Cities and Technological Waves*

Enrico Berkes† Ruben Gaetani‡ Martí Mestieri§

23rd July 2021

Abstract

New technological opportunities often coincide with major transformations in the economic geography. What are the reasons behind this empirical regularity? In this paper, we argue that spatial and sectoral frictions in idea diffusion make the growth trajectory of cities sensitive to "technological waves", defined as long-term changes in the centrality of knowledge fields in the innovation landscape. We develop a spatial, endogenous growth model with innovation and frictional idea diffusion, that we quantify using a new dataset of historical geolocated U.S. patents. The model accounts for most of the empirical relation between local exposure to technological waves and population growth, and implies large and heterogeneous geographical effects of future technological scenarios.

Keywords: Cities, Innovation, Technology Diffusion, Patents.
JEL Classification: R12, O10, O30, O33, O47.

*We thank Mike Andrews, Paco Buera, Ben Jones, Joel Mokyr, Nicola Persico, Frédéric Robert-Nicoud, Bruce Weinberg, our discussants Klaus Desmet, Ed Glaeser, and Xian Jiang, and seminar attendees at University of Colorado Boulder, University of Toronto, Ohio State University, Conference of Swiss Economists Abroad, NBER Summer Institute (Urban Economics), VMACS Junior Conference, 2021 AEA Meetings, 2021 European Meeting of the Urban Economics Association, 2021 Canadian Summer Conference in Real Estate and Urban Economics, 2021 Barcelona GSE Summer Forum (Firms in the Global Economy), and 2021 SED Meeting for helpful comments and discussions. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System. All errors are our own.

†Ohio State University, berkes.8@osu.edu.
‡University of Toronto, ruben.gaetani@utoronto.ca.
§Federal Reserve Bank of Chicago, Northwestern University and CEPR, marti.mestieri@northwestern.edu.

1 Introduction

The economic geography of countries is in perpetual evolution. In the United States, many cities and regions that have thrived in the past have progressively lost population and influence in favor of newly emerging areas. In recent decades, several cities in the Rust Belt, that had experienced extraordinary growth throughout most of the 20th century, have entered a prolonged phase of decline. At the same time, a handful of urban areas specialized in knowledge-intensive sectors, such as information technology and pharmaceuticals, have gained prominence, becoming increasingly attractive for workers and firms (Glaeser and Gottlieb, 2009; Moretti, 2012). The determinants of these rich dynamics in city growth are still a matter of debate and remain a central question in urban economics.

In this paper, we propose that frictions to knowledge diffusion across locations and fields of knowledge make the growth trajectory of cities sensitive to "technological waves", defined as long-term swings in the importance of sectors in the innovation landscape. Leveraging a new dataset of geolocated U.S. patents spanning the period 1836 through 2015, we document a robust positive relationship between a city's exposure to technological waves and its ability to attract population over the following decades. We then develop a quantitative model that formalizes the feedback between technological waves and the dynamics of city growth. The model combines an economic geography setting with a theory of economic growth that emphasizes the role of recombination, imitation, and knowledge diffusion, as recently developed by Lucas and Moll (2014), Perla and Tonetti (2014), and Buera and Oberfield (2020) among others. The quantitative results suggest that frictional knowledge diffusion accounts for most of the reduced-form relationship between exposure to technological waves and local population growth. Barriers to diffusion across technological fields and geographical areas are both important, each explaining roughly half of this relationship. We use the model to investigate how alternative scenarios of technological waves might transform the U.S. economic geography in future decades. The model predicts substantial differences in the geographical effects of scenarios such as the rise of autonomous vehicles, and the expansion of medical sciences or sustainable agriculture.

In the model, newborn agents make migration and occupational decisions after forming expectations on their lifetime productivity in the location and sector of their choice. Productivity is determined by an imitate-or-innovate decision. Agents can either imitate an idea drawn from the local knowledge distribution, or innovate by improving upon an idea drawn from the distribution of any other location and sector in the economy. The applicability of an idea is affected by frictions reflecting both geographical and technological distance. These frictions imply that knowledge drawn *within* any location-sector can be converted into new inventions more effectively than knowledge drawn from other locations and sectors. For this reason, a city's stock of knowledge determines both current productivity and future innovation possibilities, making the local growth trajectory sensitive to changes in the technological centrality of different sectors, what we refer to as "technological waves". To focus on this novel interplay between economic geography and idea diffusion, the model purposefully abstracts from other drivers of city growth such as endogenous residential amenities.

The framework remains tractable for any arbitrary number of locations, sectors, and time periods, and has a unique equilibrium with an explicit solution. Absent technological wave shocks, the model features a unique balanced growth path (BGP). The productivity distribution for each location-sector endogenously retains a Frećhet structure, and implies an intuitive equation for the law of motion of its scale parameter. This also allows us to characterize knowledge flows in closed form through a gravity representation that can be estimated using patent citation data.

Despite the relative parsimony of the model, the linkages across geographical areas and fields of knowledge that it generates imply non-trivial population dynamics. Before turning to the quantitative analysis, we study the mechanics of the model by log-linearizing the equilibrium conditions around the BGP. We use the log-linear dynamics to derive intuitive theoretical predictions on the relationship between technological waves, the evolution of local productivity, and city growth. First, the growth rate of productivity in each location-sector can be expressed as the sum of sectoral shocks weighted by the reliance of local innovation on ideas from each perturbed sector. This implies that cities specialized in expanding (declining) sectors will experience higher (lower) local productivity growth. Second, combining these productivity dynamics with individual migration decisions, we show that a measure of local exposure to technological waves relative to the overall economy is a sufficient statistic to predict local population growth. In the special case of knowledge flows across sectors being of second-order importance relative to flows within sectors, this measure of exposure resembles a common shift-share variable, and implies that a city grows if and only if the average of sectoral shocks weighted by the incidence of each sector in the city is larger than the corresponding weighted average for the rest of the economy.

We then turn to the quantitative assessment of the role of technological waves – and their interaction with frictional knowledge diffusion – in explaining the evolution of the U.S. economic geography in the last century. We show that the model has a recursive structure that allows us to calibrate the parameters and to recover the unobserved disturbances – including the technological wave shocks – by imposing a small set of transparent assumptions.

The calibrated model is successful in capturing key features of the data that are not directly targeted. In particular, the endogenous mechanism of knowledge creation and diffusion can closely replicate the empirical relationship between exposure to technological waves and population growth, which accounts for roughly 20% of the variation in city growth over the last century. A decomposition exercise reveals that frictions to knowledge diffusion across geographical areas and technological fields equally contribute to the observed relationship. The model is also successful in accounting for two of the most prominent transformations of the U.S. economic geography of the last century: the rise of manufacturing-intensive cities in the early decades of the 20th century, their later decline, and the subsequent emergence of knowledge hubs specialized in information technology.

The mechanism of knowledge creation and diffusion implies that the degree of local diversification plays a central role in determining a city's resilience to technological waves. Simulations of counterfactual paths of sectoral shocks reveal that more diversified cities experience significantly less volatile growth trajectories. There are two factors behind this relationship, that reflect the existence of frictions in the knowledge and geographical space, respectively. First, frictions to diffusion across fields imply that, in response to technological wave shocks, productivity growth is higher in some sectors compared to others. As a result, more diversified cities have a lower chance of large swings (either on the positive or negative side) in their productivity growth, since negative shocks to some sectors are likely to be compensated by positive shocks to other sectors. Second, frictions to diffusion across geographical areas imply that more diversified cities have a broader availability of ideas to draw from, so that (positive or negative) shocks to individual sectors have a weaker impact on the evolution of local productivity.

Finally, we use the quantitative model to predict city dynamics in the coming decades under different scenarios for the evolution of the technological landscape. In particular, we study which cities benefit – and which do not – compared to the status quo, in the following scenarios: (1) a rise in the importance of transportation-related technologies, due to the emergence of new modes of transportation such as autonomous vehicles; (2) an increase in the centrality of pharmaceuticals and biotech in response to new challenges in global health; (3) a comeback of agriculture as a pivotal sector in the innovation landscape as a result of regulatory changes and increasing demand for sustainable farming. We find that cities in the Rust Belt benefit from the first scenario, at the expense of cities in the North-East and the Pacific. The second scenario penalizes knowledge hubs specialized in IT-related innovation, favoring more diversified areas such as Boston and the cities in California outside the Silicon Valley. The third scenario prompts a reallocation of economic activity towards the agricultural areas in the Central states.

Related Literature This paper contributes to multiple strands of the literature. First, the theory is based on idea flows at the location-sector level, with technological and geographical frictions in knowledge diffusion playing a key role in explaining city dynamics. While a rich body of literature has documented the strength and geographical span of localized knowledge spillovers (among others, Jaffe et al., 1993; Audretsch and Feldman, 1996; Greenstone et al., 2010) there has been no attempt to perform a quantitative assessment of the importance of these externalities for understanding long-run city dynamics. One of the main obstacles for providing such an assessment is the complexity of modeling idea diffusion in a spatial setting. In recent years, two flourishing bodies of literature have provided major methodological advances in this direction. First, a number of papers have developed tractable endogenous growth models that emphasize recombination, imitation, and knowledge diffusion as major drivers of aggregate productivity growth (e.g., Perla and Tonetti, 2014; Lucas and Moll, 2014; Buera and Oberfield, 2020). Second, a rich body of work on quantitative spatial economics has developed tools for studying the distribution of economic activity in space, both within cities (e.g., Ahlfeldt et al., 2015; Heblich et al., 2020) and in a system of locations (e.g., Allen and Arkolakis, 2014; Desmet et al., 2018b).[1] This paper combines insights from these two strands of the literature and develops a dynamic, multi-sector, endogenous growth model in a spatial economy that is highly tractable and can be quantitatively disciplined using data on population and patents over a long time period. While a number of papers have used detailed data on patenting to study innovation and knowledge flows in firm and industry dynamics (e.g., Akcigit and Kerr, 2018; Cai and Li, 2019), or developed static models that emphasize localized knowledge spillovers as the main determinant of the economic geography (e.g., Davis and Dingel, 2019), this paper is, to the best of our knowledge, the first attempt at quantitatively assessing the importance of frictions in knowledge diffusion for city dynamics.

An extensive literature has investigated the forces governing the long-run evolution of the economic geography, specifically in its propensity to display path dependence and occasional reversal of fortune (e.g., Brezis and Krugman, 1997; Davis and Weinstein, 2002; Bleakley and Lin, 2012; Kline and Moretti, 2014), as well as in its responsiveness to aggregate shocks such as rising sea-level (e.g., Desmet et al., 2018a), and regional or sectoral shocks (e.g., Caliendo et al., 2018; Hornbeck and Moretti, 2018; Adao et al., 2020). The working hypothesis in this paper is that aggregate changes in the technological landscape, combined with frictional knowledge transmission, have a first-order impact on the geographical distribution of economic activity. The framework can account simultaneously for path dependence and reversal of fortune in city dynamics. While the focus on innovation and idea diffusion is new to this literature, there is a rich body of work that has analyzed the historical dynamics of the U.S. geography, both from an empirical perspective (e.g., Bostic et al., 1997; Simon and Nardinelli, 2002; Michaels et al., 2012; Desmet and Rappaport, 2017) and from a structural and quantitative viewpoint (e.g., Duranton, 2007; Desmet and Rossi-Hansberg, 2014; Nagy, 2017; Allen and Donaldson, 2018; Eckert and Peters, 2019).

This paper also contributes to the long-standing debate between the returns to local spe-

[1] Comprehensive reviews of these bodies of literature are provided by Buera and Lucas (2018) for models of endogenous growth with idea flows, and by Redding and Rossi-Hansberg (2017) for quantitative spatial equilibrium models.

cialization (Marshall, 1890) and urban diversity (Jacobs, 1969), and their effect on city growth. Notable contributions in this literature include Glaeser et al. (1992), whose empirical assessment finds evidence supporting Jane Jacob's view of urban variety as a key driver of local employment growth, and Duranton and Puga (2001), who develop a model in which diversified and specialized cities coexist in equilibrium.[2] This paper suggests and quantifies a new channel through which urban diversification affects long-run city growth, namely, by shaping the responsiveness of a city to changes in the surrounding technological landscape.[3] In this sense, the model provides a new lens for interpreting the effect of local policies directed at increasing local diversification.

The remainder of the paper is organized as follows: Section 2 introduces the data and presents historical trends and motivational facts on the relationship between city growth and the technological landscape. Section 3 introduces the model and derives the main theoretical predictions. Section 4 describes the model calibration and Section 5 presents the quantitative results. Section 6 discusses avenues for further research and concludes.

2 Data and stylized facts

Technological change is a slow-moving secular process. To study how the rise and fall of technologies determines the growth and decline of cities, we therefore need to consider a time period long enough to capture multiple episodes of technological replacement. In this paper, we exploit a recently assembled dataset of historical geolocated patents spanning the years 1836 through 2015. We approximate U.S. cities by 1990 commuting zones (CZs), that we keep fixed throughout the analysis.[4]

2.1 Data sources

To measure innovative activities at the city level, we collect patents data from the Comprehensive Universe of U.S. Patents (CUSP). The CUSP contains information on the near-universe of patents issued by the U.S. Patent and Trademark Office (USPTO) between 1836 and 2015, with an estimated coverage above 90% in each year.[5] From the CUSP, we gather information

[2]A comprehensive overview of the patterns of specialization across U.S. locations is provided by Holmes and Stevens (2004).

[3]Consistently with this interpretation, Balland et al. (2015) find that cities with more diverse knowledge bases are less sensitive to technological crises, defined as sustained declines in patenting activity.

[4]Although commuting flows have changed over time, assuming a stable geography allows us to abstract from annexations and redefinition of town borders that have been pervasive phenomena throughout the 19th and 20th century.

[5]Berkes (2018) provides details about the data collection procedure, as well as summary statistics and stylized facts related to the underlying data. Andrews (2019), in a comparison of historical patents data, describes it as

on the technology classes and location of the first inventor listed on each patent, as well as their filing date. The CUSP assigns patents to the city of the inventor's residence and does not rely on the county reported in the patent's text. This allows us to build geographically consistent measures of innovation at the commuting zone level over the long time span covered by our study.

Data on population, human capital, and industry composition at the commuting zone level are assembled, for each decade between 1870 and 2010, from the corresponding decennial censuses,[6] as provided by the Integrated Public Use Microdata Series (IPUMS, Ruggles et al., 2021) and the National Historical Geographic Information System (NHGIS, Manson et al., 2021).[7] We build a consistent measure of the local density of human capital that combines available information on local literacy and education. To make this measure comparable across decades, we rank cities in terms of the relevant measure for each decade and use the resulting ranking for the analysis.[8] In each decade, we also collect information on local employment by industry. Appendix D provides further details on the construction of the data.

Throughout the empirical and structural analysis, observations correspond to 20-year periods between 1870 and 2010. Patent counts by sector are obtained by adding patents filed in the two decades around the focal year (for example, patents in the 1990 observation correspond to the total patent count between 1980 and 1999). We restrict the sample to the subset of commuting zones in the contiguous United States that accounted for at least 0.02% of the total population for each decade since 1890. This delivers a sample of 373 commuting zones, that jointly account for roughly 87% of the U.S. population in 2010.[9] Sectors are defined as the technological class-groups obtained by grouping 3-digit International Patent Classification (IPC) categories into 11 class-groups, as detailed in Appendix Table A.1.[10]

"currently the gold standard, in terms of the patent- and inventor-level information included in the published datasets". Some slices of the data have already been used in Berkes and Nencka (2020) who study the effect of Carnegie libraries on the local patenting activity, Clemens and Rogers (2020) who study how procurement policies affect the characteristics of medical innovation, and Babina et al. (2020) who study the effect of the Great Depression on innovative activities in the U.S.

[6]For the 2010 observation, we use multi-year averages of the American Community Survey (ACS).

[7]Since data from the 1890 decennial census are not available, we construct the 1890 observations by linearly interpolating the observations from the 1880 and 1900 decennial censuses.

[8]The relevant measure is a summary index that includes several indicators of the local density of human capital. The specific indicators we use change over time depending on the availability of information in the historical Census. In the early decades, the measure focuses on indicators of literacy and schooling, while in later decades it emphasizes the local density of workers with high educational attainment.

[9]As a reference point, this rule requires that cities had a population of at least 10,711 people in 1890 and 60,387 people in 2010.

[10]Patents listing multiple 3-digit IPC classes are assigned fractionally to class-groups, in proportion to the frequency of appearance of each class-group in the list of 3-digit IPC classes.

2.2 Historical trends

The last 150 years has witnessed major shifts in the technological landscape. The bottom-right panel of Figure 1 shows how the distribution of the national patenting output across the 7 main IPC classes has evolved since 1870.[11] The share of patents in class A ("Human necessities") – that includes innovation related to both agriculture and medical sciences – declined in the first part of the century, as agriculture lost its centrality to classes complementary to the heavy manufacturing industry, such as B ("Transportation") and F ("Mechanical Engineering"). Class A rebounded in recent decades as innovation in medicine gained prominence. In the second part of the century, classes G ("Physics") and H ("Electricity") became more central in the national shares, making up more than 50% of the overall innovation output in 2010.[12]

The composition of patenting not only changes significantly over time, but also varies considerably across cities at any point in time. The top panels of Figure 1 depict two of the archetypal examples of this heterogeneity. Detroit (top-left) has been specialized in the production of patents of class B and F since the early 1900s. In 1930, these two classes made up about 70% of its patenting portfolio. This pattern has remained broadly unchanged throughout the century, with a slight shift towards patents of classes G and H since the 1990s. Austin (top-right) exhibits fairly diversified innovation activities until the 1970s, when the share of patents of classes G and H started expanding, reaching 90% of the portfolio by 2010. By contrast, Boston (bottom-left) displays a diversified patenting output that, throughout the decades, has closely tracked the national trends.

