Income Segregation and Rise of the Knowledge Economy

Enrico Berkes Ruben Gaetani*

6th July 2018

Abstract

We analyze 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 14% 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.

JEL Classification: D3, O11, O15, O33, R11

*Enrico Berkes: Northwestern University, 2211 Campus Dr, Evanston IL, 60208, enrico.berkes@u.northwestern.edu; Ruben Gaetani: University of Toronto, 105 St. George Street, Toronto ON (Canada), M5S3E6, ruben.gaetani@rotman.utoronto.ca. We thank Treb Allen, Matthias Doepke, Richard Florida, Ben Jones, Lorenz Kueng, Marti Mestieri, Matthew Notowidigdo, Jeffrey Zabel, as well as seminar participants at Northwestern University, University of Toronto, McMaster University, Ryerson University, 2016 Conference of Swiss Economists Abroad (Bern), 2017 European Meeting of the Urban Economic Association (Copenhagen), 2017 North American Meeting of the Urban Economic Association (Vancouver), EIEF, Bocconi, CREI, INSEAD, Toulouse School of Economics, Ohio State University, NBER Trade and Geography Conference (Cambridge), 2018 North American Summer Meeting of the Econometric Society (Davis), 2018 Meeting of the Society of Economic Dynamics (Mexico City), and attendees at the 2017 Workshop of the Kauffman Foundation (Chicago) for their helpful comments. This research was funded in part by the Ewing Marion Kauffman Foundation. The contents of the paper are solely our responsibility. Comments are welcome.

1 Introduction

Over the past 40 years, the economic activities the rely on non-manual and non-routine technical skills, scientific knowledge, and intellectual creativity have become the main engine of economic prosperity in advanced countries (Powell and Snellman, 2004; Moretti, 2012). 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 1.1). 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. Proposed 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, 1995; Jones, 2009).

This trend is believed to be associated with major social and cultural changes. Individuals with different educational 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. Florida (2002) proposed that the rise of the "creative class" 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, such as thick labor markets and knowledge spillovers (Glaeser and Gottlieb, 2009), as well as demand factors, such as preferences for local amenities (Baum-Snow and Hartley, 2017; Couture and Handbury, 2017), 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, closely tracking the evolution of overall inequality over the same decades (Table 1.1). 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 indeed exists. First, innovation and other creative jobs crucially depend on knowledge transmission, which has been shown to be strongly localized (Glaeser et al. 1992; Jaffe

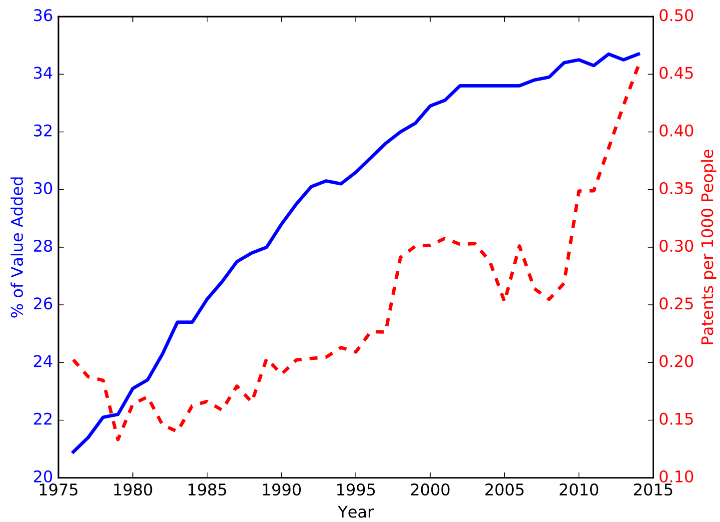


Figure 1.1: The blue line is the contribution to U.S. GDP (value added) of computer and electronic products, electrical equipment, appliances and components, information, finance and insurance, professional and business services, educational services, health care and social assistance, arts, entertainment 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.

et al., 1993; Carlino and Kerr, 2015). An increase in the returns of accessing new ideas makes geographical clustering more convenient. Second, workers in the knowledge economy tend to be disproportionately sensitive to urban and social dimensions, such as quality of schooling and social interactions, 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 outcomes, including schooling (Katz, Kling and Liebman, 2001; 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 jointly affect 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 14% 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 sizable part of the

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

Table 1.1: The overall Gini is obtained from the FRED website. The data sources and methodology for the segregation measure are explained in the text.

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 possible 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 1.17 Gini points, equal to 40% of the overall increase in segregation over the same period of time. Educational and occupational segregation, which capture 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. The direction of the 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.

The 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, measured segregation can increase even in the absence of any change in overall inequality, provided that residents relocate 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 latter as the primary cause of the innovation-driven increase in urban segregation, with the former 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 relocate their redidence 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 those cities that experience 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 observed in the data, we build a general equilibrium model of the internal structure of cities in the spirit of Ahlfeldt et al. (2015) – ARSW hereafter – that embeds endogenous amenities and productivities. We extend the ARSW model 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 receive 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 instrumented patenting growth. More precisely, 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. The estimated parameters confirm 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 simulate 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 on the lakefront in all four scenarios, the location of the campus has a sizable effect on the development of the surrounding neighborhoods, and on the overall increase in income segregation. The impact on segregation is the smallest when the campus is located in the southern part of the city, as it attracta 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, 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. More recently, 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 the 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 2000. On the contrary, we focus 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 where financing of public schools is very localized, such as the United States. 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. Our strategy extends the analysis in Acemoglu, Akcigit and Kerr (2016) in exploiting a predetermined network of knowledge diffusion to build a credible instrument for future patenting activity.

On the theory side, we augment the model developed by Ahlfeldt et al. (2015) by allowing for agents with heterogeneous backgrounds (specifically workers in creative and non-creative occupations).[1] While their strategy exploits 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 2 introduces the data and defines the measures of inequality, segregation, and knowledge intensity. Section 3 describes the empirical strategy and results. Section 4 introduces the model setting, discusses the structural estimation, and presents the quantitative results. Section 5 concludes.

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.[2] Given that our objective is to assess how innovation shocks affect residential and employment concentration within local labor markets, 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] An alternative economic geography model with heterogeneous types and different strength of agglomeration externalities across types is developed by Fajgelbaum and Gaubert (2018), who use their framework to study the optimality of the spatial equilibrium and the optimal system of place-based transfers.

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

2.1 Patents data

We proxy knowledge intensity through patenting at the CZ level. 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,[3] 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.

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 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 households in a CZ ordered from the poorest to the richest, the Gini index of city cz is defined as:

$$Ineq_{cz} = 100 \times \left[1 - 2 \times \sum_{i=1}^{N_{cz}} \sum_{i'=1}^{i} \frac{x_{i'}}{x_{cz}} \right] \tag{2.1}$$

where x_i is the income of household i, whereas x_{cz} is 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:

$$Segr_{cz} = 100 \times \left[1 - 2 \times \sum_{ct=1}^{M_{cz}} \left(\frac{N_{ct}}{N_{cz}} \sum_{ct'=1}^{ct} \frac{x_{ct'}}{x_{cz}} \right) \right] \tag{2.2}$$

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

[3] Although each patent is associated to multiple classes, the USPTO assigns a single main class to each grant. The main class is only available only in the US classification system, although our analysis is based on 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.

variation within neighborhoods has been removed.[4] In the extreme case in which average income of each neighborhood is the same, our measure is equal to 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).[5] The NHGIS collects 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.[6] 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 A.1 discusses them and provides a detailed description of the procedure we use to approximate the income distribution.[7]

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 CTs. The distribution of residence by occupation is obtained by matching information from the NHGIS and the Integrated Public Use Microdata Series (IPUMS).[8] 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.[9] Compared to the County Business Pattern, this dataset has the key advantages of providing more geographically disaggregated information, as well as 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.

[4]In the implementation of (2.2) 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.

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

[6]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$.

[7]To validate our procedure further, we compute segregation in (2.2) 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.1).

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

[9]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 establishments for which the geographical information is only available at a state level. More details on this procedure and the NETS data can be found in Appendix A.2.

Bilateral commuting times across CTs are computed using the Open Source Routing Machine (OSRM).[10] 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.[11] Finally, bilateral commuting flows are collected at the Census Block level from the Longitudinal Employer-Household Dynamics (LEHD) dataset.[12] Data at a block level are then aggregated to obtain commuting flows at our preferred level of geographical aggregation (CTs).

Appendix A.2 provides summary statistics and further details on the construction the main variables.

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 2.1, 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 conjuction 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).

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

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

[11] 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.

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

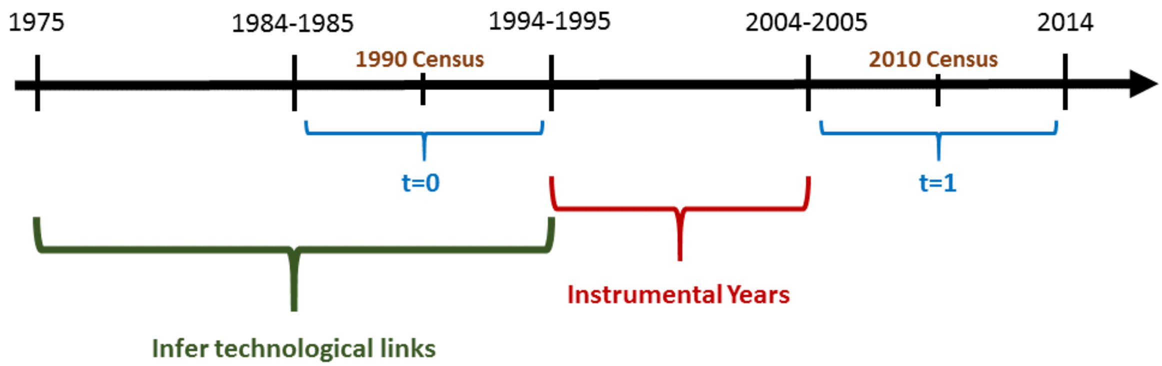


Figure 2.1: The $t = 0$ observation corresponds to 1985-1994 data for patenting, and the 1990 Census for economic and demographic variables. The $t = 1$ observation corresponds to 2005-2014 data for patenting, and the 2008-2012 ACS for economic and demographic variables.

between those phenomena and empirically investigate its features. We then use a quantitative model to infer the relative importance of the 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 the change in income segregation and the growth of patenting activity at the city level between 1990 and 2010:

$$\Delta\, Segr_{cz} = \alpha + \gamma\, \Delta\, \log(1 + Patents_{cz}) + \delta\, \Delta\, X_{cz} + \epsilon_{cz} \tag{3.1}$$

where X_{cz} is a set of controls for city cz. 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.[13] We also estimate (3.1) including the set of controls at their 1990 level. Results are robust and reported in the Appendix.

3.1 Correlations and OLS

Figure 3.1 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

[13] 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).

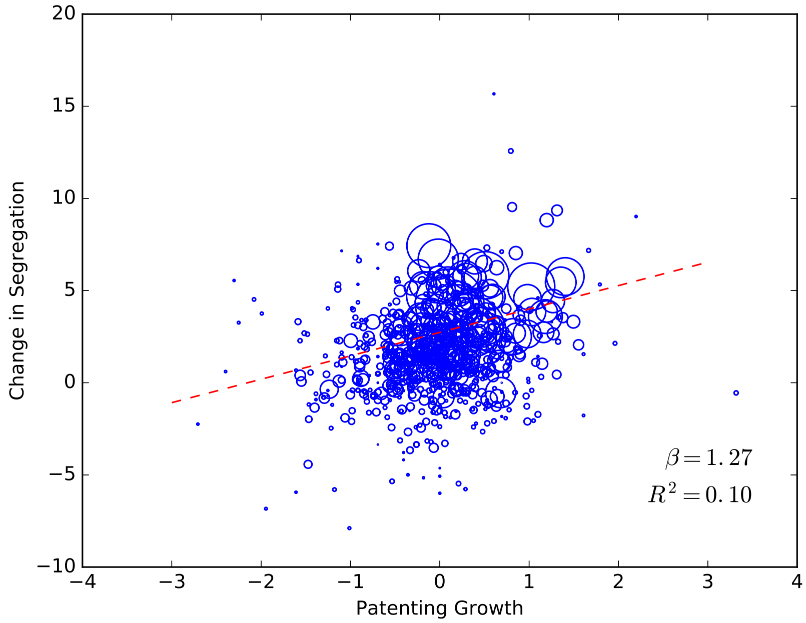


Figure 3.1: Unconditional correlation between growth in patenting and change in income segregation between 1990 and 2010, weighted by total number of households in 1990.

