Cities and Technology Cycles*

Enrico Berkes† Ruben Gaetani‡ Martí Mestieri§

21st July 2020

Abstract

What determines the success and decline of cities over time? In this paper, we propose that geographical and technological frictions in the diffusion of ideas make the growth trajectory of cities sensitive to technology cycles, defined as long-term shifts in the centrality of some fields in the knowledge space. Using a novel dataset of historical US patents that spans the period 1836-2010, we show that cities whose innovative activities are more central in the technology network grow more over the following decades. We also show that diversification makes cities more resilient to technology cycles by guaranteeing a broad spectrum of ideas to draw from as specific fields gain or lose importance. We formalize these notions through a spatial, dynamic theory of innovation and frictional knowledge diffusion across city-sector pairs. In our model, the heterogeneous effects of technology cycles across cities accounts for 45% of the variation in city growth in the last century. We show that the changes over time in importance of fields of knowledge induced the rise and fall of manufacturing-intensive cities in the Rust Belt and the recent emergence of modern knowledge hubs. Finally, we use our model to speculate how future changes in the technological landscape may affect city growth in the next decades.

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, and seminar attendees at Northwestern, University of Toronto, Ohio State University and Conference of Swiss Economists Abroad 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). What determines these rich dynamics in the growth and decline of cities is still a matter of debate and one of the central questions in urban economics.

In this paper, we propose that frictions in the diffusion of ideas across regions and fields of knowledge make cities sensitive to "technology cycles", defined as long-term swings in the centrality of technological sectors in the innovation space. To measure technology cycles and their impact on the economic geography, we leverage on a new dataset of geolocated historical US patents over the period 1836-2010. We document a strong positive relationship between shocks to a city's centrality in the innovation landscape, induced by a favorable exposure to the current technology cycle, and its ability to attract population over the following decades. We further show that diversification of local innovation activities makes cities more resilient to technology cycles by decreasing their exposure to shocks and by improving their ability to reallocate resources towards expanding fields.

We then formalize the nexus between innovation, knowledge diffusion and migration in a dynamic, spatial-equilibrium model. We propose a parsimonious and tractable model that allows to characterize migration and knowledge flows in closed form. Leveraging on measures of innovation centrality and patent citations, we use our model to quantify the effect of technology cycles on US city growth over the twentieth century. We find that our proposed channel accounts for 45% of the observed changes in city growth. We also zoom in two episodes of radical transformation of the US economic geography: the growth of manufacturing-intensive cities in the first half of the 20th century, and their later decline accompanied by the rise of modern technological hubs. We show that knowledge diffusion played a substantive role in both episodes. We conclude by simulating the evolution of the US economic geography over the next decades under different possible scenarios of technological trends.

Our dataset of historical patents provides detailed information on the geographical origin, technological classes, and backward and forward citations on the universe of US patents granted since 1836. Using the frequency of co-appearance of each pair of technology classes in the same patent grant to infer technological connections between fields, we construct PageRank-based measures of centrality of each technological field in the innovation space over time. We show that the relative importance of fields has evolved significantly since the late 19th century, with innovation in agriculture giving way as the most central field to improvements in transportation in the earlier decades of the 20th century, followed by the expansion of IT-related sectors in later decades.

Using this measure of centrality, we document novel facts on how changes in the innovation landscape affected the US economic geography since the late 19th century. First, in the spirit of a shift-share analysis, we assess the relationship between exposure to favorable technology cycles and local population growth. We find that cities whose innovation activities are more heavily skewed towards expanding sectors experience higher population growth in the following decade. This relationship is statistically and economically significant, and is robust to a wide range of specifications, controls, and time windows.

Second, we explore the effect of a city's degree of diversification, measured as the Euclidean distance between the local and the national shares of patents across fields, and its sensitivity to technology cycles. We find that diversification reduces the likelihood of being exposed to large shocks, since it guarantees that negative shocks to some knowledge fields will be compensated by positive shocks to other fields. Moreover, we show that diversified cities respond more effectively to technology cycles by reallocating their portfolios away from declining fields and towards expanding ones.

The empirical facts that we document can have both positive and normative implications, as they suggest that the local composition of innovation can be a source of sustained urban growth in case of favorable technology cycles, but also lead to a reversal of fortune when the evolution of the innovation landscape changes direction. One hypothesis to rationalize these patterns is that the existence of geographical frictions in the transmission of ideas makes cities sensitive to changes in the centrality of their fields of specialization. In the second part of the paper, we develop a dynamic model of migration, innovation, and knowledge diffusion that we use to quantitatively assess the importance of this channel.

In our setting, overlapping generations of individuals make migration and occupational choices to maximize expected lifetime utility in a spatial economy. Upon entering adulthood, individuals are exposed to both local knowledge, that they can imitate and directly use in production, and to ideas learned from external sources, which they can combine with their current knowledge to generate an innovation. The local distribution of ideas endogenously retains a Fréchet structure over time, allowing us to characterize the knowledge flows through a simple gravity equation that can be estimated using bilateral citation probabilities. Moreover, these idea flows naturally imply a notion of technological importance of different fields in the network of knowledge, that can be disciplined using our empirical measures of centrality. While our framework is flexible in accounting for the incidence of other possible mechanisms behind these correlations, such as the existence of specialized physical and human capital or frictions in the reallocation of factors across space,[1] its structure allows us to accurately quantify to what extent the interaction between technology cycles and frictions in knowledge diffusion affected the dynamics of US cities over the last century.

The model remains tractable for any arbitrary number of sectors, locations, and time periods, and has a unique equilibrium that can be solved in closed form. We also show that the model has a recursive structure that allows us to calibrate the parameters and to back up the unobserved disturbances, including the shocks controlling the technology cycles, by making a small set of transparent assumptions. The calibrated model is successful in capturing key features of the data, and suggests that the endogenous mechanism of knowledge creation and diffusion, interacted with the estimated technology cycles, can account for 45% of the variation in population growth across US cities between 1910 and 2010.

The model delivers an intuitive representation for the evolution of a city's stock of knowledge as a function of the local and external spillovers and the identified technology cycles shocks. This allows us to isolate a structural residual that captures all the factors affecting the evolution of local innovation that cannot be directly ascribed to the endogenous mechanism of knowledge creation and diffusion. These factors include a variety of forces that can be either exogenous (e.g., natural events) or endogenous (e.g., opening of new research facilities) with respect to the local exposure to technology cycles. Our setting does not require to make any assumptions on the nature of this structural residual, but allows us to tell apart cases in which residual factors amplify or dampen the direct effect of technology cycles.

We use the model to analyze the quantitative importance of technology cycles in the context of two of the most prominent transformations of the US economic geography of the last century: The rise of manufacturing-intensive cities in the early decades of the 20th century, and their later decline to the benefit of emerging knowledge hubs specialized in information and bio technology. We find that the mechanism of endogenous knowledge creation and diffusion can explain a significant portion of the growth (and subsequent decline) of the major centers of heavy manufacturing, with residual forces amplifying the oscillations in their growth trajectory. This experience was mirrored in recent decades by some of the most rapidly expanding innovation centers in the US.

Finally, we use the quantitative model to predict transformations in the US economic geography in the coming decades under different scenarios for the evolution of the technological landscape. In particular, we ask which cities would benefit and lose under the following scenarios: (1) A comeback of transportation-related technologies to their 1950 peak, due to a rise in new modes of transportation such as autonomous vehicles; (2) An increase in the centrality of pharmaceuticals and biotech as the most pivotal technologies (at the expense of IT and

[1] Adão et al. (2020) provide evidence that technology-specific skills imply slow and unequally distributed adjustments to the emergence of new technologies.

electronics) in response to new challenges in global health; (3) A comeback of agriculture to its 1890 peak as a result of extensive regulatory change and increasing demand for sustainable farming. We find that cities in the Rust Belt would benefit from the first scenario, at the expense of cities in the North-East and the Pacific. The second scenario would penalize knowledge hubs specialized in IT-related innovation, favoring more diversified areas such as Boston, and Southern California. The third scenario would prompt a reallocation of economic activity towards the agricultural areas in the Central states.

Related Literature This paper contributes to multiple strands of literature. First, our growth theory is based on idea flows at the city-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, and Buera and Oberfield, 2016). Second, a rich body of work on quantitative spatial economics has developed tools for studying the determinants of economic activity in space, both within cities (e.g., Ahlfeldt et al., 2015, Heblich et al., 2018) and in a system of locations (e.g., Allen and Arkolakis, 2014, Desmet et al., 2018b).[2] 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

[2] 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.

reversal of fortune (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 (Desmet et al., 2018a), and regional or sectoral shocks (Caliendo et al., 2018). Our working hypothesis is that aggregate changes in the technological landscape, combined with frictional knowledge transmission, have a major impact on the geographical distribution of economic activity. Our framework can account simultaneously for path dependence and reversal of fortune in city dynamics, with a focus on understanding the evolution of the US economic geography in response to the main technological transformations in the last century. 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 US geography, both from an empirical perspective (e.g., Bostic et al., 1997, Simon and Nardinelli, 2002, Desmet and Rappaport, 2017) and from a structural and quantitative viewpoint (Desmet and Rossi-Hansberg, 2014, Nagy, 2017, Allen and Donaldson, 2018, Eckert and Peters, 2019).