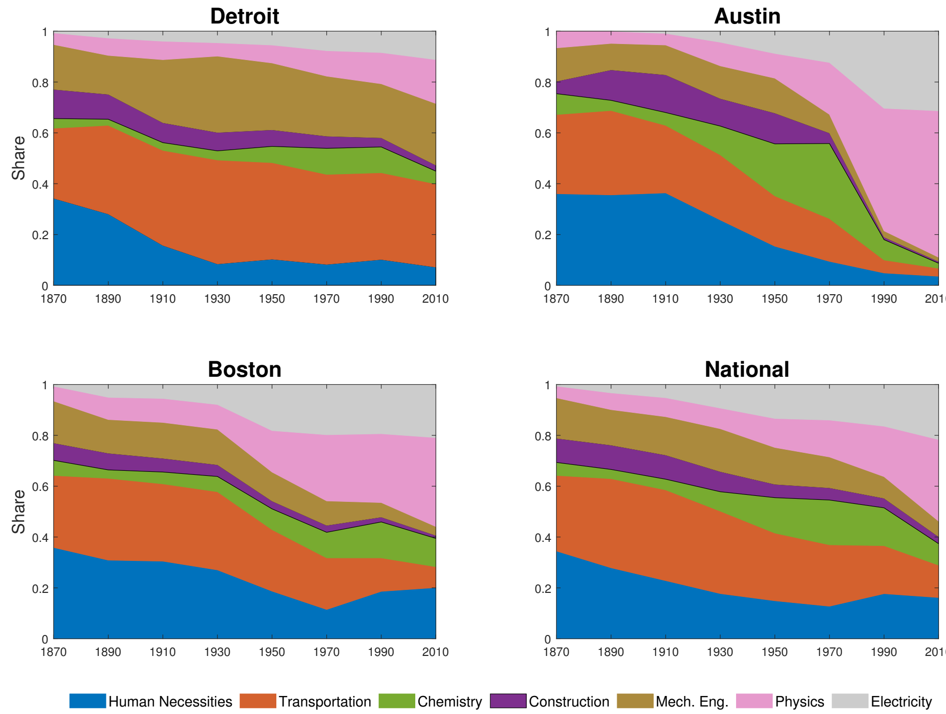
In this paper, we argue that the heterogeneity in the composition of local patenting makes cities unevenly positioned to take advantage of new innovation opportunities. This makes cities' trajectories sensitive to changes in the technological landscape, and contributes to explaining the irregular historical dynamics of U.S. urban and regional growth. The experiences of Detroit, Austin, and Boston since the late 1800s exemplify this point. Figure 2 shows the 20-year population growth of those three commuting zones since 1890, after controlling for Census Division-time fixed effects. Detroit displays the most striking growth rates in the decades after the advent of the automobile industry around 1910, followed by a long-lasting decline that resulted in a steady loss of population since the 1980s. The commuting zone of Austin experienced a specular trajectory. The city declined in relative terms in the first half of the 20th century, as the Texas Oil Boom favored areas of the state that were rich of oil, making Austin slip from the 4th to the 10th place among Texas's largest cities.[13] However, in recent decades Austin has

[11] Class names are abbreviated for clarity. The full description of each class can be found at https://www.wipo.int/classifications/ipc/en/.

[12] Classes G and H include the bulk of innovation related to computers, electronics, and information and communication technology.

[13] https://tshaonline.org/handbook/online/articles/hda03

Figure 1: **Composition of the technological output**



Notes: Composition of patenting output across the 7 main IPC classes in Appendix Table A.1. Patent count for year t is constructed as the sum of patents filed between $t-10$ and $t+9$. Class name are abbreviated. The full description of each class is available at `https://www.wipo.int/classifications/ipc/en/`.

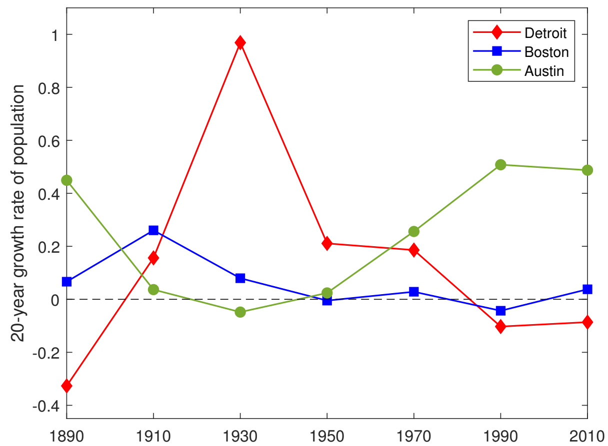
emerged as one of the leading innovation hubs in the country, leveraging its richness of science-based firms and a large college-educated population. Finally, the commuting zone of Boston had a still different experience: Throughout the last century, it has retained a considerably less volatile path, characterized by moderate relative growth interrupted by occasional periods of modest relative decline. The persistent diversification of Boston's patenting output could have made the city less sensitive to changes in the technological landscape, explaining the stability of its growth path.[14]

2.3 Technological waves and the growth and decline of cities

Taken together, Figures 1 and 2 suggest that changes in the importance of technological fields might differentially affect the growth trajectory of cities. We now show that this pattern holds systematically over the long time period covered by our data, and is robust to controlling for possible confounders such as the local density of human capital and the local industrial

[14] Glaeser (2005) provides an overview of the causes of the slow decline of Boston between 1920 and 1980, and the subsequent re-emergence of the city. The high density of human capital is proposed as the major factor behind its resilience.

Figure 2: **City dynamics**



Notes: Residuals of a regression of 20-year growth rate of population on Census Division-time fixed effects, 1890-2010.

composition.

Throughout the paper, we refer to changes in the technological landscape, here captured by shifts in the composition of national patenting by class-group, as *technological waves*. We argue that cities whose patenting portfolio is concentrated in expanding fields are in a better position to take advantage of new innovation possibilities, and will experience higher productivity and population growth. In the spirit of Bartik (1991), we construct a shift-share measure of local exposure to the technological wave:

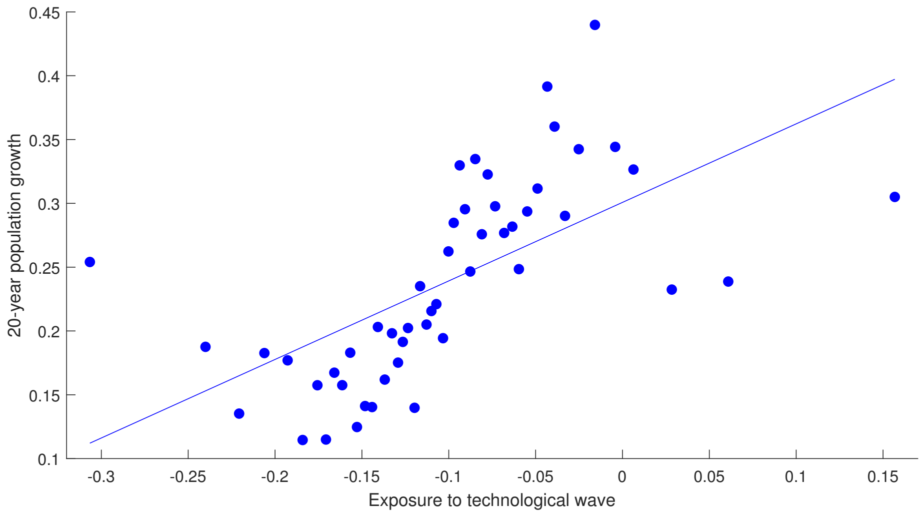
$$Exp_{n,t} \equiv \sum_{s \in S} Share_{n,s,t-1} \times g_{s,t}, \tag{1}$$

where $Share_{n,s,t-1}$ is the share of patents filed in commuting zone n belonging to class-group s at time $t-1$, and $g_{s,t}$ is the growth rate in the national share of patents of class-group s between $t-1$ and t. A city whose portfolio of patents is concentrated in expanding (declining) class-groups will record a positive (negative) value of $Exp_{n,t}$, reflecting a favorable (adverse) exposure to the current technological wave.

Figure 3 shows a bin-scatter plot of the relationship between the measure of exposure, $Exp_{n,t}$, and the 20-year growth rate of local population, between 1910 and 2010. Both measures are residualized with respect to two lags of log-population and Census Division-time fixed effects, to account for size, convergence, and persistence effects, and for the differential growth rates of commuting zones across space explained by factors such as the Westward expansion or the Great Northward Migration.[15] The scatter plot reveals a strong positive correlation, imply-

[15]The earliest period corresponds to population growth between 1890 and 1910, and controls for two lags of

Figure 3: **Technological waves and city growth**



Notes: Bin-scatter plot of exposure to the technological wave, as defined in Equation (1), and 20-year population growth, 1910-2010. The bin-scatter plot is residualized with respect to Census Division-time fixed effects and two lags of log-population.

ing that, over the period considered, cities with a more favorable exposure to the technological wave have experienced systematically higher population growth compared to cities in the same Census Division.

Table 1 reports the corresponding regression results. The estimate in column 2 (that controls for Census Division-time fixed effects) implies that an increase in the measure of exposure of one residual standard deviation is associated with an increase of 12.5% of a residual standard deviation in population growth. In column 3, we further control for the historically-consistent measure of local density of human capital. This indicator is correlated with population growth, but has a negligible effect on the estimated coefficient of the exposure measure. This suggests that the effect of the exposure measure on population growth is driven by the composition of knowledge across fields, rather than by the overall availability of skills and ideas.

In our analysis, we do not take a stance on what drives changes in the national patenting shares by class-group. Technological waves could result from genuine scientific and technological developments, such as advances in computing and bio-technology. Alternatively, they could be triggered by political and environmental factors, such as regulation, trade agreements, or changes in consumer preferences. What is critical for our analysis is that, whatever their origin, technological waves differentially affect the returns from innovating in different fields and, as a result, they influence the evolution of patenting shares across class-groups. However, factors

log-population (1870 and 1890). The latest period corresponds to population growth between 1990 and 2010, and controls for two lags of log-population (1970 and 1990).

Table 1: **Technological waves and city growth**

	Growth rate of population			
	(1)	(2)	(3)	(4)
Exposure to tech. wave	0.428***	0.400***	0.370***	0.279***
	(0.082)	(0.067)	(0.071)	(0.071)
Log-population (lag 1)	0.278***	0.264***	0.255***	0.203***
	(0.056)	(0.045)	(0.045)	(0.048)
Log-population (lag 2)	-0.301***	-0.272***	-0.269***	-0.242***
	(0.053)	(0.038)	(0.038)	(0.039)
Human capital (ranking)			0.082*	0.034
			(0.047)	(0.047)
Industry composition				0.660***
				(0.116)
Fixed effects	T	CD×T	CD×T	CD×T
# Obs.	2,238	2,238	2,238	2,228
R^2	0.39	0.50	0.51	0.53

Notes: CZ level regression, 1910-2010. Dependent variable defined as growth rate of population over 20 years. "T" denotes time fixed effects, and "CD×T" denotes Census Division-time fixed effects. Standard errors clustered at the CZ level in parenthesis. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

that differentially impact patenting across fields might be correlated with other industry-level shocks that drive differences in population growth across cities, confounding our estimates. To address this concern, in column 4 we directly control for a shift-share variable built using employment by industry and otherwise analogous to Equation (1).[16] As expected, this Bartik variable is a strong predictor of contemporaneous population growth, and its inclusion slightly reduces the estimated coefficient on the exposure measure, that remains large and significant at the 1% level.

In Appendix Table A.3, we show that results are consistent when splitting the sample into an early (1910-1950) and late (1970-2010) sub-samples, confirming that this correlation is a stable regularity throughout the period considered.

[16]See Appendix D for details on the construction of the data on employment by industry at the commuting zone level. Industries correspond to the 12 main industries in the 1950 Census Bureau industrial classification system.

2.4 Frictions to knowledge diffusion across locations and fields of knowledge

While the systematic evidence in Table 1 is novel, a rich literature has observed that changes in the technological environment often coincide with episodes of transformation in the economic geography, and has suggested explanations for this pattern. Notably, Brezis and Krugman (1997) propose that, because of knowledge accumulated over time, established cities enjoy an advantage in old technologies. This advantage implies a high opportunity cost of local factors such as land and labor, which prevents established cities from embracing new technologies, and leads those cities to decline as new technologies mature.[17]

In this paper, we explore an alternative, possibly complementary, hypothesis. We argue that the robust correlation in Table 1 is partly explained by the existence of frictions to the diffusion of ideas across space and fields of knowledge, that prevent cities from optimally reallocating resources to take advantage of technological waves. Cities whose innovation portfolio is skewed towards expanding fields are better positioned to embrace new innovation opportunities and will become more attractive for workers and firms. In the following section, we formalize this hypothesis by embedding frictional knowledge diffusion in a spatial equilibrium model of endogenous growth.

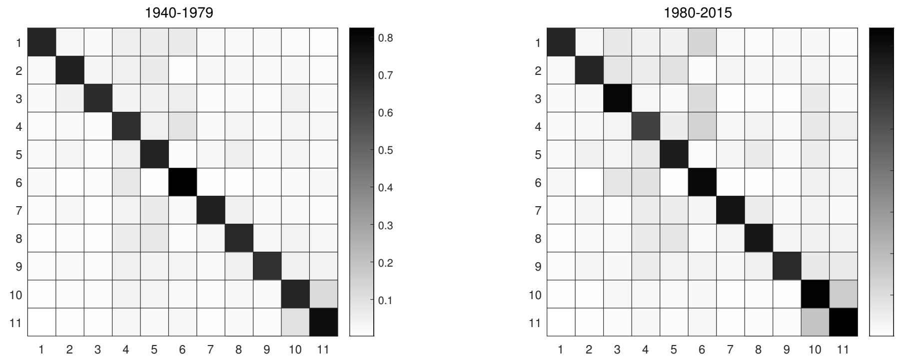
The fact that knowledge diffusion is highly localized has been widely documented in the literature on the geography of innovation. Within this literature, a rich body of work, starting with Jaffe et al. (1993), has provided evidence of this localization by studying the patterns of patent citations (Murata et al., 2014; Kerr and Kominers, 2015). This evidence of localization is confirmed in our citation data and, furthermore, it does not appear to weaken over time. For patents filed since 1940, citations to the same commuting zone account for 16.8% of all citations.[18] When we split the sample between an early (1940-1979) and a late sub-sample (1980 onwards) we find that, if anything, the evidence of localization becomes stronger. The share of citations to the same commuting zone of the citing patent is 14.5% in the early sub-sample, and increases to 18.1% in the late sub-sample.

Analogously, we find strong evidence of localization of patent citations in the technological space. The heatmaps in Figure 4 display, for the early (left panel) and late (right panel) sub-samples, the probability that a citation from each technology class-group on the vertical axis is directed towards each of the class-groups on the horizontal axis. Both heatmaps show that

[17] This argument was first proposed by Brezis et al. (1993) (and later refined by Desmet, 2002) in the context of leapfrogging among leading and lagging nations.

[18] Since patent citations are not consistently available in the earlier decades, when considering citations we restrict the sample to all the patents filed since 1940. A separate section containing referenced patents was formally introduced in patent documents only in 1947. In constructing these statistics, we only consider citations to and from commuting zones included in our sample, and assign a weight of one to each citing patent.

Figure 4: **Patent citations across fields**



Notes: Probability that patents from the class-group on the vertical axis cite patents from the class-group on the horizontal axis. Probabilities are computed using patents filed since 1940 whose first inventor is in one of the 373 commuting zones in the main sample. Each citing patent is assigned a total weight of one.

citations are strongly concentrated along the diagonal, suggesting a high degree of technological localization in the diffusion of ideas, that does not vanish over time.

In the model calibration, we use the same data to estimate a gravity equation for knowledge flows and discipline the parameters controlling the strength of frictions to idea diffusion.

3 Model

In this section, we develop a quantitative model that embeds endogenous growth through innovation and frictional idea diffusion into a spatial equilibrium framework. The theory formalizes the feedback between changes in the innovation landscape and the evolution of the economic geography over time, and rationalizes the reduced-form relationship between population growth and exposure to technological waves. To provide a transparent illustration of the mechanisms at play, in what follows we present the model in the most parsimonious form. As we show in Appendix F, the model can be extended, without a prohibitive loss in tractability, to allow for overlapping generations, costly migration, trade, and local agglomeration or congestion forces.

3.1 Environment

We consider an economy comprising a finite set N of locations and a finite set S of sectors. In what follows, we refer as N and S to both the sets of locations and sectors, and their cardinality. Time is discrete and indexed by t. At each point in time, the economy is populated by a mass L_t of individuals.

3.1.1 Preferences, endowments, and demographics

In each period, a new generation of individuals is born in the location of their parents and makes migration and occupational decisions. Individuals live for one period and, at the end of the period, have f_t children. There are no moving costs.

Migration and occupational choices are made to maximize expected utility, subject to idiosyncratic utility draws that affect the individual desirability of each location-sector pair. Specifically, at the beginning of the period, each individual i receives a full set of stochastic utility draws, one for each location-sector in the economy:

$$\mathbf{x}_i = \{x_{n,s,i}\}_{(n,s)\in N\times S}\,.$$

Each value $x_{n,s,i}$ is a random draw from a Fréchet distribution with shape parameter $\zeta > 1$. Individuals then choose the location-sector pair (n, s) that provides them with the highest expected utility, given by:

$$U_{n,s,t}(\mathbf{x}_i) = u_n\, x_{n,s,i}\, c_{n,s,i,t}, \tag{2}$$

where u_n is the level of time-invariant amenities in city n and $c_{n,s,i,t}$ denotes consumption of the final good by individual i in location-sector (n, s) at time t.

Since we calibrate the time period to be 20 years, the assumptions on the absence of moving costs and the demographic structure should be interpreted with this time horizon in mind.

3.1.2 Production and innovation technologies

Each agent i is endowed with one unit of labor that she supplies inelastically with productivity q_i. Total output in the economy is given by a linear aggregator over individual productivity across all locations and sectors:

$$Y_t = \sum_{n\in N}\sum_{s\in S} L_{n,s,t}\mathbb{E}[q_{n,s,t}],$$

where $L_{n,s,t}$ denotes the mass of agents in location-sector (n, s) and $\mathbb{E}[q_{n,s,t}]$ denotes their average productivity.

