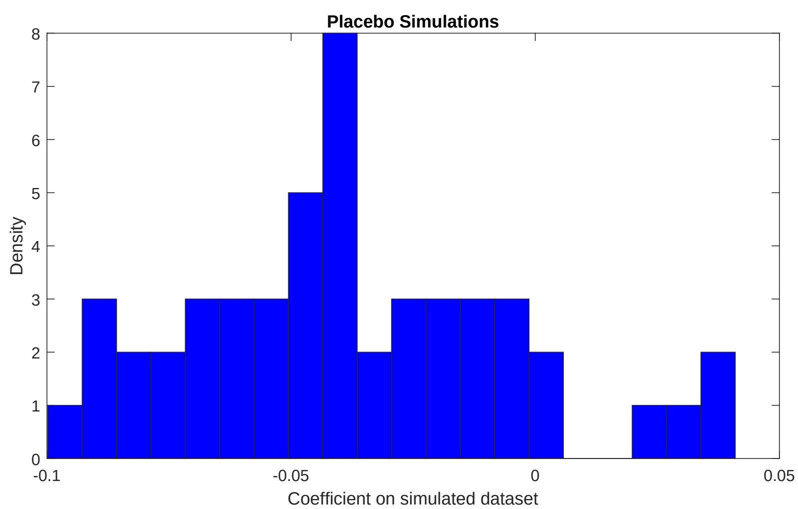


Figure A.34. Placebo experiment: Estimated coefficients from 50 regressions of log-density on concentration index on simulated patent networks.

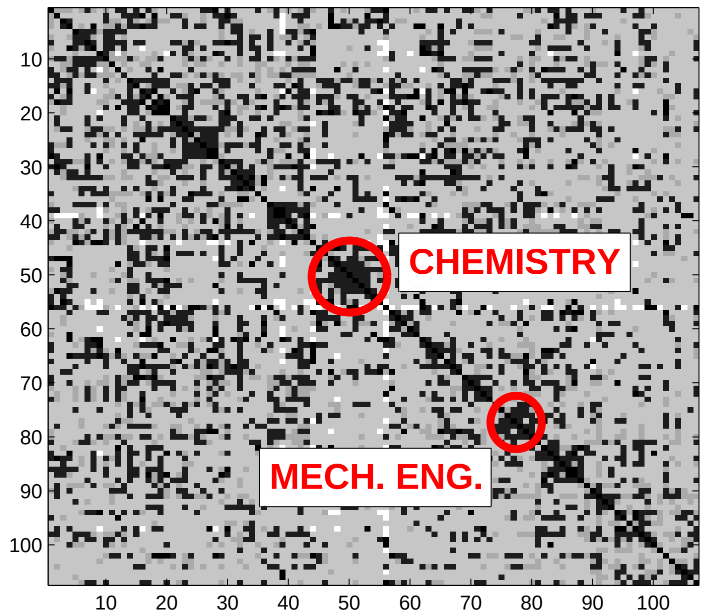
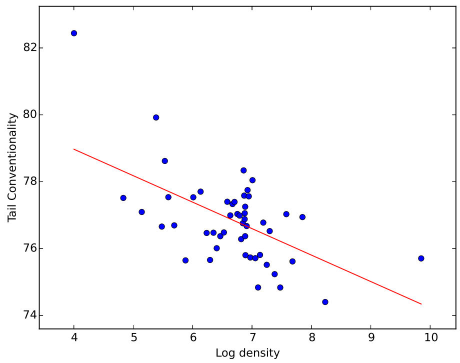
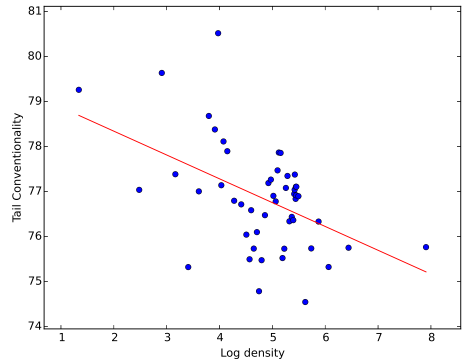


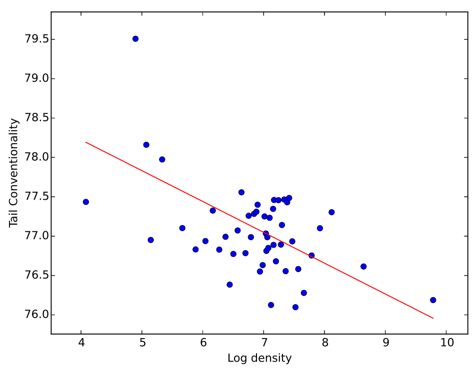
Figure A.35. Every pixel in the matrix indicates a patent class pair. The darker the pixel the higher the c-score assigned to that class pair, the lighter the lower the c-score. Diagonal elements of the matrix show a clear red tendency compared to the rest of the matrix. The “class-clusters” of Chemistry and Mechanical Engineering, among the others, are clearly visible around the diagonal.



(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 knowledge 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 residence of the first inventor.

Figure A.36. All the bin-scatter plots are weighted by total patents, and control for year and state fixed effects.

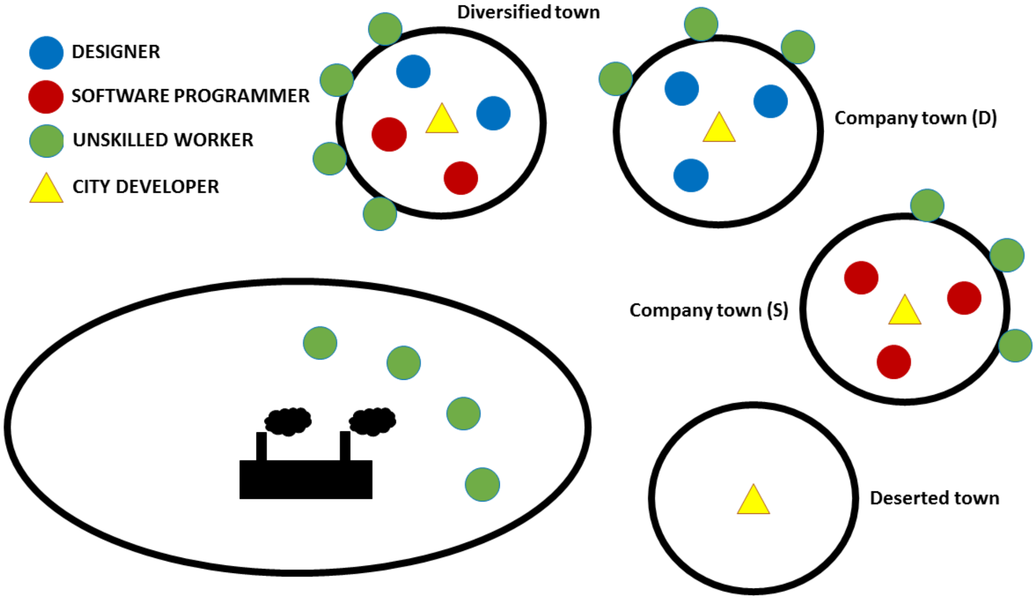
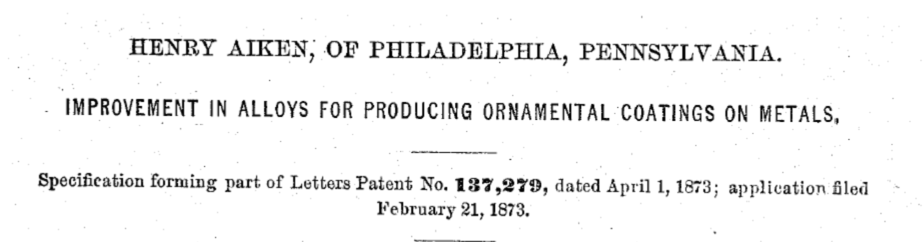


Figure A.37. Spatial economy: Illustration. Innovators from background $\mathcal{S}$ and $\mathcal{D}$ (programmers and designers) sort themselves into the downtown areas of cities. Unskilled labor lives in the outskirts of cities and in the rural areas. Production takes place in rural areas between cities.



HENRY AIKEN, OF PHILADELPHIA, PENNSYLVANIA.

IMPROVEMENT IN ALLOYS FOR PRODUCING ORNAMENTAL COATINGS ON METALS.

Specification forming part of Letters Patent No. **137,279**, dated April 1, 1873; application filed February 21, 1873.

Figure A.38. Header of patent number 137,279, the first patent that reports the year in which the application was filed.

planes.

I am aware of the Letters Patent No. 43,881 granted August 16, 1864, to Ralph Graham, of Brooklyn, Kings county, New York, for a hand fire-arm adapted to projecting grenades or small bombs, and I do not claim the invention therein shown; but

What I do claim as new and of my invention, and for which I desire Letters Patent, is—

1. Constructing a mortar with a hollow sleeve projecting from its base, instead of trunnions or cheeks, substantially as above described, for the purpose of receiving the elastic cushion, or any equivalent spring, and the end of a stake, as above set forth.

2. The combination of the slot E and pin D with the aforesaid mortar A, sleeve B, and spring C, as and for the purposes specified.

WM. F. GOODWIN.