This paper also contributes to the long-standing tension between the returns to local specialization (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 the key driver of local employment growth, and Duranton and Puga (2001), who develop a model in which diversified and specialized cities coexist in equilibrium.[3] Our paper suggests and quantifies a new channel through which urban diversification affects long-run city growth, namely, the responsiveness of a city to changes in the surrounding innovation landscape. It also implies a tradeoff between larger growth opportunities during favorable cycles, but more severe oscillations during adverse ones.[4] In this sense, our model provides a new lens for interpreting the effect of local policies directed at increasing local diversity.

The remainder of the paper is organized as follows: Section 2 introduces the data used for the empirical analysis, calibration, and quantitative exercises. Section 3 presents the stylized facts on the relationship between city growth and technological centrality and diversification. Section 4 introduces the model setting and Section 5 describes the methodology used to bring the model to the data. Section 6 presents the quantitative results. Section 7 discusses avenues for further research and concludes.

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

[4]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.

2 Data

Technological change is a slow-moving secular process. To study how the rise and fall of technologies determines the success 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 patents spanning a period of almost two centuries to measure technological cycles, as well as the centrality of cities in the innovation space and their level of specialization. We approximate cities using a full partition of the United States, namely the 1990 commuting zones (CZ), that we keep fixed throughout our analysis. Although commuting flows are likely to have changed over time, assuming a stable geography has the advantage of allowing us to abstract from annexations and redefinition of town borders that have been pervasive phenomena throughout the 19th and 20th century.

2.1 Patents Data

To measure innovative activities at the city level, we leverage on the Comprehensive Universe of US Patents (CUSP),[5] that contains information about patents filed (and subsequently issued) by the US Patent and Trademark Office (USPTO) between 1836 and 2010, with an estimated coverage above 90% in each year. Particularly, CUSP provides information about the technology classes,[6] name and location of each inventor (and assignee) listed on a patent, as well as their filing and issue dates. For our purposes, it has the advantage of assigning a set of coordinates to the residence of each inventor listed on a patent, instead of relying on the reported county. This allow us to build geographically consistent measures of innovation across decades.

2.2 Other Data

We collect economic and demographic information at the county level from the decennial historical Census (for the decades 1890-1940) and from the NHGIS (for the remaining decades). We then aggregate the data at the level of 1990 commuting zones, and we keep this geography fixed throughout the empirical and structural analysis. Assigning historical values to a recent geographical partition represents a challenge, since throughout the period considered the

[5]Berkes (2018) provides details about the data collection procedure, as well as some summary statistics and stylized facts related to the underlying data. Andrews (2019), in a comparison of historical patents data, describes it as "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 successfully employed 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 by comparing prosthetic patents during the Civil War and WWI, and Babina et al. (2020) who study the effect of the Great Depression on innovative activities in the United States.

[6]Retrieved from the USPTO's website in June 2016.

geography of the United States has undergone major modifications, with new territories being annexed and county boundaries being sometimes completely re-drawn, particularly in the West.

To construct consistent population measures over this stable geography, we follow a three step procedure. First, we assign each unique location in the historical decennial Censuses – in terms of town, county, and state – their latitude and longitude.[7] 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. Each town in this sub-sample is weighted so that the population at the county level matches the aggregate one.[8] Third, the total county population is assigned to 1990 counties depending on the distribution of people within that county in the considered decade.[9]

Following the same approach, we construct consistent measures of human capital that combine the available information on local literacy and education over time. To make this measure comparable across decades, we rank cities in terms of the relevant measure for that decade and use the resulting ranking for our analysis. To the best of our knowledge, this paper is the first to construct consistent local measures of human capital over a time span of over a century.

In both the empirical analysis and the model, we restrict the sample to the subset of commuting zones that accounted for at least 0.02% of the U.S. population for each decade since 1880. This delivers a sample of 425 commuting zones, that jointly account for 94.1% of the US population in 2010.[10]

3 Reduced-Form Evidence on Technology Cycles and City Growth

In this section, we document novel facts on the connection between the evolution of the technological landscape and the growth and decline of cities in the United States since 1860. We start by describing our PageRank-based measure of importance of technology classes over the period 1860-2010. This measure allows us to track episodes of expansion and decline of each

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

[8]Some towns in newly annexed territories are sometimes reported with generic names such as *Township 43*. These observations are dropped from our sample and the remaining ones weighted to match the county population reported in NHGIS.

[9]To fix ideas, in 1890 Denver, CO, 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. Our first two steps reveal that a large portion of Arapahoe County's population in 1890 was located in Denver. Thanks to this observation, the third step correctly assigns 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 the whole population is assigned to that 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 counties based on the overlapping of their areas.

[10]To fix ideas, this rule requires that cities had a population of at least 6,500 people in 1880 and 61,500 people in 2010.

field of knowledge over time. We then show that cities' exposure to those episodes is strongly correlated with local population growth over the following decade. Finally, we explore how diversification of the local portfolio of innovation activities affects a city's sensitivity to technology cycles by reducing their exposure to shocks and by increasing their ability to reallocate resources towards the fields that offer the best innovation possibilities.

3.1 Measuring the importance of technology classes over time

To track changes in the importance of technology classes over time, we implement a weighted variant of the PageRank algorithm (Brin and Page, 1998) to a network whose nodes are given by patent classes, and the strength of the connection between each pair of nodes is given by the frequency of co-appearance of that pair in the same patent grant. Intuitively, the PageRank of a given node captures the probability that a hypothetical "network-explorer" walking through the network at random eventually reaches that node. In our implementation, the "network-explorer" decides which edge to follow assigning a higher probability to edges with stronger connection. To avoid boundary problems with nodes without outbound edges, we assume that at each step the explorer has a probability p to jump to any node in the network. We set $p = 0.85$ consistently with the prevailing value in the literature (Brin and Page, 1998). Formally, the weighted PageRank (WPR) for technology class s in decade t is recursively defined as follows:

$$WPR_{s,t} = (1-p) + p \sum_{r \in \mathcal{S}} \frac{w_{r,s,t}}{\sum_k w_{r,k,t}} WPR_{r,t} \tag{1}$$

where $\mathcal{S}$ is the set of technology classes, and $w_{r,s,t}$ is the strength of the link between node r and node s in decade t. We compute the PageRank for each decade t using all the patents filed between decade t and $t+1$ (not included).[11] Since the PageRank captures the probability of visiting a given node in the network at any point in time, it adds up to one by construction for each decade.

Figure 1 illustrates that the PageRank captures substantial changes in the importance of technology classes over time. The graph uses a chromatic scale in which darker pixels correspond to higher PageRank. At the beginning of the sample (1860), Agriculture (class-group A1) is the one with the highest PageRank. At the turn of the century, class-groups B2 ("Shaping", that includes several IPC classes related to manufacturing technologies) and B4 ("Transporting", that includes the bulk of the automobile industry) emerge as the most central classes. Towards the end of the century, class-groups G1 ("Physics", that includes most of the inventions related to computers and information technology) and H1 ("Electricity", that includes

[11] For example, the weighted PageRank for decade 1960 takes into account all the patents filed between 1960 and 1969.

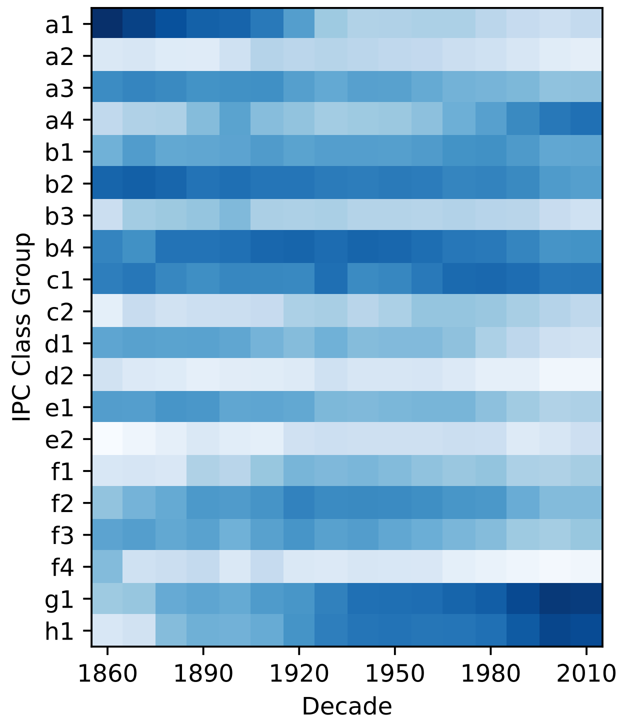


Figure 1: **Importance of patent class-groups over time.**

Notes: Heatmap representing the PageRank of each patent class-group over time. Darker pixels correspond to higher PageRank. The description of patent class-groups can be found in Table 8 in the Appendix.