Individual productivity is determined endogenously by a process of knowledge diffusion that subsumes a choice on whether to imitate or innovate.[19] At the beginning of each period every agent i in the new generation receives a full set of idiosyncratic, independently distributed draws:

[19] Our description of the process of innovation and knowledge diffusion builds on the model developed by Buera and Oberfield (2020), who study an environment in which the distribution of ideas endogenously converges to a Fréchet distribution.

$$\mathbf{z}_{n,s,i} = \left\{ z^l_{n,s,i}, \left\{ z^x_{m,r,i} \right\}_{m,r \in N \times S} \right\}. \tag{3}$$

The first term, $z^l_{n,s,i}$, represents a random draw from the distribution of productivity among agents employed in location-sector (n,s) in the previous generation, whose cumulative distribution is denoted by $F_{n,s,t-1}(q)$. This draw can be interpreted as knowledge that individual i learns from their teacher, mentor, or manager, and can be imitated and adopted directly in production.[20] If the agent chooses to adopt this idea in production, their lifetime productivity is

$$q_{n,s,i,t} = z^l_{n,s,i}.$$

The second set of terms, $\{z^x_{m,r,i}\}_{m,r \in N \times S}$, represents a full vector of random draws from each productivity distribution in all locations and sectors in the previous generation, with corresponding cumulative distributions $\{F_{m,r,t}(q)\}_{m,r \in N \times S}$. Note that this full set of draws includes local ones (i.e., $m = n$ and $r = s$). These draws can be interpreted as knowledge that the agent acquires by various channels of transmission, such as books, radio, television, internet, or via casual interactions with local or non-local individuals. Although these ideas cannot be imitated and adopted directly in production, they can be used as an input for innovation. In particular, an agent employed in (n,s) can use an idea $z^x_{m,r,i}$ to innovate and achieve productivity

$$q_{n,s,i,t} = \frac{\epsilon_{n,s,t}\, \alpha_{r,t}\, z^x_{m,r,i}}{d_{(m,r)\to(n,s)}}. \tag{4}$$

In Equation (4), the term $\alpha_{r,t}$ represents the centrality of sector r in the innovation landscape. The higher the value of $\alpha_{r,t}$, the more effectively can knowledge in sector r be developed into innovation for any sector. We refer to changes in $\alpha_{r,t}$ as *technological wave* shocks. The term $d_{(m,r)\to(n,s)}$ captures the geographical and technological frictions that discount the effectiveness of knowledge transmission between the idea origin (m,r) and the idea destination (n,s). The term $\epsilon_{n,s,t}$ is a structural residual that captures the current effectiveness of innovation in (n,s) and is common to all innovators in the location-sector pair. It accounts for all the residual factors that affect the productivity of the local sector but are not otherwise included in (4), such as the opening of production facilities, universities, and research centers.

There is no market to smooth consumption across generations. Thus, agents live hand-to-mouth, with consumption of final good given by own production:

[20] De la Croix et al. (2018) develop a model in which the institutions controlling the effectiveness of knowledge transmission between journeymen and apprentices contribute to explain differences across countries in long-run growth.

$$c_{n,s,i,t} = q_{n,s,i,t}.$$

3.2 Equilibrium

3.2.1 Diffusion of knowledge

Agent i in location-sector (n, s) chooses whether to imitate or innovate to maximize her productivity given her vector of idiosyncratic idea draws $\mathbf{z}_{n,s,i}$:

$$q_{n,s,i,t} = \max \left\{ z^l_{n,s,i}, \; \max \left\{ \frac{\epsilon_{n,s,t} \, \alpha_{r,t} \, z^x_{m,r,i}}{d_{(m,r)\to(n,s)}} \right\}_{m,r \in N \times S} \right\} \tag{5}$$

Equation (5) shows how this process can be divided into two steps. First, the agent chooses the best innovative idea available to her. Then she compares this idea with her imitation draw, and picks the one that yields higher productivity for her.

The following assumption, that we maintain throughout the paper, will play an important role in keeping the theory tractable:

Assumption A1. *The initial productivity distribution $F_{n,s,0}(q)$ in all location-sector pairs (n, s) is Fréchet with shape parameter $\theta > 1$ and scale parameter $\lambda_{n,s,0} > 0$:*

$$F_{n,s,0}(q) = e^{-\lambda_{n,s,0} q^{-\theta}}. \tag{6}$$

A multivariate Fréchet distribution with common shape parameter is max-stable. This implies that, under Assumption A1, the resulting distribution over the max of Fréchet draws is also Fréchet with the same shape parameter.[21] Combining (5) with (6), we find that individual productivity at any time $t \geq 0$ is distributed Fréchet with shape parameter $\theta > 1$ and with scale parameter evolving according to the following law of motion:

$$\lambda_{n,s,t} = \underbrace{\lambda_{n,s,t-1}}_{\text{Imitation}} + \underbrace{\sum_{m \in N} \sum_{r \in S} \lambda_{m,r,t-1} \left(\frac{\epsilon_{n,s,t} \, \alpha_{r,t}}{d_{(m,r)\to(n,s)}} \right)^{\theta}}_{\text{Innovation}}. \tag{7}$$

Equation (7) summarizes the productivity dynamics implied by the model. The scale parameter of the new generation in location-sector (n, s) is equal to the scale parameter of the previous generation augmented by a second term which captures inventive activities. This second term is composed by the sum of scale parameters across all location-sectors weighted by their applicability to location-sector (n, s). This applicability term includes the importance of

[21] The same degree of tractability can be achieved without assuming independence, as in Lind and Ramondo (2019).

each field of knowledge ($\alpha_{r,t}$) and the local effectiveness of innovation ($\epsilon_{n,s,t}$), and is discounted by technological and physical distance between location-sector pairs ($d_{(m,r)\to(n,s)}$).

Equation (7) also implies that, conditional on innovating, the probability that an inventor in location-sector (n,s) builds upon an idea from any location-sector (m,r) at time t can be expressed as follows:

$$\eta^t_{(m,r)\to(n,s)} = \frac{\lambda_{m,r,t-1}\left(\frac{\alpha_{r,t}}{d_{(m,r)\to(n,s)}}\right)^{\theta}}{\sum_{l,p}\lambda_{l,p,t-1}\left(\frac{\alpha_{p,t}}{d_{(l,p)\to(n,s)}}\right)^{\theta}}. \tag{8}$$

3.2.2 Migration and occupational choice

At the beginning period t, agents in the new generation observe sectoral and local shocks ($\alpha_{r,t}$ and $\epsilon_{n,s,t}$) but do not know their idiosyncratic idea draws, so they have to form expectations about productivity before making their migration and occupational decisions. Agent i moving to location-sector pair (n,s) has expected utility equal to

$$\mathbb{E}\left[U_{n,s,t}(\mathbf{x}_i)\right] = u_n\, x_{n,s,i}\, \mathbb{E}\left[q_{n,s,t}\right]. \tag{9}$$

In equilibrium, $q_{n,s,t}$ is distributed Fréchet with shape parameter θ and scale parameter $\lambda_{n,s,t}$, which can be inferred at time t via the law of motion (7), so that

$$\mathbb{E}\left[q_{n,s,t}\right] = \Gamma\left(1-\frac{1}{\theta}\right)\lambda_{n,s,t}^{\frac{1}{\theta}}, \tag{10}$$

where $\Gamma(\cdot)$ denotes the gamma function. This implies that the probability that any newborn individual selects location-sector (n,s) is

$$\pi_{n,s,t} = \frac{\left(u_n\,\lambda_{n,s,t}^{\frac{1}{\theta}}\right)^{\zeta}}{\sum_{m,r}\left(u_m\,\lambda_{m,r,t}^{\frac{1}{\theta}}\right)^{\zeta}}. \tag{11}$$

Thus, the mass of agents in location-sector (n,s) at time t is equal to

$$L_{n,s,t} \equiv \pi_{n,s,t}\, L_{t-1}\, f_t. \tag{12}$$

For notational convenience, we define $\pi_{n,t} \equiv \sum_{s\in S}\pi_{n,s,t}$ and $L_{n,t} \equiv \sum_{s\in S} L_{n,s,t}$ as, respectively, the share and mass of individuals living in location n.

3.2.3 Equilibrium Definition

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

Definition 1. *For a given set of initial conditions*

$$L_0,\ \{u_n\}_{n\in N},\ \{\lambda_{n,s,0}\}_{n,s\in N\times S},$$

and a given path for the exogenous variables

$$\{f_t\}_{t\geq 0},\ \{\alpha_{r,t}\}_{r\in S,t\geq 0},\ \{\epsilon_{n,s,t}\}_{n,s\in N\times S,\,t\geq 0},$$

an equilibrium is a path for the endogenous variables

$$\{\lambda_{n,s,t},\ \pi_{n,s,t},\ L_{n,s,t}\}_{n,s\in N\times S,\,t\geq 0}$$

that satisfies the following conditions:

1. *Migration and occupational probabilities* $\{\pi_{n,s,t}\}_{n,s\in N\times S,\,t\geq 0}$ *satisfy equation* (11).
2. *The path for* $\{\lambda_{n,s,t}\}_{n,s\in N\times S,\,t\geq 0}$ *satisfies the law of motion of equation* (7).
3. *Population by location-sector* $\{L_{n,s,t}\}_{n,s\in N\times S,\,t\geq 0}$, *satisfies the transition identity* (12).

All equilibrium conditions have an explicit solution. Hence, a unique equilibrium exists and can be written in closed form for any given set of initial conditions and any given path for the exogenous variables.

3.2.4 Existence and uniqueness of a balanced growth path

We define a balanced growth path (BGP) as an equilibrium in which sectoral importance $\alpha_{r,t}$ and structural residuals $\epsilon_{n,s,t}$ are constant, and the scale parameters $\lambda_{n,s,t}$ grow at the same rate for all location-sectors. Incidentally, these conditions also imply that migration and occupational choices (and, as a result, the distribution of people across locations and sectors) are constant over time.

Notice that Equation (7) can be rewritten in matrix form as

$$\vec{\lambda}_{t+1} = A_t\vec{\lambda}_t, \tag{13}$$

where $\vec{\lambda}_t$ is a $N\times S$ vector of all scale parameters $\lambda_{n,s,t}$ and A_t is the $(N\times S)^2$ diffusion matrix implied by Equation (7). In BGP, the matrix A_t is constant, and is denoted by A^* (in what follows, we use upper-stars to denote variables at their BGP value).

From Equation (13), it is immediate to see that, in BGP, $\vec{\lambda}_t$ must be an eigenvector of A^*, with the associated eigenvalue equal to its gross growth rate $1+g^*_\lambda$. The Perron-Frobenius theorem guarantees that A^* has a unique positive eigenvector (and corresponding eigenvalue), provided that all entries in A^* are positive. A sufficient condition for A^* to have only positive entries is that frictions to knowledge diffusion $d_{(m,r)\to(n,s)}$ are positive and finite for each combination of idea origin and destination. This proves the following:[22]

Proposition 1. *Let $0 < d_{(m,r)\to(n,s)} < +\infty$ for all $(m,r),(n,s) \in N \times S$. Then, for each set of constant sectoral importance $\{\alpha_r^*\}_{r\in S}$ and structural residuals $\{\epsilon_{n,s}^*\}_{(n,s)\in N\times S}$, there exists a unique balanced growth path in which $\{\lambda_{n,s,t}\}_{(n,s)\in N\times S,t\geq 0}$ grow at constant rate g_λ^*.*

While scale parameters grow at the same rate, in general the BGP implies persistent variation in productivity across locations and sectors. In particular, in BGP, the following relationship holds for each location-sector (n,s):

$$g_\lambda^* = (\epsilon_{n,s}^*)^\theta \sum_{m,r} \left(\frac{\lambda_{m,r}}{\lambda_{n,s}}\right)^* \left(\frac{\alpha_r^*}{d_{(m,r)\to(n,s)}}\right)^\theta . \tag{14}$$

This equation illustrates that higher BGP productivity can result from higher structural residuals ($\epsilon_{n,s}^*$) or from higher proximity (as dictated by frictions $d_{(m,r)\to(n,s)}$) to location-sectors with higher knowledge stock ($\lambda_{m,r}$) and sectoral importance (α_r^*).

3.3 Log-linearized model dynamics

We now study the dynamics of the model by log-linearizing the equilibrium conditions around the BGP. We assume that at time $t-1$ the economy is in a BGP in which the average productivity in each location-sector grows at the same rate and, as a result, the distribution of people across locations is constant. At time t, the economy is hit by technological wave shocks $\{\hat{\alpha}_{r,t}\}_{r\in S}$, where hats denote log-deviations from BGP values.

First, consider the dynamics of the scale parameter of the local distribution of productivity, $\lambda_{n,s,t}$. Log-linearizing Equation (7) yields

$$\hat{\lambda}_{n,s,t} = \frac{\theta(\epsilon_{n,s}^*)^\theta}{1+g_\lambda^*} \sum_{m,r} \left(\frac{\lambda_{m,r}}{\lambda_{n,s}}\right)^* \left(\frac{\alpha_r^*}{d_{(m,r)\to(n,s)}}\right)^\theta \hat{\alpha}_{r,t}. \tag{15}$$

Multiplying and dividing the right-hand side of (15) by g_λ^*, and using (8) and (14), we derive the following proposition that links changes in local sectoral productivity to technological wave shocks via the strength of the knowledge diffusion link between the perturbed sector and the receiving location-sector:

[22] Huang and Zenou (2020) is another paper that studies the BGP properties of an endogenous growth model with idea diffusion across multiple sectors. While the setting for idea diffusion in Huang and Zenou (2020) is different from ours, in both models the Perron-Frobenius theorem is central to establishing the existence and uniqueness of a BGP equilibrium.

Proposition 2. *The log deviation of the scale parameter of the productivity distribution of each location-sector (n, s) from the BGP, $\hat{\lambda}_{n,s,t}$, is equal to the sum over all sectors $r \in S$ of the sectoral shock to r, $\hat{\alpha}_{r,t}$, weighted by the reliance of innovation in (n, s) on ideas from sector r, $\eta^*_{r\to(n,s)} \equiv \sum_{m\in N} \eta^*_{(m,r)\to(n,s)}$:*

$$\hat{\lambda}_{n,s,t} = \frac{\theta g^*_\lambda}{1 + g^*_\lambda} \sum_{r\in S} \eta^*_{r\to(n,s)} \hat{\alpha}_{r,t}. \tag{16}$$

The existence of geographical frictions in idea diffusion implies that the reliance on ideas from any given sector r, $\eta^*_{r\to(n,s)}$, primarily depends on the local stock of knowledge in the same sector, $\lambda_{n,r}$. From Equation (11) it is also immediate to see that this stock of knowledge is tightly related to the local share of population employed in the same sector.[23] Hence, Proposition 2 implies that, other things being equal, the sensitivity of local productivity to shocks to any given sector is increasing in the prevalence of the sector in the local economy.

Second, consider the population shares $\pi_{n,s,t}$. Combining Equation (11) with the definition $\pi_{n,t} \equiv \sum_{s\in S} \pi_{n,s,t}$ and log-linearizing the resulting expression for any arbitrary deviation of $\lambda_{m,s,t}$ from their BGP values yields

$$\hat{\pi}_{n,t} = \frac{\zeta}{\theta} \sum_{s\in S} \left\{ (1 - \pi^*_n)\pi^*_{s|n} \hat{\lambda}_{n,s,t} - \sum_{m\neq n} \pi^*_{m,s} \hat{\lambda}_{m,s,t} \right\}, \tag{17}$$

where $\pi^*_{s|n}$ denotes the probability of being employed in sector s conditional on living in location n. Equation (17) contains an intuitive condition that controls whether a city grows or shrinks relative to the rest of the economy: A location grows if and only if changes of local sectoral productivities, weighted by the incidence of each sector in the city, are larger than the average corresponding changes for the rest of the economy:

$$\hat{\pi}_{n,t} > 0 \iff \sum_{s\in S} \pi^*_{s|n} \hat{\lambda}_{n,s,t} > \sum_{s\in S} \sum_{m\neq n} \frac{\pi^*_{m,s}}{1 - \pi^*_n} \hat{\lambda}_{m,s,t}.$$

By combining Equations (16) and (17), we derive the following proposition, that summarizes the population dynamics implied by the model in response to technological wave shocks:

Proposition 3. *The log deviation of the population share of location n from the BGP, $\hat{\pi}_{n,t}$, can be written as:*

$$\hat{\pi}_{n,t} = \frac{\zeta g^*_\lambda}{1 + g^*_\lambda} \sum_{r\in S} \sum_{s\in S} \left\{ (1 - \pi^*_n)\pi^*_{s|n} \eta^*_{r\to(n,s)} - \sum_{m\neq n} \pi^*_{m,s} \eta^*_{r\to(m,s)} \right\} \hat{\alpha}_{r,t}. \tag{18}$$

To interpret Equation (18) and better illustrate the economic mechanism at play, we first

[23] To see this, note that in the limit case of $\theta = \zeta$, $\alpha^*_r = \alpha^*_s$, and $d_{(n,r)\to(n,s)} = \bar{d}$ for all $r, s \in S$, the reliance of (n, s) on ideas from r, $\eta^*_{r\to(n,s)}$, is exactly equal to local sectoral share, $\pi^*_{r|n}$.

consider a simplified version of the model in which knowledge flows across sectors are of second-order importance relative to flows within sectors. In particular, we impose the following:

Assumption A2. *Frictions to knowledge diffusion across sectors are large enough, so that:*

$$\eta^*_{s\to(n,s)} \approx 1, \qquad \forall s \in S. \tag{19}$$

Under Assumption A2, we can combine Equations (16) and (17) to obtain the following expression for the change in population shares:

$$\hat{\pi}_{n,t} \overset{A2}{=} \frac{\zeta g^*_\lambda}{1+g^*_\lambda} \sum_{s\in S} \left\{ (1-\pi^*_n)\pi^*_{s|n} - \sum_{m\neq n} \pi^*_{m,s} \right\} \hat{\alpha}_{s,t}. \tag{20}$$

Equation (20) rationalizes the reduced-form relationship between exposure to technological waves and population growth documented in Section 2.3. In particular, if the size of any given city is negligible compared to the overall economy,[24] the variation in the right-hand side of Equation (20) is driven entirely by the term $\sum_{s\in S} \pi^*_{s|n}\hat{\alpha}_{s,t}$, which mirrors the shift-share measure of exposure in Equation (1).[25]

Consider now the general case in which knowledge flows across fields are non-negligible (i.e., $\eta^*_{s\to(n,s)} < 1$). In this case, Equation (20) captures only part of the total effect of technological waves on population growth described in Proposition 3. In general, because of geographical frictions to idea diffusion, cities display different degrees of reliance of local innovation on ideas from each of the sectors ($\eta^*_{r\to(n,s)}$). This implies that productivity growth in *all* sectors will be larger (smaller) in cities where expanding (shrinking) fields are more prominent. In other words, in response to technological wave shocks, localized knowledge flows *across* fields amplify fluctuations in productivity growth and, via Equation (18), in population dynamics.