Figure A.39. The figure shows an extract from patent number 46,101. The patent references another patent in the text. This piece of information is used to build a data set of citations prior to 1947.

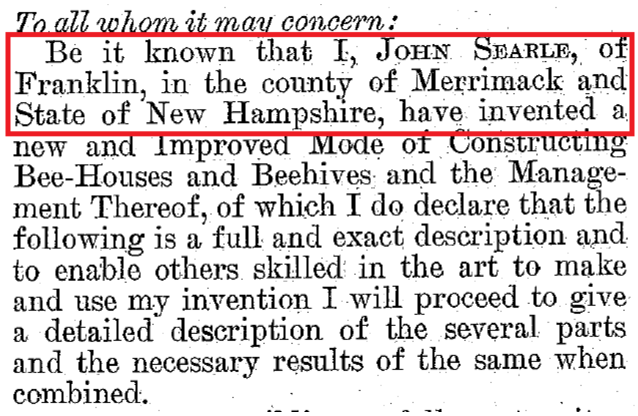
What I claim as my invention and desire to secure by Letters Patent is—

The construction of the spout, the balcony and its appendages, the ventilator, the construction of the feeder, and the method of constructing the double top of the hive, and the cement floor of the house; these I claim separately and in combination, the aforesaid invention being the best mode of producing artificial swarms of bees.

JOHN SEARLE.

Witnesses:
GEO. M. PHELPS,
JOSHUA FIFIELD.

Figure A.40. The figure shows an extract from the end of patent number 580. There the name of the inventor is listed in capital letters together with the name of two witnesses.


To all whom it may concern:

Be it known that I, JOHN SEARLE, of Franklin, in the county of Merrimack and State of New Hampshire, have invented a new and Improved Mode of Constructing Bee-Houses and Beehives and the Management Thereof, of which I do declare that the following is a full and exact description and to enable others skilled in the art to make and use my invention I will proceed to give a detailed description of the several parts and the necessary results of the same when combined.

Figure A.41. The figure shows an extract from the of end patent number 580. There the name of the inventor is listed in capital letters together with the name of two witnesses.

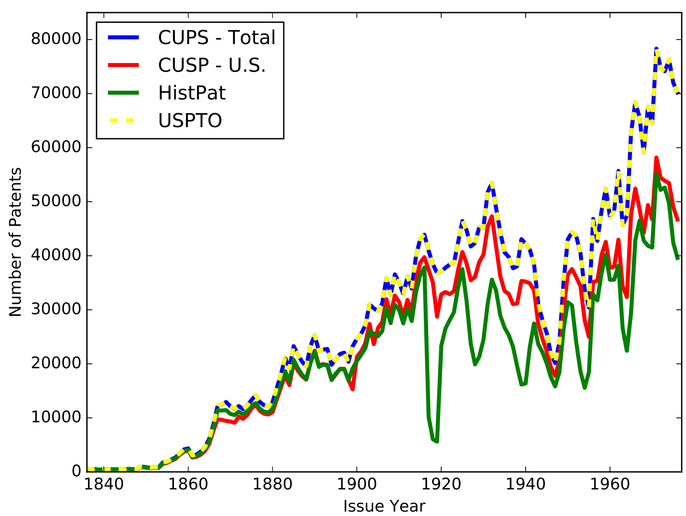


Figure A.42. The figure compares the number of patents reported in HistPat and CUSP. For CUSP I only selected the patents for which at least one inventor or one assignee are U.S. residents.

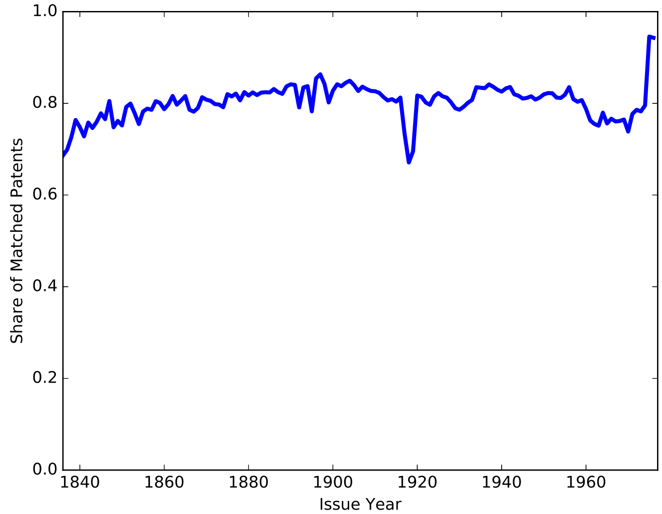


Figure A.43. The figure shows the share of patents for which all the locations that appear in HistPat are also contained in CUSP. The denominator is given by the number of patents available in HistPat in a given year.

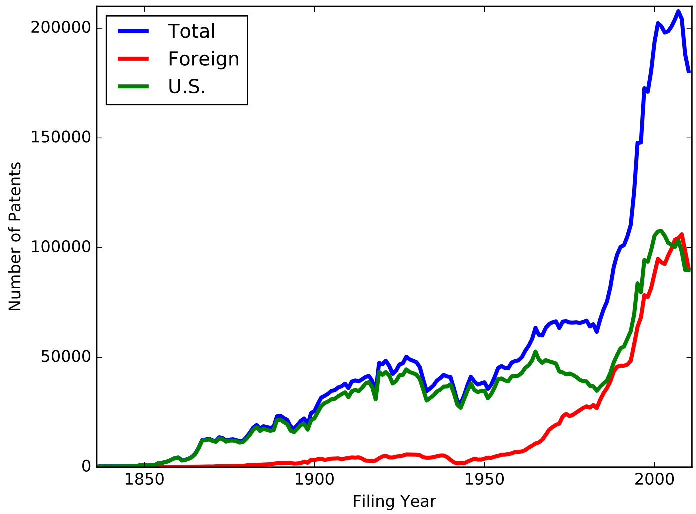


Figure A.44. The graph plots the number of patents granted by the USPTO by filing year and country of residence of their inventors. The blue line represents the total number of patents issued by the USPTO. The green line shows the number of patents whose inventors are U.S. residents. The red line shows the number of patents whose inventors are foreign residents.

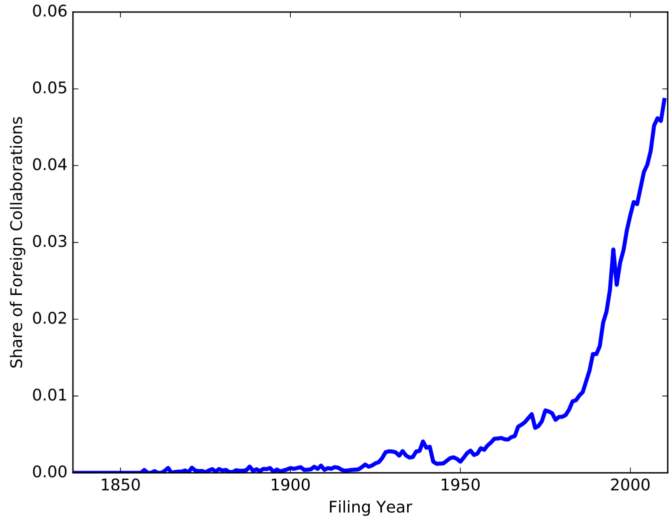


Figure A.45. The figure shows the share of patents that were the result of an international collaboration by filing year. A grant is considered an international collaboration if at least one inventor is a U.S. resident and at least another one has her residence outside the United States.

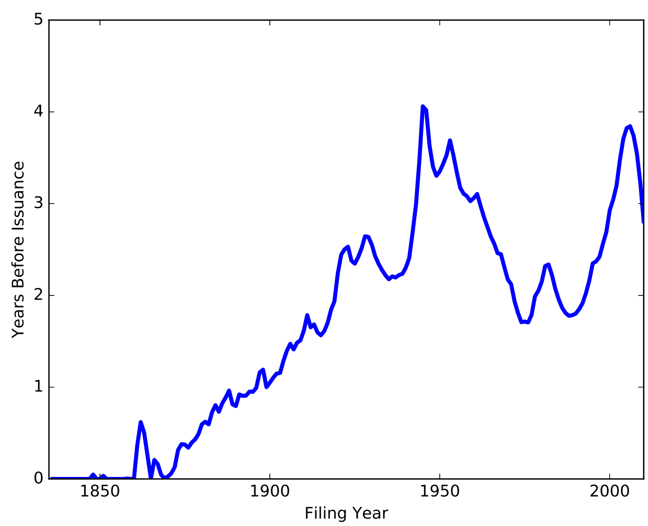
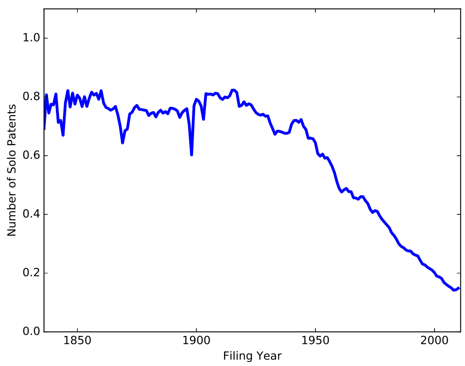
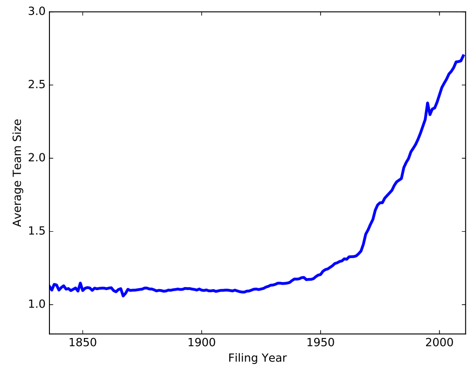


Figure A.46. The figure shows the average time (in years) that a patent application filed in a certain year had to wait before being granted.