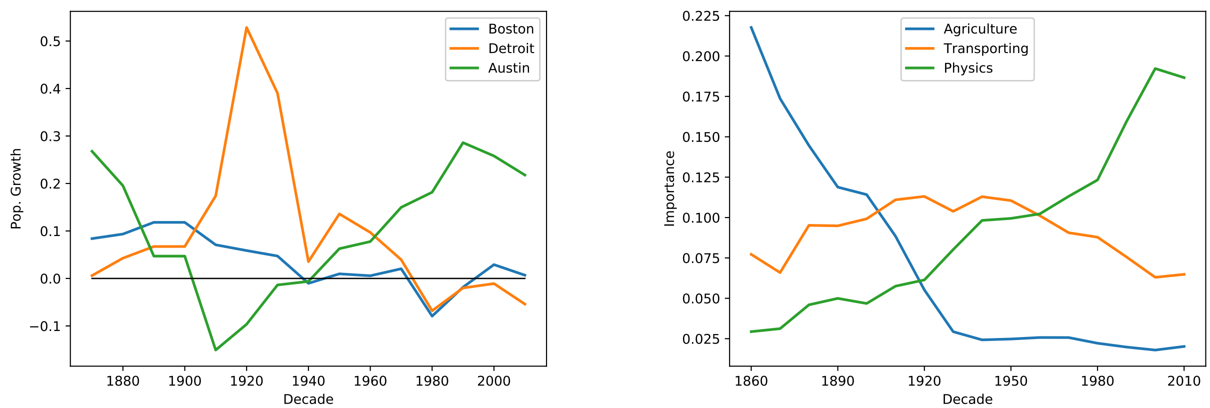
most technologies around electronics and semi-conductors) rise to the top, together with class-group A4 ("Health; Life-Saving; Amusement", that includes the majority of innovation related to Medical Sciences) and C1 ("Chemistry").[12]

3.2 A tale of three cities

Our hypothesis of interest is that the swings in technological importance depicted in Figure 1 can contribute to explaining the evolution of the US economic geography since the late 19th century. Before turning to our econometric analysis, we illustrate three archetypal examples of cities whose urban history is commonly narrated as intertwined with the history of their innovation activities and fields of specialization.

The left panel of Figure 2 shows the population growth of the commuting zones of Detroit, Austin, and Boston over the past 130 years. Detroit displays the most striking growth rates in

[12]These findings are robust to using other measures of network centrality, such us eigenvector centrality.



(a) Population growth

(b) Centrality of top class-groups

Figure 2: **Technology and city dynamics**

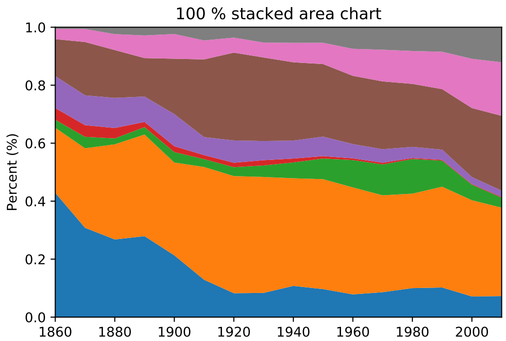
Notes: Panel (a): Residuals of a regression of decade-by-decade population growth on a set of Census Division-decade fixed effects. Panel (b): Evolution of the PageRank of the patent class-groups that have had the highest centrality for at least one decade since 1860 (the labels of the patent classes are abbreviated).

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. A specular trajectory is visible in the commuting zone of Austin. The city lost population in the early part of the 20th century as a consequence of the Texas Oil Boom that made Austin slip from the 4th to the 10th place among Texas's largest cities.[13] However, in recent decades Austin has emerged as one of the leading innovation hubs in the country, leveraging its base of high-tech start-ups and a large college-educated population. Finally, a still different experience is observed for the commuting zone of Boston, that, throughout the last century, has retained a significantly less volatile growth path, characterized by moderate population growth, interrupted by occasional periods of modest population decline.

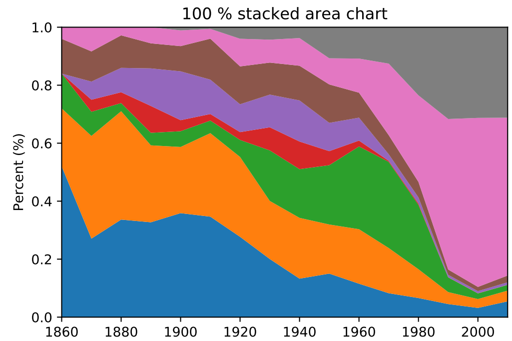
The left panel of Figure 2 reproposes a snapshot of the evolution of the centrality of the three patent class-groups that have occupied, for at least one decade since 1860, the top spot in terms of PageRank. The timing of these technology cycles suggests a connection with the city dynamics of the right panel of Figure 2. The rise and fall of transportation as the most central class-group coincides with the rapid expansion and decline of Detroit in the first half of the 20th century. Similarly, the ascent of computing-related technologies at the end of the 20th century closely tracks the expansion of Austin over the same period.

Those cities varied significantly not only in their fields of specialization, but also in the

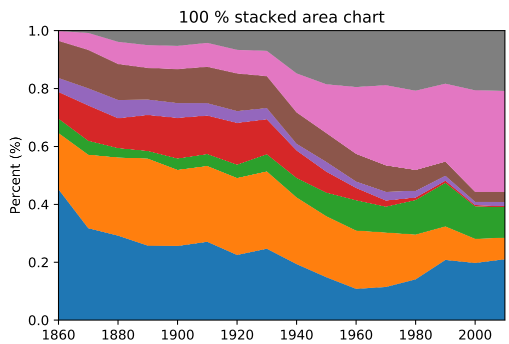
[13] https://tshaonline.org/handbook/online/articles/hda03



(a) Detroit



(b) Austin



(c) Boston

Figure 3: **Composition of the Technological Output**

Notes: The composition of the technological output is heterogeneous over time and across cities. The plots show the patenting output in Detroit, Austin, and Boston by technology class. The areas in the figure represent the 8 main classes in the International Patent Classification System. The blue area (bottom) represents class A (Human Necessities), whereas the gray area (top) represents the share of patents of class H (Electricity) in each city. The remaining colors are in descending order from H to A.

degree of diversification of their innovation activities. Figure 3 shows how the technological composition of patenting output changed over time across the cities considered in Figure 2. The areas in the Figure represent the 8 main classes in the International Patent Classification System ordered from A to H going from the bottom to the top.[14] Detroit's patenting output has specialized since the 1920s in patents of class B (Performing Operations; Transporting) and F (Mechanical Engineering), that made up about 80% of the technology output. Its portfolio has remained broadly unchanged since then with a slight shift towards patents of classes G (includes Computing) and H (includes Microchips) in the past decade. Austin, on the other hand, shows a fairly diversified patenting activity up until the 1960s when patents of classes

[14]The full description of each class is available at https://www.wipo.int/classifications/ipc/en/.

G and H started gaining importance to constitute about 90% of the technology portfolio in 2010. Interestingly, this shift towards specialization started around the same time Austin's population started its rapid increase. Boston displays a diversified technology portfolio which includes about 20% of patents in class A (includes Agriculture and Medicine) and about half of patents split between classes G and H. The consistent diversification of Boston's innovation output could make the city less sensitive to technology cycles, and contribute to explain the stability of its growth rate shown in Figure 2.[15]

In the remainder of this section we provide a systematic assessment of these patterns.

3.3 Technology cycles and the growth and decline of cities

Changes in the PageRank over time capture the long-run evolution of the innovation landscape. A positive (negative) change in the PageRank for a given knowledge field reflects an increase (decrease) in the centrality of that field in the patents network. In what follows, we explore to what extent those long-run changes in the relative importance of technological fields ("technology cycles") can be responsible for the growth and decline of cities. To this end, we first construct for each commuting zone a measure of local exposure to changes in the PageRank. We then verify whether commuting zones with more favorable exposure to technology cycles display systematically higher growth rates of population.

For any commuting zone c, we define exposure to the technology cycle in decade t as the sum over all the patent class-groups, $s \in \mathcal{S}$, of the change in the PageRank of s between $t-1$ and t, interacted with local patenting per capita in class s in decade $t-1$:

$$K_{c,t} = \sum_{s \in \mathcal{S}} \left(WPR_{s,t} - WPR_{s,t-1} \right) \frac{Pat_{c,s,t-1}}{Pop_{c,t-1}}. \tag{2}$$

The measure in (2) reflects both the distribution of patenting activity across classes within a commuting zone, which determines the sign of the shock, and the local patenting intensity, which controls the magnitude of the shock. Since the measure allows for negative values, corresponding to cities that are mostly exposed to declining fields, it cannot be regularized by taking logarithms.[16] Hence, in order to obtain a well-distributed variable, in what follows we work with the within-decade ranking of $K_{c,t}$, that we denote by $\tilde{K}_{c,t}$ and that we rescale so that a value of 0 reflects the largest negative shock, and a value of 1 the largest positive shock.[17]

[15] 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.

[16] This is particularly problematic in this case, since, as it is shown in Figure 13 in the Appendix, $K_{c,t}$ has thick tails.

[17] Note that this transformation corresponds to the c.d.f. of the exposure measure in each decade t. Results are consistent when using the non-transformed measure winsorized to regularize the fat tails.

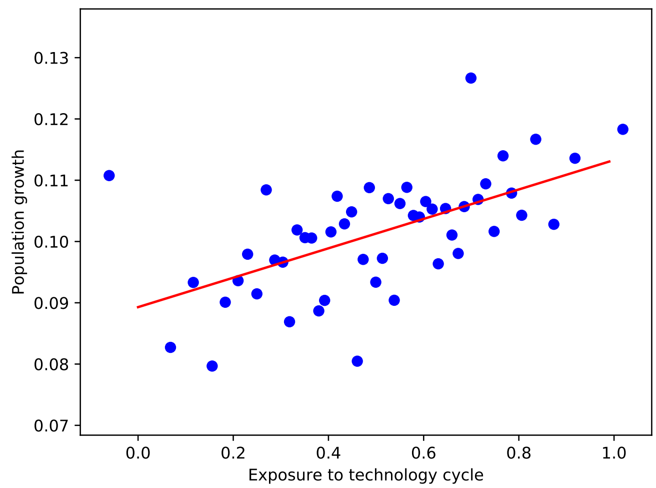


Figure 4: **Cities Exposure to Technology Cycles**

Notes: Cities exposure to technology cycles is correlated with local population growth in the following decade. The bin-scatter plot is residualized with respect to the full set of controls of equation (3).