(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 a standard deviation in segregation in the cross-section of CZs.

In Table 3.1, we include a set of control variables that might naturally confound this correlation. First, since the number of CTs changes between 1990 and 2010, a dimensionality bias in the construction of the segregation measures could lead to a mismeasurement of the increase in segregation in cities where the number of CTs has changed significantly. To account for this possibility, in column (2) we directly control for the growth in the number of CTs within the city.[14] In columns (3)-(4), we include the growth rate of population and income, respectively. The expansion of local innovation activities is likely to be correlated with a change in the composition of the local population towards highly educated individuals, raising the concern that the intensity of human capital, rather than the prevalence of knowledge-intensive activities, is ultimately responsible for the change in segregation. To control for this, in column (5) we include the growth in the number of residents with a

[14]Controlling for the growth rate in the number of CTs may not be enough to account for the potential dimensionality bias in the construction of the segregation measure. To address this concern, we run a set of simulations in which we reassign CTs to CZs under the constraints 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.

post-graduate degree. 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 (6) we control for trade shocks using the measure of exposure to import from China developed by Autor et al. (2013).[15] 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 (7), we control for the growth rate of local public spending, provided by the Census at the County level.[16] Although some of the controls attenuate the size, the coefficient for patent growth remains positive, statistically significant and economically large.[17]Table C.1 reports the results for the OLS regressions when the controls are included in levels at their 1990 value, instead of growth rates. The estimated coefficients are smaller, but the qualitative results are unchanged.

As shown in Appendix C.6, we uncover a similar pattern when we consider segregation along an educational or occupational dimension. To measure 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 A, which consitutes the basis for our structural model in Section 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.

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

[15]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.

[16]This data is available for download at http://www2.census.gov/pub/outgoing/govs/special60/.

[17]Data for the last two controls is not available for all the commuting zones in our sample. As a result, 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.

	Dep. Variable: Change in Segregation, 1990-2010						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Patenting Growth	1.27***	0.84***	0.93***	0.63**	0.48**	0.49**	0.49**
	(0.21)	(0.34)	(0.28)	(0.28)	(0.23)	(0.22)	(0.23)
# CT		2.21*	3.19**	4.34***	4.16***	4.15***	4.06***
		(1.13)	(1.46)	(1.15)	(1.00)	(1.01)	(1.07)
# of Household			-2.11	-3.42***	-7.96***	-7.99***	-8.23***
			(1.42)	(1.22)	(1.78)	(1.82)	(1.90)
Income				8.21***	4.57***	7.59***	7.74***
				(2.14)	(0.97)	(1.88)	(1.93)
# post-grad degree					4.57***	4.62***	4.86***
					(0.97)	(0.98)	(1.06)
Import Exposure						0.01	-0.01
						(0.03)	(0.03)
Local Govt Spending							-0.08
							(0.37)
# obs.	703	703	703	703	703	687	579
R^2	0.10	0.14	0.16	0.23	0.28	0.27	0.28

Table 3.1: All regressions are weighted by total number of households in 1990. Controls are in growth rates, 1990-2010. Missing observations in columns (6) and (7) reflect data availability at the source and are concentrated in low population regions. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

factors include short-run phenomena such as housing shocks and financial shocks, or long-run trends such as technological obsolence of local industries. Inverse causality is also a possible concern, 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 innovative 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 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 3.4 discusses conditions and evidence for its validity.

3.2.1 Construction of the instrument

The idea behind the instrument is that local patenting is determined, at least in part, 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 inferrable from the network of patent citations. To be valid, 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 innovation and segregation at the same time. We extensively discuss the first point in the next subsection, 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; (2) There are no 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 3.4 discusses the conditions for and the evidence in support of the validity of the instrument.

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 $t-\Delta$. Formally, we calculate the coefficient of diffusion as:

$$d^{75-94}_{r,s,\mu,\nu,\Delta} = \begin{cases} \dfrac{\sum_{t=1985}^{1994} \sum_{p \in (\mathcal{S},\mathcal{N},\mathcal{T})} s_{p \to (r,\mu,t-\Delta)}}{\sum_{t=1985}^{1994} \sum_{q} \mathbb{1}_{\{q \in (r,\mu,t-\Delta)\}}} & r \neq s \\ 0 & r = s \end{cases} \qquad \text{for } \Delta \in \{1,\dots,10\} \tag{3.2}$$

where $s_{p \to (r,\mu,t-\Delta)}$ is the share of citations that 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 $d^{75-94}_{r,s,\mu,\nu,\Delta}$ 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 (μ, r) - the *origin* class-CZ pair - Δ years after filing. Note that since $\Delta \in \{1, \ldots, 10\}$, the entire early sample (1975-1994) must be used to compute the coefficients of diffusion. This approach implicitly assumes an input-output model for the production of ideas, in which existing patents are perfectly substitutable building blocks for future innovation. 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),[18] and time lag, Δ. 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.[19]

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 $\Delta = 1$ 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 $\Delta = 2$ to the actual patenting activity that occured in 2003, and so on until $\Delta = 10$. 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 insures that the total

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

[19] 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:

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

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.

number of patents we estimate nationwide is the same as the one we observe in the data. The same strategy is used to predict patenting activity in the subsequent years, with the only exception that when predicting total patents for 2006, the network with $\Delta = 1$ is applied to the predicted patents in 2005, instead of the acutal ones (and similarly for all the years between 2006 and 2014).[20] We do this to avoid endogeneity concerns that might arise when using contemporaneous patenting activity. Table C.2 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 directed. 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 exclusively relies on the 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, the Bartik approach carries some undesirable properties, most notably the inability to separate innovation shocks from nationwide industry or technological trends that ultimately affect innovation, but also have an impact on the dependent variable. As we extensively discuss in Section 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 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 into the set of controls.

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 is 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

[20] The role of c_{2005} is now evident: It prevents the national predicted number of patents in the later years to be altered by the use of predicted patents alongside actual patents.

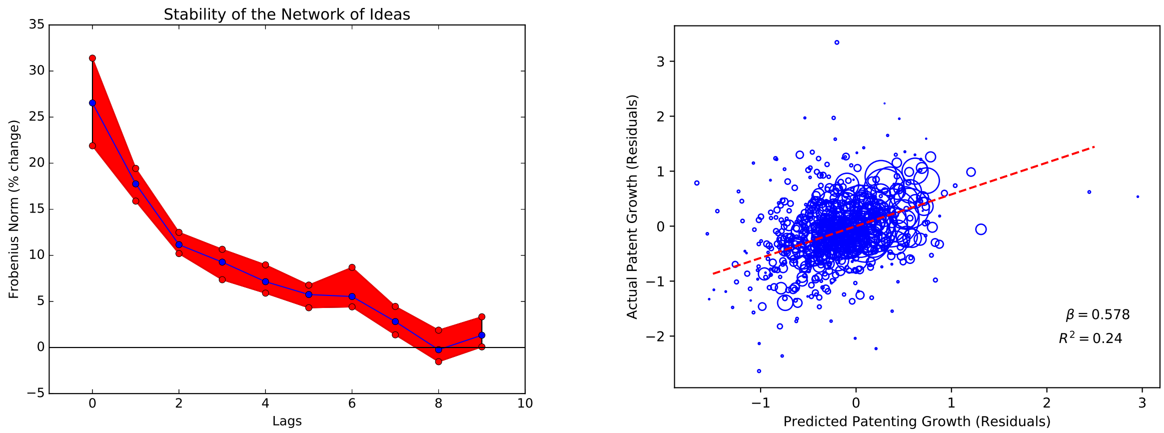


Figure 3.2: Left-panel: Comparison between the Frobenius norm of the difference between the real diffusion matrices in the early and in the late samples, and the Frobenius norm of the difference between the reshuffled diffusion matrices in the early and in the late samples. Right-panel: Scatter plot 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.

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.[21] We repeat 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

[21]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.

are more similar than the reshuffled ones.[22] Figure 3.2 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 about 26% larger 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 the citation patterns are distinguishable from links that are generated at random. This result is intuitive. With time a new technology becomes widespread knowledge and is embedded in patents produced in areas that do not have any direct link with the origin CZ-class pair.

Consistently with the results in the left-panel of Figure 3.2, the right-panel shows a scatter plot of the first stage relationship between predicted and actual growth rate of patenting. We plot the residuals of a regression of patent growth on the full set of controls. The two variables are strongly but not perfectly correlated. The residual R^2 is 0.24, while the coefficient of the regression is 0.58. The Cragg-Donald Wald F statistics in the benchmark regression is 216.56, which rules out weak instrument concerns.

Figure C.3 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.

3.3 IV Results

Our identification strategy captures local shocks to patenting that are due to knowledge created in *other* geographical areas, linked to the destination CZ through the pre-determined channels of diffusion computed in (3.2). In this section, we explore the effects of these shocks on income segregation.

Table 3.2 shows the 2SLS estimates of the relationship between innovation and segregation, outlined in (3.1). All regressions are weighted by total number of households in 1990. The coefficient on patent growth is positive and statistically significant. Columns (2)-(7) introduce the same 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. Table C.5 in Appendix reports 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

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

	Dep. Variable: Change in Segregation, 1990-2010						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Patenting Growth	2.88***	2.87***	2.84***	2.35***	2.40***	2.49***	2.49***
	(0.55)	(0.70)	(0.69)	(0.686)	(0.64)	(0.62)	(0.61)
# CT		0.03	1.83	2.83*	2.55*	2.40	2.29
		(1.46)	(1.65)	(1.60)	(1.52)	(1.55)	(1.63)
# of Household			-3.47	-4.17**	-7.61**	-7.68***	-7.684***
			(2.40)	(2.11)	(2.63)	(2.69)	(2.84)
Income				5.65***	5.11***	4.59**	4.62**
				(1.85)	(1.78)	(1.83)	(1.93)
# post-grad degree					3.39***	3.48***	3.67***
					(1.22)	(1.29)	(1.37)
Import Exposure						-0.01	-0.01
						(0.01)	(0.01)
Local Govt Spending							-0.24
							(0.33)
# obs.	703	703	703	703	703	687	579
	First-stage estimates						
Predicted	0.72***	0.60***	0.60***	0.57***	0.58***	0.57***	0.58***
Patenting Growth	(0.10)	(0.11)	(0.12)	(0.10)	(0.10)	(0.09)	(0.09)
Wald F stat.	388.33	233.96	247.03	205.77	216.56	203.27	184.21
R^2	0.36	0.40	0.43	0.43	0.45	0.46	0.47

Table 3.2: 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 (6) and (7) reflect data availability at the source and are concentrated in low population regions. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

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 $0.24 - 0.28$ Gini points, depending on the specification. Since the (population weighted) average growth rate of patents is 16.7% and the average increase in segregation 2.94, the effect is also economically large. In term of cross-sectional variation, according to our benchmark specification (column 5), one residual standard deviation increase in patenting growth increases segregation by 58% 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 C.6 in Appendix shows that a similar effect is observed for segregation defined in terms of educational achievement and occupation type (as defined in Section 3.1), instead of income level. Patent intensity has 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 0.12, whereas the corresponding figure for educational segregation is only 0.02.