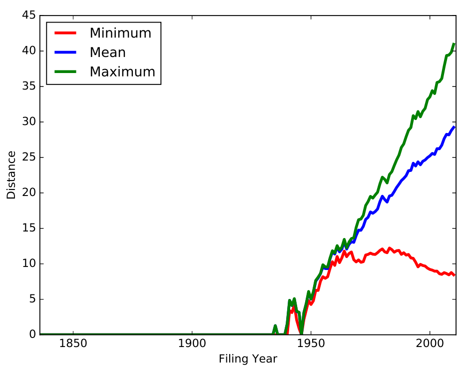


(a)

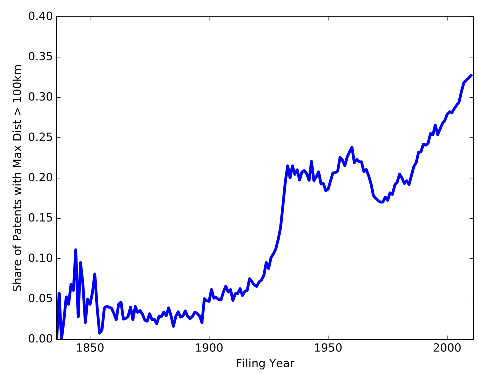


(b)

Figure A.47. The figure shows the decline of single-authored patents over time. Panel a reports the share of patents filed by a single inventor by filing year. Panel b the average number of inventor for each filing year.



(a)



(b)

Figure A.48. This figure analyzes the distance patterns among the inventors of a same patent. The left panel shows the minimum, mean, and maximum distance across the inventors of the median patent. The right panel reports the share of patents with at least two inventors that live more than 100 km apart.

Figure A.49. The figure shows the share of patents for which at least two inventors live in the same city.

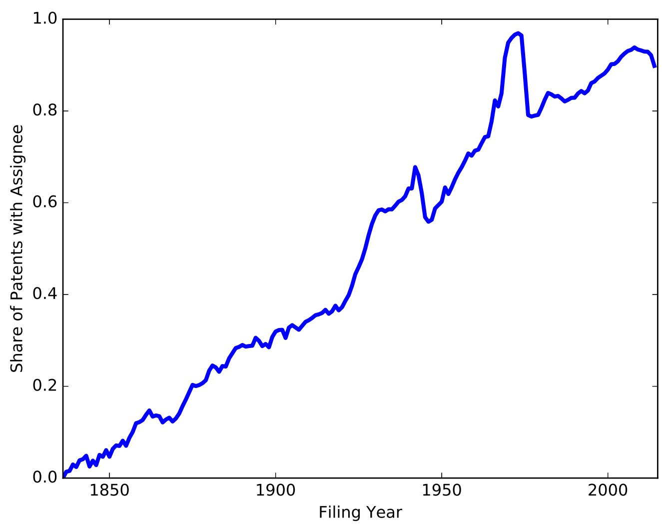
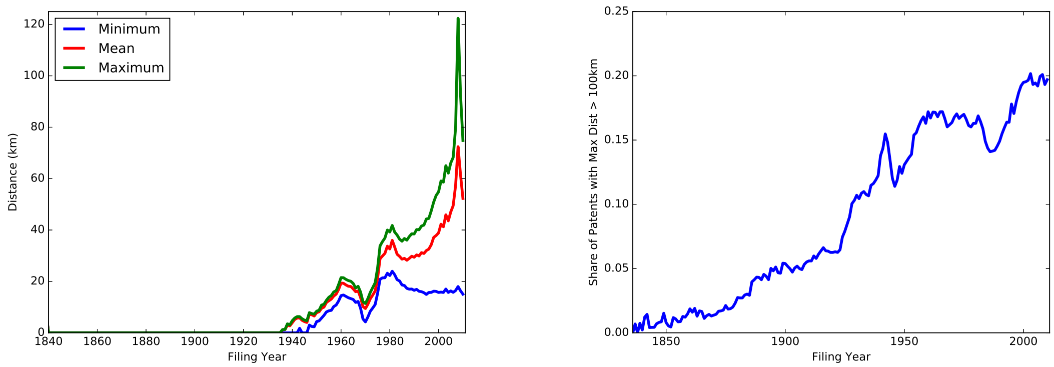
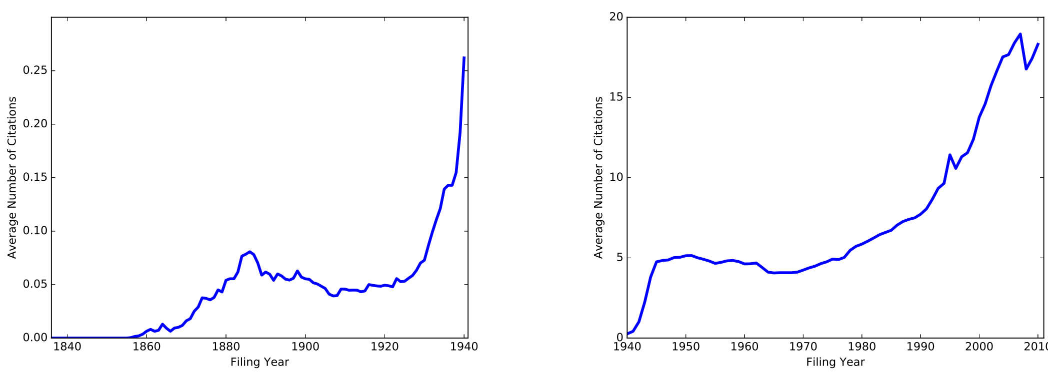


Figure A.50. The figure shows the share of patents by filing year that were assigned, in full or in part, to at least one person (or company) different from the inventors.



(a) (b)

Figure A.51. The figure analyzes the distance patterns between assignees and inventors. The left panel shows the minimum, mean, and maximum distance between the inventors and assignee fo the median patent. The right panel reports the share of patents for which the distance between the assignee and at least one of the inventors is larger than 100 kilometers.



(a) Pre-1940 (b) Post-1940

Figure A.52. The two figures report the average number of citation by filing year. The left panel shows the series for the years between 1836 and 1940, whereas the right panel for the years after 1940.

Figure A.53. The figure shows the share of patents that have not received any citation at any point in time after being filed.

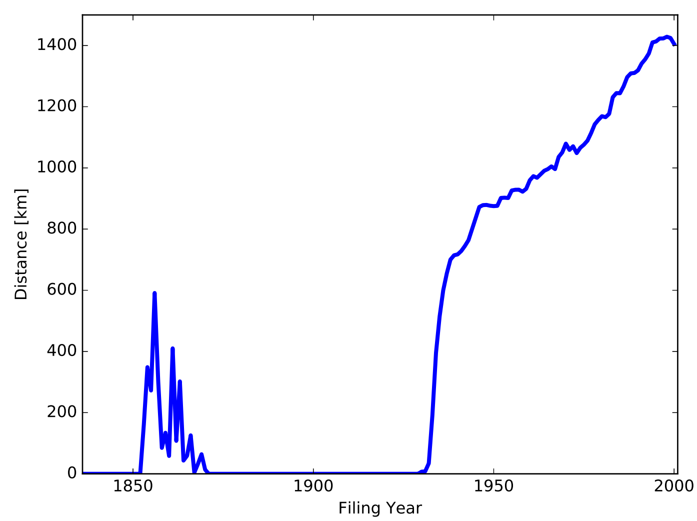


Figure A.54. The figure shows the average distance of citations received by the median patent in the 10 years following its filing.

APPENDIX B

Tables

	1990	2000	2010
Overall Gini	**42.8**	**46.2**	**47.0**
Across CTs - Within CZs (Segregation)	**19.5**	**20.6**	**22.5**

Table B.1. The overall Gini is obtained from the FRED website. The data sources and methodology for the across-CT and segregation measures are explained in the text.