We next show that exposure to technology cycles is positively correlated with local population growth. Specifically, we run a regression of the following form:

$$\Delta \log(Pop_{c,t+1}) = \beta \tilde{K}_{c,t} + \sum_{j=0}^{N} \delta_j \log(Pop_{c,t-j}) + \gamma h_{c,t} + \mu_{d,t} + \nu_c + \zeta_{c,t}. \quad (3)$$

In (3), $\Delta \log(Pop_{c,t+1})$ represents the growth rate of population between decade t and $t+1$. The full empirical model includes Census Division times decade fixed effects, $\mu_{d,t}$. A positive estimate of the parameter of interest, β, implies that cities that are more positively affected by changes in the technological environment will attract more population than other cities in the same Census division. We cluster standard errors two-way at the commuting zone and Census Division-decade level.

We control for current and lagged (up to $N = 2$ lags) log-population to account for size and convergence effects, and for the persistence of past shocks to population growth. We also control for a consistent measure of local density of human capital, $h_{c,t}$, assembled by combining different indicators from the historical Censuses.[18] Figure 4 shows a bin-scatter plot of the relationship between exposure to technology cycles ($\tilde{K}_{c,t}$) and local population growth ($\Delta \log(Pop_{c,t+1})$),

[18]This measure corresponds to a within-decade ranking of commuting zones along a summary index that encapsulates several local measures of human capital. The specific indicators we use change over the decades 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.

Table 1: **OLS Regressions of Technology Dynamics and City Growth**

	Growth rate of population					
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure to tech. cycle	.048***	.049***	.043***	.040***	.051***	.039***
	(.013)	(.013)	(.012)	(.012)	(.012)	(.011)
Human capital		.063***		.066**		.14***
		(.021)		(.024)		(.026)
Lags of population (0-2)	yes	yes	yes	yes	yes	yes
Decade FE	yes	yes	yes	yes	yes	yes
Census Division × Decade FE	no	no	yes	yes	yes	yes
Commuting Zone FE	no	no	no	no	yes	yes
# Obs.	5,525	5,503	5,525	5,503	5,525	5,503
R^2	0.42	0.41	0.50	0.49	0.68	0.68

Notes: CZ-level regression, 1860-2010. Standard errors clustered two-way at the Census Division × decade and commuting zone level in parenthesis. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

where both variables are residualized with respect to the complete set of controls.

Table 1 reports the Pooled-OLS and within estimates of (3) when the complete set of fixed effects and controls is progressively introduced in the regression model. Exposure to technology cycles is systematically correlated with higher population growth over the following decade. The magnitude of the estimated coefficient is similar across specifications and implies that cities with the highest positive exposure experience growth rates of population that are between 3.9 and 5.1 percentage points higher than cities with the most negative exposure. The effect of human capital on population growth is also consistently positive and statistically significant. In the appendix, we show that these results are robust to considering various sub-samples of decades (Figure 14) or to dropping from the sample one commuting zone at a time (15).

Within estimates are unbiased for large T under the assumption of weak exogeneity (i.e., conditional on controls, current and past values of the independent variable are uncorrelated with current and future population shocks).[19] Although this assumption is not implausible since past population trajectories are likely to capture various unobserved local factors that affect population growth, including persistence in the propensity to innovate and the innovation composition of a city, it is not completely innocuous. Unobserved shocks, such as local

[19]See Acemoglu et al. (2019) for a discussion of this point. The asymptotic bias of the within estimator (Nickell bias) is of order $1/T$.

Table 2: **System GMM Regressions of Technology Dynamics and City Growth**

	Growth rate of population			
	(1)	(2)	(3)	(4)
Exposure to tech. cycle	.124***	.110***	.132***	.108***
	(.016)	(.011)	(.020)	(.017)
Human capital		.067***		.054***
		(.011)		(.018)
Lags of population (0-2)	yes	yes	yes	yes
Decade FE	yes	yes	yes	yes
Census Division × Decade FE	no	no	yes	yes
# Obs.	5,525	5,503	5,525	5,503
Arellano-Bond AR(1) test	-4.34***	-4.38***	-4.53***	-4.44***
Arellano-Bond AR(2) test	1.74*	1.70*	1.63	1.54

Notes: CZ-level System GMM regression, 1860-2010. Two-step robust standard errors in parenthesis corrected for small-sample bias as described Windmeijer (2005). ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

financial disruptions, might contemporaneously affect our measure of exposure (for example, by changing local patenting rates) as well as population growth. To attenuate these concerns, we complement the analysis with the GMM system estimator proposed by Arellano and Bover (1995), and Blundell and Bond (1998).[20] The main idea behind this approach is to use past values of the endogenous variables to instrument the regression model in its version in levels and differences.[21]

Table 2 reports the estimates obtained with the system estimator, with standard errors computed following the two-step procedure proposed by Arellano and Bond (1991) and corrected for finite-sample bias as recommended in Windmeijer (2005). Results are qualitatively in line with the ones obtained via OLS, but the point estimates are larger. They imply that moving from the most negatively to the most positively exposed city increases subsequent city growth between 10.8% and 13.2%. Since the average city growth in the sample is 12% (see Table 7) this effect is economically meaningful. An increase in exposure of one residual standard deviation leads to an increase of 24.3% to 29.7% of one residual standard deviation of population growth.

[20] Examples of papers adopting this empirical approach include Beck and Levine (2004), Arcand et al. (2015), and Acemoglu et al. (2019).

[21] It can be shown that in the absence of second-order correlation for the error terms in the first differences equation, these are valid instruments. Under similar assumptions, lagged differences are valid instruments for the model in levels.

3.4 Diversification and resilience to technology cycles

As previously shown in Figure 3, there is a significant amount of variation in the degree of diversification of local portfolios of innovation activities across cities and over time. In what follows, we explore to what extent diversification can make cities more resilient to changes in the technological environment. We first show that more diversified cities have significantly less exposure to technology cycles, since diversification makes it more likely that negative shocks in some sectors are counterbalanced by corresponding positive shocks in other sectors. Moreover, we show that local diversification increases local resilience to technology cycles by facilitating the reallocation of activities away from declining, and towards expanding, fields. Specialized cities, on the other hand, are at the same time more exposed to technology cycles, and more sensitive to them, as they are less effective at reallocating innovation activities.

We start by defining our measure of local specialization as the Euclidean distance between the local and the national vector of patenting shares across patent class-groups:

$$S_{c,t} = \left[\sum_{s \in \mathcal{S}} \left(share_{c,s,t} - share_{US,s,t} \right)^2 \right]^{\frac{1}{2}}. \tag{4}$$

The expression in (4) assumes that a city is perfectly diversified whenever its distribution of innovation activities across all patent class-groups, $s \in \mathcal{S}$, is identical to the nationwide distribution, captured by the aggregate shares. To flexibly control for dimensionality biases in the construction of the measure, in the following regressions, we include a full set of 10 indicators for the total local patent count. Also in this case, to regularize the distribution of the measure, we work with the within-decade ranking of $S_{c,t}$, denoted by $\tilde{S}_{c,t}$, that we rescale so that a value of 0 represents the most diversified and a value of 1 the most specialized commuting zone.

To study the relationship between diversification and economic performance, we proceed in two steps. First, we show that more specialized cities are significantly more exposed to technology cycles. We construct a commuting zone-decade level dummy variable, $Exp_{c,t}$ that is equal to one if a given commuting zone has recorded in decade t an extraordinary large exposure (either positive or negative) to the technology cycle. In particular, this is captured by $\tilde{K}_{c,t}$ being either close to zero (for negative shocks) or to one (for positive shocks). In our benchmark, we define the variable $Exp_{c,t}$ as equal to one if $\tilde{K}_{c,t}$ is either in the bottom quartile ($\tilde{K}_{c,t} \in (0,\ 0.25)$) or in the top quartile ($\tilde{K}_{c,t} \in (0.75,\ 1)$) of the distribution of exposure across commuting zones.[22].