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. The main identifying assumption is that shocks that affect innovation in the origin CZs should not be correlated with other factors (e.g., long-term trends or nationwide industry shocks) that jointly affect innovation and segregation in the destination CZ.

To verify to what extent the instrumented growth rate of patenting reflects pre-existing trends in innovation and segregation, we run a number of falsification tests. We start by regressing predicted patenting growth (1990-2010) on past changes in segregation (1980-1990). Figure C.7 and columns (1)-(2) of Table C.8 show the correlation between our instrument and the pre-sample trend in segregation. This correlation is indistinguishable from zero.[23] Then, we check whether the instrument is correlated with previous trends in innovation, and to what extent this could affect the second stage results. Figure C.2 shows the correlation between the residuals of the regressions of predicted patenting growth and past trends in patenting growth (1980-1990) on the basic set of controls. The two variables are weakly correlated (the coefficient of the regression is 0.14), and the R^2 of the regression is just 0.03. Column (2) in Table C.3 shows the 2SLS regression with the basic set of controls once the past trend in innovation is explicitely controlled for. The coefficient on patenting growth remains positive and significant, and is slightly larger in magnitude. This suggests that the

[23] 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 the benchmark regressions only using the CZs available in 1980. Columns (3)-(4) of Table C.8 report the results, which remain virtually unchanged.

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 disentagle the genuine effect of knowledge shocks from the effect of other factors that have an impact on innovation in the origin CZ and 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, considering 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 C.3 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, confirming that industry performance tends to operate on income segregation and innovation output in opposite directions.

To provide further evidence that the instrument is not capturing correlated industry trends across technologically linked CZs, column (4) of Table C.3 replicates the main 2SLS, with a version of the instrument 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.[24] This version of the instrument displays a weaker correlation with observed patenting growth (the R^2 of the first stage regression drops from 46% to 40%) 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 factors by introducing state fixed effects in the 2SLS estimation of (3.1). 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 C.3. The estimated coefficient is smaller, but the share of explained within-state

[24]In other words, we set $d_{r,s,\mu,\nu,\Delta}^{75-94} = 0$ whenever either $r = s$ or $\mu = \nu$.

variation is still sizeable. One residual standard deviation in patenting growth explains 38% of a residual standard deviation in 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.[25]

3.5 Which technologies are driving the effect?

Our analysis can be decomposed 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 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 given a causal interpretation.

The International Patent Classification (IPC) classifies patents into 8 main technological areas (each one hierarchically 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 3.3, with the exception that patenting growth is measured (and instrumented for) only within a given technological area.

Results are shown in Table 3.3. The positive effect of patenting on segregation seems to be entirely driven by 4 out of 8 technological areas: class A (Human Necessities), which includes Medicine and Pharmaceuticals among the others; class C (Chemistry); class G (Physics) which includes all IT and Computer patents; 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 exclusively capture 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 at first. Two explanations are the most likely candidates. On the one hand, learning-

[25] In Appendix B, we run a falsification test to rule out the possibility that the links of knowledge diffusion used for the instrument capture a demand pull from the destination CZ, rather than a supply push from the origin CZ. To get at this, we exploit the fact that the network is asymmetric, and predict patenting in 1995-2004 using, alternatively, backward links from 1985-1994 and forward links from 2005-2014. We show that only backward links have a predictive power on actual patenting.

Dep. Variable: Change in Segregation, 1990-2010			
Human Necessities (A)	3.49*** (1.10)	Performing Op. Trasporting (B)	5.13 (3.30)
Chemistry; Metallurgy (C)	1.62** (0.69)	Textiles; Paper (D)	-1.12** (0.49)
Fixed Constructions (E)	-2.01 (4.11)	Mech Eng.; Heating; Weapons (F)	-1.42 (1.06)
Physics (G)	4.37*** (1.57)	Electricity (H)	2.79*** (0.90)

Table 3.3: 2SLS estimates. All regressions are weighted by total number of households in 1990. Baseline controls are included in growth. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

intensive sectors benefit more from learning spillovers and the proximity that such spillovers require. This implies that, in areas where returns from learning are higher, incentives to cluster geographically are stronger. On the other hand, people employed in those sectors might be disproportionally sensitive to residential amenities, which amplify their incentives to cluster in space. Later in this section, we provide reduced-form evidence in support of this mechanism.

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. Net of migration, the measure of segregation can increase for two reasons. First, starting from a city with positive segregation (i.e. a condition in which average 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. We refer to this phenomenon as *inequality effect*. Second, 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 polar cases 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:

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

	Dep. Variable:				
	$\Delta Ineq_{cz}$	$\Delta Segr_{cz}$		$\Delta CT - Ineq$	
	(1)	(2)	(3)	(4)	(5)
Patenting Growth	0.95**	2.40***	1.48***	-0.86**	0.40
	(0.36)	(0.64)	(0.46)	(0.39)	(0.31)
$\Delta Ineq_{cz}$			0.96***		
			(0.08)		
# obs.	703	703	703	703	703
Controls (Growth)	✓	✓	✓	×	✓

Table 3.4: 2SLS estimates. All regressions are weighted by total number of households in 1990. Controls are included in growth. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

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 follows, we discipline how much of the observed effect can be due to inequality and how much to sorting effects.

In Table 3.4, we provide a comparison of the impact of patenting on segregation and inequality within-city. Specifically, we estimate (3.1) 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 (3.1) using $\Delta Segr_{cz}$ as dependent variable, and including $\Delta Ineq_{cz}$ as a control. The coefficient of $\Delta Ineq_{cz}$ is 0.96, 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, and 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 in 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.[26] 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 3.4 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 sorting effect completely offsets the increase in the dispersion of income within-CT that stems from the inequality effect.

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 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 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 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 in 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.

[26] 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 contrary, if people were perfectly sorted along the income dimension, an increase in city-level inequality would have no impact on within-CT inequality.

	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 3.5: 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.

3.7.1 Clustering of employment

One possible mechanism behind the results described in Section 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, since 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.

To confront this intuition, we 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 neigborhood level is adapted from Ahlfeldt et al. (2015), and is based on the structural model outlined in Section 4.3. The index captures the concentration of knowledge workers surrounding a given neighborhood.[27] Specifically, for each CT j in city cz, knowledge externalities in 1990 are computed as:

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

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 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 distri-

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

bution of Λ^{kk} within cz in 1990, we estimate via 2SLS the following equation:

$$\Delta s^{k}_{j,cz} = \alpha_{cz} + \beta\, rank_{j,cz} + \gamma\, rank_{j,cz} \times \Delta \log(1 + Patents_{cz}) + \epsilon_{j,cz}. \tag{3.3}$$

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 3.3 are displayed in Table 3.5. 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 signficantly smaller (1.40).

3.7.2 The role of residential amenities

The clustering of high-knowledge firms 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 sensitive.

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), who 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 valuable and persistent amenities should have a dampening effect to the agglomeration forces documented in the previous sections, since the endogenous spillovers would play a more marginal role.

	Dep. Variable: $\Delta Segr_{cz}$, 1990-2010			
	(1) - OLS	(2) - OLS	(3) - IV	(4) - IV
Patenting Growth	0.48*	0.56**	2.06***	2.16***
	(0.28)	(0.26)	(0.59)	(0.57)
Persistent Amenities		0.32		0.28
		(0.22)		(0.19)
Persistent Amenities ×		-0.36		-0.61**
× Patenting Growth		(0.30)		(0.31)
Controls (Growth)	✓	✓	✓	✓
# obs.	337	337	337	337

Table 3.6: 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. Standard errors clustered at the state level in paranthesis. ***p<0.01, **p<0.05, *p<0.1.

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.[28] 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 3.6 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 the 95% 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 1.16 Gini points, significantly less than the marginal impact in a city at the 5th percentile of the distribution (3.17). This suggests that residential amenities play indeed an important role in amplifying the effect of innovation shocks on income segregation.

One possible concern with the interpretation of the interaction term in Table 3.6 is that cities with more persistent residential amenities are also cities with a higher rigidity of housing supply, determined by the presence of natural constraints (e.g., ocean, mountains, etc...). Ex-ante, we expect the rigidity of housing supply to amplify, rather than dampen, the effect of an innovation shock, since it makes the pecuniary channel on the rental price of housing more responsive. To verify the extent to which a differential elasticity of housing supply can explain the results in Table 3.6, we run an anologous regression, in which we include

[28] 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.

the index of persistent residential amenities alongside with the index of land unavailability developed in Saiz (2010) and adopted by Diamond (2017).[29] The estimates, reported in Table C.4, confirm that persistent amenities, rather than the rigidity of housing supply, are the main driver of the heterogeneous effect.

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). According to our estimates, less than one third 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. Further, we provide evidence that (1) innovation shocks induce an increase in the geographical concentration of employment in 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 works 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.

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.

[29] This index is only available at the MSA level. We construct a crosswalk to map each CZ in our sample to the MSA that contains that highest share of overalapping land. This reduces the number of observations to 250.

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

$$U_{ijo}^{x} = \frac{z_{ijo}}{d_{ij}} B_i^x \left(\frac{c_{ijo}^x}{\beta}\right)^{\beta} \left(\frac{h_{ijo}^x}{1-\beta}\right)^{1-\beta} \tag{4.1}$$

where c_{ijo} is a tradable consumption good (the numeraire), h_{ijo} is consumption of housing of price q_i, B_i^x 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 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

$$u_{ijo}^x = B_i^x \frac{z_{ijo}}{d_{ij}} w_j^x (q_i)^{\beta-1}. \tag{4.2}$$

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 (4.2) and the properties of the Frechet distribution, we can calculate the share of workers of type x choosing to live in CT i and work in CT j:

$$\pi_{ij}^x = \frac{\left(B_i^x w_j^x\right)^{\varepsilon} \left(d_{ij} q_i^{1-\beta}\right)^{-\varepsilon}}{\sum\limits_{l,m \in \mathcal{S}\times\mathcal{S}} \left(B_l^x w_m^x\right)^{\varepsilon} \left(d_{lm} q_l^{1-\beta}\right)} \equiv \frac{\Phi_{ij}^x}{\Phi^x}. \tag{4.3}$$

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

$$\pi^x_{Ri} = \sum_{j\in\mathcal{S}} \pi^x_{ij} = \frac{\sum_{j\in\mathcal{S}} \Phi^x_{ij}}{\Phi^x}. \tag{4.4}$$

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

$$\pi^x_{Wj} = \sum_{i\in\mathcal{S}} \pi^x_{ij} = \frac{\sum_{i\in\mathcal{S}} \Phi^x_{ij}}{\Phi^x}. \tag{4.5}$$

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

$$\pi^x_{ij|i} = \frac{\left(w^x_j / d_{ij}\right)^\varepsilon}{\sum_{l\in\mathcal{S}} \left(w^x_l / d_{il}\right)^\varepsilon} \tag{4.6}$$

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

$$W^x_j = \sum_{l\in\mathcal{S}} \frac{\left(w^x_j / d_{lj}\right)^\varepsilon}{\sum_{m\in\mathcal{S}} \left(w^x_m / d_{lm}\right)^\varepsilon} \pi^x_{lj} R^x, \tag{4.7}$$

where R^x is the amount of residents of type x living in the city.[30]

Using the conditional probability derived in (4.6), we can calculate the expected wage of type x conditional on living in neighborhood i:

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

As shown by ARSW, the expected utility for type x upon moving to the city is equal to:

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

where $\Gamma(\cdot)$ is the Gamma function. In equilibrium, the expected utility must be equal to the reservation utility $\bar{U}^x$, that is constant across cities.

[30] Note that $R^x = \sum_j W^x_j = \sum_i R^x_i$, where R^x_i is the mass of residents of type x in CT i.