	Dep. Variable: Change in Segregation (Gini), 1990-2010					
	(1)	(2)	(3)	(4)	(5)	(6)
Patenting Growth	1.27***	0.84***	0.93***	0.63**	0.62**	0.64**
	(0.23)	(0.32)	(0.30)	(0.26)	(0.25)	(0.28)
# CT		2.22**	3.21**	4.36***	4.35***	4.29***
		(0.99)	(1.33)	(1.07)	(1.09)	(1.14)
# of Household			-2.13	-3.43**	-3.39**	-3.43**
			(1.58)	(1.33)	(1.35)	(1.45)
Income				8.20***	8.02***	8.21***
				(1.91)	(1.92)	(1.97)
Import Exposure					0.01	0.01
					(0.03)	(0.03)
Local Govt Spending						-0.06
						(0.39)
# obs.	703	703	703	703	687	579
R^2	0.10	0.14	0.16	0.23	0.23	0.23

Table B.2. All regressions are weighted by total number of households in 1990. Controls are in growth rates, 1990-2010. Missing observations in columns (5) and (6) reflect data availability at the source and are concentrated in low population regions. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable: Change in Segregation (Gini), 1990-2010					
	(1)	(2)	(3)	(4)	(5)	(6)
Patenting Growth	2.88***	2.87***	2.84***	2.34***	2.41***	2.40***
	(0.47)	(0.70)	(0.69)	(0.68)	(0.71)	(0.70)
# CT		0.05	1.85	2.45*	2.72*	2.65*
		(1.39)	(1.65)	(1.47)	(1.53)	(1.58)
# of Household			-3.49*	-4.19**	-4.14**	-3.96**
			(1.93)	(1.70)	(1.73)	(1.84)
Income				5.65***	5.20**	5.54**
				(2.11)	(2.23)	(2.28)
Import Exposure					-0.02	-0.03
					(0.04)	(0.04)
Local Govt Spending						-0.21
						(0.34)
# obs.	703	703	703	703	687	579
	First-stage estimates					
Predicted	0.72***	0.60***	0.60***	0.57***	0.56***	0.57***
Patenting Growth	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
Wald F stat.	388.38	233.96	247.03	205.88	192.35	173.07
R^2	0.36	0.41	0.43	0.44	0.44	0.46

Table B.3. 2SLS estimates. All regressions are weighted by total number of households in 1990. First-stage estimates include all the controls specific to the model. Controls are in growth rates, 1990-2010. Missing observations in columns (5) and (6) reflect data availability at the source and are concentrated in low population regions. Robust standard errors in parentheses. ***p $<$ 0.01, **p $<$ 0.05, *p $<$ 0.1.

Dep. Variable: Change in Segregation, 1990-2010

Area A	3.39***	Area B	4.18	Area C	1.52**	Area D	-1.27***
	(0.99)		(2.77)		(0.64)		(0.46)
Controls	✓	Controls	✓	Controls	✓	Controls	✓
Area E	-3.95	Area F	-1.59	Area G	4.37***	Area H	2.88***
	(5.64)		(1.13)		(1.33)		(0.92)
Controls	✓	Controls	✓	Controls	✓	Controls	✓

Table B.4. 2SLS estimates. All regressions are weighted by total number of households in 1990. Controls are included in growth. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable:				
	$\Delta Ineq_{cz}$	$\Delta Segr_{cz}$		$\Delta CT - Ineq$	
	(1)	(2)	(3)	(4)	(5)
Patenting Growth	0.93**	2.34***	1.48***	-0.86**	0.36
	(0.46)	(0.68)	(0.44)	(0.35)	(0.35)
$\Delta Ineq_{cz}$			0.96***		
			(0.08)		
# obs.	703	703	703	703	703
Controls (Growth)	✓	✓	✓	×	✓

Table B.5. 2SLS estimates. All regressions are weighted by total number of households in 1990. Controls are included in growth. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable: Δs_j^k	
	(1) - OLS	(2) - IV
$rank_j$ $\times$ Patenting Growth	0.87**	1.75***
	(0.41)	(0.60)
$rank_j$	2.29***	2.12***
	(0.21)	(0.21)
CZ Fixed effects	✓	✓
# obs.	57,285	57,285

Table B.6. Regressions are weighted by total number of workers in 1990. Standard errors are clustered at the CZ level. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable: $\Delta Segr$			
	(1) - OLS	(2) - OLS	(3) - IV	(4) - IV
Patenting Growth	0.65**	0.78***	1.99***	2.11***
	(0.30)	(0.30)	(0.65)	(0.68)
Persistent Amenities		0.32		0.29
		(0.23)		(0.21)
Persistent Amenities $\times$		-0.45		-0.63*
$\times$ Patenting Growth		(0.29)		(0.35)
Controls (Growth)	✓	✓	✓	✓
# obs.	337	337	337	337

Table B.7. All regressions are weighted by total number of households in 1990. Controls are included in growth. Number of observations reflect data availability from Lee and Lin (2017). The index of persistent amenities is normalized to have a mean of zero and a standard deviation of one. Robust standard errors in parenthesis. ***p < 0.01, **p < 0.05, *p < 0.1.

Assigned Parameters		*Structural Estimation*			
α	0.80	ω_{nn}	0.04	ω_{nk}	0.18
β	0.75	ω_{kn}	-0.02	ω_{kk}	0.32
κ	0.01	λ_{kn}	0.12	λ_{kk}	0.31
ν_c	Figure A.15	ρ_n	0.467	ρ_k	0.497
		δ_k	0.055	θ_0	-0.004
		θ_1	0.001		

Table B.8. Parameter values

	Dep. Variable:					
	Δ Occ-Gini		Δ *Segr*			
	Model	Data	Model		Data	
Bin	0.60***	0.51***	0.27***	0.18***	0.22***	0.14***
	(0.22)	(0.15)	(0.04)	(0.02)	(0.04)	(0.04)
Δ Occ-Gini				0.14***		0.15***
				(0.65)		(0.45)
# obs.	663	663	663	663	663	663
R^2	0.02	0.24	0.12	0.68	0.10	0.14

Table B.9. All regressions are weighted by total number of households in 1990. Robust standard errors in parentheses. ***p $<$ 0.01, **p $<$ 0.05, *p $<$ 0.1.

	# Obs	Min	Max	Mean	SD
Patents per 1,000p (CZ, 1985-1994)	703	0	22.3	2.05	1.50
Patents per 1,000p (CZ, 2005-2014)	703	0	29.4	2.41	2.97
Average HH income (CZ, 1990)	703	17,776	64,369	39,688	8,940
Average HH income (CZ, 2010)	703	39,021	140,656	77,108	17,770
Average HH income (CT, 1990)	59,525	5,000	558,810	39,687	20,638
Average HH income (CT, 2010)	72,236	5,000	640,456	77,107	45,419
Number of CTs (CZ, 1990)	703	1	2,728	603.3	760.5
Number of CTs (CZ, 2010)	703	1	3,890	728.7	948.2
Average Rent (CZ, 1990)	703	129.2	688.2	396.4	124.8
Average Rent (CZ, 2010)	703	231.6	2,020.4	903.6	334.8
Average Rent (CT, 1990)	59,383	99.5	1,500	398.5	183.6
Average Rent (CT, 2010)	72,007	99.5	2336.7	916.4	521.2
Segregation (CZ, 1990)	703	0	27.0	19.5	5.2
Segregation (CZ, 2010)	703	0	32.3	22.5	6.0
Inequality (CZ, 1990)	703	36.2	49.8	44.3	2.1
Inequality (CZ, 2010)	703	36.2	52.4	45.6	2.3

Table B.10. Summary statistics (weighted by total HH in respective year).

	Dep. Variable: Change in Segregation (Theil), 1990-2010					
	(1)	(2)	(3)	(4)	(5)	(6)
Patenting Growth	1.26***	0.75**	0.81***	0.50*	0.52*	0.51*
	(0.24)	(0.32)	(0.30)	(0.27)	(0.29)	(0.30)
# CT		2.60**	3.35**	4.59***	4.56***	4.54***
		(0.95)	(1.30)	(1.06)	(1.08)	(1.13)
# of Household			-1.61	-3.01**	-3.04**	-3.22**
			(1.57)	(1.33)	(1.34)	(1.45)
Income				8.80***	8.64***	8.97***
				(2.07)	(2.10)	(2.17)
Import Exposure					-0.02	-0.02
					(0.03)	(0.03)
Local Govt Spending						0.05
						(0.40)
# obs.	703	703	703	703	687	579
R^2	0.10	0.16	0.17	0.25	0.25	0.25

Table B.11. All regressions are weighted by total number of households in 1990. Controls are in growth rates, 1990-2010. Missing observations in columns (5) and (6) reflect data availability at the source and are concentrated in low population regions. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	1995	1996	...	2003	2004	$\widehat{2005}$	$\widehat{2006}$	...	$\widehat{2013}$	$\widehat{2014}$
$\widehat{2005}$	d_{10}	d_9	...	d_2	d_1			...		
$\widehat{2006}$		d_{10}	...	d_3	d_2	d_1		...		
...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
$\widehat{2013}$			...	d_{10}	d_9	d_8	d_7	...		
$\widehat{2014}$			...		d_{10}	d_9	d_8	...	d_1	

Table B.12. Structure and timing of the instrument. Years with a hat are predicted, years without a hat are actual.