3.4 Taking stock

Propositions 2 and 3 show that frictions to knowledge diffusion across geographical areas and technological fields imply rich and heterogeneous effects of technological waves on the evolution of local productivity and on the distribution of population across cities. In the following section,

[24]Formally, this requires to impose that $(1-\pi^*_n) \approx 1$ and $\sum_{s\in S, m\neq n} \pi^*_{m,s}\hat{\alpha}_{s,t}$ is approximately the same for all $n \in N$.

[25]Equation (20) also implies that, under Assumption A2, n grows (shrinks) if and only if the average local exposure to the technological wave is larger (smaller) than the average exposure for the rest of the economy:

$$\hat{\pi}_{n,t} > 0 \overset{A2}{\iff} \sum_{s\in S} \pi^*_{s|n}\hat{\alpha}_{s,t} > \sum_{s\in S} \pi^*_{s|-n}\hat{\alpha}_{s,t}, \tag{21}$$

where $\pi_{s|-n}$ is the probability of being employed in sector s conditional on living outside of location n.

we show how we bring the model to the data to infer the key parameters and unobserved variables, that we then use to obtain the quantitative results of Section 5.

4 Model calibration

The model has a recursive structure that allows us to calibrate parameters and unobserved variables sequentially by making a limited set of transparent assumptions on how to map the model's objects into data on population, income, and patenting. Throughout the quantitative exploration, we set the model period to 20 years, we let N be the set of 1990 commuting zones that accounted for at least 0.02% of the total population for each decade since 1890, and we define sectors as the 11 class-groups detailed in Appendix Table A.1.

The purpose of the quantitative analysis is to explore the impact of technological waves on city dynamics for a given set of initial conditions. Explaining the initial patterns of local specialization, while interesting, is beyond the scope of this paper. For this reason, we run our quantitative analysis starting from two sets of starting conditions, corresponding to the outset of two of the most striking episodes of technological and geographical transformation in the last century. In particular, we split the 1910-2010 period into two long intervals: an early interval ($\mathcal{E}$), starting in the 20-year period around 1910 and ending in the 20-year period around 1950, characterized by the rise of manufacturing-related fields and the decline of agriculture; and a late interval ($\mathcal{L}$), starting in the 20-year period around 1970 and ending in the 20-year period around 2010, characterized by the decline of manufacturing and the rise of IT and medicine-related fields. We analyze these two intervals separately by resetting the initial conditions in the first period of each interval.

One challenge for the quantitative experiment is that the process of knowledge diffusion described by Equation (7) implies a high degree of inertia as the economy converges towards the BGP.[26] To simplify the exposition and isolate the impact of technological waves from the model's inertia, in our baseline exercise we assume that the economy is in BGP in the first period of each of the two intervals (1910 and 1970) so that, in the absence of shocks, all cities grow at the same rate. In Appendix E we show that results are robust to relaxing this assumption, once we properly normalize the outcomes to control for the model's inertia.

[26] Kleinman et al. (2021) consider a model where capital accumulation and frictional mobility give rise to slow convergence towards the steady state. They find that a significant part of the evolution of the U.S. economic geography since 1965 (including the decline of the Rust Belt and the rise of the Sun Belt) can be accounted for by convergence towards the steady state.

4.1 Overview of the calibration

To calibrate the time-invariant parameters, we focus on the late interval (1970-2010) and use the empirical moments from the 1990 observation, for which we have the most recent and complete data on population, income, and patenting.[27]

We proceed in three steps. In the first step, we infer exogenous amenities $u_{n,\mathcal{L}}$,[28] the path of local productivities $\lambda_{n,s,t}$, and aggregate fertility f_t, and simultaneously pin down the structural parameters ζ and θ by matching moments on the dispersion of income and population across cities. We show that the model accurately reproduces the relationship between city size and income despite not being directly targeted. In the second step, we infer the costs of knowledge transmission $d_{(m,r)\to(n,s)}$ by deriving and estimating a gravity equation for idea flows using patent citations data. In the third step, we recover technological wave shocks $\alpha_{s,t}$ and structural residuals $\epsilon_{n,s,t}$ via the law of motion for local productivities. Given the time-invariant parameters, we then apply an analogous procedure to infer exogenous amenities and the path of time-varying variables in the early interval (1910-1950).

4.2 Amenities and productivity

As a first step, we jointly calibrate the shape parameters of the Frećhet distributions of utility draws, ζ, and the initial distribution of productivity, θ. Here, we also recover the values of local amenities $u_{n,\mathcal{L}}$, and the full path of scale parameters $\lambda_{n,s,t}$ and aggregate fertility f_t.

4.2.1 Productivity distribution

Consider first the scale parameters of the productivity distribution of each location-sector, $\lambda_{n,s,t}$. These objects are at the core of the quantitative analysis: Higher values of $\lambda_{n,s,t}$ imply higher local income, higher ability to attract population, and higher potential to innovate and grow more in the future. In this step of the calibration, we postulate (and later validate) a direct mapping between the discounted stock of patents in a given location-sector and the value of $\lambda_{n,s,t}$. Specifically, we assume that, at any point in time, $\lambda_{n,s,t}$ is equal to a function of current and past patenting:

$$\lambda_{n,s,t} = G_t \times \left[1 + \sum_{\tau=0}^{\tau_{max}} \gamma^{\tau} Pat_{n,s,t-\tau}\right]^{\sigma}, \tag{22}$$

[27]Since we assign patents according to their filing year, patents data and citations in the most recent observation (2010) might suffer from truncation issues. Similarly, data on income and population for the 2010 observation are only available from the ACS, that offers a less complete picture than the 1990 census.

[28]We assume exogenous amenities to be time-invariant within each interval, but to vary between the two intervals. We denote them accordingly as $u_{n,\mathcal{E}}$ for the early interval, and as $u_{n,\mathcal{L}}$ for the late interval.

where $Pat_{n,s,t}$ denotes the total number of patents filed at time t in location-sector (n, s) and G_t is a time-variant factor.[29] We set $\gamma = 0.5$ and $\tau_{max} = 2$, which assumes that the contribution of past patents to variation in local productivity halves every 20 years and vanishes after 60 years. The parameter σ represents the elasticity of $\lambda_{n,s,t}$ with respect to the observed stock of patents. This elasticity converts the variation in the local stock of patents into meaningful variation in the average productivity across location-sectors.

We calibrate σ and θ to jointly match the standard deviation of log-income per capita across cities (in the sample of 373 CZs) and in the overall population in 1990, that are equal to 0.19 and 0.67, respectively.[30] The constant G_t is set to induce an aggregate growth in income per capita of 2% per year.[31]

4.2.2 Amenities, preference draws, and fertility

Consider now local amenities $u_{n,\mathcal{L}}$ and the shape parameter of the distribution of utility draws ζ. Given any guess for ζ, θ, and $\lambda_{n,s,t}$, we calibrate local amenities to exactly match population by city in the first period of the interval (1970).[32] The value of ζ is then calibrated to match the standard deviation of log-population across cities in 1990. The intuition for the identification is that a higher value of ζ implies lower dispersion in the utility draws among newborn agents, so that differences in the desirability of locations, given by amenities and productivity, are more strongly reflected in migration choices.[33] While we assume time-invariant amenities (within each interval) and do not match population by city in each period, the joint calibration of θ, ζ, and σ guarantees that the equilibrium geography reproduces a realistic dispersion of income and population across locations.

We calibrate the path of fertility f_t to match total population by period in the included commuting zones. Notice that, in the absence of moving costs, this is equivalent to assuming that the aggregate increase in population occurs through migration from abroad, fertility, or a combination of the two.

4.2.3 Discussion

Table 2 shows the values of θ, ζ, and σ calibrated through this procedure. The corresponding data moments are matched exactly by construction. In Appendix Figure B.2 we show compu-

[29] We add one to the stock of patents in each sector-city pair to assign a meaningful value to cases in which patenting is zero.

[30] The standard deviation of log-income in the overall population is taken from Krueger and Perri (2006).

[31] We choose units of the final good so that the geometric average of $\lambda_{n,s}^{\frac{1}{\sigma}}$ is equal to one in the first period of the interval.

[32] We normalize amenities to have a geometric mean of one.

[33] This identification of the dispersion of idiosyncratic preference draws follows a similar intuition as Peters (2019).

Table 2: **Parameter values and targets**

Parameter	Value	Target	Model	Data
σ	0.22	S.d. log-income (across CZs), 1990	0.19	0.19
θ	2.00	S.d. log-income (overall), 1990	0.67	0.67
ζ	5.50	S.d. log-population (across CZs), 1990	1.07	1.07

Notes: S.d. of log-income for the overall population is taken from Krueger and Perri (2006). S.d. of log-income and log-population across CZs are author's calculations from the NHGIS.

tationally that there are unique values of the three parameters that jointly match those data moments.

There are two key aspects of this calibration strategy that are worth further discussion. First, the mapping of $\lambda_{n,s,t}$ to the stock of patenting (Equation (22)) includes a size effect in which larger cities have, other things being equal, higher average productivity. The existence of a correlation between size and productivity is a well-known empirical regularity (see e.g. Glaeser and Gottlieb, 2009) that can emerge as a result of a range of theoretical mechanisms (e.g., sorting, variety, local learning productivity spillovers, higher availability of productive inputs, etc...). While the model is silent on the underlying mechanism behind this correlation (besides the fact that more productive cities will *attract* more population) what is crucial for the quantitative performance of the model is that the resulting elasticity of population with respect to income per capita is empirically accurate. Figure 5 shows a bin-scatter plot of the relationship between log-population and log-income in 1990, both in the model and in the data. Although this correlation is not directly targeted in the calibration, the model captures it closely.[34]

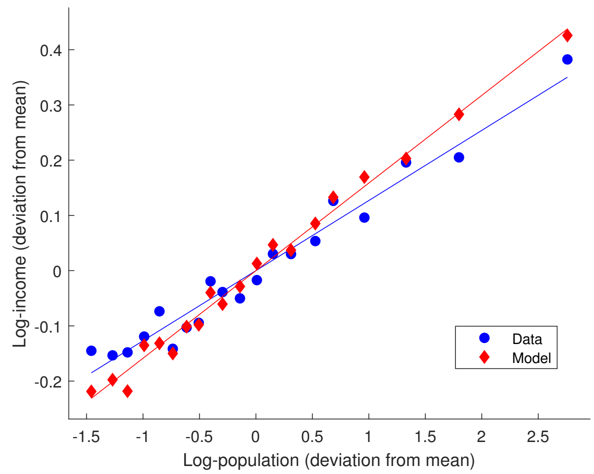
Second, in quantifying the model we assume that residential amenities are time-invariant within each interval. This assumption is crucial for the identification of the shape parameter ζ but comes at the cost of not matching population by city exactly after the first period. As we show in Section 5, even without time-varying amenities, the model goes a long way in fitting population growth by city over the last century. But, as we show in Appendix F, given a value for ζ, allowing for time-varying residential amenities would be an immediate extension of the model.

4.3 Gravity equation for knowledge flows

In the second step of the calibration, we recover the parameters controlling knowledge transmission costs, $d_{(m,r)\to(n,s)}$. To this end, we derive a gravity representation for knowledge flows that

[34]The slope of the regression line is equal to 15.9 in the model and 12.7 in the data.

Figure 5: **Population and Income: Data vs. Model (untargeted)**



Notes: Bin-scatter plot of the relationship between log-population and log-income per capita in the data (blue) and the model (red) in 1990. All variables are displayed as deviations from the mean.

we estimate using data on patent citations. We parametrize frictions to knowledge diffusion as multiplicatively separable between a geographical and a technological component:

$$d_{(m,r)\to(n,s)} = e^{\delta^G \mathbf{1}_{m\neq n} + \delta^K_{r\to s}}, \tag{23}$$

where δ^G controls the effectiveness of knowledge flows *across* locations relative to flows *within* locations, and $\delta^K_{r\to s}$ controls the applicability of ideas from sector r for innovation in sector s.

Combining Equations (8) and (23) and taking logs on both sides yields

$$\log(\eta^t_{(m,r)\to(n,s)}) = \phi^0_{m,r,t} + \phi^1_{n,s,t} - \theta\delta^G \mathbf{1}_{m\neq n} - \theta\delta^K_{r\to s}, \tag{24}$$

where ϕ^0 and ϕ^1 represent idea origin and idea destination-time fixed effects, respectively.

Equation (24) illustrates that bilateral citation probabilities $\eta^t_{(m,r)\to(n,s)}$ depend on the composite parameters $\theta\delta^G$ and $\theta\delta^K_{r\to s}$. To recover those composite parameters, we estimate Equation (24) using data on patent citations across location-sector pairs from the 1990 period (i.e., using patents filed between 1980 and 1999). We compute $\eta^t_{(m,r)\to(n,s)}$ as the share of citations given by patents in (n,s) and directed to patents in (m,r).[35]

Table 3 shows OLS estimates of the composite parameter $\theta\delta^G$. In the baseline specification

[35]Note that the direction of the arrow from (m,r) to (n,s) denotes knowledge flows going from the *cited* patent to the *citing* patent. Every citing patent in our regression has a total weight of one. In other words, every observation is weighted by the inverse of the total number of citations given by (n,s). To account for the fact that local knowledge transmission is more likely to be tacit and less likely to be captured by citations, we further assume that every grant cites one patent from its own location and sector. This also guarantees that all patents, including the ones with no backward citations, are included in the estimation.

Table 3: **Gravity equation for knowledge flows**

	Log share of citations	
	(1)	(2)
Origin CZ $\neq$ Destination CZ	-8.677***	-2.892***
	(.046)	(.022)
Origin location-sector FE	yes	yes
Destination location-sector FE	yes	yes
Origin-Destination sector FE	yes	yes
# Obs.	16,834,609	1,267,834
R^2	0.32	0.69
Zero values	Set to min	No

Notes: OLS estimates. The sample includes patents filed between 1980 and 1999. Observations are all the combinations of pairs of location-sectors. The dependent variable is the logarithm of the share of citations given by each destination location-sector to each origin location-sector, where each citing patent is given a weight of one. We assume that every grant cites one patent from its own location and sector. Standard errors clustered at the destination location-sector level in parenthesis. $***p < 0.01$; $**p < 0.05$; $*p < 0.1$.

of column 1 we replace zero-valued outcomes with the minimum among positive values.[36] Using column 1 as baseline, in combination with the estimate of θ, we obtain a value for $\delta^G = 4.34$. The coefficient implies highly localized knowledge flows, with the effectiveness of transmission across locations, defined as $e^{-\delta^G}$, estimated at around 1.31% of the effectiveness of transmission within locations. Notice that, despite the apparent low effectiveness of transmission, the overall weight of ideas from outside locations may still be large in determining innovation in n, since transmission can happen from *all* the other locations $m \neq n$. Column 2 shows the same regression when only positive values of $\eta^t_{(m,r)\to(n,s)}$ are used in the estimation. The estimate still reveals highly localized knowledge flows, but the coefficient declines in absolute value, suggesting that, as expected, zero values are concentrated among pairs of different locations.

The same regression also delivers a full set of bilateral transmission costs across sectors ($\delta^K_{r\to s}$), that we show in a heatmap in Appendix Figure B.1. As expected, these costs are estimated to be lower within sectors (along the diagonal of the heatmap), although all pairs of sectors display some degree of knowledge exchange that, in some cases, is far from negligible, such as in the cluster of class-groups 10 (G1, "Physics") and 11 (H1, "Electricity").

[36]Notice that, because of the fractional nature of the assignment of patents to class-group, $\eta^t_{(m,r)\to(n,s)}$ cannot be written as a count divided by an exposure variable, so Equation (24) cannot be estimated via Poisson Pseudo-Maximum Likelihood.

4.4 Technological waves and structural residuals

In the third step of the calibration, we use the estimates of θ and $\delta^K_{r\to s}$ and the values of $\lambda_{n,s,t}$ in combination with the law of motion (7) to recover technological wave shocks ($\alpha_{s,t}$) and structural residuals ($\epsilon_{n,s,t}$).

For all periods t, we first guess the full vector of technological wave shocks $\{\alpha_{s,t}\}_{s\in S}$. Given this guess, we use Equation (7) to recover the full set of structural residuals. By construction, this step rationalizes the path of $\lambda_{n,s,t}$ for any initial guess of $\{\alpha_{s,t}\}_{s\in S}$. Hence, to complete the identification, we need to impose S additional conditions. We assume that, in each model period, variation in average growth in productivity across sectors is fully explained by technological waves and their interaction with the endogenous process of knowledge creation and diffusion. In this context, structural residuals explain the remaining variation in productivity growth across locations. Specifically, we impose that adjusted structural residuals, $\epsilon^{\theta}_{n,s,t}$, have a common weighted average for each sector and time period, that we normalize to one:

$$\mathbb{E}\left[\epsilon^{\theta}_{n,s,t}\right] = 1, \qquad \forall s \in S,\, t \geq 0. \tag{25}$$

It is important to emphasize that we do not make any assumption on the nature and properties of the structural residuals, including on whether they are stochastic or deterministic, their spatial and temporal correlation, and whether they are systematically correlated with the other terms in Equation (7).

4.5 Taking stock

Given estimates for the time-invariant parameters, we apply the same procedure to the early interval (1910-1950) to recover the corresponding values of local amenities, sectoral productivities, technological waves, and structural residuals. This step completes the calibration of the model.

In the following section, we use the calibrated model to quantify the role of technological waves, interacted with the endogenous mechanism of knowledge creation and diffusion, in explaining the variation in population growth across U.S. cities throughout the last century, to study how local diversification mediates the impact of technological waves, and to speculate on the future evolution of the economic geography under different plausible scenarios of technological trends.

5 Quantitative results

We now investigate the quantitative implications of the model. We run separate experiments for the early interval (1910-1950), characterized by the rise of manufacturing-intensive cities, and the late interval (1970-2010), that witnesses their decline and the emergence of modern knowledge hubs. We show that, across the two intervals, the model yields comparable results in terms of its ability to reproduce the empirical patterns and its implications for the role of technological waves and frictional knowledge diffusion.

In our baseline experiments, we assume that the economy is in BGP in the first period of each interval, and simulate the model forward under different assumptions on the evolution of the exogenous shocks and the nature of knowledge flows. In Appendix E, we show that the main quantitative results remain consistent when we relax this assumption, once we properly normalize outcomes to account for the model's inertial dynamics of convergence towards the BGP.