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

$$(1-\alpha) A_j^x \left(\frac{W_j^x}{H_j^x}\right)^{\alpha} = q_j, \qquad \alpha A_j^x \left(\frac{H_j^x}{W_j^x}\right)^{1-\alpha} = w_j^x. \tag{4.8}$$

Combining the FOCs with the zero profit condition yields the following equation for rents:

$$q_j^x = (1-\alpha) \left(\frac{\alpha}{w_j^x}\right)^{\alpha/(1-\alpha)} \left(A_j^x\right)^{1/(1-\alpha)}. \tag{4.9}$$

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

$$\Omega_i^{x_1 x_2} = \sum_{l \in \mathcal{S}} e^{-\rho_{x_1} \tau_{il}} \frac{R_l^{x_2}}{K_l}, \tag{4.10}$$

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.[31] Then, residential amenities for type x_1 in location i are

$$B_i^{x_1} = b_i^{x_1} \left(\Omega_i^{x_1 x_1}\right)^{\omega_{x_1 x_1}} \left(\Omega_i^{x_1 x_2}\right)^{\omega_{x_1 x_2}}, \tag{4.11}$$

where $\omega_{x_1 x_1}$ ($\omega_{x_1 x_2}$) represents the elasticitiy of residential externalities from residents of

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

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

$$\Lambda_j^{kx} = \sum_{l \in \mathcal{S}} e^{-\delta_k \tau_{jl}} \frac{W_l^x}{K_l}, \tag{4.12}$$

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

$$A_j^k = a_j^k \left(\Lambda_j^{kk}\right)^{\lambda_{kk}} \left(\Lambda_j^{kn}\right)^{\lambda_{kn}}, \tag{4.13}$$

where λ_{kk} (λ_{kn}) represents the elasticitiy 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 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 mantain 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 4.6, our quantitative results support this interpretation.

4.4 Equilibrium

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

Definition 4.1. Given quantities $\left\{\tau_{ij}\right\}_{i,j \in \mathcal{S} \times \mathcal{S}} \in (0, \infty)$ and $\left\{L_i, K_i, \left\{a_i^x, b_i^x\right\}_{x \in \{k,n\}}\right\}_{i \in \mathcal{S}} \in (0, \infty)$ and reservation utilities $\left\{\bar{U}^k, \bar{U}^n\right\}$, an **equilibrium** is a set of quantity and prices $\left\{\left\{\pi_{Ri}^x, \pi_{Wi}^x, R_i^x, W_i^x, w_i^x, A_i^x, B_i^x\right\}_{x \in \{k,n\}}, q_i\right\}_{i \in \mathcal{S}}$, so that, for each type $x \in \{k, n\}$:

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

$$\Gamma\left(1 - \frac{1}{\varepsilon}\right) \left[\sum_{l \in \mathcal{S}} \sum_{m \in \mathcal{S}} \left(d_{lm} \left(q_l\right)^{1-\beta}\right)^{-\varepsilon} \left(B_l^x w_m^x\right)^{\varepsilon}\right]^{1/\varepsilon} = \bar{U}^x \tag{4.14}$$

- The share of population living in i is given by (4.4)
- The share of population working in j is given by (4.5)
- Land markets clear for each $i \in \mathcal{S}$:

$$\sum_{x\in\{k,n\}}\left(\frac{(1-\alpha)\,A_i^x}{q_i}\right)^{1/\alpha}W_i^x+(1-\beta)\sum_{x\in\{k,n\}}\left[\sum_{l\in\mathcal{S}}\frac{\left(w_l^x/d_{il}\right)^{\varepsilon}}{\sum_{m\in\mathcal{S}}\left(w_m^x/d_{im}\right)^{\varepsilon}}w_l\right]\frac{R_i^x}{q_i}=L_i \quad (4.15)$$

- Productivity and residential externalities are determined by (4.11) and (4.13), respectively
- Factor prices satisfy (4.8), so that firms make zero profits
- Labor markets clear:

$$R_i^x=\pi_{Ri}^x\sum_{l\in\mathcal{S}}R_l^x, \qquad W_j^x=\pi_{Wj}^x\sum_{l\in\mathcal{S}}W_l^x,$$

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

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.

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, $\{R_i^k, R_i^n\}_{i\in\mathcal{S}}$, 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, $\{\bar{w}_c^k, \bar{w}_c^n\}$, 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 $\{B_i^k, B_i^n\}_{i\in\mathcal{S}}$, 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 A.

Estimating sensitivity to commuting times We allow the parameter that controls the sensitvity of the utility function to commuting times to vary by city. Taking logs of (4.3) yields a gravity equation for commuting flows from CT i to CT j:

$$\log\left(\pi_{ij}^{x}\right) = \alpha^{x} + \psi_{i}^{x} + \zeta_{j}^{x} + \nu_{c}\tau_{ij} + \eta_{ij}^{x} \tag{4.16}$$

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:

$$\log\left(\pi_{ij}\right) = \alpha + \psi_{i} + \zeta_{j} + \nu_{c}\tau_{ij} + \eta_{ij}. \tag{4.17}$$

We show in the Appendix (Figure C.5) 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 (4.16) 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 C.4. The median value is -0.041, which implies that one additional minute of commuting time decreases commuting probability by 4.1%.[32]

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:

$$W_{j}^{x} = \sum_{i \in \mathcal{S}} \frac{\left(w_{j}^{x}\right)^{\varepsilon_c} / e^{\nu_c \tau_{ij}}}{\sum_{l \in \mathcal{S}} \left(w_{l}^{x}\right)^{\varepsilon_c} / e^{\nu_c \tau_{il}}} R_{i}^{x}, \tag{4.18}$$

where $\left\{W_i^x,\, R_i^x\right\}_{i\in\mathcal{S}}$ 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 arithmetic mean of the non-creative sector's wage in the CZ with the first index (Memphis) is equal to one.[33]

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

[33]See Lemma S.7 in the Supplement to ARSW for a proof that the system of equations in (4.18) determine a unique (up to a normalization) vector of wages $\left\{w_j^x\right\}_{j\in\mathcal{S}}$.

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 (4.9). 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 (4.12) and (4.13).

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 (4.4) and (4.14) allows us to recover residential amenities $\left\{B_i^k,\, B_i^n\right\}_{i\in\mathcal{S}}$:

$$B_i^x = \left(\frac{R_i^x}{R^x}\right)^{\frac{1}{\varepsilon_c}} \left(\frac{\bar{U}^x}{\Gamma\left(1-\frac{1}{\varepsilon_c}\right)}\right) \frac{q_i^{1-\beta}}{\left(\tilde{w}_i^x\right)^{1/\varepsilon_c}} \qquad x \in \{k, n\}\,, \tag{4.19}$$

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 (4.19) for the remaining cities.[34]

4.6 Structural estimation

We follow ARSW and set $\alpha = 0.8$, $\beta = 0.75$ and $\kappa = 0.01$ in our calibration, which implies $\varepsilon_c = \nu_c/0.01$.[35] 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.

[34] 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.

[35] 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 (4.16). 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. 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 Appendix E.

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:

$$\lambda_{kk,c}^{10} = \lambda_{kk}^{90} + \theta_0 + \theta_1 \cdot bin_c \tag{4.20}$$

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 3.2.1:

$$\hat{g}_c = \log \ \hat{pat}_{c,05-14} - \log \ pat_{c,85-94}. \tag{4.21}$$

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 (4.20). The orthogonality condition can then be expressed as

$$\begin{cases} \mathbb{E}_{c \in \mathcal{C}_{bin}} \left[\Delta_{10-90} \mathbb{E}_{i \in \mathcal{S}_c} \log \left(a_i^x \right) \right] = \mathbb{E}_{c \in \mathcal{C}} \left[\Delta_{10-90} \mathbb{E}_{i \in \mathcal{S}_c} \log \left(a_i^x \right) \right] \\ \mathbb{E}_{c \in \mathcal{C}_{bin}} \left[\Delta_{10-90} \mathbb{E}_{i \in \mathcal{S}_c} \log \left(b_i^x \right) \right] = \mathbb{E}_{c \in \mathcal{C}} \left[\Delta_{10-90} \mathbb{E}_{i \in \mathcal{S}_c} \log \left(b_i^x \right) \right] \end{cases} \tag{4.22}$$

for all $bin \in \{0, ..., 9\}$ and $x \in \{k, n\}$. In (4.22), 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 (4.18)), residential and productivity fundamentals can be recovered by combining (4.19) with (4.11) and (4.9) with (4.13), respectively.

Condition (4.22) 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 change in the strength of productivity externalities of the creative sector induced by the knowledge shock, and the endogenous agglomeration forces in the model.

The system in (4.22) delivers 3×10 moment conditions for a set of 11 parameters to

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 C.4	ρ_n	0.467	ρ_k	0.497
		δ_k	0.055	θ_0	-0.004
		θ_1	0.001		

Table 4.1: Parameter values

estimate:[36]

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

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:

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

where $\mathbb{W}$ is the optimal weighting matrix. Details on the estimation algorithm can be found in Appendix D.2.

The results of the estimation are displayed in the right panel of Table 4.1. 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 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.[37] The estimated value of δ_k implies that a 10 minutes commuting time reduces the strength of the externality by roughly 42%.

It is worth it to emphasize two additional points. 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

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

[37]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.

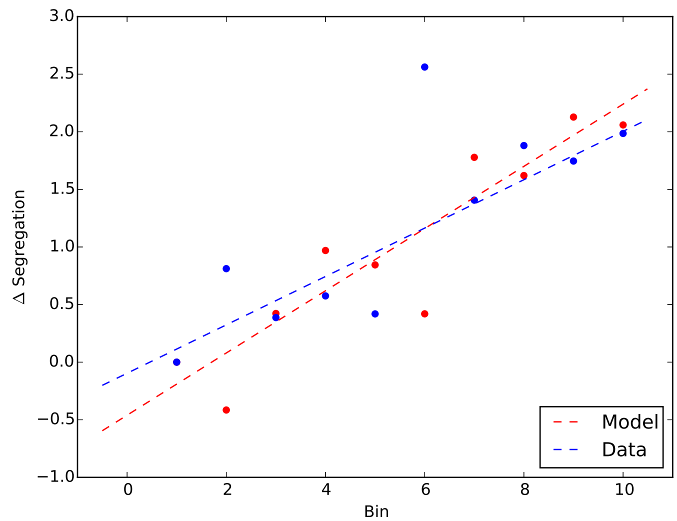


Figure 4.1: Knowledge shock (bin) and change in segregation, 1990-2010: Data and Model.

sector receive very steep productivity externalities from other knowledge workers, and less powerful externalities from non-creative workers.

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.

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, $\left\{a_i^x, b_i^x\right\}$, as described in Section 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 (4.20). The algorithm used to find the new equilibrium (adapted from ARSW) can be found in Appendix D.3. 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 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 exercise. 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 4.1 shows the model performance in replicating the empirical relationship between the estimated knowledge shock (i.e., the predicted patenting growth, as in (4.21)) 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 4.2). The model coefficient (0.60) is not significantly different from the empirical one (0.51). Table 4.2 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 4.2, 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 C.6 in Appendix replicates the results in Table 3.5 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 4.1 and Table 4.2 and C.6 appear as a target in the structural estimation.

Figure 4.2 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 4.2. The slope of

	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 4.2: 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.

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.

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 shocks. 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 4.6, the estimated values for residential elasticities suggest 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 residents - generated, for example, by a reshuffling of the distribution of employment - can be significantly amplified by the endogenous response of residential amenities.

Figure 4.3 shows the results of a counterfactual experiment, in which residential amenities are exogenously given ($B_i^x = b_i^x$). 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 is 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. Hence, roughly two thirds of the overall estimated impact of knowledge shocks on income segregation can be attributed to the amplifying effect of localized, occupation-specific residential amenities.