	Dep. Variable: Change in Segregation, 1990-2010				
	(1)	(2)	(3)	(4)	(5)
Patenting Growth	1.27***	0.97***	0.70***	0.61**	0.56**
	(0.23)	(0.23)	(0.25)	(0.24)	(0.25)
# CT		0.51***	-2.12**	-2.25**	-2.64**
		(1.13)	(1.06)	(1.06)	(1.07)
# of Household			2.55**	2.54**	3.19***
			(1.05)	(1.08)	(1.07)
Income				1.09	1.55
				(1.21)	(1.22)
Local Govt Spending					-0.28*
					(0.16)
# obs.	703	703	703	703	643
R^2	0.10	0.22	0.26	0.27	0.28

Table B.13. All regressions are weighted by total number of households in 1990. Missing observations in columns (4) and (8) reflect data availability at the source and are concentrated in low population regions. Controls are included as the log value in 1990. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable: Change in Segregation, 1990-2010					
	(1)	(2)	(3)	(4)	(5)	(6)
Patenting Growth (1990-2010)	2.34***	2.59***	3.28***	2.73***	1.69***	1.91*
	(0.47)	(0.72)	(1.10)	(0.73)	(0.54)	(0.50)
Patenting Growth (1980-1990)		-1.08**				-0.50*
		(0.45)				(0.30)
Bartik-like variable			-1.38			-0.01
			(1.26)			(1.02)
Constrained Instrument	No	No	No	Yes	No	No
State-year fixed effects	No	No	No	No	Yes	Yes
Baseline controls (Growth)	✓	✓	✓	✓	✓	✓
# obs.	703	703	690	703	703	690
	First-stage estimates					
Predicted Patenting Growth	0.57***	0.55***	0.55***	0.48***	0.54***	0.46***
	(0.07)	(0.06)	(0.11)	(0.07)	(0.06)	(0.09)
Wald F-stat	205.88	188.22	57.40	135.34	151.75	38.88
R^2	0.44	0.44	0.44	0.39	0.61	0.61

Table B.14. 2SLS estimates. All regressions are weighted by total number of households in 1990. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable: Change in Segregation, 1990-2010				
	(1)	(2)	(3)	(4)	(5)
Patenting Growth	2.88***	1.96***	1.62***	1.62***	1.49***
	(0.47)	(0.45)	(0.50)	(0.52)	(0.50)
# CT		0.44***	-1.31	-1.31	-1.70
		(0.13)	(1.28)	(1.24)	(1.20)
# of Household			1.70	1.70	2.30*
			(1.28)	(1.25)	(1.20)
Income				-0.01	0.49
				(1.29)	(1.32)
Local Govt Spending					-0.24
					(0.16)
# obs.	703		703	703	643
	First-stage estimates				
Predicted Patenting Growth	0.72*** (0.07)	0.76*** (0.08)	0.71*** (0.08)	0.65*** (0.08)	0.64*** (0.08)
Wald F-stat	388.38	348.74	269.11	196.85	173.07
R^2	0.36	0.36	0.38	0.39	0.40

Table B.15. 2SLS estimates. All regressions are weighted by total number of households in 1990. Missing observations in columns (4) and (8) reflect data availability at the source and are concentrated in low population regions. Controls are included as the log value in 1990. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable:							
	Δ Edu-Gini				Δ Occ-Gini			
	OLS		IV		OLS		IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Patenting	1.38***	0.86**	2.34***	1.61**	1.81***	1.97***	4.41***	5.91***
Growth	(0.31)	(0.33)	(0.56)	(0.70)	(0.35)	(0.34)	(0.57)	(0.90)
# obs.	703	703	703	703	703	703	703	703
Controls	×	✓	×	✓	×	✓	×	✓

Table B.16. 2SLS estimates. All regressions are weighted by total number of households in 1990. Controls are included in growth. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable: $\Delta Segr_{cz}$			
	Past Trend		1990-2010	
	(1)	(2)	(3)	(4)
Predicted	0.02	0.27		
Patenting Growth	(0.29)	(0.30)		
Patenting Growth			2.31***	1.41**
			(0.63)	(0.64)
# obs.	309	309	309	309
Controls (Growth)	×	✓	×	✓

Table B.17. All regressions are weighted by total number of households in 1990. Controls are in growth rates, 1990-2010. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

	Dep. Variable: Δs_j^k	
	(1) - Model	(2) - Data
$rank_j$ × Bin	2.89***	0.160**
	(0.31)	(0.063)
$rank_j$	-20.39***	1.52***
	(2.39)	(0.39)
CZ Fixed effects	✓	✓
# obs.	58,050	57,197

Table B.18. Regressions are weighted by total number of workers in 1990. Standard errors are clustered at the CZ level. ***p $<$ 0.01, **p $<$ 0.05, *p $<$ 0.1.

	Median Tail-conventionality					
	(1)	(2)	(3)	(4)	(5)	(6)
Log population density	-1.10***	-1.10***			-0.78***	-0.81***
	(0.38)	(0.31)			(0.19)	(0.14)
Log college-graduate density			-0.87***	-0.92***		
			(0.33)	(0.29)		
State/year f.e.	no	yes	no	yes	no	no
Weighted	Pat	Pat	Pat	Pat	no	Pop
N. Obs	18,095	18,095	18,095	18,095	18,095	18,095
R^2	0.02	0.08	0.013	0.13	0.003	0.01

Table B.19. The dependent variable is defined as the tail-conventionality of the median patent in the CSD-year observation. All regressions, except for (5) and (6), are weighted by the total number of patents filed in the CSD/Year observation. Standard errors in all the regressions are clustered at the CSD level. U-scores are winsorized (1%) at the patent level. ***p $<$ 0.01, **p $<$ 0.05, *p $<$ 0.1.

	Median Tail-conventionality			
	(1)	(2)	(3)	(4)
Log population density	-1.10***	-0.83***	-0.80***	-0.56**
	(0.31)	(0.26)	(0.27)	(0.27)
Log median income		2.01***	2.7***	1.82*
		(0.67)	(0.99)	(1.0)
% College Graduates			-0.028	-0.0173
			(0.022)	(0.0234)
Gini				-0.14
				(0.105)
State/year f.e.	yes	yes	yes	yes
Weighted	Pat	Pat	Pat	Pat
N. Obs	18,095	18,095	18,095	17,995
R^2	0.08	0.09	0.10	0.10

Table B.20. 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.

	Unconventional Tail		
	(1)	(2)	(3)
Log population density	0.0087**	0.0074**	0.0105***
	(0.0037)	(0.0033)	(0.0038)
Publicly Traded		-0.0161**	-0.0109
		(0.0068)	(0.0080)
Log total patents			-0.0038*
			(0.0022)
State/year/class f.e.	yes	yes	yes
N. Obs	1,059,999	706,469	706,469
Pseudo R^2	0.007	0.007	0.008

Table B.21. 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.

Percentage of citations to class $\mathcal{A}$ from class $\neq \mathcal{A}$

	(1)	(2)	(3)	(4)
Arrival of new firm of class $\mathcal{A}$	0.49*** (0.0002)	0.43*** (0.0002)	0.92*** (0.0011)	0.65*** (0.0002)
Class-CSD f.e.	yes	yes	yes	yes
Class-Year f.e.	no	yes	no	yes
Shock arrival year	2001	2001	2005	2005
Average $\bar{S}$	0.43	0.43	0.43	0.43
N. Obs	682,116	682,116	682,116	682,116
within R^2	0.003	0.006	0.001	0.005

Table B.22. This table reports the coefficients of a regression of the share of citations received by patent class $\mathcal{A}$ from patents of classes other than $\mathcal{A}$ in a given CSD at a given time on time/class and class/CSD fixed effects and the cumulative normalized arrival of new firms of class $\mathcal{A}$ in that CSD. Columns 2 and 4 include time/class fixed effects. Columns 3 and 4 only include incoming firms on or after 2005. Standard errors clustered at the CSD/class level are reported in parenthesis. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

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.23. 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	Mean	Min	Max	Winsor	Weight	# Obs.
Tail Conventionality	Patent	85	40	164	1%	No	1,058,999
Core Conventionality	Patent	108	61	183	1%	No	1,058,999
Median Tail Conventionality	CSD	77	40	164	No	Pat	18,095
Patents per capita	CSD	5.e-4	3.e-6	0.641	No	Pop	18,095
Patents per capita (winsor)	CSD	5.e-4	1.e-4	0.011	1%	Pop	18,095
Density of population ($/km^2$)	CSD	1966	0.931	26821	No	Pop	18,095

Table B.24. Summary statistics for the main variables used in the analysis.

	Median Tail Conventionality	
Log population density	-1.31***	-1.36***
	(0.32)	(0.35)
Chicago		1.13
		(1.00)
Boston		-3.48**
		(1.72)
New York		1.18
		(0.97)
San Francisco		1.45***
		(0.55)
State/year f.e.	yes	yes
N. Obs	18,095	18,095
R^2	0.14	0.14

Table B.25. 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. ***p < 0.01, **p < 0.05, *p < 0.1.

	HistPat	CUSP
Years Covered	1836-1976	1836-2015
Inventor First Name	N	Y
Inventor Last Name	N	Y
Inventor Town	N	Y
Inventor County	Y	Y
Inventor State	Y	Y
Full Patent Text	N	Available upon request
Patent Number	Y	Y
Application Date	N	Y
Grant Date	Y	Y
Names of Multiple Inventors	N	Y
Names of Assignees	N	Y
Assignee Town	N	Y
Assignee County & State	Y	Y
Patent Class	N	Y
Backward and Forward Citations	N	Y

Table B.26. The table shows a schematic comparison of the variables available in HistPat and CUSP.

Decade	Leading Class	Description
1836-1845	E02	Hydraulic Engineering; Foundations
1846-1865	F16	Engineering Elements or Units
1866-1875	A01	Agriculture
1876-1885	D05	Sewing
1886-1895	B41	Printing
1896-1905	D03	Weaving
1906-1945	F16	Engineering Elements or Units
1946-1995	A61	Medical or Veterinary Science; Hygiene
1996-2015	G06	Computing; Calculating; Counting

Table B.27. The table reports the leading technology for each decade from 1836 to 2015. A leading technology is defined as the most frequent technological class in the top percentile of the distribution of citations received.