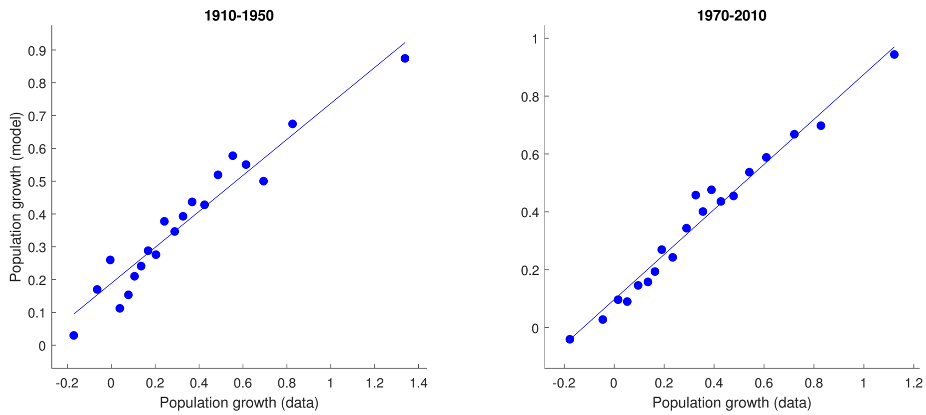
5.1 City growth in the model and the data

To start, we consider the model with the full set of shocks (technological waves and structural residuals). Figure 6 shows bin-scatter plots of population growth across cities in the data (horizontal axis) and the model (vertical axis), for both intervals. The model reproduces the data closely, with the R^2 of the underlying regressions equal to 0.48 and 0.52 for the early and late intervals, respectively. Since the model abstracts from other sources of variation in city growth, such as time-varying amenities and endogenous housing supply, population growth in Figure 6 only reflects the evolution of local sectoral productivity, $\lambda_{n,s,t}$, as inferred by the mapping in Equation (22). Figure 6 illustrates that, even in the absence of time-varying amenities, the full model goes a long way in accounting for the empirical variation in population growth across cities.

We next consider a counterfactual experiment in which, starting from the BGP, we only feed the path of calibrated technological wave shocks, $\alpha_{s,t}$, while keeping the structural residuals constant at their initial BGP values. In this version of the model, the evolution in local productivity is dictated entirely by the interaction between the initial stock of ideas in each city and the path of technological wave shocks, via the law of motion in Equation (7). In other words, this counterfactual isolates the direct effect of technological waves on city growth through the endogenous mechanism of innovation and knowledge diffusion.

Figure 7 shows, for both intervals, bin-scatter plots of population growth in the data (horizontal axis) and the counterfactual experiment (vertical axis). A comparison between the range on the vertical axes in Figures 6 and 7 reveals that the variation in population growth generated

Figure 6: **Population growth: Data vs. Full Model**



Notes: Bin-scatter plots of the 1910-1950 (left panel) and 1970-2010 (right panel) population growth across CZs in the data (horizontal axis) and the model with the full set of shocks (vertical axis).

by the counterfactual is significantly smaller than the one generated by the full model. The standard deviation across cities declines from 0.28 to 0.07 in the early interval, and from 0.34 to 0.07 in the late interval. This is not surprising, since the counterfactual abstracts not only from time-varying amenities, but also from other determinants of local productivity captured by the structural residual. However, population growth in the counterfactual still displays a strong correlation with the data, with the underlying regressions delivering significant (at the 1% level) coefficients, and R^2 equal to 0.06 for both intervals.

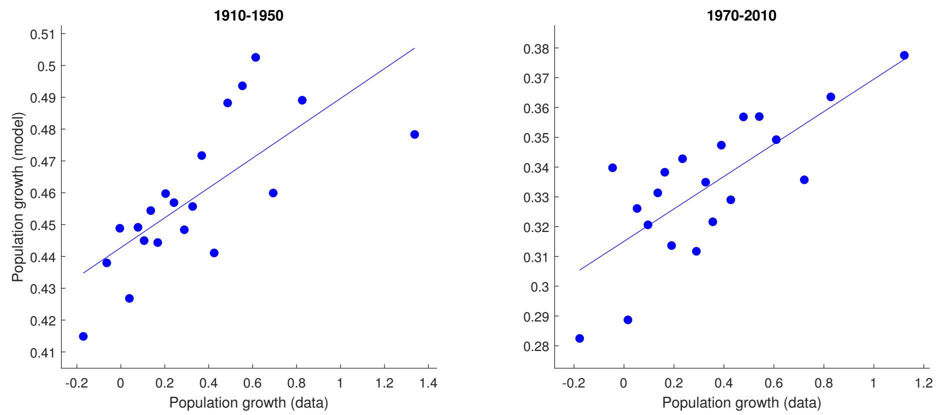
5.2 City growth, technological waves, and knowledge diffusion

Figure 7 shows that the endogenous mechanism of innovation and knowledge diffusion accounts for a small, albeit significant, portion of the overall empirical variation in population growth across cities. We now show that the same mechanism closely replicates the reduced-form empirical relationship between exposure to technological waves and population growth. We then decompose this relationship into a contribution from frictions to idea diffusion across cities and across fields of knowledge.

To define exposure to technological waves in the quantitative model, we adapt the empirical measure in Equation (1) using the calibrated local sectoral shares and aggregate technological wave shocks. In particular, we define measures of exposure separately for the two intervals $\mathcal{I} \in \{\mathcal{E}, \mathcal{L}\}$, as

$$Exp_{n,\mathcal{I}} \equiv \sum_{s \in S} \pi^*_{s|n,\mathcal{I}} \, \hat{\alpha}_{s,\mathcal{I}}, \tag{26}$$

Figure 7: **Population growth: Data vs. Model with technological wave shocks**



Notes: Bin-scatter plots of the 1910-1950 (left panel) and 1970-2010 (right panel) population growth across CZs in the data (horizontal axis) and a counterfactual in which we feed the path of technological wave shocks, $\alpha_{s,t}$, and keep the structural residuals constant at their initial BGP values. (vertical axis).

where $\pi^*_{s|n,\mathcal{I}}$ is the BGP share of sector s in location n at the beginning of interval $\mathcal{I}$, and $\hat{\alpha}_{s,\mathcal{I}}$ is the log-difference of $\alpha_{s,t}$ between the first and the last period of interval $\mathcal{I}$. As robustness, we also report results where exposure is defined using the empirical shift-share measure of Equation (1).[37]

Column 1 of Table 4 reports regressions of actual population growth on the exposure measure in Equation (26), in the early (Panel A) and late (Panel B) intervals. In line with the reduced-form estimates of Section 2.3, the coefficients are statistically significant and economically large. An increase of one standard deviation in the measure of exposure is associated with an increase in population growth equal to 25.1% of a standard deviation in the early interval, and 22.0% of a standard deviation in the late interval. Column 2 shows the corresponding coefficients in the model with the full set of shocks. While population growth in the full model is only imperfectly correlated with actual population growth (as shown in Figure 5) the coefficients of columns 1 and 2 are similar, suggesting that the sources of variation in city growth from which the full model abstracts (e.g., time-varying amenities) do not correlate systematically with cities' exposure to technological waves.

Column 3 of Table 4 reports the corresponding estimates for the counterfactual in which we

[37]This measure is adapted to account for patenting growth throughout each of the two long intervals. In particular, for $\mathcal{I} \in \{\mathcal{E}, \mathcal{L}\}$, we define:

$$Exp_{n,\mathcal{I}} \equiv \sum_{s \in S} Share_{n,s,\mathcal{I}} \times g_{s,\mathcal{I}},$$

where $Share_{n,s,\mathcal{I}}$ is the share of patents filed in commuting zone n belonging to class-group s in the first period of interval $\mathcal{I}$, and $g_{s,\mathcal{I}}$ is the growth rate in the national share of patents of class-group s between the first and the last period of interval $\mathcal{I}$.

Table 4: **Population growth and technological wave shocks**

	Population growth			
	Data	Model		
	(1)	(2)	(3)	(4)
	Panel A: Early interval, 1910-1950			
Exposure to tech. wave	6.25***	6.10***	4.72***	2.27***
	(1.25)	(0.97)	(0.03)	(0.01)
# Obs.	373	373	373	373
R^2	0.06	0.10	0.99	0.99
	Panel B: Late interval, 1970-2010			
Exposure to tech. wave	4.30***	4.84***	4.27***	2.48***
	(0.99)	(1.07)	(0.03)	(0.02)
# Obs.	373	373	373	373
R^2	0.05	0.05	0.98	0.96
Idea flows across fields	-	Yes	Yes	No
Structural residuals	-	Yes	No	No

Notes: OLS estimates. Exposure to the technological wave is defined as in Equation (26). The dependent variable is defined as population growth in the data (column 1), in the full model (column 2), in the model with technological wave shocks and constant structural residuals (column 3), and in the model with technological wave shocks, constant structural residuals, and knowledge flows restricted to within-field flows only (column 4). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

only feed the path of technological wave shocks, and keep structural residuals constant at their BGP values. In the absence of shocks to the structural residuals, Proposition 3 implies that city growth is strongly correlated with exposure to technological waves, driving the R^2 of the regressions close to one. Crucially, in both intervals, the slope of the relationship stays remarkably stable between columns 1 and 3 (in fact, coefficients are statistically indistinguishable), even though this slope is not targeted in the calibration. This implies that the endogenous mechanism of innovation and frictional knowledge diffusion accounts for most of the empirical relationship between exposure to the technological wave and city growth.

Our theory embeds two separate channels behind the impact of a city's exposure to technological waves on local population growth, as summarized by Equation (18). First, frictions to knowledge diffusion *across fields* imply that productivity will increase more in sectors that receive favorable technological wave shocks. As a result, cities in which expanding fields are more prominent will experience higher productivity and population growth. This channel is emphasized by Equation (20), that is derived under the approximation $\eta^*_{s\to(n,s)} \approx 1$ (Assumption A2). Second, frictions to knowledge diffusion *across locations* imply that cities where expanding fields are more prominent will experience higher productivity growth in *all* sectors, because of localized knowledge spillovers between fields.

To decompose the relative importance of these two channels, we re-calibrate technological wave shocks and structural residuals under Assumption A2, that is, by imposing that knowledge flows only happen within fields. We then run, for both intervals, counterfactual experiments in which we predict city growth by feeding the path of (re-calibrated) technological wave shocks, while keeping structural residuals constant at their BGP values. Column 4 of Table 4 reports estimates of the resulting relationship between exposure to the technological wave and population growth. The magnitude of the coefficient declines by 52% and 42% for the early and late intervals, respectively. This suggests that around half of the overall impact of technological waves on city growth is driven by the existence of localized knowledge flows across fields that amplify the direct impact of shocks to sectoral productivity.

Appendix Table A.4 shows that results are fully consistent when the empirical shift-share measure of exposure is used as explanatory variable. Estimated coefficients in columns 1 and 3 are statistically indistinguishable, and the coefficient declines by 50% and 42% (in the early and late intervals, respectively) in the model without idea flows across fields.

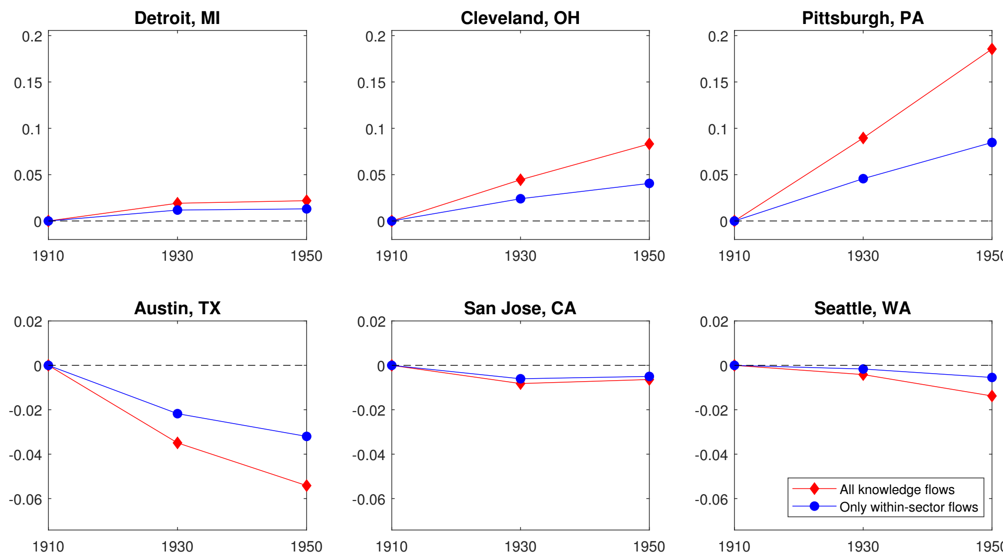
5.3 The rise of manufacturing-intensive cities

We now show that the interaction between technological waves and frictional knowledge diffusion can account for two of the most striking episodes of transformation of the economic geography of the U.S. in the last century: the extraordinary rise of manufacturing-intensive cities in the early 20th century, followed by their decline and simultaneous rise of knowledge-intensive urban areas.

Figure 8 shows, for a selected subset of cities in the U.S., population growth in the early interval (1910-1950) in the counterfactual experiments with knowledge flows across fields (red line) and within-field flows only (blue line). We normalize data so that the horizontal axis corresponds to a scenario in which all cities grow at the same rate, as dictated by aggregate fertility.[38] The plots in the top row show that Detroit, Cleveland, and Pittsburgh, three of the largest urban areas of what would later be known as the Rust Belt, were favorably exposed to the technological wave in the early part of the century, and, by 1950, they received a net inflow of population. Among them, Detroit records the weakest impact on population growth (+2.2%),

[38] This corresponds to a scenario in which there are no technological wave shocks, and the economy remains on the BGP.

Figure 8: **Growth Decomposition: Early interval, 1910-1950**



Notes: The red and blue lines show log-population in deviation from a trajectory of constant population growth across cities, as dictated by aggregate fertility. The red line corresponds to a counterfactual experiment in which we feed the path of technological wave shocks and keep structural residuals constant at their initial BGP values. The blue line shows an analogous counterfactual in which we restrict knowledge flows to within-field flows only.

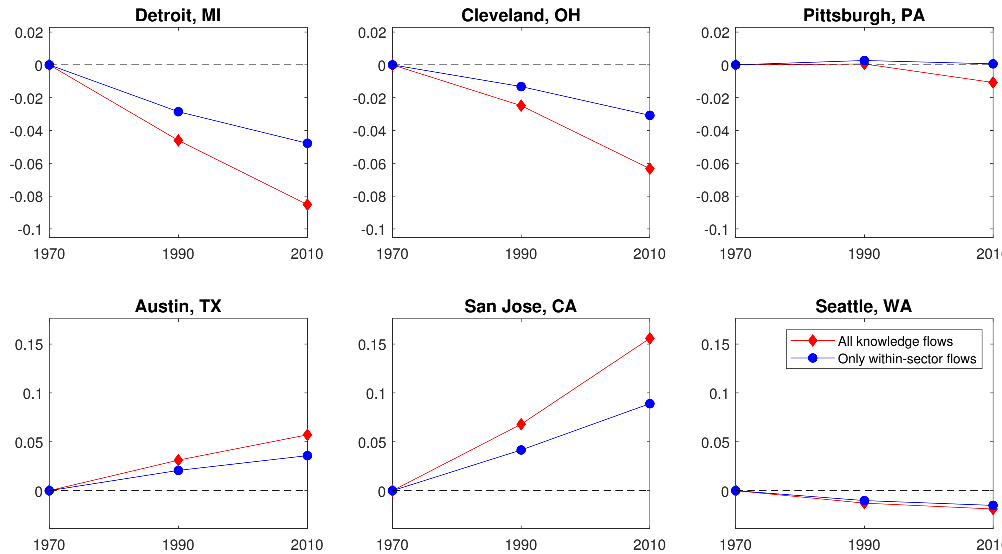
compared to Cleveland (+8.3%) and Pittsburgh (+18.6%). This small, albeit positive, impact, might be explained by the fact that major events that raised Detroit to its position of major center of car manufacturing only materialized after the first period (1910) and are captured by the structural residual (which is kept constant in this experiment). For all the three cities, knowledge flows across fields (red line) significantly amplified the impact of the shocks (blue line).

During the same period, cities did not uniformly benefit from this transformation in the technological landscape. As described in Section 2.2, the commuting zone of Austin declined over this period. The blue and red lines show that part of this decline was due to an unfavorable exposure to the technological wave, as Austin, highly specialized in agriculture-related fields, lost population (-5.4%). Both San Jose and Seattle were negatively but weakly affected.

5.4 The emergence of modern knowledge hubs

The experiences of Detroit, Cleveland, and Pittsburgh in the first half of the 20th century were not isolated cases. Several other cities that were specialized in heavy manufacturing and were mostly concentrated in what is now known as the Rust Belt witnessed exceptional growth in population. The model suggests that part of this growth can be explained by the availability of local ideas in fields that were complementary to the prevailing technological wave. We now

Figure 9: **Growth Decomposition: Late interval: 1970-2010**



Notes: The red and blue lines show log-population in deviation from a trajectory of constant population growth across cities, as dictated by aggregate fertility. The red line corresponds to a counterfactual experiment in which we feed the path of technological wave shocks and keep structural residuals constant at their initial BGP values. The blue line shows an analogous counterfactual in which we restrict knowledge flows to within-field flows only.

explore whether the same factors that led to the remarkable growth of manufacturing-intensive cities contributed to their later decline to the benefit of emerging knowledge hubs specialized in information technologies.