Figure 5 shows a bin-scatter plot of the relationship between local specialization at the beginning of the period $(t-1)$, $\tilde{S}_{c,t-1}$, and the likelihood of experiencing high exposure to

[22]Results are robust to using more stringent cutoffs, e.g. $0.1 - 0.9$ or $0.05 - 0.95$

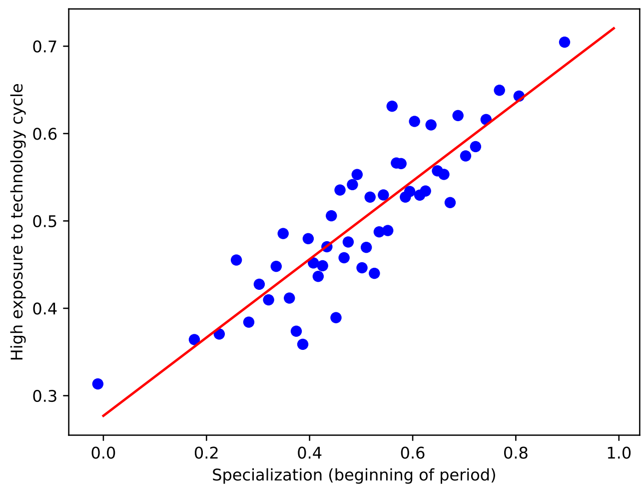


Figure 5: **Cities specialization**

Notes: Cities specialization is positively correlated to high exposure to technology cycle. The bin-scatter plot is residualized with respect to the full set of controls of equation (3)

the technology cycle, $Exp_{c,t}$, once the full set of controls of equation (3) and the indicators for the local count of patents are included. The scatter reveals a strong positive correlation between specialization and large exposure to technology cycles. This correlation reflects the fact that a higher local diversification makes it more likely that negative shocks to some sector are counterbalanced by positive shocks to other sectors, making diversified cities more insulated from changes in the aggregate technological environment.

Second, we explore whether diversification of the local portfolio of innovation makes cities more resilient to technology cycles by facilitating the reallocation of resources away from declining sectors and towards newly expanding fields. If this intuition is correct, we should observe that patenting per capita in a given class-group responds more strongly to changes in the corresponding PageRank in diversified cities compared to specialized ones. To test this intuition, we estimate a regression model of this form:

$$\Delta \log \left(\frac{Pat_{c,s,t}}{Pop_{c,t}} \right) = \beta \tilde{S}_{c,t-1} + \gamma \tilde{S}_{c,t-1} \times \Delta WPR_{s,t} + \delta X_{c,t} + \mu_{d,t} + \nu_c + \omega_{s,t} + \zeta_{c,s,t}. \quad (5)$$

In (5), a negative estimate of the parameter of the interaction between local specialization ($\tilde{S}_{c,t-1}$) and the technology shock (measured as the change in the PageRank, $\Delta WPR_{s,t}$) implies that more specialized cities respond to technology cycles by reallocating less resources towards expanding fields, and by shifting fewer resources away from declining sectors.

The results of the OLS regression are reported in Table 3. The estimate of γ is consistently

Table 3: **Exposure to Technology Cycles**

	Growth rate of patents per capita			
	(1)	(2)	(3)	(4)
Specialization (lagged)	.016	.014	.075***	.073***
	(.020)	(.021)	(.025)	(.027)
Specialization (lagged) × $\Delta WPR_{s,t}$	-4.96***	-4.97***	-5.00***	-5.01***
	(1.43)	(1.46)	(1.45)	(1.47)
Human capital		.022		.004
		(.043)		(.048)
Lags of population (0-2)	yes	yes	yes	yes
Patent count indicators	yes	yes	yes	yes
Class-group × decade FE	yes	yes	yes	yes
Census Division × Decade FE	yes	yes	yes	yes
Commuting Zone FE	no	no	yes	yes
# Obs.	119,000	118,560	119,000	118,560
R^2	0.28	0.28	0.29	0.28

Notes: CZ × class-group level regression, 1860-2010. Standard errors clustered at the Census Division × decade, commuting zone, and class-group × decade level in parenthesis. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

negative and statistically significant across specifications. The magnitude of the coefficient in column (4) implies that cities with the highest degree of specialization (i.e., $\tilde{S}_{c,t-1} = 1$) respond by increasing patenting per capita towards expanding fields (at the 90th percentile of the distribution of $\Delta WPR_{s,t}$) by 5.54 percentage points less than cities with the highest degree of diversification (i.e., $\tilde{S}_{c,t} = 0$). The corresponding figure for declining fields (at the 10th percentile of the distribution of $\Delta WPR_{s,t}$) implies that the most specialized cities decrease patenting by 5 percentage points less than diversified cities.

Taking stock Taken together, the results of this section paint a picture consistent with the view that the technology cycle affects the evolution of population growth. Both the extent of city specialization and diversification across sectors appear to have a significant and sizable effect on subsequent city growth. To interpret these facts, we develop in the next section a model with location choice, frictional idea diffusion and endogenous specialization patterns. The model allows us to make explicit the role of the technology cycle on workers' productivity and its effect on migration decisions. Moreover, it allows us to quantify the importance of our mechanism for city growth and conduct counterfactual analysis.

4 Model

In this section, we develop a quantitative, spatial, endogenous growth model to rationalize the relationship between city size and their innovative activity documented in Section 3. We build on the recent literature of idea diffusion and growth (e.g., Buera and Oberfield, 2016) and extend them into a spatial setting.[23] The central theme of our theory is that the evolution of the network of knowledge and idea diffusion across cities and sectors drive cities' productivity and population.

4.1 Environment

We consider a discrete-time overlapping-generations economy of innovators and workers, who make migration and occupational choices over locations and technological sectors. The economy comprises a finite set N of locations and a finite set S of sectors. In what follows, we refer to N and S as both the sets of locations and sectors, and their cardinality.

4.1.1 Preferences, endowments and demographics

Individuals live for three periods. We refer to an individual in the first, second, and third periods of their life as "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). We discuss the evolution of agents productivity in detail below.

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. In other words, denoting by $L^y_{n,s,t}$ and $L^o_{n,s,t}$ the mass of young and old agents, the following identity holds:

$$L^o_{n,s,t} \equiv L^y_{n,s,t-1}. \tag{6}$$

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

$$L^k_{n,t} \equiv f_t \sum_{s \in S} L^y_{n,s,t}. \tag{7}$$

[23]See Buera and Lucas (2018) for an excellent review. These models either consider an economy without space or are embedded in a trade setting.

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

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

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 lifetime utility. Utility of individual i born in location m is given by

$$U^m_{n,s,t}(\boldsymbol{x}_i) = u_n \, \frac{x_{n,s,i}\,(c^y_{n,s,i,t})^{\beta}\,(c^o_{n,s,i,t+1})^{1-\beta}}{\mu^m_{n,t}}, \tag{8}$$

where u_n is the level of amenities in city n, $\mu^m_{n,t}$ represents moving costs (expressed in utility terms) of moving from m to n at time t, $c^y_{n,s,i,t}$ 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.

4.1.2 Production and Innovation Technologies

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 will 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.

During their youth period, individuals acquire skills, knowledge, and ideas that they later convert into their productivity when old. Specifically, every young agent before becoming old receives a full set of idiosyncratic, independently distributed draws:

$$\boldsymbol{z}_{n,s,i} = \left\{ z^{l}_{n,s,i}, \{z^{x}_{m,r,i}\}_{m,r \in N \times S} \right\}. \tag{9}$$

The first term in Equation (9), $z^{l}_{n,s,i}$, represents a random draw from the distribution of productivity among the old in the same location-sector pair of the youth, whose cumulative distribution is denoted by $F_{i,s,t}(q)$. 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.[24] If the agent chooses to adopt this idea in production, their productivity in the old-period is

$$q^{o}_{n,s,i,t+1} = z^{l}_{n,s,i}.$$

The second set of terms in Equation (9), $\{z^{x}_{m,r,i}\}_{j,r \in N \times S}$, represents a full vector of random draws from each productivity distribution among the old of all locations and sectors in the economy, all with corresponding cumulative distributions $\{F_{m,r,t}\}_{m,r \in N \times S}$. 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 currently employed in (n, s) can use an idea $z_{m,r,i}$ drawn from (m, r) to innovate and achieve old-period productivity

$$q^{o}_{n,s,i,t+1} = q^{y}_{n,s,i,t} \frac{\xi_{n,s,t} \alpha_{r,t} z^{x}_{m,r,i}}{d^{n,s}_{m,r,t}}. \tag{10}$$

In (10), the term $\alpha_{r,t}$ represents an economy-wide technological shock to 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. The term $d^{n,s}_{m,r,t}$ captures the geographical and technological frictions that discount the effectiveness of knowledge transmission between the idea origin (j, r) and the idea destination (i, s). We discuss in the next section how we parametrize these costs.

The term $\xi_{n,s,t}$ is a term that captures the current effectiveness of innovation in (n, s) and is common to all innovators in this location-sector pair. This term is the product of two

[24] de la Croix et al. (2017) 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.

components:

$$\xi_{n,s,t} = \underbrace{\frac{\mathbb{E}[q^o_{n,s,t}]}{\mathbb{E}[q^o_{n,s,t-1}]}}_{\text{Absorptive capacity}} \times \underbrace{\epsilon_{n,s,t}}_{\text{Structural residual}} . \tag{11}$$

The "Absorptive capacity" term captures the fact that places with higher rates of innovation in the past period (resulting in higher productivity growth) are better equipped at receiving knowledge and developing new ideas in the current period. That is, conditional on having the same idea quality z, being in a location with higher absorptive capacity improves the final productivity of an agent's innovation. For example, the same idea draw z in IT yields today a more productive innovation in the commuting zone of San Jose, CA, whose patenting rate in class-group G1 was 11.2 grants per 1,000 people in 1990, than in Lincoln, NE, whose patenting rate in the same class-group was 0.62 per 1,000 people. The "Structural residual" term includes all the residual factors that affect the productivity of the local sector but are not explicitly included in (10), such as the opening of production facilities, universities, and research centers.[25]

4.1.3 Markets

We assume that agents live hand-to-mouth. That is, there is no market to smooth consumption intertemporally. Thus, each agents' consumption of final good is given by her production of final good at each point in time.[26]

4.2 Equilibrium

4.2.1 Diffusion of knowledge

Entering the old age, agent i in (n, s) chooses whether to imitate or innovate to maximize her old-period productivity given her set of idiosyncratic idea draws $\boldsymbol{z}_{n,s,i}$.

$$q^o_{n,s,i,t+1} = \max\left\{ z^l_{n,s,i},\ \max\left\{ q^y_{n,s,i,t} \frac{\xi_{n,s,t}\alpha_{r,t} z^x_{m,r,i}}{d^{n,s}_{m,r,t}} \right\}_{m,r \in N \times S} \right\} \tag{12}$$

Equation (12) shows how this process can be divided in two steps. First, the agent chooses the best innovative idea available to her. Then she compares this best innovative idea with her imitation draw, and picks the one that yields higher productivity for her.