APPENDIX C

Appendices

C.1. Appendix to Chapter 1

C.1.1. Data description

C.1.1.1. Income distribution at the CT level. The NHGIS provides information on yearly household income at the CT level by dividing residents into 15 income bins. The lower bounds of each income bin are: 0$, 15,000$, 20,000$, 25,000$, 30,000$, 35,000$, 40,000$, 45,000$, 50,000$, 60,000$, 75,000$, 100,000$, 125,000$, and 150,000$. In order to measure inequality and segregation, we need to approximate the income distribution. For each bracket except for the top one, we assume that all households have income equal to the midpoint of the bracket. The top bin is unbounded, with an average that potentially varies substantially across CTs, and our measures will depend on the assumptions made on the distribution of income in the top bracket. The literature has dealt with this issue by either fitting the parameters of an income distribution (usually assumed to be Pareto) or assuming that the average is a fixed percentage above the amount reported in top coded data (usually 40-50% more).[1] These two methods have been subject to several critics.[2]

[1]See for example Autor, Katz, and Kearney (2008) and Lemieux (2006).

[2]Critics of the former approach have argued that if the underlying distribution is far from the assumed one, a researcher would obtain better results by taking the bin averages. Critics of the latter have pointed to the fact that the assumption of the average income for the last bin is somewhat

For our analysis, we design an alternative approach to assign a value to the top bin, and validate our procedure by comparing the resulting segregation index with the corresponding index we obtain by using information on average personal income, that does not require to make arbitrary assumptions. First, the 5-year 2008-2012 ACS provides CT-level Gini indices using households as basic unit of analysis. For each census tract in 2010, we set the average of the top bin so that the resulting Gini matches the one reported in the ACS.[3] Second, we use the time series of individual-level Gini data at a state level computed by Frank (2009). From there we collect estimates for the Gini index for all the states in 1990 and 2010 and calculate the percentage change. Assuming that the state trends for individual-level Gini are mirrored by the corresponding CT trends for household-level Gini, we set the average income in the top bin so that the percentage change in the Gini index is equal to the one in Frank (2009).[4]

To validate our procedure further, in Figure A.9 we show the correlation between segregation in 1990 and 2010, respectively, using the household income distribution

arbitrary. Different methods to deal with binned income data have been reviewed by Von Hippel et al. (2014).

[3]Note that in 3609 out of 98032 CTs (3.7%) there is no value that allows us to exactly match the Gini reported in the ACS. This might be due to measurement errors or the approximation that all the households earn the average of the income braket. In this case, our algorithm diverges, either assigning values that are too low (i.e., smaller than 150,000$ which is the lower bound of the top bin) or too high (i.e., bigger than 1,000,000$). When this happens we assign to the CTs in question a default value of 200,000$ which is in line with the 1.4-rule. We experimented with different default values and the main results are robust. Another 908 CTs (or 0.9%) appear in the income data but not in the Gini data. In that case, we try to match the 2010 national Gini (0.48).

[4]We are not able to match 20,966 (or 21%) of the 1990 CTs with the 2010 data. In this case, we assume that their Gini is the same as the national one in 1990 (0.43). As we did in 2010, when the algorithm diverges or estimates an implausible value, we assign to the top bin a default value of 200,000$.

approximated using the procedure described above, and the same measure computed using average personal income at the CT level, which does not require to make arbitrary assumptions on the distribution of income within brackets. The correlation between the two variables is equal to 90% in 1990 and 91% in 2010.

C.1.1.2. Other data sources.

Distribution of residents and workers by occupation. The distribution of residents by occupation at the CT level is constructed as follows. First, from the NHGIS we obtain information on the CT-level distribution of residents according to a coarse definition of occupations, comprising 13 occupations in 1990 and 25 occupations in 2010. Then, using the IPUMS, we construct a CZ-specific crosswalk that maps the coarse definition of occupation into the fine one (386 occupations in 1990 and 454 in 2010). To this end, we exploit the CZ-specific frequency of each fine occupation code in each coarse category. Occupations are then categorized in two classes: knowledge intensive and non-knowledge intensive. These two categories are defined according to Florida (2017) definition of creative class: "*The creative class is made up of workers in occupations spanning computer science and mathematics; architecture and engineering, the life, physical, and social sciences; the arts, design, music, entertainment, sports, and media; management, business, and finance; and law, health care, education, and training.*" (p. 217).[5]

We assign workers to workplaces using the National Establishment Time Series (NETS). This data set contains information about employment for the universe of

[5]The precise list of occupations that fall into the knowledge intensive category for 1990 and 2010 is available upon request.

establishments between 1990 and 2010, as well as their location and NAICS code. The latitude and longitude is provided at 5 geographical levels (namely block face, block group, census tract centroid, ZIP code centroid or street level). We allocate workers to each census tract according to the following procedure. First, we assign to a census tract those establishments whose geographical coordinates are provided at a block face, block group or census tract centroid level.[6] Second, we assign the workers of each establishment geo-located at ZIP code level based on the area of the census tracts it contains.[7] We discard all those establishments whose coordinates are missing or are at a street level.[8] This gives us an estimate of workers per NAICS at a census tract level.

Since the NETS is a relatively new data set in the literature and there might be some concerns related to its validity, before assigning each NAICS to a distribution of occupations, we compare our employment estimates with the distribution of workers obtained from the ZIP Code Business Patterns (ZBP) provided by the Census Bureau. We first aggregate the employment data obtained from the NETS data at a ZIP code level and we then check whether they systematically differ in the two data sets. Note that we do expect them to somewhat differ for various reasons. For example, the ZBP does not consider workers that are employed by the public sector. Therefore, the number of workers in ZIP codes that contain public universities

[6] 7,573,637 establishments were assigned this way in 1990; 28,111,455 in 2010.

[7] For example, if a certain ZIP code cotains two census tracts that cover 40% and 60% of its area, respectively, we assign 40% of the employment of an establishment assigned to that ZIP code to the first census tract and 60% to the second one. In 1990, 3,002,490 establishments were assigned this way; 2,457,796 in 2010.

[8] 156,185 establishments were discarded in 1990; 332,091 in 2010.

or government buildings is likely to be significantly lower in the ZBP.[9] Figure A.10 shows the correlation between the workers estimated using the ZBP (x-axis) and the NETS (y-axis) in 1994 (left panel) and 2010 (right panel).[10] As we expected, the NETS systematically reports more workers than the ZBP, although the two measures are very close. Interestingly and in line with our prior expectations, the difference between the two employment estimates is highest in ZIP codes that contain public universities or government buildings. For example, the three largest differences in 1994 come from ZIP codes 90012, 43215 and 77002 (92,662 vs. 20,667; 159,815 vs. 80,413; and 159,847 vs. 77,565, respectively). ZIP code 90012 contains the Los Angeles City Hall as well as other government buildings (e.g., the California Department of Transportation's offices), the Ohio Statehouse is located in ZIP code 43215, and ZIP code 77002 contains the Houston City Administration. In 1994 the NETS reports an estimate of 16,336 workers for ZIP code 94720 (UC Berkeley), whereas the ZBP of only 1,028.

Finally, we use the Occupational Employment Statistics (OES) provided by the Bureau of Labor Statistics (BLS) to get an estimate of the occupational distribution of workers in each census tract. The OES reports the percentage of workers active in a certain occupation for each NAICS (SIC90 for 1990) code.[11] Similarly to what

[9]Some other NAICS codes, as for example agriculture, are excluded from the ZBP and the sampling frame differs in the two data sets. For more details, see http://www.exceptionalgrowth.org/downloads/NETSvsBLS_DataCollectionDifferences.pdf

[10]We used 1994 instead of 1990, since this is the first year for which the ZIP Code Business Patterns was made available.

[11]Note that since in the 90s only certain industry codes were reported in different years, we built the crosswalk for 1990 using OES data from 1990 to 1993. Also, since the data are provided for SIC (instead of NAICS) codes, we first build a crosswalk from NAICS to SIC and we then use the appropriate distributions reported in the OES.

we did for the residents, the occupations are then categorized in the two classes according to their knowledge intensity.

Rent. Housing rent at the CT level is computed as the average rent for a one bedroom apartment. The NHGIS provides rent data in brackets, as numer of apartments leased for less than \$200, \$300, \$500, \$750, \$1,000 and for more than \$1,000. We assign to all the apartments in each bin except for the top one the midpoint value of the bracket. For the top bin, we set it to \$1,500 in 1990 and \$2,250 in 2010, assuming an approximate growth of rent in the top bin of 2% per year.

Data on rent are not available for 6,535 CTs out of 61,258 in 1990, and for 12,862 CTs out of 74,001 in 2010. To complete the dataset, we extrapolate the missing values by running a regression of log average rent on log income, a third-degree polynomial of density and log median house prices, and applying the estimated coefficients to the observations with missing rent. This reduces the number of missing observations to 1,874 in 1990 and 1,993 in 2010. All the missing observations are concentrated in low population CTs.