The top panels of Figure 9 track the response of the three manufacturing-intensive cities of Figure 8 in the late interval (1970-2010). Both Detroit and Cleveland receive a negative impact on city population (red line) of -8.5% and -6.3%, respectively. However, Pittsburgh experiences a smaller decline (-1.1%) which can be explained by the fact that, in 1970, the city had already planted the initial seed for what would later transform Pittsburgh into a fast-growing center for robotics and artificial intelligence.[39]

Throughout the same decades, a handful of cities emerged as leading modern technological hubs. The commuting zones of Austin, TX and San Jose, CA are archetypal examples of this phenomenon. Population in Austin and San Jose increases, relative to the baseline, by 5.7% and 15.6%, respectively, as an effect of the technological wave. Also in this case, localized knowledge flows across fields amplify the response significantly. By contrast, the commuting zone of Seattle appears to be weakly but negatively affected by the technological wave. This is not surprising since most of the recent growth in the IT sector in Seattle is a consequence of

[39]In 1970, the calibrated model delivers combined sectoral shares for fields G1 (Physics) and H1 (Electricity) for Pittsburgh equal to 24.1%. The corresponding shares for Detroit and Cleveland were 17.4% and 18.7%.

local events that happened *after* 1970 (such as the relocation of Microsoft to Bellevue in 1979 and the establishment of Amazon in 1994).[40]

5.5 Diversification and resilience to technological waves

The process of innovation through frictional idea diffusion implies a role for local diversification in making cities resilient to changes in the technological environment. In particular, the same two channels that control local population dynamics in response to technological wave shocks, reflecting respectively frictions to knowledge diffusion across fields and across locations, make the growth trajectory of diversified cities less volatile than the one of specialized cities. First, frictions to knowledge diffusion *across fields* imply that the path of productivity of any given sector is mainly driven by technological wave shocks to the sector itself. As a consequence, diversified cities, whose sectoral composition is dispersed across multiple sectors, experience a less volatile path of average productivity. Second, frictions to knowledge diffusion *across locations* imply that the reliance of each location-sector on ideas from any given field is an increasing function of the *local* availability of ideas from that field. For this reason, innovators in diversified cities rely on ideas from a broader set of fields, and the local path of productivity is less sensitive to shocks to individual sectors. Overall, the less volatile path for average productivity in more diversified cities also implies a less volatile trajectory in population growth.

Exploring this link formally requires to define the correct measure of local specialization. Proposition C.1 in the Appendix shows that, under Assumption A2 and intuitive conditions on the distribution of shocks, the variance of local population growth is approximately equal to the Euclidean distance between the local and nationwide vectors of sectoral shares. For both intervals $\mathcal{I} \in \{\mathcal{E}, \mathcal{L}\}$, we use this distance as measure of local specialization:

$$Spec_{n,\mathcal{I}} \equiv \sum_{s \in S} \left(\pi^*_{s|n,\mathcal{I}} - \pi^*_{\cdot,s,,\mathcal{I}} \right)^2, \tag{27}$$

where $\pi^*_{\cdot,s,\mathcal{I}}$ is the share of the national population employed in sector s. According to this measure, cities are perfectly diversified if their local sectoral shares are exactly equal to the national ones.

To quantify the effect of diversification on the volatility of city growth, we perform simulations in which we randomly draw shocks $\{\hat{\alpha}_{r,t}\}_{r \in S}$ and compute the corresponding counterfactual equilibria for the economy. We then correlate the standard deviation of population growth across all the simulations with the measure of local specialization in Equation (27).

Columns 1 and 3 of Table 5 report results for the early and late intervals, respectively. The simulations are obtained by keeping structural residuals at their BGP values, and randomly

[40] In fact, we do find a positive direct effect of the technological wave if we consider the model with the full set of shocks until 1990, and only provide the technological wave shocks in 2010.

Table 5: **Specialization and volatility of population growth**

	Standard deviation across simulations			
	1910-1950		1970-2010	
	(1)	(2)	(3)	(4)
$Spec_{n,\mathcal{I}}$	3.43***	1.90***	1.88***	1.59***
	(0.06)	(0.03)	(0.03)	(0.03)
# Obs.	373	373	373	373
R^2	0.91	0.92	0.92	0.91
Idea flows across fields	Yes	No	Yes	No
Structural residuals	No	No	No	No
$Spec_{n,\mathcal{I}}$ 90-10 perc.	0.008	0.008	0.019	0.019

Notes: OLS estimates. Specialization is defined as in Equation (27). The dependent variable is defined as the city-level standard deviation of population growth across 10,000 simulations. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

drawing technological wave shocks from a normal distribution with mean zero and standard deviation equal to the one of the calibrated shocks.[41] We run 10,000 simulations for each interval. The results reveal that, as predicted by the theory, specialized cities have higher volatility of population growth across simulations. The magnitude of the coefficients implies that the standard deviation for cities at the 90th percentile of the specialization distribution is 2.7% higher than for cities at the 10th percentile of the distribution, in the early interval, and 3.6% higher in the late interval.

To disentangle the importance of the first channel (frictions across fields) and the second channel (frictions across locations) in explaining this relationship, we run the same experiment again in the version of the model in which we restrict knowledge diffusion to within-field flows only. Results for the two intervals are reported in columns 2 and 4 of Table 5. The coefficient on the measure of specialization drops by about 45% in the early interval and 15% in the late interval. This implies that the direct effect of technological wave shocks, interacting with the local sectoral shares, explains most the impact of specialization on local volatility. However, a smaller but still significant portion of the impact is accounted for by the channel of localized knowledge flows across fields, that attenuate the fluctuations in productivity growth in more diversified cities.

[41] This standard deviation is equal to 0.126 for the early interval, and 0.116 for the late interval.

5.6 The U.S. economic geography under future technological waves

The quantitative model can be used to predict the evolution of the U.S. economic geography in the coming decades in response to transformations in the technological environment. In this Section, we propose plausible scenarios for future technological waves and look at which commuting zones will be most positively and negatively affected by those changes. In particular, we project population growth across cities until 2050 under different assumptions on the evolution of the importance of different sectors ($\alpha_{s,t}$), and compare the outcome with a baseline in which the importance for all sectors is kept constant at their 2010 values.

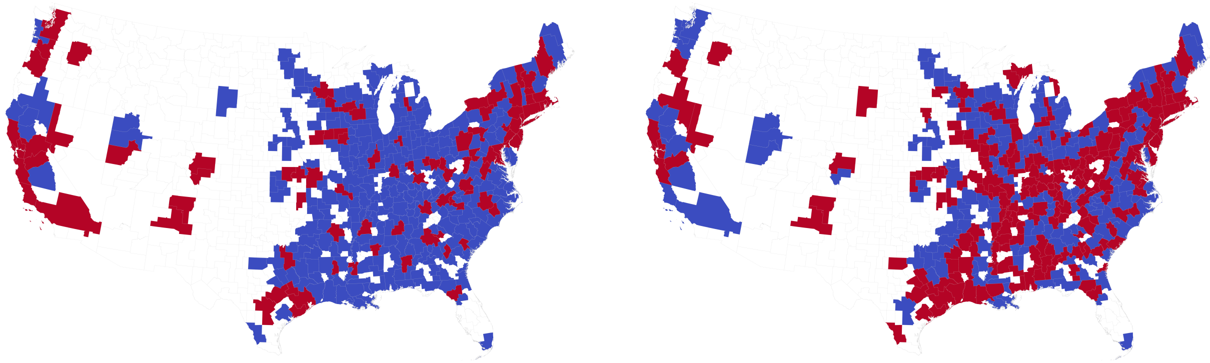
In the first scenario, we assume that sector B2 ("Transporting") experiences a technological wave shock of magnitude equal to twice the standard deviation of technological wave shocks throughout the late interval (+23.2%). This scenario could emerge as new advances in transit technologies and autonomous vehicles induce innovation in transportation to return to a pivotal role. The left map in Figure 10 visually illustrates the results. Commuting zones in blue (red) experience a net gain (loss) of population compared to the baseline. Cities in the Rust Belt are the areas that are best positioned to take advantage from this transformation. Detroit would experience a 10.0% increase in population compared to the baseline. Other centers of manufacturing related to (but not specialized in) transportation, would benefit, too, albeit to a lesser extent. For example, population in Cleveland and Gary would increase by 0.7% and 5.2%, respectively. The knowledge hubs of Austin (-4.8%), San Jose (-5.7%), and Seattle (-2.8%), would all experience a relative loss of population.

An alternative way of modeling transportation-related technologies regaining prominence is to assume that ideas from B2 become more relevant for innovation in either G1 ("Physics") or H1 ("Electricity"), and vice versa. An example of the increasing inter-dependence of those sectors is the gradual integration of IT components in electric and autonomous vehicles. We model this strengthening connection as a drop in the cost of knowledge transmission ($\delta_{s\to r}$) by assuming a 20% decline in composite knowledge frictions ($d^{\theta}_{(m,r)\to(n,s)}$) from (to) B2 to (from) both G1 and H1.[42] In this case, we keep sectoral importance ($\alpha_{s,t}$) at its 2010 value. The right map in Figure 10 displays the results. Also in this case, Detroit gains population (+4.2%), while the knowledge hubs of Austin (-0.6%) and San Jose (-1.1%) experience a net loss of population. The reason is that, while the economy of Detroit has recently diversified, to some extent, towards fields G1 and H1, Austin and San Jose have increasingly specialized, preventing them from leveraging cross-field spillovers. On the other hand, Seattle (+1.4%) experiences a relative gain in population, leveraging its more diversified base in IT and transportation.

In the second scenario, we simulate a large positive technological wave shock to sector A3

[42] While the assumption of a proportional 20% decline is arbitrary, this choice only affects the magnitude of the results but does not alter the qualitative patterns.

Figure 10: **Future scenarios: Autonomous vehicles**



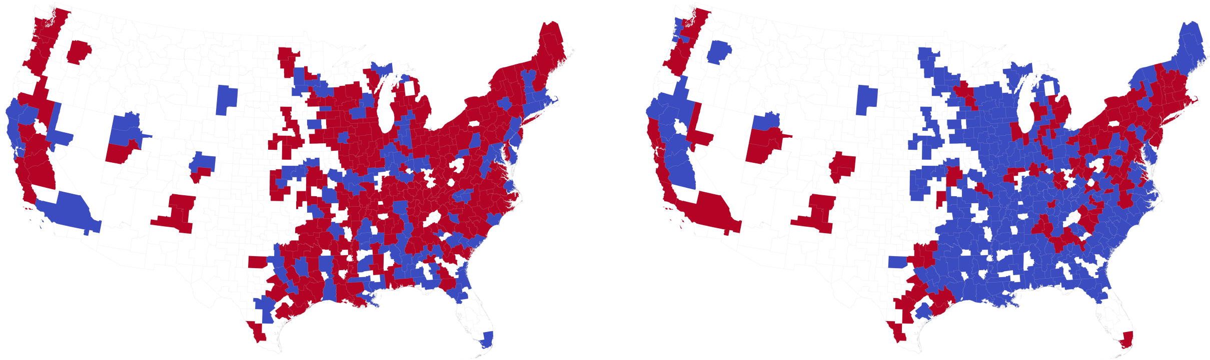
Notes: The maps show log-population in 2050 after technological wave shocks of magnitude +23.2% to B2 (left map), and a 20% decline in composite knowledge frictions ($d^{\theta}_{(m,r)\rightarrow(n,s)}$) from (to) B2 to (from) both G1 and H1 (right map), in deviation from a status quo in which $\alpha_{s,t}$ are kept at their 2010 values. Blue CZs correspond to a net population gain, red CZs correspond to a net population loss.

("Health; Life-Saving; Amusement", that includes the bulk of innovation related to pharmaceuticals and medical sciences) possibly in response to new challenges in global health such as the COVID-19 pandemic.[43] The results are depicted in the left map of Figure 11. The experiment suggests that major commuting zones in the North-East, such as Boston (+5.4%) and Providence (+15.0%), and in California, such as Los Angeles (+4.2%) and San Francisco-Oakland (+2.2%) would experience a net inflow of population, at the expense of IT clusters such of Austin (-7.6%), San Jose (-3.5%) and Seattle (-3.9%).

In the third scenario, we assume that sector A1 ("Agriculture"), regains centrality by experiencing an analogous large technological wave shock. This is a plausible scenario that can emerge as a result of tightening regulatory constraints and shifting demand towards sustainable farming, possibly in response to global challenges such as climate change. Results are in the right map of Figure 11. Under this scenario, the economic geography of the U.S. experiences a pronounced shift away from the East and West coast and the Rust Belt, towards the Central States. Among the major commuting zones, Des Moines (IA) receives the highest net gain (+18.3%). This scenario would represent a significant convergence force in relative population across commuting zones. A regression of log-population in 2010 on the predicted growth rate between 2010 and 2050 delivers a coefficient of -1.1%, implying that population would mostly relocate away from larger commuting zones and towards less-populated ones.

[43] Also in the second and third scenario, we input a shock of magnitude 23.2%, equal to twice the standard deviation of technological wave shocks throughout the late interval.

Figure 11: **Future scenarios: Pharmaceuticals and Agriculture**



Notes: The maps show log-population in 2050 after technological wave shocks of magnitude +23.2% to A3 (left map), and to A1 (right map), in deviation from a status quo in which $\alpha_{s,t}$ are kept at their 2010 values. Blue CZs correspond to a net population gain, red CZs correspond to a net population loss.

6 Conclusions

The economic geography of countries is characterized by rich and heterogeneous dynamics, alternating path dependence to occasional reversal of fortunes. Some cities remain large and important throughout long time spans, while others experience episodes of sharp growth and decline. In this paper, we explore and quantify the hypothesis that these rich dynamics result in part from cities' patterns of specialization across sectors, coupled with the continuous evolution of the technological landscape. We develop a parsimonious framework that combines elements from quantitative spatial equilibrium models and theories of endogenous growth through innovation and idea diffusion. The model remains tractable for any arbitrary number of sectors, locations, and time periods – making it suitable for quantitative analysis – and delivers a wide range of predictions on how the economic geography of countries responds to changes in the technological environment. The quantitative results support the idea that the interaction of frictional knowledge diffusion with technological waves played a major role in shaping the evolution of the U.S. economic geography in the last century, explaining the rise of manufacturing-intensive cities in the Rust Belt and the recent emergence of modern knowledge hubs. We use the model to speculate on future transformations of the U.S. economic geography under different technological scenarios, such as a come back of transportation and agriculture and a further rise in the centrality of medical sciences.

While we consistently find a large impact of technological waves through frictional knowledge diffusion, there are some moderating channels from which the model abstracts but could be embedded without a prohibitive loss of tractability. Extensions can include trade and migration frictions across locations, production activities that are geographically separated from innovation, the inclusion of local non-tradable goods, and partially endogenous residential amenities and structural residuals. In particular, the quantitative results suggest that residual factors contribute significantly to the dynamics of local innovation and to the variation in city growth. The framework allows us to isolate the direct effect of the technological wave via innovation and knowledge diffusion, and does not require to make specific assumptions on the nature of this residual. A possible way to endogenize this error term is to allow innovators to exert effort to improve their ideas, in the spirit of an endogenous growth theory with expanding varieties (as in Jones, 2005). An alternative route to unpack the residual term is to account for the granularity of the locational choices of individual firms. Events such as Microsoft's relocation to the Seattle area, or Amazon's selection of a site for its second headquarters, can have a major impact in shaping the destiny of cities. Other endogenous forces that enter the residual include congestion, pecuniary externalities on local assets, and the response of policy to local shocks. Understanding how these factors contribute to amplifying or dampening the effects of technological waves is the next step in our agenda.

References

Adao, R., C. Arkolakis, and F. Esposito (2020): "General Equilibrium Effects in Space: Theory and Measurement," Tech. rep., Department of Economics, Tufts University.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Akcigit, U. and W. R. Kerr (2018): "Growth through heterogeneous innovations," *Journal of Political Economy*, 126, 1374–1443.

Allen, T. and C. Arkolakis (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 129, 1085–1140.

Allen, T. and D. Donaldson (2018): "Geography and path dependence," Tech. rep.

Andrews, M. (2019): "Historical Patent Data: A Primer and Some Pitfalls," mimeo, available at: https://sites.google.com/site/michaeljeffreyandrews/research.

Audretsch, D. B. and M. P. Feldman (1996): "R&D spillovers and the geography of innovation and production," *The American economic review*, 86, 630–640.

Babina, T., A. Bernstein, and F. Mezzanotti (2020): "Crisis Innovation," *Available at SSRN*.

Balland, P.-A., D. Rigby, and R. Boschma (2015): "The technological resilience of US cities," *Cambridge Journal of Regions, Economy and Society*, 8, 167–184.

Bartik, T. J. (1991): "Who benefits from state and local economic development policies?" .

Berkes, E. (2018): "Comprehensive Universe of U.S. Patents (CUSP): Data and Facts," mimeo, available at: https://sites.google.com/view/enricoberkes/work-in-progress.

Berkes, E. and P. Nencka (2020): "Knowledge Access: The Effects of Carnegie Libraries on Innovation," Available at SSRN: https://ssrn.com/abstract=3629299.

Bleakley, H. and J. Lin (2012): "Portage and path dependence," *The quarterly journal of economics*, 127, 587–644.

Bostic, R. W., J. S. Gans, and S. Stern (1997): "Urban productivity and factor growth in the late nineteenth century," *Journal of urban economics*, 41, 38–55.

Brezis, E. S. and P. R. Krugman (1997): "Technology and the life cycle of cities," *Journal of Economic Growth*, 2, 369–383.

Brezis, E. S., P. R. Krugman, and D. Tsiddon (1993): "Leapfrogging in international competition: A theory of cycles in national technological leadership," *The American economic review*, 1211–1219.

Buera, F. J. and R. E. Lucas (2018): "Idea flows and economic growth," *Annual Review of Economics*, 10, 315–345.

Buera, F. J. and E. Oberfield (2020): "The global diffusion of ideas," *Econometrica*, 88, 83–114.

Cai, J. and N. Li (2019): "Growth through inter-sectoral knowledge linkages," *The Review of Economic Studies*, 86, 1827–1866.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2018): "The impact of regional and sectoral productivity changes on the US economy," *The Review of economic studies*, 85, 2042–2096.

Clemens, J. and P. Rogers (2020): "Demand Shocks, Procurement Policies, and the Nature of Medical Innovation: Evidence from Wartime Prosthetic Device Patents," Tech. rep., National Bureau of Economic Research.

Davis, D. R. and J. I. Dingel (2019): "A spatial knowledge economy," *American Economic Review*, 109, 153–70.

Davis, D. R. and D. E. Weinstein (2002): "Bones, bombs, and break points: the geography of economic activity," *American Economic Review*, 92, 1269–1289.

De la Croix, D., M. Doepke, and J. Mokyr (2018): "Clans, guilds, and markets: Apprenticeship institutions and growth in the preindustrial economy," *The Quarterly Journal of Economics*, 133, 1–70.

Desmet, K. (2002): "A simple dynamic model of uneven development and overtaking," *The Economic Journal*, 112, 894–918.

Desmet, K., R. E. Kopp, S. A. Kulp, D. K. Nagy, M. Oppenheimer, E. Rossi-Hansberg, and B. H. Strauss (2018a): "Evaluating the economic cost of coastal flooding," Tech. rep., National Bureau of Economic Research.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018b): "The geography of development," *Journal of Political Economy*, 126, 903–983.