The following assumption will play an important role in keeping our theory tractable:

[25] Part of these factors can also capture endogenous research effort, which we plan to show in a model extension.

[26] We have also used an alternative setup with linear preferences over consumption across periods where we allow for intertemporal allocation of consumption, obtaining very similar results.

Assumption 1. *The initial productivity distributions* $\{F_{n,s,0}(q)\}_{(n,s)\in N\times S}$ *across all sector-city pairs are given by independently distributed Fréchet distributions with shape parameter* $\theta > 1$ *and scale parameters* $\lambda_{n,s,0} > 0$,

$$F_{n,s,0}(q) = e^{-\lambda_{n,s,0}q^{-\theta}}. \tag{13}$$

A multivariate Fréchet distribution with common shape parameter is max-stable. This implies that, under Assumption 1, the resulting distribution over the max of Fréchet draws is also Fréchet with the same shape parameter.[27] Combining (12) with (13), we find that the old-period 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:

$$\lambda_{n,s,t+1} = \underbrace{\lambda_{n,s,t}}_{\text{Imitation}} + \underbrace{\lambda_{n,s,t} \sum_{m\in N} \sum_{r\in S} \lambda_{m,r,t} \left(\frac{\xi_{n,s,t}\alpha_{r,t}}{d^{n,s}_{m,r,t}}\right)^{\theta}}_{\text{Innovation}}. \tag{14}$$

Equation (14) summarizes the growth dynamics implied by our model. The scale parameter of the old generation at $t+1$ corresponds to the previous generation scale parameter augmented by a second term which captures innovation in the city-sector (n, s). This second term in Equation (14) is composed of the previous period scale parameter augmented by the sum of scale parameters across all sector-locations weighted by their applicability to city-sector (n, s). This applicability term comprises technological and physical distance $d^{n,s}_{m,r,t}$ between city-sector pairs, changes in the importance of each field of knowledge $\alpha_{r,t}$ and idiosyncratic effectiveness of innovation in (n, s), $\xi_{n,s,t}$.

Moreover, Equation (14) 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:

$$\eta^{n,s}_{m,r,t} = \frac{\lambda_{m,r,t}\left(\frac{\alpha_{r,t}}{d^{n,s}_{m,r,t}}\right)^{\theta}}{\sum_{l,z} \lambda_{l,z,t}\left(\frac{\alpha_{z,t}}{d^{n,s}_{l,z}}\right)^{\theta}}. \tag{15}$$

Knowledge and Productivity Dynamics across Cities and Sectors Before turning to the analysis of the rest of the equilibrium, we discuss the implications of our model for knowledge and productivity dynamics. Rearranging (14), we have that the growth in the Fréchet scale parameter is

$$g_{n,s,t} \equiv \frac{\lambda_{n,s,t+1} - \lambda_{n,s,t}}{\lambda_{n,s,t}} = \sum_{m\in N} \sum_{r\in S} \lambda_{m,r,t}\delta^{n,s}_{m,r,t} \tag{16}$$

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

where $\delta_{m,r,t}^{n,s} \equiv \left(\frac{\xi_{n,s,t}\alpha_{r,t}}{d_{m,r,t}^{n,s}}\right)^{\theta}$ summarizes the applicability of knowledge from city-sector (m, r) to (n, s). This implies that the "step-size" increase in labor productivity in city-sector (n, s) is

$$\frac{\mathbb{E}[q_{n,s,t+1}^{o}]}{\mathbb{E}[q_{n,s,t}^{o}]} = (1 + g_{n,s,t})^{1/\theta}. \tag{17}$$

These results underscore the central role of knowledge and idea diffusion in our model. The growth rate of the the Fréchet scale parameters also governs the growth rate of city-sector productivity and, ultimately, migration and city growth.

To gain intuition on the model dynamics, we begin by discussing two extreme cases. First, suppose that applicability considerations are irrelevant and $\delta_{m,r,t}^{n,s} = 1$, so that all ideas in all locations are equally applicable in all city-sectors. We have that

$$\lambda_{n,s,t+1} = \lambda_{n,s,t}(1 + g_t), \tag{18}$$

where the growth rate is constant across all city-sector pairs and satisfies the autonomous equation $g_{t+1} = g_t(1 + g_t)$, where $g_0 = \sum_{s\in S, n\in N} \lambda_{n,s,0}$.[28] In this case, knowledge grows at a constant rate that is common in all city-sectors. As a result, relative productivity would be constant across cities, implying full persistence of initial conditions.

Next, consider the case in which knowledge from one field of knowledge is not useful for any other field, but there are no applicability concerns across cities within a field of knowledge. In other words, suppose that $\delta_{m,r,t}^{n,s}$ is one if $r = s$ and zero otherwise. In this case, we have that

$$\lambda_{n,s,t+1} = \lambda_{n,s,t} \left(1 + g_{s,t}\right), \tag{19}$$

where $g_{s,t+1} = g_{s,t}(1+g_{s,t})$ and $g_{s,0} = \sum_{n\in N} \lambda_{n,s,0}$. At first sight, this case appears similar to the first example for the diffusion of knowledge. Each sector grows at a common rate. Thus, the relative knowledge across cities within a field of knowledge is again constant and determined by the initial conditions. However, the growth rate of knowledge can be heterogeneous across fields of knowledge. To the extent that cities differ in their pattern of specialization on fields of knowledge, this heterogeneity in sectoral growth implies dynamics for city productivity and population growth that are substantially different from the first example. Indeed, if some sector grows at a faster rate than another, more agents will choose to specialize in the sector. As a result, productivity growth is going to be driven by the faster-growing sectors and cities with a higher initial stock on knowledge in the sector are going to receive population inflows.

In practice, the stark assumptions of the previous two examples are not realistic, and the

[28]ote that if the initial growth rate is sufficiently small, i.e. $g_t \ll 1$ so that $g_{t+1} = g_t + g_t^2 \simeq g_t$, we have that the system grows at an approximately constant rate and $\lambda_{n,s,t} = \lambda_{n,s,0} \left(1 + \sum_{m\in N} \sum_{r\in S} \lambda_{m,r,0}\right)^t$.

applicability term, $\delta^{n,s}_{m,r,t}$, will capture the other forces controlling innovation and productivity dynamics. First, although there is some degree of geographical transmission and cross-fertilization, the flow of ideas across cities and fields of knowledge is not frictionless. This effect is captured through the "distance" term $d^{n,s}_{m,r,t}$. Second, the degree to which knowledge in a given field can be used as an input in innovation varies over time and across fields and is captured by the term $\alpha_{r,t}$. Finally, sector-city pairs differ in their ability to absorb new ideas and effectively convert them into productive uses. This variation is encapsulated in the idiosyncratic term $\xi_{n,s,t}$, which incorporates "absorptive capacity" (places with higher innovation rates in the previous period are in a better position to innovate in the current period) as well as a residual term that may subsume exogenous or endogenous forces, such as the endogenous effort in research and development that agents may do to further develop their ideas.

Substituting the structure of the effectiveness of innovation $\xi_{n,s,t}$ from Equation (11) in (14), we can express the growth rate of $\lambda_{i,s}$ as

$$\frac{\lambda_{n,s,t+1}-\lambda_{n,s,t}}{\lambda_{n,s,t}} = \frac{\lambda_{n,s,t}}{\lambda_{n,s,t-1}}\tilde{\epsilon}_{n,s,t}\sum_{m,r}\lambda_{m,r,t}\left(\frac{\alpha_{r,t}}{d^{i,s}_{m,r,t}}\right)^{\theta}, \tag{20}$$

where we have used the fact that $\mathbb{E}[q^o_{i,s,t}] \propto \lambda^{\frac{1}{\theta}}_{i,s,t}$ and defined $\tilde{\epsilon}_{i,s,t} \equiv \epsilon^{\theta}_{i,s,t}$.