Commuting time and flows. Commuting flows are collected from the Longitudinal Employer-Household Dynamics (LEHD) dataset.[12]. The LEHD collects data about bilateral commuting flows from and to each Census Block starting from 2002.[13] These data are used to estimate the commuting flows/commuting times semi-elasticities using the gravity equation (1.22) obtained from the structural model. Since we

[12]https://lehd.ces.census.gov/

[13]See https://lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.2.pdf for more details. Note that some years are missing for some states.

assume in our model that the semi-elasticities of commuting are stable over the period 1990-2010 and given the data availability, we collect commuting flows for 2010 for all the states (with the exception of Massachusetts for which data are only available from 2011 onward). Data at a block level are then aggregated to obtain commuting flows at our preferred level of geographical aggregation (i.e., 1990 CTs).

Commuting times between each pair of CTs are calculated using driving times between the centroids of each Census Tract. Because of the high number of possible combinations we were unable to use commercial routing services (e.g., Google Maps) and we relied on the Open Source Routing Machine (OSRM).[14] The advantage of using the OSRM is that it is possible to run it locally. This allows us to send queries without limits and in parallel. In particular, it was possible to collect data on commuting times for each pair of neighborhoods withing each city (for a total of 32.4 million pairs) in just few hours. The disadvantage is that the OSRM does not contain any data on traffic (and in particular traffic during rush hours) which might underestimate the actual commuting times/costs faced by workers.[15]

C.1.2. Derivations

C.1.2.1. Derivation of (1.9) and (1.12). The probability that an agent of type x commutes from neighborhood i to neighborhood j can be derived as follows:

[14]http://project-osrm.org/

[15]Note that, because of the lack of traffic data, commuting from A to B always takes the same time as commuting from B to A. The commuting matrices are therefore symmetric which reduces the number of queries necessary to populate them to 16.2 million.

$$
\begin{aligned}
\pi_{ij}^{x} &= P\left(u_{ijo}^{x} \geq \max_{l,m\in\mathcal{S}\times\mathcal{S}\setminus\{i,j\}} u_{lmo}^{x}\right) \\
&= \int_{0}^{\infty} \varepsilon u^{-\varepsilon-1}\left(s_{ij}^{x}\right)^{\varepsilon} e^{-\left(s_{ij}^{x}\right)^{\varepsilon}u^{-\varepsilon}} P\left(u \geq \max_{l,m\in\mathcal{S}\times\mathcal{S}\setminus\{i,j\}} u_{lmo}^{x}\right) du \\
&= \int_{0}^{\infty} \varepsilon u^{-\varepsilon-1}\left(s_{ij}^{x}\right)^{\varepsilon} e^{-\left(s_{ij}^{x}\right)^{\varepsilon}u^{-\varepsilon}} P\left(\bigcap_{l,m\in\mathcal{S}\times\mathcal{S}\setminus\{i,j\}} u \geq u_{lmo}^{x}\right) du \\
&= \int_{0}^{\infty} \varepsilon u^{-\varepsilon-1}\left(s_{ij}^{x}\right)^{\varepsilon} e^{-\left(s_{ij}^{x}\right)^{\varepsilon}u^{-\varepsilon}} \prod_{l,m\in\mathcal{S}\times\mathcal{S}\setminus\{i,j\}} P\left(u \geq u_{lmo}^{x}\right) du
\end{aligned}
$$

which, using the expression for the Frechet distribution, yields:

$$
\begin{aligned}
\pi_{ij}^{x} &= \int_{0}^{\infty} \varepsilon u^{-\varepsilon-1}\left(s_{ij}^{x}\right)^{\varepsilon} e^{-\left(s_{ij}^{x}\right)^{\varepsilon}u^{-\varepsilon}} \prod_{l,m\in\mathcal{S}\times\mathcal{S}\setminus\{i,j\}} e^{-\left(s_{lm}^{x}\right)^{\varepsilon}u^{-\varepsilon}} du \\
&= \int_{0}^{\infty} \varepsilon u^{-\varepsilon-1}\left(s_{ij}^{x}\right)^{\varepsilon} e^{-\sum_{l,m}\left(s_{lm}^{x}\right)^{\varepsilon}u^{-\varepsilon}} du \\
&= \frac{\left(s_{ij}^{x}\right)^{\varepsilon}}{\sum_{l,m\in\mathcal{S}\times\mathcal{S}} \left(s_{lm}^{x}\right)^{\varepsilon}}
\end{aligned}
$$

which implies:

$$
\pi_{ij}^{x} = \frac{\left(\frac{B_i^x w_j^x \left(q_i^x\right)^{\beta-1}}{d_{ij}}\right)^{\varepsilon}}{\sum_{l,m\in\mathcal{S}\times\mathcal{S}} \left(\frac{B_l^x w_m^x \left(q_l^X\right)^{\beta-1}}{d_{lm}}\right)^{\varepsilon}} \equiv \frac{\Phi_{ij}^{x}}{\Phi^{x}}
$$

where $s_{ij}^{x} \equiv \frac{u_c^x B_i^x w_j^x \left(q_i^x\right)^{\beta-1}}{d_{ij}}$.

The probability that an agent of type x commutes to neighborhood j, conditional on living in neighborhood j, can be derived as follows:

$$\begin{aligned}
\pi^x_{ij|i} &= P\left(u^x_{ijo} \geq \max_{m\in S\backslash\{j\}} u^x_{imo}\right) \\
&= \int_0^\infty \varepsilon u^{-\varepsilon-1}\left(s^x_{ij}\right)^\varepsilon e^{-\left(s^x_{ij}\right)^\varepsilon u^{-\varepsilon}} P\left(u \geq \max_{m\in S\backslash\{j\}} u^x_{imo}\right) du \\
&= \int_0^\infty \varepsilon u^{-\varepsilon-1}\left(s^x_{ij}\right)^\varepsilon e^{-\sum_m \left(s^x_{im}\right)^\varepsilon u^{-\varepsilon}} du \\
&= \int_0^\infty \varepsilon u^{-\varepsilon-1}\left(s^x_{ij}\right)^\varepsilon e^{-\sum_m \left(s^x_{im}\right)^\varepsilon u^{-\varepsilon}} du \\
&= \frac{\left(w^x_j/d_{ij}\right)^\varepsilon}{\sum\limits_{m\in\mathcal{S}} \left(w^x_m/d_{im}\right)^\varepsilon}.
\end{aligned}$$

C.1.3. Details on the Structural Estimation

We estimate the structural parameters of the model using the moment conditions described in (1.28). In particular, we need to estimate the parameter set:

$$\boldsymbol{p} \equiv \{\rho_n,\, \rho_k,\, \delta_k,\, \omega_{nn},\, \omega_{nk},\, \omega_{kn},\, \omega_{kk},\, \lambda_{kn},\, \lambda_{kk},\, \theta\}$$

given the data matrix:

$$\boldsymbol{X} = \{\boldsymbol{R},\, \boldsymbol{W},\, \boldsymbol{Q},\, \boldsymbol{K},\, \boldsymbol{\tau}\}$$

as well as the parameters $\{\alpha,\, \beta,\, \nu_c,\, \kappa\}$.

To do this, we use a N-step GMM approach, where the loss function is given by:

$$L \equiv \boldsymbol{m}\left(\boldsymbol{X},\, \boldsymbol{p}\right)^{'}\, \boldsymbol{W}\, \boldsymbol{m}\left(\boldsymbol{X},\, \boldsymbol{p}\right)$$

where$\boldsymbol{m}\left(\boldsymbol{X},\, \boldsymbol{p}\right)$ is the value of the moment condition given the data matrix $\boldsymbol{X}$ and parameters $\boldsymbol{p}$, whereas $\boldsymbol{W}$ is a weight matrix which is updated at each step. In the first step, we set $\boldsymbol{W}$ equal to the identity matrix and estimate the parameters $\boldsymbol{p}$ that minimize L. Formally,

$$\boldsymbol{p}_{first} \equiv \arg\min_{\boldsymbol{p}} \boldsymbol{m}\left(\boldsymbol{X},\, \boldsymbol{p}\right)^{'}\, \boldsymbol{m}\left(\boldsymbol{X},\, \boldsymbol{p}\right).$$

The parameters estimated in the first step are used to estimate the optimal weighting matrix. The optimal weighting matrix, $\boldsymbol{W}$, is the White (1980) heteroskedasticity consistent matrix of standard errors:

$$\boldsymbol{W} = \boldsymbol{m}\left(\boldsymbol{X},\, \boldsymbol{p}_{first}\right)\, \boldsymbol{m}\left(\boldsymbol{X},\, \boldsymbol{p}_{first}\right)^{'}.$$

The process is repeated until convergence.

C.1.4. Recursion to Find Equilibrium After Shocks

We define the share of land commercially used by the firm of type x in neighborhood j as

$$\theta_j^x \equiv \frac{H_j^x}{L_j},$$

where L_j is the total amount of floor space available for (commercial or residential) construction in neighborhood j, that we take as exogenous.