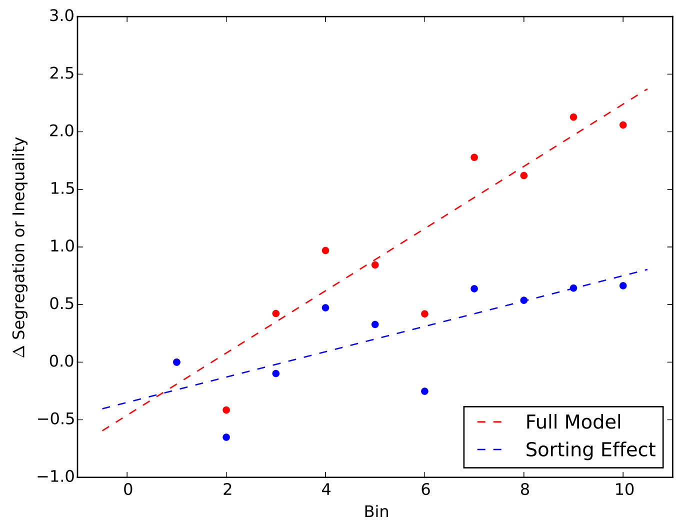


Figure 4.2: Knowledge shock (bin) and change in segregation, 1990-2010: Full effect (red line) and Sorting effect (blue line) computed as segregation with 1990 distribution of average wages by CT/occupation and 2010 (model based) distribution of residents by CT/occupation.

4.7.2 Chicago Bids for Amazon HQ2

At the beginning of September 2017, Amazon announced that it was seeking a location for its second North American headquarter and accepting bids. By the end of the month, more than 50 cities across the United States and Canada declared their interest, some of them submitting multiple bids. In total, city developers worked on more than 200 proposals and submitted more than 100 projects.[38] In this Section, we illustrate how our model can be used for policy experiments by assessing the differential impact of Chicago's bids (or potential bids) on its structure. City developers worked on a total of six projects. The one located furthest North plans to develop the area by the river that was occupied by the A. Finkl & Sons steel plant demolished in 2011. The second proposal is located closer to the Loop (Chicago's historical core and current business center), and it would be composed of four new buildings overlooking the river in the property owned by Tribune Media. Three additional projects were proposed just South-West of the Loop: The first one in the Old Main Post Office; The second one in a completely redesigned Union Station; The third one in a currently unused area just South of the Post Office.[39] Finally, the last bid locates the campus in the South Side on the premises of the Michael Reese Hospital, which caesed activity in 2009. Figure F.1 in

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

[39]Since these 3 projects are located in a radius of less than one mile from each other, for the purpose of this exercise, we only consider the one in the Old Main Post Office.

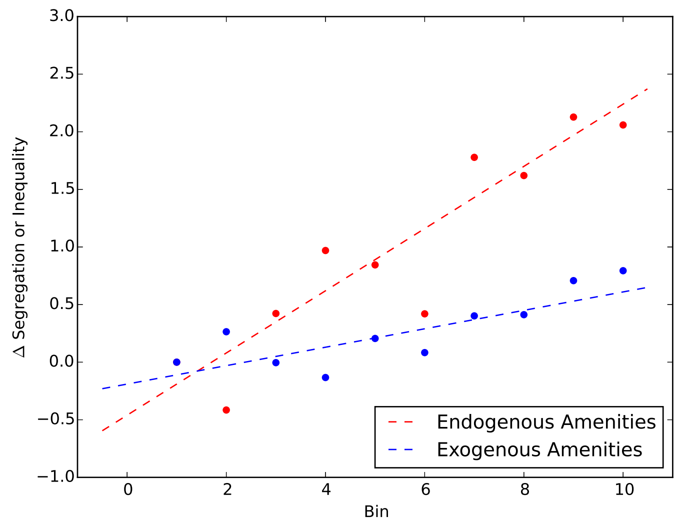


Figure 4.3: Knowledge shock and change in segregation 1990-2010: Exogenous vs Endogenous residential amenities.

the Appendix shows the exact location of the four projects considered in this simulation.

For this experiment, we first estimate the equilibrium quantities in 2010, and shock the exogenous term of productivities. The shock is calibrated to attract 50,000 high-knowledge workers in the considered neighborhood. This number matches the number of workers Amazon expects to employ in its second headquarter. Figure 4.4 shows the forecasted change in high-knowledge residents in the four scenarios. Census tracts colored in red denote areas with a net influx of high-knowledge residents, whereas those colored in blue denote a net outflow. The four panels are arranged clockwise showing the estimated impact of the four campuses from the one furthest North to the one furthest South.

There are two main trends that is possible to identify by comparing the counterfactuals. First, in all the scenarios knowledge workers tend to relocate their residence in high-amenities areas by the lakefront and downtown. Second, despite this general trend, the location of the headquarters plays a key role in local development. On the one extreme, in the Michael Reese Hospital scenario, knowledge workers relocate to the South Side. Although the already well-developed areas around the University of Chicago and along the coast seem to be the most attractive, some of them optimally decide to live in poorer neighborhoods very close to the HQ2. On the other extreme, in the A. Finkl & Sons steel plant scenario, the majority of the gains are concentrated in the richer North Side. This is also reflected in the estimated changes in segregation, which increases by 1.0 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) when the Amazon campus

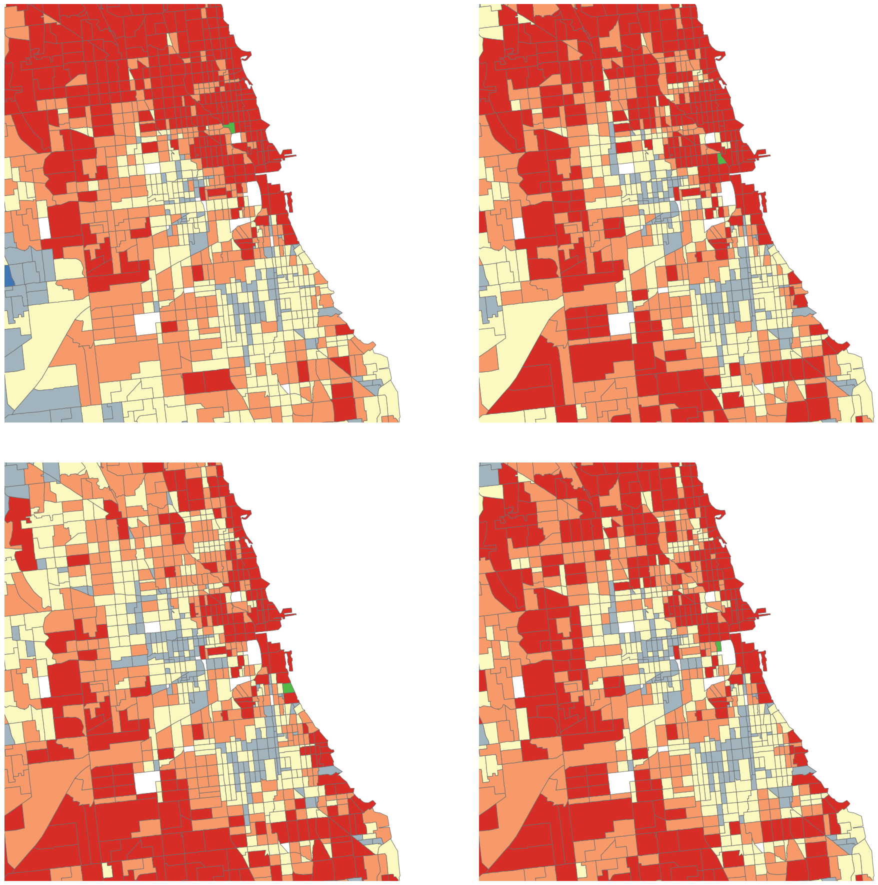


Figure 4.4: Change in knowledge residents in each Census Tract of Chicago as a result of Amazon's HQ2 locating in a specific neighborhood (colored in green). Panel (a): A. Finkl & Sons steel plant; Panel (b): Tribune Media River Front property; Panel (c): Michael Reese Hospital premises; Panel (d): Old Main Post Office. 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.

is located on the Tribune Media's property. Figure F.2 in the Appendix shows the change in the concentration of employment of knowledge workers.

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 come from the commuting zone of Chicago. Taking migration into consideration might actually amplify the segregation effect, since rents in high-demand neighborhoods would increase more than in our counterfactuals.

5 Conclusions

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 58% of the cross-sectional variation, and 14% 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. Therein lies a powerful amplification mechanism, as the endogenous response of residential amenities, valued disproportionately by knowledge workers, 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 be explained as the result of the endogenous development 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.

References

[1] Acemoglu, Daron, Ufuk Akcigit, and William R. Kerr. 2016. "Innovation network." *Proceedings of the National Academy of Sciences*, 113(41): 11483-11488.

[2] 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.

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

[4] 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.

[5] 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

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

[7] 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.

[8] 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.

[9] 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.

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

[11] Baum-Snow, Nathaniel, and Daniel Hartley. 2017. "Accounting for Central Neighborhood Change, 1980-2010." Unpublished Manuscript.

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

[13] 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.

[14] 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.

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

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

[17] Couture, Victor, and Jessie Handbury. "Urban revival in America, 2000 to 2010." NBER Working Paper No. 24084.

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

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

[20] 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.

[21] Diamond, Rebecca. 2017. "Housing supply elasticity and rent extraction by state and local governments." *American Economic Journal: Economic Policy*, 9(1): 74-111.

[22] Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741-1779.

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

[24] Florida, Richard, and Charlotta Mellander. 2015. "Segregated City. The Geography of Economic Segregation in America's Metro" Martin Prosperity Institute, February.

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

[26] Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. 1992. "Growth in cities." *Journal of political economy,* 100(6): 1126-1152.

[27] 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.

[28] Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. “Endogenous Gentrification and Housing Price Dynamics.” *Journal of Public Economics*, 100: 45-60.

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

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

[31] 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.

[32] Jones, Charles. 1995. “R&D-Based Models of Economic Growth.” *Journal of Political Economy*, 103(4): 759-784.

[33] Katz, Lawrence, Kling Jeffrey, and Jeffrey Liebman. 2001. “Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment.” *The Quarterly Journal of Economics*, 116(2): 607-654.

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

[35] 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.

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

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

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

[39] Powell, Walter, and Kaisa Snellman. 2004. “The Knowledge Economy.” *The Annual Review of Sociology*, 30: 199-220.

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

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

[42] Saiz, Albert. 2010. "The geographic determinants of housing supply." *The Quarterly Journal of Economics*, 125(3): 1253-1296.

[43] 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.

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

Appendix

A Data description

A.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).[40] These two methods have been subject to several critics.[41]

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.[42] 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-

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

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

[42]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).

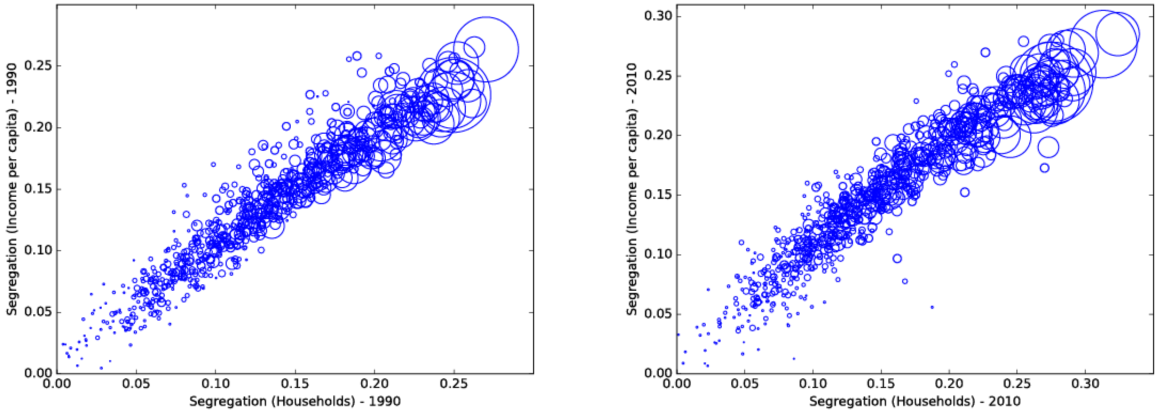


Figure A.1: Correlation between $Segr_{cz}$ computed using the approximated distribution of household income (horizontal axis) and the actual distribution of average personal income across CTs (vertical axis) in 1990 (left panel) and 2010 (right panel).