4.2.2 Migration and occupational choice

All the variables that enter lifetime utility (8), with the exception of old-period consumption $c^o_{n,s,t+1}$, are known by agents at the time in which migration and occupational choices are made (upon entry in the youth-period). In particular, an agent i born in location m moving to location-sector pair (n, s) has expected lifetime utility equal to

$$\mathbb{E}\left[U^m_{n,s,t}(\boldsymbol{x}_i)\right] = u_n A^{\beta}_y \frac{x_{n,s,i}\,\lambda^{\frac{\beta}{\theta}}_{n,s,t}\,\mathbb{E}\left[(q^o_{n,s,t+1})^{1-\beta}\right]}{\mu^m_{n,t}}. \tag{21}$$

In equilibrium, $q^o_{i,s,t+1}$ is distributed Fréchet with shape parameter θ and scale parameter $\lambda_{i,s,t+1}$, which can be inferred at time t via the law of motion (14). Hence, the term $(q^o_{i,s,t+1})^{1-\beta}$ is itself distributed Fréchet with shape parameter $\frac{\theta}{1-\beta}$ and scale parameter $\lambda_{i,s,t+1}$, so that

$$\mathbb{E}\left[(q^o_{i,s,t+1})^{1-\beta}\right] = \Gamma\left(1-\frac{1-\beta}{\theta}\right)\lambda^{\frac{1-\beta}{\theta}}_{i,s,t+1},$$

where $\Gamma(\cdot)$ denotes the gamma function. This implies that the probability of an individual born in location m to select city-sector (n, s) is

$$\pi^m_{n,s,t} = \frac{\left(u_n \frac{\Lambda_{n,s,t+1}}{\mu^m_{n,t}}\right)^{\zeta}}{\sum_{l \in N} \sum_{r \in S} \left(u_l \frac{\Lambda_{l,r,t+1}}{\mu^m_{l,t}}\right)^{\zeta}}, \tag{22}$$

where we define

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

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

$$L^y_{n,s,t} \equiv \sum_{m=1}^{N} \pi^m_{n,s,t} L^k_{m,t-1}. \tag{23}$$

4.2.3 Equilibrium Definition

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

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

$$\left\{u_n,\, \lambda_{n,s,0},\, L^k_{n,0},\, L^y_{n,s,0},\, L^o_{n,s,0}\,\right\}_{n,s \in N \times S},$$

and a given path for the exogenous variables

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

an equilibrium is a path for the endogenous variables

$$\left\{\lambda_{n,s,t},\, \{\pi^m_{n,s,t}\}_{m \in N},\, L^k_{n,t},\, L^y_{n,s,t},\, L^o_{n,s,t}\right\}_{n,s \in N \times S,\, t \geq 0}$$

that satisfies the following conditions:

1. *Migration probabilities* $\{\pi^m_{n,s,t}\}_{m,n,s \in N \times N \times S,\, t \geq 0}$ *satisfy equation (22).*

2. *The path for* $\{\lambda^m_{n,s,t}\}_{n,s \in N \times S,\, t \geq 0}$ *satisfies the law of motion of equation (14).*

3. *Population by age and sector,* $\left\{L^k_{n,t},\, L^y_{n,s,t},\, L^o_{n,s,t}\right\}_{n,s \in N \times S,\, t \geq 0}$, *satisfies the transition identities (6), (7), and (23).*

All equilibrium conditions have an explicit solution. Hence, it is straightforward to see that 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.

5 Model Quantification

In this section, we show how we bring the model to the data to infer the key structural parameters and the unobserved exogenous variables. The model has a recursive structure that allows to estimate these values sequentially by making a limited set of transparent assumptions on how to map the model's objects into data on population, income, and patenting by sector. We then show that the quantified model is successful at capturing key features of the data that have not been directly targeted.

We set a time period to be 20 years, so that agents' life is divided into three periods, corresponding to ages 0-19, 20-39, and 40-59. In the notation, we denote model periods by the central year of a 20-year window (for example, the period 1990 refers to the window 1980-1999). We interpret locations in the model as 1990 commuting zones.[29] We remind the reader that we restrict attention to commuting zones that, in all decades since 1880, have hosted at least 0.02% of the US population. This gives a subset of $N = 425$ locations that we use in the quantitative analysis. Sectors are defined by the technological class-groups introduced in Section 3, with 3-digit IPC classes grouped into $S = 20$ main sectors, as detailed in Table 8.

We assume that aggregate population growth is entirely driven by fertility. This is equivalent to assuming that, for new cohorts of immigrants, the probability of first settling in a given location is proportional to current youth-period population.[30]

To set the ground for the estimation of the model, we take 1880-1899 (which we express as $t = 1890$) to be the starting period. We assume that the economy is initially in its long-run steady state, so that $\alpha_{s,1890} = 0$ for all sectors $s \in S$. In the initial steady state, the distribution of productivity for each sector is constant. Although migration still takes place, and is governed by the transition probabilities $\pi^{m}_{n,s}$, population is determined as the fixed point of the identities (6), (7), and (23), so that the mass of individuals by age, city and sector is fixed.[31]

We start by calibrating two of the structural parameters externally. We set the weight of youth-period consumption in lifetime utility β equal to 0.60, implying an average yearly discount factor of 0.98 over a 20 years period. We set the parameter A_y to 0.80, which implies an experience premium within each location-sector equal to 1.24.[32] We summarize these parameter value choices in Table 6.

[29] Consistently with the results in Section 3, we keep this definition over decades.

[30] The model can be easily extended to account for migration from foreign countries being skewed into particular cities.

[31] Although this assumption is not critical to any of our results, it facilitates the assessment of the model's performance since it implies that, in the absence of aggregate shocks, the economy would not depart from its steady state. An alternative calibration strategy is to use two consecutive periods, e.g. 1880 and 1900 to calibrate the model parameters and not impose an initial stead-state condition. By and large, both methodologies give similar results.

[32] This figure is obtained from the Panel Study of Income Dynamics as the average premium of workers of age 45-54 over workers of age 25-34 for the year 1985-1994.

5.1 Gravity representations for migration and knowledge flows

The first step of the model's quantification entails recovering the parameters controlling migration and knowledge transmission costs.

Migration flows Recall from Section 4.1.1 that migration costs enter the lifetime utility of individuals and hence affect the probability of bilateral migration between each pair of commuting zones. We assume migration costs to be an exponential function of geographical distance and include a fixed cost of migration:

$$\mu_{n,t}^{m} = e^{\mu_t^0 \mathbb{1}_{\{i \neq j\}} + \mu_t^1 \tau_{n,m}}, \tag{24}$$

where $\tau_{n,m}$ is the distance in kilometers between locations n and m. Note that we allow the parameters that control migration costs to depend on time to account for the substantial changes in transportation costs that have occurred in the US over the 20th century.

Denoting the sum over all sectors in each location $\pi_{n,t} \equiv \sum_{s \in S} \pi_{s,n,t}$, we can combine the transition probabilities in Equation (22) with the functional form in (24) to obtain a linear gravity representation for bilateral migration flows:

$$\log(\pi_{n,t}^{m}) = \psi_{m,t}^{0} + \psi_{n,t}^{1} - \zeta \mu_t^0 \mathbb{1}_{\{i \neq j\}} - \zeta \mu_t^1 \tau_{j,i}, \tag{25}$$

where ψ^0 and ψ^1 represent origin and destination - time fixed effects, respectively. This equation illustrates that the logarithm of each bilateral migration probability depends linearly on geographical distance and on the fixed cost of migration, with the coefficients of this relationship corresponding to the composite parameters $\zeta \mu_t^0$ and $\zeta \mu_t^1$.

We estimate (25) using data on lifetime migration from the 1940 historical Census and the 1990 IPUMS. In particular, we focus on individuals of age 20-59, for which we observe both the state of birth and either the commuting zone (for the 1940 sample) or the PUMA[33] (for the 1990 sample) of residence. Under the assumption that out-of-state migration probabilities are equal for all commuting zones within each state, we can construct full matrices of bilateral migration flows for both samples. We then use these matrices to estimate the gravity equation in (25) by OLS, separately for the 1940 and 1990 samples. The results are displayed in Table 4. Our estimates indicate that migration costs decline slightly over time but remain in the same order of magnitude. In the calibrated model, we use the estimates in the 1940 column for all periods up and including 1950, and the estimates in the 1990 column for the following periods.[34]

[33] We map PUMAs to commuting zones using a crosswalk based on the intersecting areas.

[34] We set observations where $\pi_{n,t}^{m} = 0$ equal to the minimum positive value in the sample.

Table 4: **Gravity equations for migration flows**

	$\log(\pi^{m}_{n,t})$	
	1940	1990
Fixed cost of migration ($\zeta\mu^0_t$)	-3.257***	-2.867***
	(.067)	(.084)
Geo distance, 1000 km ($\zeta\mu^1_t$)	-1.916***	-1.473***
	(.006)	(.007)
origin CZ FE	yes	yes
destination CZ FE	yes	yes
# Obs.	180,625	180,625
R^2	0.65	0.51

Notes: ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

Knowledge flows In addition to a gravity representation for migration flows, our model also delivers a gravity representation for knowledge flows, that we can estimate using patent citation data across cities and technological fields. Recall from Section 4.1.2 that the effectiveness of knowledge diffusion between each combination of origin and destination location-sectors is limited by a transmission cost $d^{n,s}_{m,r,t}$. We assume this transmission cost to have the following exponential form:[35]

$$d^{n,s}_{m,r,t} = e^{\nu^0_{n,s,t} + \nu^1_t \mathbb{1}_{\{n \neq m\}} + \nu^2_t \tau_{n,m} + \nu^D_{n,m,t} + \nu^S_{s,r,t}}, \tag{26}$$

where $\tau_{n,m}$ represents geographical distance in km and ν^1_t is a fixed cost of drawing ideas outside of an individual's own location, that can be interpreted as the cost of not having access to serendipitous interactions. The expression also includes origin-destination Census Division fixed effects ($\nu^D_{n,m,t}$), to account for geographical non-linearities in the frictions to knowledge diffusion,[36] as well as origin-destination technological sector fixed effects ($\nu^S_{s,r,t}$) that flexibly capture frictions in knowledge diffusion across and within sectors. Finally, we include a destination fixed effect ($\nu^0_{n,s,t}$) that we will set so that average transmission costs have a

[35]The assumption that diffusion frictions decay exponentially with distance is also used by Desmet and Rossi-Hansberg, 2014.