Desmet, K. and J. Rappaport (2017): "The settlement of the United States, 1800–2000: the long transition towards Gibrat's law," *Journal of Urban Economics*, 98, 50–68.

Desmet, K. and E. Rossi-Hansberg (2014): "Spatial development," *American Economic Review*, 104, 1211–43.

Duranton, G. (2007): "Urban evolutions: The fast, the slow, and the still," *American Economic Review*, 97, 197–221.

Duranton, G. and D. Puga (2001): "Nursery cities: Urban diversity, process innovation, and the life cycle of products," *American Economic Review*, 91, 1454–1477.

Eckert, F. and M. Peters (2019): "Spatial Structural Change," mimeo, available at: https://mipeters.weebly.com/research.html.

Glaeser, E. L. (2005): "Reinventing Boston: 1630–2003," *Journal of Economic Geography*, 5, 119–153.

Glaeser, E. L. and J. D. Gottlieb (2009): "The wealth of cities: Agglomeration economies and spatial equilibrium in the United States," *Journal of economic literature*, 47, 983–1028.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992): "Growth in cities," *Journal of political economy*, 100, 1126–1152.

Greenstone, M., R. Hornbeck, and E. Moretti (2010): "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings," *Journal of Political Economy*, 118, 536–598.

Heblich, S., S. J. Redding, and D. M. Sturm (2020): "The making of the modern metropolis: evidence from London," *The Quarterly Journal of Economics*, 135, 2059–2133.

Holmes, T. J. and J. J. Stevens (2004): "Spatial distribution of economic activities in North America," in *Handbook of regional and urban economics*, Elsevier, vol. 4, 2797–2843.

Hornbeck, R. and E. Moretti (2018): "Who benefits from productivity growth? Direct and indirect effects of local TFP growth on wages, rents, and inequality," Tech. rep., National Bureau of Economic Research.

Huang, J. and Y. Zenou (2020): "Key Sectors in Endogeneous Growth," .

Jacobs, J. (1969): *The economy of cities*, Vintage international, Random House.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993): "Geographic localization of knowledge spillovers as evidenced by patent citations," *the Quarterly journal of Economics*, 108, 577–598.

Jones, C. I. (2005): "Growth and ideas," in *Handbook of economic growth*, Elsevier, vol. 1, 1063–1111.

Kerr, W. R. and S. D. Kominers (2015): "Agglomerative forces and cluster shapes," *Review of Economics and Statistics*, 97, 877–899.

Kleinman, B., E. Liu, and S. J. Redding (2021): "Suficient Statistics for Dynamic Spatial Economics," .

Kline, P. and E. Moretti (2014): "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority," *The Quarterly journal of economics*, 129, 275–331.

Krueger, D. and F. Perri (2006): "Does income inequality lead to consumption inequality? Evidence and theory," *The Review of Economic Studies*, 73, 163–193.

Lind, N. and N. Ramondo (2019): "The Economics of Innovation, Knowledge Diffusion, and Globalization," in *Oxford Research Encyclopedia of Economics and Finance*.

Lucas, R. E. and B. Moll (2014): "Knowledge growth and the allocation of time," *Journal of Political Economy*, 122, 1–51.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2021): "IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS," http://doi.org/10.18128/D050.V16.0.

Marshall, A. (1890): *The Principles of Economics*, McMaster University Archive for the History of Economic Thought.

Michaels, G., F. Rauch, and S. J. Redding (2012): "Urbanization and structural transformation," *The Quarterly Journal of Economics*, 127, 535–586.

Moretti, E. (2012): *The new geography of jobs*, Houghton Mifflin Harcourt.

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014): "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach," *Review of Economics and Statistics*, 96, 967–985.

Nagy, D. K. (2017): "City Location and Economic Development," mimeo, available at: https://sites.google.com/site/davidknagy/research.

Perla, J. and C. Tonetti (2014): "Equilibrium imitation and growth," *Journal of Political Economy*, 122, 52–76.

Peters, M. (2019): "Market Size and Spatial Growth–Evidence from Germanys Post-War Population Expulsions," *Unpublished manuscript*.

Redding, S. J. and E. Rossi-Hansberg (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

Ruggles, S., S. Flood, S. Foster, R. Goeken, J. Pacas, M. Schouweiler, and M. Sobek (2021): "IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021," https://doi.org/10.18128/D010.V11.0.

Simon, C. J. and C. Nardinelli (2002): "Human capital and the rise of American cities, 1900–1990," *Regional Science and Urban Economics*, 32, 59–96.

A Additional tables

Table A.1: **IPC Class-Groups**

Class ID	Class Group	IPC Class Range	Label
1	A1	A01-A24	Agriculture - Foodstuffs; Tobacco
2	A2	A41-A47	Personal or Domestic Articles
3	A3	A61-A99	Health; Life-Saving; Amusement
4	B1	B01-B44	Separating; Mixing - Shaping - Printing
5	B2	B60-B68	Transporting
-	B3	B81-B99	Microstructural Technology; Nanotechnology
6	C1	C01–C30	Chemistry - Metallurgy
-	C2	C40-C99	Combinatorial Technology
-	D1	D01-D07	Textiles - Paper
7	E1	E01-E99	Building - Earth or Rock Drilling; Mining
8	F1	F01-F17	Engines or Pumps - Engineering in General
9	F2	F21-F99	Lighting; Heating - Weapons; Blasting
10	G1	G01-G16	Physics
-	G2	G21-G99	Nuclear Physics; Nuclear Engineering
11	H1	H01-H99	Electricity

Notes: This table provides label and a mapping to the original IPC classes for the class-groups used for the empirical and quantitative analysis of this paper. Groups B3, C2, D1, and G2 are excluded from the sample since they are either negligible in size or they cover innovation in fields, such as nuclear physics, was acquired only in the later portion of the sample.

Table A.2: **Summary Statistics**

Variable	Obs.	Mean	Std. Dev.	Min	Max
Population	2,984	353,213	897,673	225	1.79e+07
Log-population	2,984	11.93	1.16	5.42	16.70
Population growth	2,238	.238	.307	-.543	3.24
Exposure to tech. wave	2,238	-.102	.087	-.465	.409
Industrial composition	2,228	-.150	.155	-.692	.182

Notes: Summary statistics refer to the period 1870-2010 (Population and Log-population), or the period 1910-2010 (remaining variables).

Table A.3: **Technological waves and city growth: Early vs. late samples**

	Growth rate of population			
	(1)	(2)	(3)	(4)
	Panel A: Early sample, 1910-1950			
Exposure to tech. wave	0.680***	0.555***	0.575***	0.253*
	(0.194)	(0.160)	(0.164)	(0.141)
Log-population (lag 1)	0.167**	0.183***	0.184***	0.086
	(0.064)	(0.060)	(0.059)	(0.068)
Log-population (lag 2)	-0.235***	-0.223***	-0.221***	-0.179***
	(0.047)	(0.038)	(0.038)	(0.042)
Human capital (ranking)			-0.147	-0.297***
			(0.101)	(0.104)
Industry composition				0.929***
				(0.211)
# Obs.	1,119	1,119	1,119	1,109
R^2	0.39	0.49	0.49	0.53
	Panel B: Late sample, 1970-2010			
Exposure to tech. wave	0.219***	0.246***	0.199***	0.183***
	(0.078)	(0.054)	(0.056)	(0.059)
Log-population (lag 1)	0.599***	0.550***	0.503***	0.456***
	(0.058)	(0.066)	(0.069)	(0.058)
Log-population (lag 2)	-0.612***	-0.554***	-0.518***	-0.484***
	(0.052)	(0.062)	(0.064)	(0.055)
Human capital (ranking)			0.121***	0.075**
			(0.035)	(0.031)
Industry composition				0.575***
				(0.066)
# Obs.	1,119	1,119	1,119	1,119
R^2	0.41	0.52	0.54	0.58
Fixed effects	T	CD×T	CD×T	CD×T

Notes: CZ level regression. Dependent variable defined as growth rate of population over 20 years. "T" denotes time fixed effects, and "CD×T" denotes Census Division-time fixed effects. Standard errors clustered at the CZ level in parenthesis. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

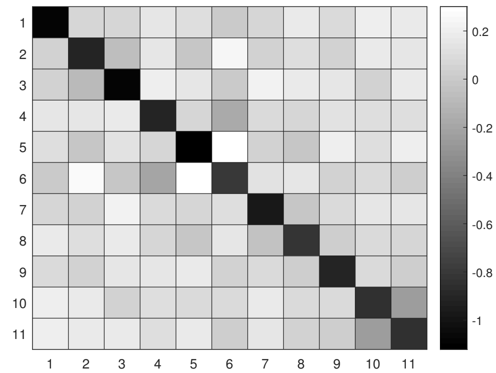
Table A.4: **Population growth and technological wave shocks: Empirical measure of exposure**

	Population growth			
	Data	Model		
	(1)	(2)	(3)	(4)
	Panel A: Early interval, 1910-1950			
Exposure to tech. wave (empirical)	.897***	.946***	.631***	.316***
	(.201)	(.156)	(.022)	(.009)
# Obs.	373	373	373	373
R^2	0.05	0.09	0.69	0.75
	Panel B: Late interval, 1970-2010			
Exposure to tech. wave (empirical)	.400***	.367***	.385***	.222***
	(.098)	(.107)	(.009)	(.006)
# Obs.	373	373	373	373
R^2	0.04	0.03	0.83	0.80
Idea flows across fields	-	Yes	Yes	No
Structural residuals	-	Yes	No	No

Notes: OLS estimates. Exposure to the technological wave is defined as $Exp_{n,\mathcal{I}} \equiv \sum_{s \in S} Share_{n,s,\mathcal{I}} \times g_{s,\mathcal{I}}$, for $\mathcal{I} \in \{\mathcal{E}, \mathcal{L}\}$. The dependent variable is defined as population growth in the data (column 1), in the full model (column 2), in the model with technological wave shocks and constant structural residuals (column 3), and in the model with technological wave shocks, constant structural residuals, and knowledge flows restricted to within-field flows only (column 4). ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

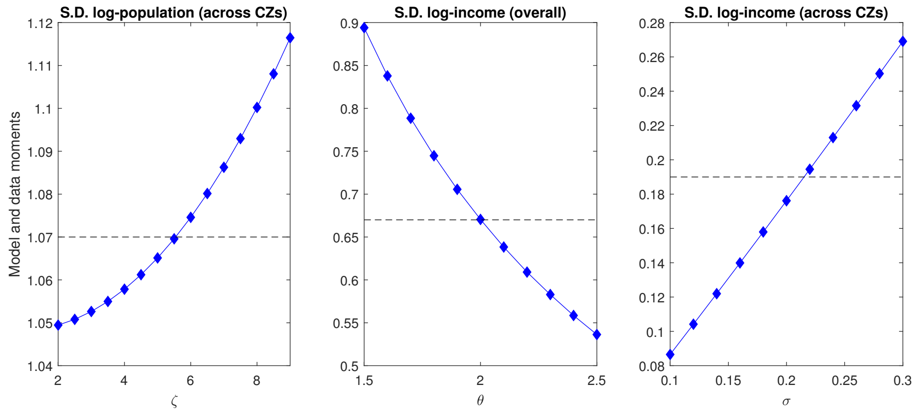
B Additional figures

Figure B.1: **Knowledge transmission costs across sectors**



Notes: Estimated (OLS) coefficients $\delta^K_{r\to s}$, from regression of Table 3, Column 1. The sample includes patents filed between 1980 and 1999. Observations are all the combinations of pairs of location-sectors. The dependent variable is the logarithm of the share of citations given by each destination location-sector to each origin location-sector, where each citing patent is given a weight of one. We assume that every grant cites one patent from its own location and sector. Rows correspond to citing (idea destination) sectors. Columns correspond to cited (idea origin) sectors. Number of observations: 16,834,609. R^2: 0.32.

Figure B.2: **Identification of ζ, θ, and σ**



Notes: Moments in the data (horizontal dotted line) and in the model (blue marked line). Each of the plots is obtained by keeping the other two parameters constant at their calibrated values. Both model and data refer to the 1990 observation.

C Derivations

To rationalize the measure of local specialization in Equation (27), we impose the following assumption on the distribution of technological wave shocks:

Assumption A3. *Technological wave shocks are uncorrelated across sectors and have a constant variance:*

1. $Cov(\hat{\alpha}_{s,t}, \hat{\alpha}_{r,t}) = 0$ *for all* $s \neq r$
2. $Var(\hat{\alpha}_{s,t}) = V$ *for all* $s \in S$.

Using Assumption A3 in combination with Assumption A2 we derive the following theoretical result, that links the volatility of local population growth to the local degree of specialization:

Proposition C.1. *Under Assumptions A2 and A3, the variance of the percentage change in the population share of location* n *satisfies:*

$$Var(\hat{\pi}_{n,t}) \propto (1 - \pi_n^*)^2 \sum_{s \in S} \left(\pi_{s|n}^* - \pi_{s|-n}^*\right)^2. \tag{28}$$

Proof. Factoring out $(1 - \pi_n^*)$ from Equation (20), and realizing that $\pi_{s|-n}^* \equiv \sum_{m \neq n} \frac{\pi_{m,s}^*}{1-\pi_n^*}$, we can rewrite:

$$\hat{\pi}_{n,t} \overset{A2}{\propto} (1 - \pi_n^*) \sum_{s \in S} \left(\pi_{s|n}^* - \pi_{s|-n}^*\right) \hat{\alpha}_{s,t}. \tag{29}$$

Under Assumption A3, the technological wave shocks $\alpha_{s,t}$ have zero covariance and common variance V. Hence, the variance of $\hat{\pi}_{n,t}$ is equal to

$$Var(\hat{\pi}_{n,t}) \overset{A2,A3}{\propto} (1 - \pi_n^*)^2 \sum_{s \in S} \left(\pi_{s|n}^* - \pi_{s|-n}^*\right)^2.$$

□

If all cities are negligible in size compared to the overall economy, the measure in Equation (28) is approximately equal to the measure of specialization in Equation (27):

$$Spec_n \equiv \sum_{s \in S} \left(\pi_{s|n}^* - \pi_{\cdot,s}^*\right)^2.$$

D Data description

In this section, we provide details on the construction of the data on population, human capital, and employment by industry at the commuting zone level. Our starting points are the publicly available data from the Integrated Public Use Microdata Series (IPUMS, Ruggles et al., 2021) and the National Historical Geographic Information System (NHGIS, Manson et al., 2021), and the restricted full-count censuses for the decades until 1940.

We use the full-count censuses until 1940 to build crosswalks from historical counties to 1990 commuting zones. We then use these crosswalks in combination with county-level data from the NHGIS to consistently assign population and human capital data at the commuting zone level. Specifically, we follow a three-step procedure. First, we attempt to assign to each unique location in the historical decennial censuses – in terms of town, county, and state – their latitude and longitude.[44] Second, we count the number of people living in each town for the subset of locations that we were able to geolocate in the previous step, and reweight each town in this sub-sample so that the overall population count matches the county level data from the NHGIS.[45] Third, we assign the resulting town population to 1990 commuting zones.[46] For the decades 1950 onwards, we build crosswalks from counties to 1990 commuting zones based on the shares of overlapping areas.

Measures of the local density of human capital are assembled starting from county-level data from the NHGIS and aggregated at the level of 1990 commuting zones using the same crosswalks. The specific variables used to construct the measure vary by decade depending on the availability. We make the measures comparable across decades by converting them into the corresponding ranking. In 1890, we interpolate measures from the 1880 and 1990 decennial census. The 1880 measure concerns the share of people who attended school. Between 1900 and 1930, the measure represents the (inverse of the) share of illiterate people. In 1950, the measure reflects the median years of schooling of the population. From 1970 onwards, the measure corresponds to the share of population with at least a college degree.

[44]We retrieve the coordinates from Google Maps or, when uniquely available, from an offline database available at https://nationalmap.gov.

[45]Some towns in newly annexed territories are occasionally reported with generic names such as *Township 43*. We drop these observations from the sample and reweight the remaining ones to match the county level population data.

[46]As an example, consider the town of Denver, CO, that in 1890 was part of Arapahoe County, a large and sparsely populated county. By 1990, the city of Denver had separated from the rest of the county to form its own. The first two steps reveal that a large portion of Arapahoe County's population in 1890 was located in Denver. The third step uses this information to correctly assign the largest share of population to the city and county of Denver. There are two special cases that are worth mentioning. First, when more than 95% of the area of the historical county falls within a 1990 county, then we assign the whole population to the 1990 commuting zone that contains the 1990 county. Second, when a historical county does not contain any town that we were able to geolocate reliably, then its population is assigned to 1990 commuting zones based on the overlapping of their areas.

Local industry composition is constructed using the full-count census (until 1940) and the IPUMS (from 1950 onwards). We consider the number of people in each of the 12 main industries in the 1950 Census Bureau industrial classification system. To allocate individuals to 1990 commuting zones, we construct area-based crosswalks from State Economic Areas (in 1950), County Groups (in 1970), and PUMAs (in 1990 and 2010).