Given starting values q_i^0, $w_j^{x,0}$, $\theta_j^{x,0}$

(1) $\pi_{ij}^x = \frac{\left(d_{ij}\left(q_i^0\right)^{(1-\beta)}\right)^{-\varepsilon}\left(B_i^x w_j^0\right)^{\varepsilon}}{\sum_{k\in\mathcal{S}}\sum_{l\in\mathcal{S}}\left(d_{kl}\left(q_k^0\right)^{(1-\beta)}\right)^{-\varepsilon}\left(B_k^x w_l^{x,0}\right)^{\varepsilon}}$

(2) $\pi_{ij|i}^x = \frac{\left(w_j^{x,0}/d_{ij}\right)^{\varepsilon}}{\sum_{l\in\mathcal{S}}\left(w_l^{x,0}/d_{il}\right)^{\varepsilon}}$

(3) $R_i^x = \sum_{l\in\mathcal{S}} \pi_{il}^x R^x$

(4) $W_j^x = \sum_{k\in\mathcal{S}} \pi_{kj}^x R^x$

(5) $Y_j^x = A_j^x \left(W_j^x\right)^{\alpha}\left(\theta_j^{x,0} H_j^x\right)^{1-\alpha}$

(6) $v_j^x = E\left(w^{x,0} \mid i\right) = \sum_{l\in\mathcal{S}} \pi_{il|i} w_l^{x,0}$

(7) $w_j^{x,1} = \frac{\alpha Y_j^x}{W_j^X}$

(8) $q_i^1 = \sum_{x\in\{k,n\}} \frac{(1-\alpha)Y_i^x + (1-\beta)v_i^x R_i^x}{L_i}$

(9) $\theta_j^{x,1} = \frac{(1-\alpha)Y_i^x}{q_j^1 L_i}$

(10) $A_j^x = a_j$

(11) $B_i^X =$

We iterate until $|q_i^0 - q_i^1|$, $\left|w_j^{x,0} - w_j^{x,1}\right|$ and $\left|\theta_j^{x,0} - \theta_j^{x,1}\right|$ are below 10^{-6} for all i, j. Otherwise, update the starting values according to:

$$q_i^2 = 0.3\, q_i^1 + 0.7\, q_i^0$$

$$w_j^{x,2} = 0.3\, w_i^{x,1} + 0.7\, w_i^{x,0}$$

$$\theta_j^{x,2} = 0.3\, \theta_i^{x,1} + 0.7\, \theta_i^{x,0}$$

C.1.5. Model-Generated Instrument for the Gravity Equation

In Section 1.4.5, we show that the equilibrium conditions of the model yield a gravity equation that can be used to estimate the semi-elasticity of commuting flows to commuting times for each city in our sample. The gravity equation has the following form:

$$\log(\pi_{ij}) = \psi_i + \zeta_j + \nu_c \tau_{ij} + \eta_{ij}$$

where $\psi_i = -\varepsilon(1-\beta)\, q_i + \varepsilon B_i$. Since, B_i is not directly observable it is not possible to use this structural identity to estimate ε. In particular, if we were trying to regress the fixed effects on the observed rents, B_i would be part of the error term and, since residential amenities also affect rents, the estimate of ε would be biased by construction. In a similar setup, Allen et al. (2017) suggest it should be possible to use the rents obtained through a model in which residential amenities are exogenous and equalized across neighborhoods as instrument for the observed rents. The rents estimated throught this procedure would be uncorrelated with B_i by construction and, if correlated with the actual rents, would constitute a valid instrument.

The 2SLS procedure

$$\psi_i = \gamma \hat{q}_i + \xi_i$$

$$q_i = \sigma q_i^{model} + \chi_i$$

gives us an unbiased estimate of $\gamma = -\varepsilon(1-\beta)$ for each city c, and since the value of β is known, from there it is possible to obtain an unbiased estimate of ε_c. Being the shape parameter of a Frechet distribution, ε_c needs to be strictly greater than 1.[16] The point estimates we obtain through these procedure are bigger than one in about 80% of cases, although values bigger than ones are included in the 0.95 confidence interval in 97% of commuting zones. The left panel of Figure A.17 shows the distribution of ε_c obtained through the 2SLS procedure after discarding the top and bottom 5% of observations. Although the distribution is clearly skewed towards the right, it is possible to see that we obtain an estimate smaller than 1 for a non-negligible share of commuting zones in our sample. The right panel of Figure A.17 shows the distribution of all the epsilons greater than 1. The majority of them (95%) is included in an interval between 1.08 and 13.06, with an average of 6.52 (weighted average: 6.00). This is consistent with the estimates obtained by Eaton and Kortum (2002) in the context of a gravity trade model. Their estimations of the shape parameter range from 3.60 to 12.86.

[16]The expected value of a Frechet distribution with shape parameters between 0 and 1 is infinity. This is a problem in our setup, since the expected utility for each agent needs to be equalized across cities.

We now calculate the value of κ_c implied by our estimates of ν_c and ε_c and see how it compares with our calibrated value of 0.01. For this exercise, we only consider commuting zones with $\varepsilon_c > 1$. The left panel of Figure A.18 shows the unweighted distribution of κ_c for the selected sample of commuting zones. All the values are contained in an interval between 0 and 0.048 with an average of 0.01 and a median of 0.007. Similarly, the right panel shows the same distribution weighted by the number of people in each city. The weighted mean and median are very close to the previous values (0.01 and 0.008, respectively).

C.2. Appendix to Chapter 2

C.2.1. C-Score: Details and Example

The c-score of the class pair $(\mathcal{A}, \mathcal{B})$ is calculated according to the following algorithm:[17]

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

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

[17]the conventionality score we also use foreign patents.

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

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

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

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

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

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

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

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

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

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

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

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

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

This generates a distribution of c-score for this specific patent (Figure A.29) from which we can extract the 10th percentile (tail-conventionality) and its median (core-conventionality).

C.2.2. Proofs and Derivations

C.2.2.1. Proof of Proposition 2.3.1. We start with the maximization problem of the developer that sets up a company town. We conjecture and verify at the end of the proof that company towns are fully specialized. We focus on the case of a $\mathcal{S}$-specialized location, as the one for $\mathcal{D}$-specialized sites is identical. Letting θ^C denote the Lagrange multiplier on the developer's participation constraint, the first order conditions of her problem can be expressed as:

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

$$\tau^C_{\mathcal{S}} = \phi.$$

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

$$N^C_{\mathcal{S}} = \left[\frac{\phi\alpha}{1-\alpha}\right]^{\frac{\alpha}{1-\alpha(\phi+1)}} \left[\frac{1}{\mathcal{W}}\right]^{\frac{\alpha}{1-\alpha(\phi+1)}}. \tag{C.1}$$

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

$$\theta_{\mathcal{S}}^{G} = N_{\mathcal{S}}^{G} \qquad \theta_{\mathcal{D}}^{G} = N_{\mathcal{D}}^{G}$$

$$\tau_{\mathcal{S}}^{G} = \phi + \left(\frac{N_{\mathcal{D}}^{G}}{N_{\mathcal{S}}^{G}}\right)^{\phi} \qquad \tau_{\mathcal{D}}^{G} = \phi + \left(\frac{N_{\mathcal{S}}^{G}}{N_{\mathcal{D}}^{G}}\right)^{\phi}$$

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

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

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

$$\tau_{\mathcal{S}}^{G} = \phi + 1 \qquad \tau_{\mathcal{D}}^{G} = \phi + 1.$$

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

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

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:

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

where

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

Plugging (C.3) into (C.1) and (C.2), and using the fact that $\mathcal{V} = \frac{b}{1-b}\kappa$, yields (2.18), where F^G and F^C are constants that only depend on the primitives of the model. In particular, define:

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

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

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

Finally, we need to show that, in equilibrium, generic cities are more denseley populated that 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 $N^k = N^k_{\mathcal{S}}$ or $N^k = N^k_{\mathcal{D}}$.$\square$

C.2.2.2. Proof of Proposition 2.3.2. In equilibrium, the rate of conventional (ψ) and unconventional (ζ) innovation can be written, respectively, as:

$$\psi = \left|\mathcal{K}^C\right| \left(N^C\right)^{\phi+1} \qquad \zeta = \left|\mathcal{K}^G\right| \left(\frac{N^G}{2}\right)^{\phi+2}.$$

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

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

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

$$\frac{\left|\mathcal{K}^G\right|}{\left|\mathcal{K}^C\right|} = \frac{b}{1-b} C^{\mathcal{K}}, \tag{C.4}$$

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

The labor market clearing condition for skilled labor is:

$$\left|\mathcal{K}^G\right| N^G + \left|\mathcal{K}^C\right| N^C = 1.$$

Using (2.18) and (C.4) to substitute for $\left|\mathcal{K}^G\right|$, N^G and N^G, we obtain:

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

The total amount of unskilled labor used in the production of housing in generic towns is equal to $\left|\mathcal{K}^G\right| \left(N^G\right)^{\frac{1}{\alpha}}$, and the total amount of unskilled labor used in the production of housing in company towns is equal to $\left|\mathcal{K}^C\right| \left(N^C\right)^{\frac{1}{\alpha}}$. 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:

$$\left|\mathcal{K}^G\right|\left(N^G\right)^{\frac{1}{\alpha}} + \left|\mathcal{K}^C\right|\left(N^C\right)^{\frac{1}{\alpha}} + x\left(\frac{\beta}{w}\right)^{\frac{1}{1-\beta}} = L. \tag{C.5}$$

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 (2.16) 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.2.2.3. Proof of Proposition 2.3.3. 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.

$\square$