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).[43]

To validate our procedure further, in Figure A.1 we show the correlation between segregation in 1990 and 2010, respectively, using the household income distribution 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.

A.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

[43] 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$.

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).[44]

We assign workers to workplaces using the National Establishment Time Series (NETS). This data set contains information about employment for the universe of 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.[45] Second, we assign the workers of each establishment geo-located at ZIP code level based on the area of the census tracts it contains.[46] We discard all those establishments whose coordinates are missing or are at a street level.[47] 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 or government buildings is likely to be significantly lower in the ZBP.[48] Figure A.2 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).[49] As we expected, the NETS sys-

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

[45] 7,573,637 establishments were assigned this way in 1990; 28,111,455 in 2010.

[46] 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.

[47] 156,185 establishments were discarded in 1990; 332,091 in 2010.

[48] 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

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

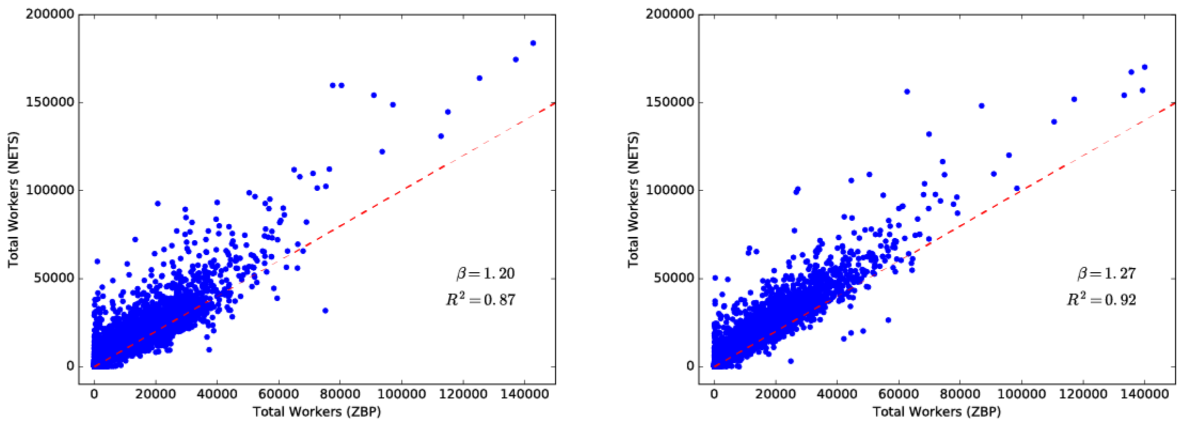


Figure A.2: Correlation between workers as reported in the ZBP (x-axis) and in the NETS (y-axis) in 1994 (left) and 2010 (right). Each point represents a ZIP code. The dashed red line is the 45-degree line.

tematically 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.[50] Similarly to what 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 apart-

[50]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.

	# 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 A.1: Summary statistics (weighted by total HH in respective year).

ments 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.[51]. The LEHD collects data about bilateral commuting flows from and to each Census Block starting from 2002.[52] These data are used to estimate the commuting flows/commuting times semi-elasticities using the gravity equation (4.16) obtained from the structural model. Since we 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

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

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

data are only available from 2011 onwards). 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).[53] The advanatage 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 understimate the actual commuting times/costs faced by workers.[54]

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

[54] 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.

B Innovation Network: Demand Pull or Supply Push?

In Section 3.2.1, we developed an instrument for local patenting activity that exploits a predetermined network of knowledge flows. Our instrument is valid as long as these knowledge links are determined by factors that are orthogonal to the local future economic activity. A possible concern that would invalidate our identification strategy is that the channels captured through the network of citations reflect demand instead of supply links. This would be problematic for the validity of the model, since demand links are likely to be informative about the state of the local economy. To fix ideas, suppose that an IT firm in San Jose supplies innovation to a car manufacturer in Detroit under commission. In this case, our network would record a strong link from San Jose to Detroit, but the associated knowledge flows would violate the orthogonality assumption, since Detroit's demand is likely to be correlated with other unobservable factors.

The structure of our network and the long time series of patents data can be used to test for the presence of demand-driven links. Formally, we proceed in three steps. First, we use the knowledge network defined in Section 3.2.1 and the observed patenting activity in the period 1985-1994 to get a forward estimate of the patenting activity between 1995 and 2004. Second, we reverse the network and use the patents filed between 2005-2014 to get an upstream estimate of the patenting activity we expect to observe in the period 1995-2004 if the citations were capturing demand links. The reversed network closely mirrors the one defined in (3.2), but instead of exploiting forward citations, it uses backward citations:

$$o^{75-94}_{r,s,\mu,\nu,\Delta} = \begin{cases} \dfrac{\sum\limits_{t=1975}^{1984} \sum\limits_{p\in(\mathcal{S},\mathcal{N},\mathcal{T})} s_{p\leftarrow(r,\mu,t+\Delta)}}{\sum\limits_{t=1975}^{1984} \sum\limits_{q} \mathbb{1}_{\{q\in(r,\mu,t+\Delta)\}}} & r \neq s \\ 0 & r = s \end{cases} \qquad \text{for } \Delta \in \{1,\ldots,10\}$$

where $o^{75-94}_{r,s,\mu,\nu,\Delta}$ represents the number of patents of class μ in commuting zone r that we expect to observe upstream if Δ years later we observe one patent of class ν in commuting zone s downstream . Finally, we compare the the two models to see which one offers the most accurate description of the innovation process. To do this, we follow Acemoglu et al. (2016) and regress the actual 1995-2004 patenting activity on the patenting activity predicted by the two procedures:

$$P^{actual}_{2000,\,cz} = \alpha + \beta \hat{P}^{down}_{2000,\,cz} + \gamma \hat{P}^{up}_{2000,\,cz} + \varepsilon_j$$

where $\hat{P}^{down}_{2000,cz}$ is patenting activity predicted by the forward model, whereas $\hat{P}^{up}_{2000,cz}$ is the upstream patenting activity predicted by the backward model. The results, reported in Table B.1, show that once we control for the downstream effect, the backward model does not have any additional predictive power.

	$P^{actual}_{2000,cz}$
$\hat{P}^{down}_{2000,cz}$	0.933***
	(0.05)
$\hat{P}^{up}_{2000,cz}$	0.037
	(0.04)
# obs.	703
R^2	0.978

Table B.1: All regressions are weighted by total number of households in 1990. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

C Additional Tables and Figures

	Dep. Variable: Change in Segregation, 1990-2010					
	(1)	(2)	(3)	(4)	(5)	(6)
Patenting Growth	1.27***	0.97***	0.70***	0.61**	0.55**	0.52*
	(0.21)	(0.20)	(0.26)	(0.25)	(0.26)	(0.27)
# CT		0.50***	-2.11*	-2.24*	-2.37*	-2.73**
		(0.12)	(1.23)	(1.26)	(1.21)	(1.25)
# of Household			2.54**	2.53*	1.88	2.62**
			(1.23)	(1.28)	(1.36)	(1.26)
Income				1.09	0.19	0.80
				(1.45)	(1.34)	(1.39)
# post-grad degree					0.76	1.61
					(0.57)	(0.50)
Local Govt Spending						-0.26
						(0.16)
# obs.	703	703	703	703	703	643
R^2	0.10	0.22	0.26	0.27	0.27	0.28

Table C.1: 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. Standard errors clustered at the state level 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 C.2: 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)	(6)
Patenting Growth (1990-2010)	2.40***	2.66***	3.10**	2.75***	1.74***	3.70*
	(0.64)	(0.61)	(1.22)	(0.83)	(0.60)	(2.06)
Patenting Growth (1980-1990)		-1.15**				-0.56
		(0.48)				(0.41)
Bartik-like variable			-1.04			-1.76
			(1.42)			(1.92)
Constrained Instrument	No	No	No	Yes	No	Yes
State 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***	0.56***	0.52***	0.48***	0.55***	0.18***
	(0.10)	(0.09)	(0.13)	(0.10)	(0.09)	(0.08)
Wald F-stat	216.56	198.85	53.53	135.39	153.69	12.16
R^2	0.46	0.46	0.46	0.40	0.61	0.59

Table C.3: 2SLS estimates. All regressions are weighted by total number of households in 1990. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

	Dep. Variable: $\Delta Segr$, 1990-2010			
	(1)	(2)	(3)	(4)
Patenting Growth	1.58***	2.08***	2.02***	2.05***
	(0.78)	(0.64)	(0.57)	(0.53)
Housing Rigidity		-0.19		-0.27
		(0.21)		(0.23)
Persistent Amenities			0.31	0.38
			(0.23)	(0.24)
Interaction × Housing Rigidity		-0.28		0.19
		(0.47)		(0.62)
Interaction × Persistent Amenities			-0.77**	-0.85*
			(0.37)	(0.49)
Controls (Growth)	✓	✓	✓	✓
# obs.	250	250	249	249

Table C.4: 2SLS estimates. All regressions are weighted by total number of households in 1990. Column (2) uses the index of unavailable land in Saiz (2010). Column (3) replicates the results in Table 3.6 for the subset of CZs in which the index is available. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

	Dep Variable: Change in Segregation, 1990-2010					
	(1)	(2)	(3)	(4)	(5)	(5)
Patenting Growth	2.88***	1.96***	1.63***	1.63***	1.47***	1.38***
	(0.54)	(0.49)	(0.44)	(0.45)	(0.37)	(0.41)
# CT		0.43***	-1.29	-1.29	-1.49	-1.84
		(0.13)	(1.40)	(1.34)	(1.26)	(1.25)
# of Household			1.769	1.69	1.41	2.08
			(1.41)	(1.37)	(1.44)	(1.30)
Income				-0.01	-0.39	0.19
				(1.52)	(1.54)	(1.64)
# post-grad degree					0.44	0.33
					(0.50)	(0.43)
Local Govt Spending						-0.23
						(0.16)
# obs.	703	703	703	703	703	643
	First-stage estimates					
Predicted Patenting Growth	0.72***	0.77***	0.70***	0.65***	0.65***	0.65***
	(0.10)	(0.11)	(0.10)	(0.09)	(0.09)	(0.09)
Wald F-stat	388.33	348.73	269.11	197.00	175.01	161.26
R^2	0.35	0.36	0.37	0.39	0.38	0.40

Table C.5: 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. Standard errors clustered at the state level 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 Growth	1.38***	0.94**	2.33***	1.56***	1.81***	1.97***	2.13***	5.85***
	(0.43)	(0.41)	(0.63)	(0.58)	(0.53)	(0.36)	(0.35)	(0.96)
# obs.	703	703	703	703	703	703	703	703
Controls	×	✓	×	✓	×	✓	×	✓

Table C.6: 2SLS estimates. All regressions are weighted by total number of households in 1990. Controls are included in growth. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

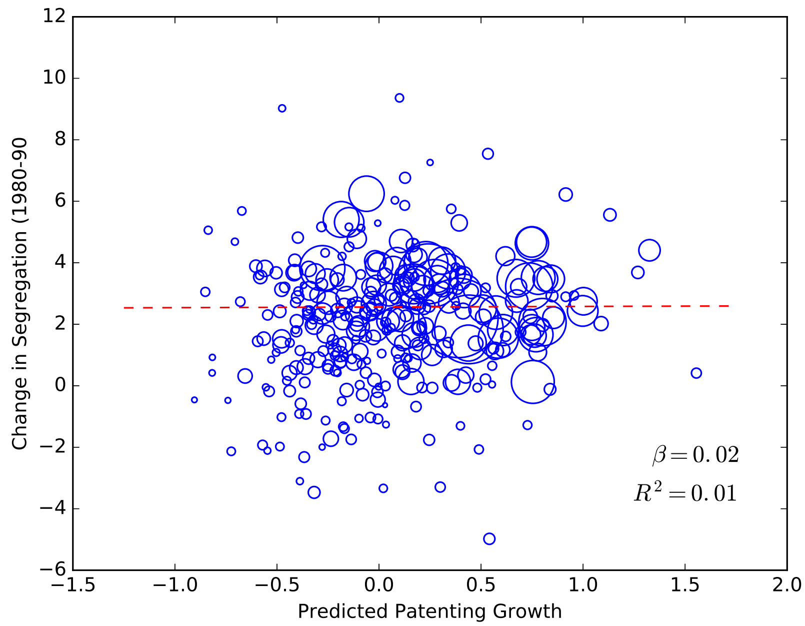


Table C.7: Scatter plot of predicted patenting growth (instrument) and pre-trend in segregation (1980-1990). The scatter plot is weighted by total households in 1990.