E Inertial dynamics and convergence to the BGP

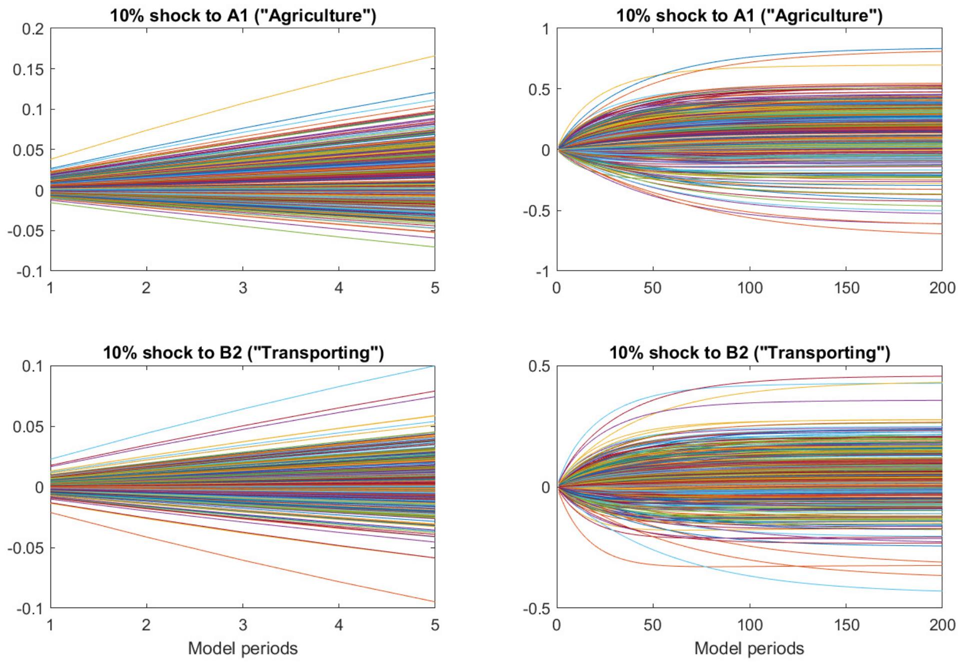
The quantitative exploration of Sections 4 and 5 assumes that the economy is in BGP in the initial period of the early and late intervals (1910 and 1970, respectively). Since any BGP can be rationalized by a proper choice of technological wave parameters (α_s^*) and structural residuals ($\epsilon_{n,s}^*$), this assumption can be justified by a (largely exogenous) reorganization of the U.S. spatial distribution of economic activities in the aftermath of WWII. Quantitatively, assuming two separate BGPs is convenient since it allows us to isolate the impact of technological wave shocks compared to a benchmark scenario in which the distribution of population in space is constant. Without this assumption, isolating the impact of technological wave shocks requires to properly normalize population growth to account for the model's inertial dynamics of convergence towards the new BGP. In this section, we start by showing that these inertial dynamics are substantial (that is, convergence towards the BGP is slow). We then show that, once a proper normalization is imposed to account for these inertial dynamics, the main results in the late interval are qualitatively consistent when relaxing the assumption of two separate BGPs.

The model's inertia is given by the fact that, following a technological wave shock, productivity growth in some location-sectors is faster (or slower) than the average. Initially, this fact results in a spatial and sectoral reallocation of population and economic activity. Eventually, new ideas diffuse to the rest of the economy, and the growth rate of productivity in all locations and sectors converges to a common value. In this new BGP, locations that were favorably exposed to the initial shock have permanently higher levels of productivity and population.

In Figure E.3, we illustrate this process by plotting the model's dynamics following a permanent 10% technological wave shock to A1 ("Agriculture", top panels) and B2 ("Transporting", bottom panels), starting from the 1910 BGP. The plotted lines represent population, in log-deviation from the initial BGP, for the 373 cities in the sample. The right panels show the dynamics of convergence towards the new BGP, while the left panels emphasize the initial response by focusing on the first 5 periods following the shock.

The left panels show that, in the periods immediately following the shock, the model's dynamics are close to linear. In the initial periods, geographical frictions to idea diffusion prevail, and local population growth is mostly controlled by the local exposure to the shock. However, after a significant number of periods, the growth rate of population in all cities converges to a common value, and the economy approaches a new BGP. This is illustrated in the right panels: As new ideas diffuse geographically over time, productivity growth in cities that were adversely exposed to the shock catches up with the locations that were more favorably exposed. In the new BGP, cities with a more favorable exposure tend to be permanently larger. The convergence towards the new BGP is slow: In the two experiments of Figure E.3, the half-

Figure E.3: **Inertial dynamics**



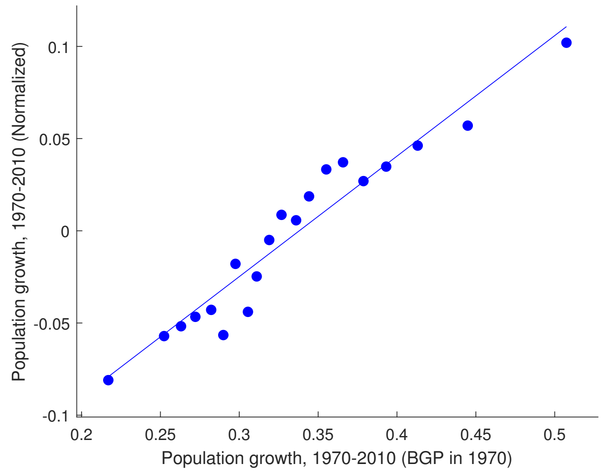
Notes: Each line represents the population of a commuting zone in log-deviation from the initial BGP (1910), following a 10% technological wave shock to A1 (top panels) and B2 (bottom panels).

life of convergence for the median city is 25 model periods (500 years) in the top panel and 22 model periods (440 years) in the bottom panel.[47]

Next, we relax the assumption of BGP at the beginning of the late interval (1970) and we verify that the patterns of population growth in this version of the model are consistent with our baseline. Since we want to isolate the effect of technological wave shocks, we look at population growth between 1970 and 2010 in deviation from the growth rate implied by the model's convergence towards the new BGP (which, in the baseline, is zero by construction). Specifically, we start by assuming the economy is in BGP in 1910, and calibrate the path of technological wave shocks and structural residuals for the entire period 1910-2010, following the same steps described in Section 4. We then look at two counterfactual experiments. In the first experiment, we feed the full path of technological wave shocks and structural residuals until 1970, and only the path of technological wave shocks in 1990 and 2010 (keeping structural residuals constant at their 1970 values). In the second experiment, we feed the full path of technological wave shocks and structural residuals until 1970, and we keep both of them constant for the remaining periods. The latter experiment is meant to isolate exclusively the inertia in population dynamics. Finally, we take the difference between the 1970-2010 population growth

[47]This measure assumes that the median city has reached the new BGP population in 200 model periods.

Figure E.4: **Population growth: baseline VS normalized model**



Notes: Bin-scatter plot of 1970-2010 population growth in the baseline model with technological wave shocks and constant structural residuals (horizontal axis), and in the experiment where population growth is normalized to account for convergence towards the BGP (vertical axis).

in the first and second experiments, to recover population growth normalized by the effect of convergence towards the BGP.

Figure E.4 shows that normalized population growth (vertical axis) is strongly correlated with population growth in the baseline model (horizontal axis). The R^2 of the underlying regression is 0.33 and the coefficient of the regression is 0.65. Table E.5 shows regressions of 1970-2010 population growth on the measure of exposure to the technological wave in Equation (26). Columns 1 and 2 show the coefficients in the data and in the baseline model (as they appear in Table 4). Column 3 reports the corresponding coefficient in the model normalized by the effect of convergence. The coefficient remains fairly stable and is still statistically indistinguishable from the estimate in column 1.

Table E.5: **Population growth and technological wave shocks: baseline VS normalized model**

	Population growth 1970-2010		
	Data	Model	
	(1)	(2)	(3)
Exposure to tech. wave	4.30***	4.27***	4.91***
	(0.99)	(0.03)	(0.13)
# Obs.	373	373	373
R^2	0.05	0.98	0.80
Idea flows across fields	-	Yes	Yes
Structural residuals	-	No	No
BGP in 1970	-	Yes	No

Notes: OLS estimates. Exposure to the technological wave is defined as in Equation (26). The dependent variable is defined as 1970-2010 population growth in the data (column 1), in the baseline model with technological wave shocks and constant structural residuals (column 2), and in the experiment where population growth is normalized to account for convergence towards the BGP (column 3). ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

F Model extensions

In this section, we illustrate how the model can be extended, without a prohibitive loss in tractability, to allow for overlapping generations, costly migration, local agglomeration or congestion forces, and trade among cities.

F.1 Overlapping generations and costly migration

The assumptions on the demographic structure can be easily reformulated to allow for overlapping generations and costly migration decisions. Individuals live for three periods (”child”, ”youth”, and ”old”). An agent i in her youth and old periods is endowed with one unit of inelastically-supplied labor and has labor productivity levels (q_i^y, q_i^o).

Every child is born in the location of their parents. At the end of childhood, the agent makes her migration and occupational choice by selecting which location n she migrates to and which sector s she specializes in. This choice is irreversible, so each agent spends the youth and old period in the same location-sector. Denoting by $L^y_{n,s,t}$ and $L^o_{n,s,t}$ the mass of young and old agents, the following identity holds:

$$L^o_{n,s,t} \equiv L^y_{n,s,t-1}.$$

After the migration and occupational choices have been made, each youth in period t has f_t children. Denoting by $L^k_{n,t}$ the mass of children born in location n at time t, we have that

$$L^k_{n,t} \equiv f_t \sum_{s \in S} L^y_{n,s,t}.$$

Migration and occupational choices are made to maximize expected lifetime utility, subject to migration costs and idiosyncratic utility draws. Utility of individual i born in location m and choosing location-sector (n, s) is given by

$$U_{m \to (n,s),t}(\mathbf{x}_i) = u_n \, \frac{x_{n,s,i} \left(c^y_{n,s,i,t}\right)^{\beta} \left(c^o_{n,s,i,t+1}\right)^{1-\beta}}{\mu_{m \to n}},$$

where u_n is the level of amenities in city n, $\mathbf{x}_i$ is a vector of idiosyncratic utility draws from a Fréchet distribution with shape parameter $\zeta > 1$, $\mu_{m \to n}$ represents moving costs (expressed in utility terms) of moving from m to n, $c^y_{n,s,i,t}$ and $c^o_{n,s,i,t+1}$ denote consumption in the youth and the old period, and $\beta \in (0, 1)$ is the weight on consumption during youth in lifetime utility. There are no markets to smooth consumption across time and generations. Hence, individual consumption is equal to individual production in every period.

Young and old agents produce the final good using their unit of time according to their idionsyncratic productivity q. Thus, total output in the economy is given by a linear aggregator over individual productivity across all locations and sectors

$$Y_t = \sum_{n \in N} \sum_{s \in S} \left(L^y_{n,s,t} \mathbb{E}[q^y_{n,s,t}] + L^o_{n,s,t} \mathbb{E}[q^o_{n,s,t}] \right),$$

where $\mathbb{E}[q^y_{n,s,t}]$ and $\mathbb{E}[q^o_{n,s,t}]$ denote the average productivity of young and old agents in location-sector (n, s).

How individual productivity is determined differs between young and old agents. Young agents benefit from a local learning externality that makes their productivity depend linearly on the average productivity of old agents:

$$q^y_{n,s,t} = A_y \, \mathbb{E}[q^o_{n,s,t}],$$

where A_y is a positive constant that can be set to replicate a measure of the experience premium. This formulation can be interpreted as reflecting the prevailing conditions in a local segmented labor market, in which the wage is an increasing function of the average quality of local ideas.

To determine productivity in the old period, young agents go through the imitate or innovate process described in Section 3.2.1. Since this process happens in the youth period, the relevant variables ($\epsilon_{n,s,t}$ and $\alpha_{n,s,t}$) are known at the time of the migration and occupational decision. Under Assumption A1, the local distribution of productivity among old agents remains Fréchet. The corresponding scale parameter, $\lambda_{n,s,t}$, follows the law of motion in Equation (7).

The probability of an individual born in location m to select city-sector (n, s) is

$$\pi_{m \to (n,s),t} = \frac{\left(u_n \frac{\Lambda_{n,s,t+1}}{\mu_{m \to n}} \right)^{\zeta}}{\sum_{l,r} \left(u_l \frac{\Lambda_{l,r,t+1}}{\mu_{m \to l}} \right)^{\zeta}},$$

where we define

$$\Lambda_{n,s,t+1} \equiv \lambda^{\frac{\beta}{\theta}}_{n,s,t} \lambda^{\frac{1-\beta}{\theta}}_{n,s,t+1}.$$

Thus, the following accounting identity between children and youth holds for all cities and sectors:

$$L^y_{n,s,t} \equiv \sum_{m=1}^{N} \pi_{m \to (n,s),t} L^k_{m,t-1}.$$

Consider now the dynamics of the migration shares, $\pi_{n,t}$, in response to an arbitrary deviation of $\Lambda_{n,s,t+1}$ from the BGP (denoted by $\hat{\Lambda}_{n,s,t+1}$). Denoting by $\pi_{m \to n,t}$ the probability of migrating from m to n, and log-linearizing this probability around the BGP, yields

$$\hat{\pi}_{m\to n,t} = \zeta \sum_{s\in S} \left\{ (1-\pi^*_{m\to n})\pi^*_{s|n}\hat{\Lambda}_{n,s,t+1} - \sum_{l\neq n} \pi^*_{m\to(l,s)}\hat{\Lambda}_{l,s,t+1} \right\}, \tag{30}$$

where $\pi^*_{s|n}$ is the BGP probability of choosing sector s conditional on migrating to n (note that this probability does not depend on the city of origin m). The total migration probability to location n can be written as

$$\pi_{n,t} = \sum_{m\in N} \pi_{m\to n,t}\pi_{m,t-1}.$$

Assuming the economy is in BGP at $t-1$, the response of the local migration share, $\pi_{n,t}$, can be written as

$$\hat{\pi}_{n,t} = \sum_{m\in N} \frac{\pi^*_{m\to n}\pi^*_m}{\pi^*_n}\hat{\pi}_{m\to n,t},$$

where $\hat{\pi}_{m\to n,t}$ is given by Equation (30). The term $\frac{\pi^*_{m\to n}\pi^*_m}{\pi^*_n}$ corresponds to the probability that a youth living in n was born in m. This probability can be interpreted as the "reliance" of n on migration from m.

As a last step, using Proposition 2, we can rewrite the shocks $\hat{\Lambda}_{n,s,t+1}$ in terms of technological wave shocks:

$$\hat{\Lambda}_{n,s,t+1} = \frac{1-\beta}{\theta}\hat{\lambda}_{n,s,t+1} = \frac{(1-\beta)g^*_\lambda}{1+g^*_\lambda}\sum_{r\in S}\eta^*_{r\to(n,s)}\hat{\alpha}_{r,t}.$$

F.2 Local agglomeration and congestion forces

The model can be augmented to account for local agglomeration and congestion forces, such as endogenous amenities and disamenities from density or local productivity externalities. In this case, the equilibrium conditions can no longer be written in closed form. While, as in Ahlfeldt et al. (2015), model inversion can be used to rationalize any collection of data as an equilibrium outcome, computing a counterfactual equilibrium in this extension requires an iterative procedure.

Congestion and agglomeration forces can be introduced, as in Allen and Arkolakis (2014), by postulating local amenities and productivity to be isoelastic functions of local population. Local amenities can be defined as

$$u_{n,t} = v_{n,t}\, L^{\omega}_{n,t},$$

where $v_{n,t}$ is the exogenous component of local amenities, and ω is the elasticity of local amenities to population, that can account for congestion forces in the case $\omega < 0$.

Analogously, productivity externalities can be introduced by assuming that individual productivity is augmented by an isoelastic function of local population. In this case, individual consumption can be written as

$$c_{n,s,i,t} = a_{n,t}\, L_{n,t}^{\rho}\, q_{n,s,i,t},$$

where $a_{n,t}$ is the exogenous component of local productivity, and ρ is the elasticity of local productivity to local population. Notice that, given values for the elasticities ω and ρ, inverting the model to back up $a_{n,t}$ would require an additional city-level observable variable, such as local income or consumption per capita.

The inclusion of endogenous agglomeration and congestion forces (in this formulation) does not interfere with the mechanism of innovation and knowledge diffusion. However, it affects migration and occupational probabilities (Equation (11)) that can now be rewritten as

$$\pi_{n,s,t} = \frac{\left(v_{n,t}\, a_{n,t}\, L_{n,t}^{\omega+\rho}\, \lambda_{n,s,t}^{\frac{1}{\theta}}\right)^{\zeta}}{\sum_{m,r}\left(v_{m,t}\, a_{m,t}\, L_{m,t}^{\omega+\rho}\, \lambda_{m,r,t}^{\frac{1}{\theta}}\right)^{\zeta}}.$$

F.3 Trade among cities

The model can also be extended to incorporate trade among cities. Also in this case, closed-form solutions are not available and counterfactual equilibria must be computed using an iterative procedure.

Assume that individuals consume a nested constant elasticity of substitution (CES) aggregate of output from each location and sector in the economy. Specifically, individual consumption is given by

$$c_{n,s,t,i} = \left[\sum_{r=1}^{S}\left(c_{n,s,t,i}^{r}\right)^{\frac{\gamma-1}{\gamma}}\right]^{\frac{\gamma}{\gamma-1}},$$

where $c_{n,s,t,i}^{r}$ denotes individual consumption of sector r aggregate, and $\gamma > 0$ is the elasticity of substitution across sectors. Sectoral aggregates are also CES functions of goods from each location:

$$c_{n,s,t,i}^{r} = \left[\sum_{m=1}^{N}\left(c_{n,s,t,i}^{m,r}\right)^{\frac{\beta-1}{\beta}}\right]^{\frac{\beta}{\beta-1}},$$

where $c_{n,s,t,i}^{m,r}$ denotes individual consumption of goods from location-sector (m, r).

There are iceberg shipping costs, $\{\tau_n^{m,r}\}_{m,n\in N\times N, r\in S}$, that satisfy $\tau_n^{n,r}=1$. Letting $p_{n,t}^{m,r}$ be the price of good (r,m) in city n, in equilibrium producers must be indifferent between selling to each location market:

$$p_{n,t}^{m,r} = \tau_n^{m,r} p_{m,t}^{m,r}.$$

These assumptions imply that each location n has a price index $P_n(\mathbf{p}_t)$, where $\mathbf{p}_t$ is the vector of all prices in the economy. Then, consumption for individual i in (n,s) can be written as

$$c_{n,s,t,i} = \frac{p_{n,t}^{n,s} q_{n,s,i,t}}{P_n(\mathbf{p}_t)}.$$

Market clearing requires that supply of each variety (m,r) is equal to its demand:

$$L_{m,r,t}\mathbb{E}\left[q_{m,r,t}\right] = \sum_{n,s} L_{n,s,t} p_{n,t}^{n,s} \mathbb{E}\left[q_{n,s,t}\right] \tilde{c}_n^{m,r}(\mathbf{p}_t),$$

where $\tilde{c}_n^{m,r}(\mathbf{p}_t)$ is the demand of variety (m,r) for an individual from location n with income 1.

Migration and occupational probabilities can be rewritten as

$$\pi_{n,s,t} = \frac{\left(p_{n,t}^{n,s} \lambda_{n,s,t}^{\frac{1}{\theta}} / P_n(\mathbf{p}_t)\right)^{\zeta}}{\sum_{m,r}\left(p_{m,t}^{m,r} \lambda_{m,r,t}^{\frac{1}{\theta}} / P_m(\mathbf{p}_t)\right)^{\zeta}}.$$