	Dep. Variable: $\Delta Segr_{cz}$			
	Past Trend		1990-2010	
	(1)	(2)	(3)	(4)
Predicted Patenting Growth	0.02 (0.33)	0.31 (0.30)		
Patenting Growth			2.53*** (0.50)	1.90** (0.53)
# obs.	309	309	309	309
Controls (Growth)	×	✓	×	✓

Table C.8: All regressions are weighted by total number of households in 1990. Controls are in growth rates, 1990-2010. Standard errors clustered at the state level in parentheses. ***p<0.01, **p<0.05, *p<0.1.

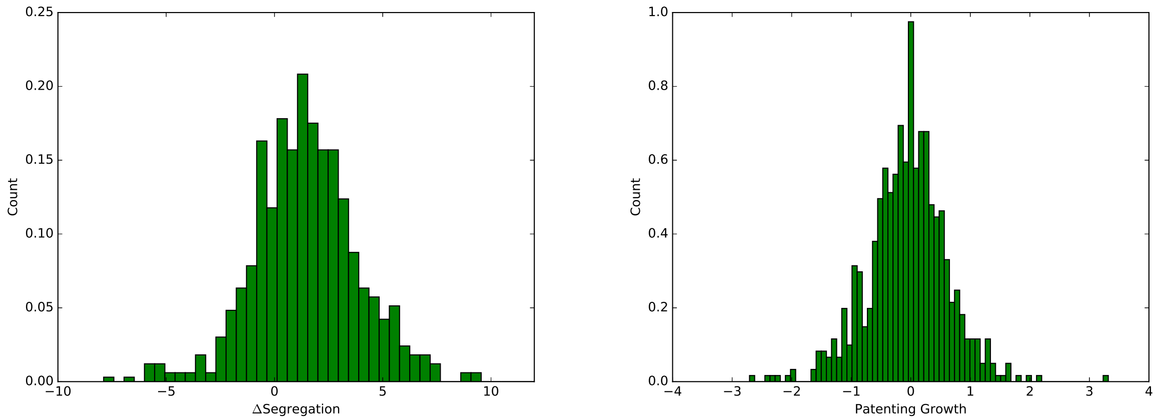


Figure C.1: Distribution of changes in measured segregation and patenting growth in the cross section of U.S. CZs, 1990-2010.

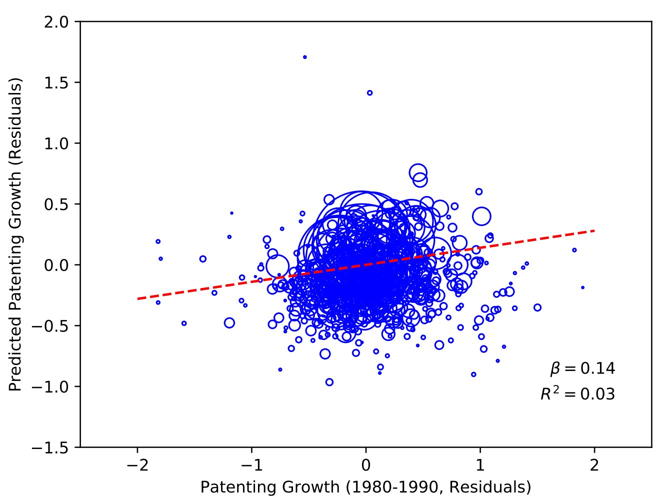
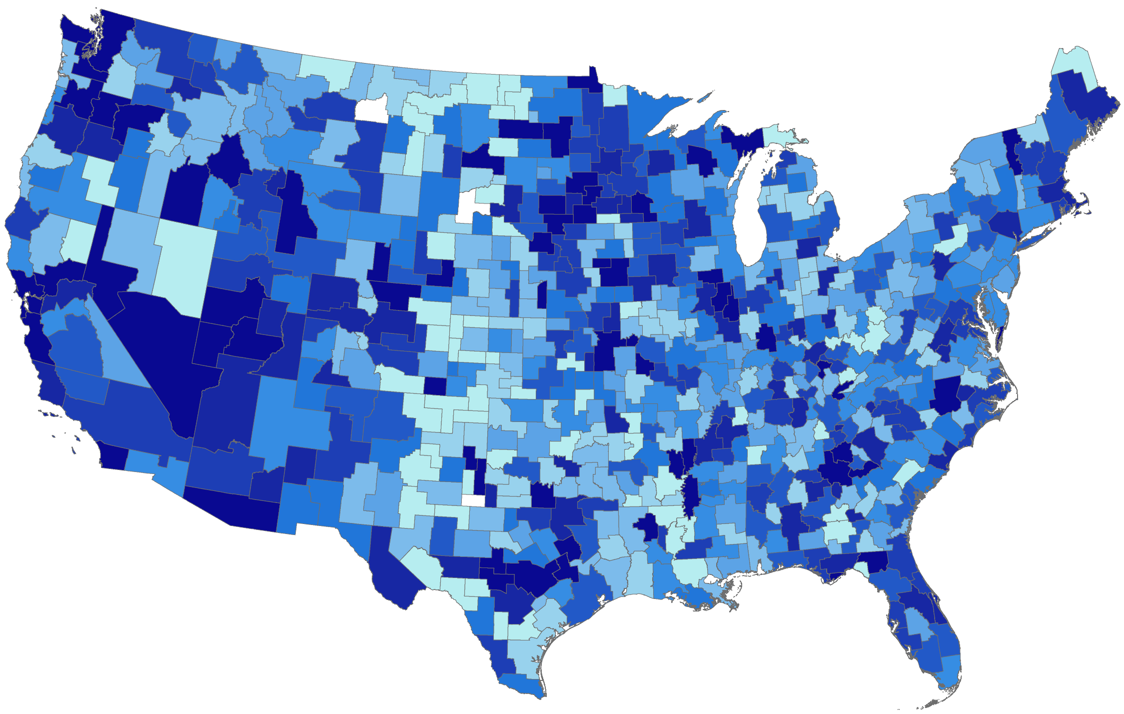


Figure C.2: Scatter plot of predicted patenting growth (instrument) and pre-trend in patenting (1980-1990). The scatter plot is weighted by total households in 1990.



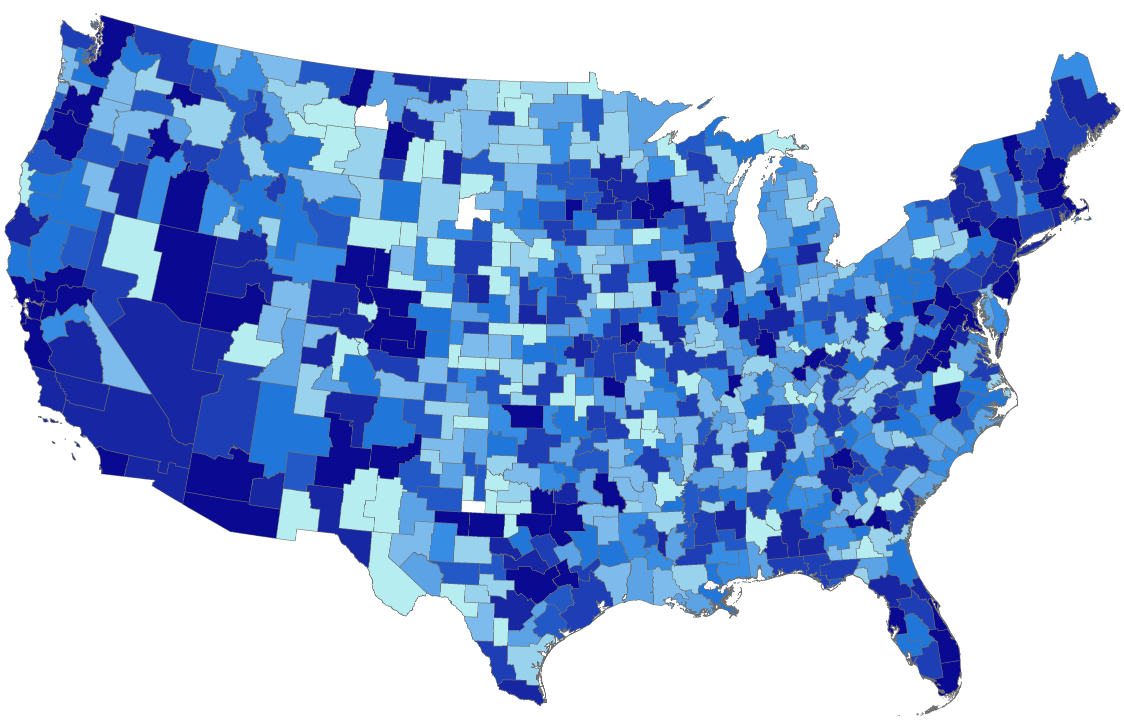


Figure C.3: Predicted (top map) and actual (bottom map) growth rate of patents in U.S. commuting zones, 1990-2010.

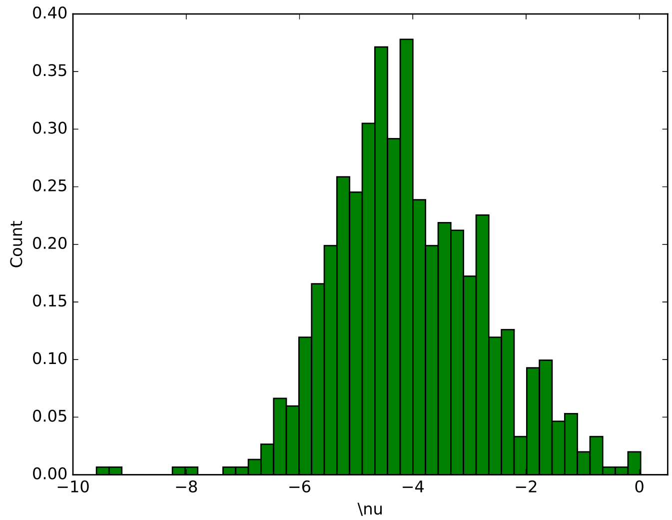


Figure C.4: Distribution of estimated values of ν_c in U.S. commuting zones

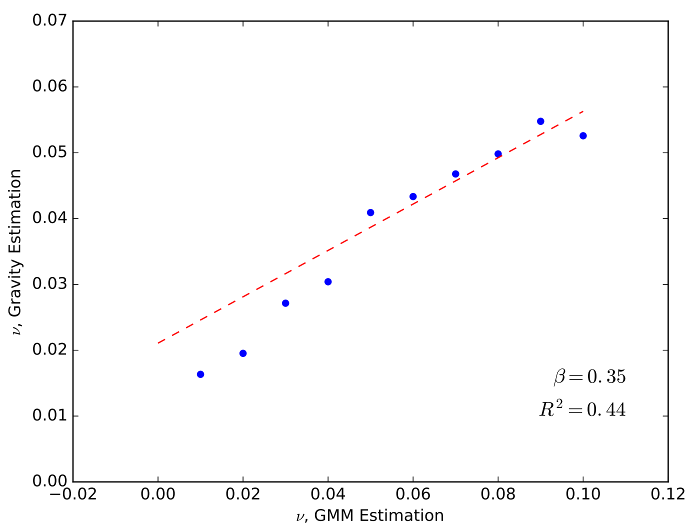


Figure C.5: The bin scatter plot compares the value of ν_c estimated using (4.16) with the value of ν_c that minimizes the difference between the share of people in city c commuting for less than 60 minutes and the model generated counterpart.

	Dep. Variable: Δs_j^k	
	(1) - Model	(2) - Data
$rank_j \times$ 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

Figure C.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.

D Derivations

D.1 Derivation of (4.3) and (4.6)

The probability that an agent of type x commutes from neighborhood i to neighborhood j can be derived as follows:

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

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:

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

D.2 Details on the Structural Estimation

We estimate the structural parameters of the model using the moment conditions described in (4.22). 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:

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

where$\boldsymbol{m}(\boldsymbol{X}, \boldsymbol{p})$ 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,

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

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) heteroscedasticity-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.

D.3 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. $\pi_{ij}^x = \dfrac{\left(d_{ij}\left(q_i^0\right)^{(1-\beta)}\right)^{-\varepsilon}\left(B_i^x w_j^0\right)^{\varepsilon}}{\sum\limits_{k\in\mathcal{S}}\sum\limits_{l\in\mathcal{S}}\left(d_{kl}\left(q_k^0\right)^{(1-\beta)}\right)^{-\varepsilon}\left(B_k^x w_l^{x,0}\right)^{\varepsilon}}$

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

3. $R_i^x = \sum\limits_{l\in\mathcal{S}}\pi_{il}^x R^x$

4. $W_j^x = \sum\limits_{k\in\mathcal{S}}\pi_{kj}^x R^x$

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

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

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

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

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

10. $A_j^k = a_j^k\left(\sum_{l\in\mathcal{S}} e^{-\delta_k\tau_{jl}}\frac{W_l^k}{K_l}\right)^{\lambda_{kk,c}}\left(\sum_{l\in\mathcal{S}} e^{-\delta_k\tau_{jl}}\frac{W_l^n}{K_l}\right)^{\lambda_{kn}}$

11. $B_i^x = b_j^x\left(\sum_{l\in\mathcal{S}} e^{-\rho_x\tau_{jl}}\frac{R_l^{x_1}}{K_l}\right)^{\omega_{xx_1}}\left(\sum_{l\in\mathcal{S}} e^{-\rho_x\tau_{jl}}\frac{R_l^{x_2}}{K_l}\right)^{\omega_{xx_2}}$

We iterate until $\left|q_i^0 - q_i^1\right|$, $\left|w_j^{x,0} - w_j^{x,1}\right|$ 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:

$$
\begin{aligned}
q_i^2 &= 0.3\, q_i^1 + 0.7\, q_i^0 \\
w_j^{x,2} &= 0.3\, w_i^{x,1} + 0.7\, w_i^{x,0} \\
\theta_j^{x,2} &= 0.3\, \theta_i^{x,1} + 0.7\, \theta_i^{x,0}
\end{aligned}
$$

E Model-Generated Instrument for the Gravity Equation

In Section 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:

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

where $\psi_i = -\varepsilon\left(1-\beta\right) q_i + \varepsilon B_i$. 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

$$\begin{aligned} \psi_i &= \gamma \hat{q}_i + \xi_i \\ q_i &= \sigma q_i^{model} + \chi_i \end{aligned}$$

gives us an unbiased estimate of $\gamma = -\varepsilon\left(1-\beta\right)$ 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.[55] 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 E.1 shows the distribution of ε_c obtained throught 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 E.1 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

[55] 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.

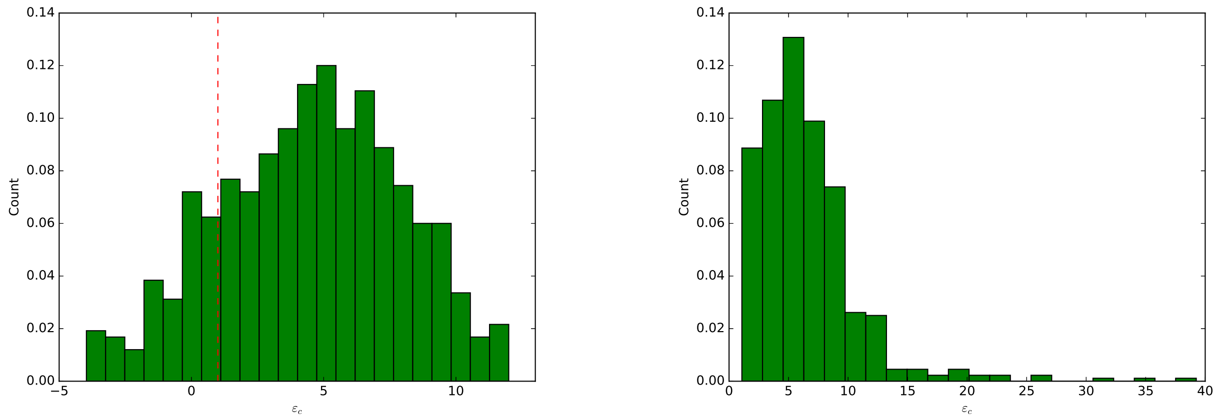


Figure E.1: The two histograms show the distribution of ε_c estimated through a 2SLS procedure that uses a model-generated instrument. The left histogram reports the entire distribution after dropping the top and bottom 5% of the values. The right panel reports the distribution for $\varepsilon_c > 1$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

3.60 to 12.86.

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 E.2 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).

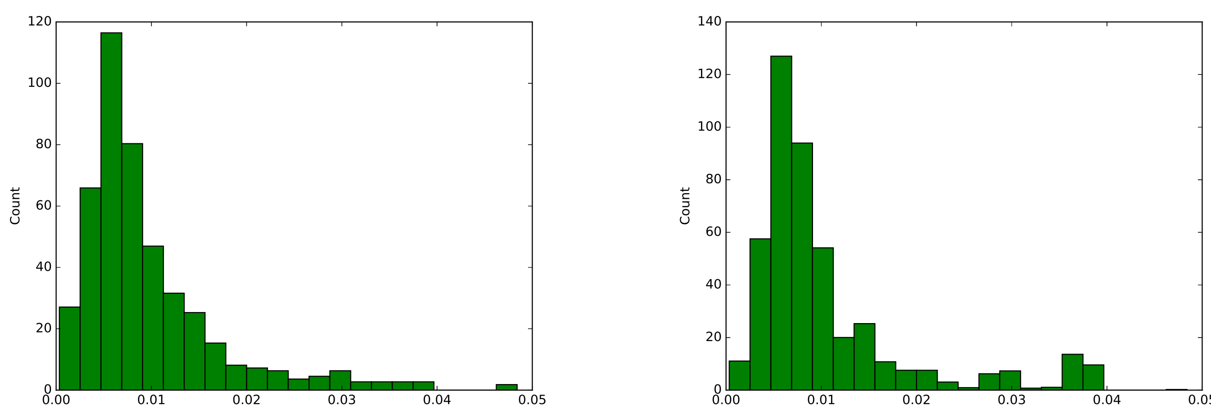


Figure E.2: The two histograms report the distribution of κ_c for those commuting zones with $\varepsilon_c > 1$. The left and right histograms show the unweighted and weighted distribution of this variable, respectively.

F Amazon's HQ2 Counterfactual

Figure F.1: 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.

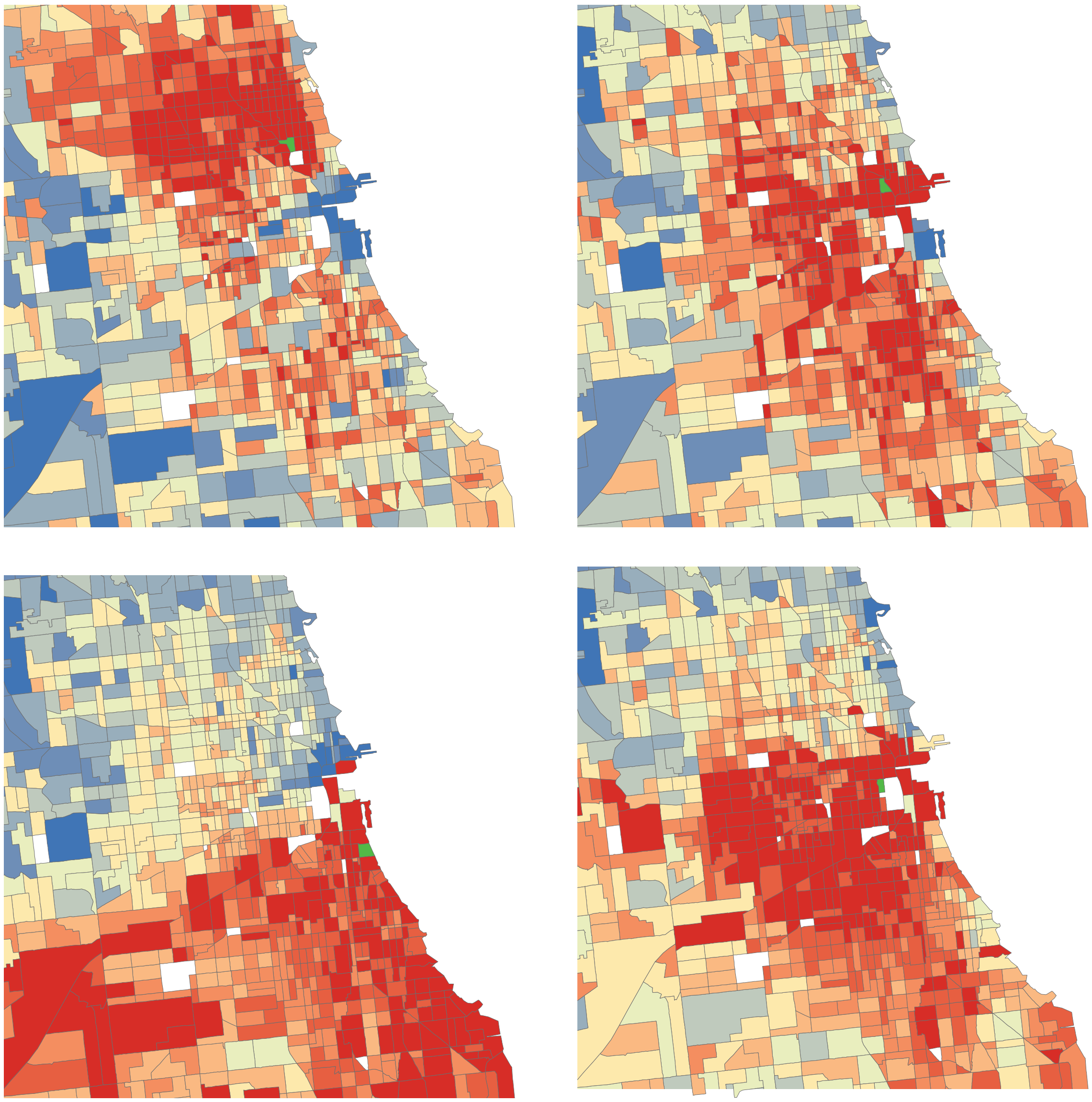


Figure F.2: Change in high-knowledge workers in each census tract of Chicago as a result of Amazon's new headquarter locating in a specific neighborhood (colored in green). Panel (a): A. Finkl & Sons steel plant; Panel (b): Tribune Media River Front property; Panel (c): Michael Reese Hospital premises; Panel (d): Old Main Post Office. 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.