[36]For example, it is reasonable to assume that, despite their exceptional geographical distance, knowledge transmission links between New England and Southern California can be stronger than links between other areas of the country.

constant average across locations and sectors.[37] This normalization is inconsequential for our purposes, since it does not rule out the possibility of systematic differences across receiving location-sectors in their ability to acquire external ideas for innovation, but rather it bundles those differences with the structural error term $\epsilon_{n,s,t}$. As in the expression for migration costs, we allow knowledge transmission costs to depend on time.

Using the functional form for $d^{n,s}_{m,r,t}$ in combination with the expression for the bilateral probability of acquiring ideas (15), we derive a linear gravity equation for citation flows for any idea origin (m, r) - destination (n, s) pair:

$$\log\left(\eta^{n,s}_{m,r,t}\right) = \phi^0_{m,r,t} + \phi^1_{n,s,t} - \theta\nu^1_t \mathbb{1}_{\{i\neq j\}} - \theta\nu^2_t \tau_{n,s} - \theta\nu^D_{n,m,t} - \theta\nu^S_{s,r,t} \tag{27}$$

where ϕ^0 and ϕ^1 represent idea origin and idea destination - time fixed effects, respectively. This equation illustrates that the logarithm of each bilateral citation probability depends linearly on geographical distance, on the of fixed cost of migration, and on the set of fixed effects, with the coefficients of this relationship corresponding to the composite parameters $\theta\nu^1_t$, $\theta\nu^2_t$, $\theta\nu^D_{n,m,t}$, and $\theta\nu^S_{s,r,t}$.

We estimate (27) by OLS using data on bilateral citation probabilities for each combination of origin and destination location-sectors. Since citations in patents are rare before 1950, we estimate (27) separately for all patents up to 1959, and then for the periods 1960-1979, 1980-1999, and 2000-onward.[38]

Table 5 reports the OLS estimates of (27) for the three time periods. Although the effect of distance appears to be stronger in later periods compared to 1970, the estimates are not directly comparable because of the inclusion of origin-destination Census division fixed effects, that capture a significant amount of the variation in geographical distance. The heatmaps of Figure 16 display a graphical illustration of the estimated bilateral fixed effects across each pair of sectors in the four sample periods.[39]

5.2 Fertility, productivity, and amenities

The second step of the quantification of the model requires to back up the shape parameters of the Fréchet distributions of utility draws, ζ, and productivity, θ, as well as the variables that control the evolution of population over the period covered by the quantitative analysis. The evolution of population is determined by aggregate fertility f_t and by the transition probab-

[37]Specifically, we impose $\sum_{m,r}\left(\frac{1}{d^{n,s}_{m,r,t}}\right)^{\theta} = 1$ for all $(n, s) \in N \times S$, $(m, r) \in N \times S$, and all t

[38]We set observations where $\eta^{n,s}_{m,r,t} = 0$ equal to the minimum positive value in the sample.

[39]Note that the inclusion of origin and destination fixed effects, $\phi^0_{m,r,t}$ and $\phi^1_{n,s,t}$, implies that only *relative* values of $\nu^D_{n,m,t}$ and $\nu^S_{s,r,t}$ are identified in (27). Since the expression for transmission costs includes a term $\nu^0_{n,s,t}$, that will be normalized to a constant average across all sector-city pairs, these relative values are sufficient for our quantification.

Table 5: **Gravity equations for knowledge flows**

	$\log\left(\eta^{n,s}_{m,r,t}\right)$			
	pre-1950	1970	1990	2010
Fixed cost $(\theta\nu^1_t)$	-0.443***	-0.786***	-1.668***	-1.708***
	(.001)	(.002)	(.003)	(.005)
Geo distance, 1000 km $(\theta\nu^2_t)$	-.016***	-.027***	-.041***	-.033***
	(.001)	(.001)	(.001)	(.001)
Class-CZ-cited FE	yes	yes	yes	yes
Class-CZ-citing FE	yes	yes	yes	yes
Division-division FE	yes	yes	yes	yes
Class-class FE	Heatmap	Heatmap	Heatmap	Heatmap
# Obs.	72,250,000	72,250,000	72,250,000	72,250,000
R^2	0.10	0.13	0.20	0.20

Notes: Gravity equations estimated with the 20 main IPC classes. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

ilities $\pi^m_{n,s,t}$. These probabilities only depend on the path of $\lambda_{n,s,t}$, the value of time-invariant residential amenities u_n, as well as the value of Fréchet shape parameters ζ and θ.

Fertility Consider first the fertility shocks f_t. These shocks can be backed up sequentially using aggregate data on population. Specifically, given population by age in period $t-1$, fertility at time t can be written as the following identity:

$$f_t \equiv \frac{L_t - L_{t-1} + L^o_{t-1}}{L^k_{t-1}}, \tag{28}$$

where L_t is total population, which is observed, and L_{t-1}, L^o_{t-1}, and L^k_{t-1} denote total, old-period, and childhood-period population in the previous period.

Productivity distribution Consider now the scale parameters of the productivity distribution of each location-sector, $\lambda_{n,s,t}$. These objects are at the core of our quantitative analysis: Higher values of $\lambda_{n,s,t}$ imply higher local income, higher ability to attract population (due to the migration probabilities (22)), and higher potential to innovate and grow more in the future (due to the law of motion (14)).

In this step of the quantification, we postulate a direct mapping between the relative stock of patents in a given location-sector and the relative value of $\lambda_{n,s,t}$. Specifically, we assume that, at any point in time, $\lambda_{n,s,t}$ is equal to a geometric function of current and past patenting:

$$\lambda_{n,s,t} = G_t \times \left[(Pat_{n,s,t})^{\gamma} (Pat_{n,s,t-1})^{1-\gamma} \right]^{\sigma}, \tag{29}$$

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.[40] The parameter γ controls the weight of current patenting on the local stock of knowledge. We set this weight equal to 0.5. Note that none of the parameters appearing in Equation (29) is sector specific. In principle, richer functional forms and parametrizations of (29) are possible. However, we find that this parsimonious functional form provides a good fit.

The parameter σ represents the elasticity of $\lambda_{n,s,t}$ to the observed patenting stock. The purpose of this elasticity is to convert the variation in the local stock of patenting into meaningful variation in average productivity across cities. The identification of σ is complicated by the fact that average local productivity is proportional to $\lambda_{n,s,t}^{\frac{1}{\theta}}$, and θ has not yet been estimated.

However, using (29) in combination with the transition probabilities (22), it is immediate to see that predicting the evolution of population only requires knowledge of $\lambda_{n,s,t}$ and θ via the following composite variables:

$$\tilde{\lambda}_{n,s,t} = \lambda_{n,s,t}^{\frac{1}{\sigma}}$$
$$\tilde{\theta} = \frac{\theta}{\sigma},$$

The first composite term, $\tilde{\lambda}_{n,s,t}$, is observed, via (29), up to the proportionality constant $\tilde{G}_t = G_t^{\frac{1}{\sigma}}$. We calibrate this constant for each t to induce an aggregate growth in income per capita of 2% per year. As for the second composite term, $\tilde{\theta}$, we calibrate it at this stage to match the standard deviation of income per capita across cities in 1990, which is equal to 0.18 in our sample of 425 cities. We choose units of the final good so that the geometric average of $\tilde{\lambda}_{n,s}$ in the initial steady state is equal to one.

Amenities and preference draws To calibrate local amenities u_n and the shape parameter of the distribution of utility draws ζ, we proceed as follows. First, given values for ζ, $\tilde{\theta}$, and $\tilde{\lambda}_{n,s,t}$, we calibrate local amenities to exactly match population by city in the initial steady state (1890).[41] The value of ζ is calibrated to match a ratio of lifetime migration over total population in 1990 of 82.3%. The identification of ζ using total migration probabilities is

[40] We add one to patenting in each sector-city pair to assign a meaningful value to cases in which patenting is zero.

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

simple: The higher the value of ζ, the lower the dispersion of utility draws among potential movers. Hence, a higher value of ζ will induce lower aggregate gross migration flows.[42]

Discussion This procedure delivers values for $\tilde{\theta}$ and ζ of 8.75 and 6.3, respectively. Panels (a) and (b) of Figure 17 in Appendix show graphically how matching the dispersion of income across cities and the ratio of lifetime movers to total population identifies these parameters uniquely. Using the moving costs $\mu^m_{n,t}$, the estimated parameters $\tilde{\theta}$ and ζ, and the aggregate shocks f_t and $\tilde{G}_t$, in combination with the inferred amenities u_n and composite productivity levels $\tilde{\lambda}_{n,s,t}$, we can simulate the model forward starting from the initial steady state (1890) and predict the evolution of local population in all the following periods until 2010. We will present and discuss these results in Section 6.

There are three key aspects of this calibration strategy that are worth further discussion. First, the mapping of $\lambda_{n,s,t}$ to the stock of patenting 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 the result of a range of theoretical mechanisms (e.g., sorting, variety, local learning productivity spillovers, higher availability of productive inputs, etc...). While our 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 our model is that the resulting relationship between population and income per capita is empirically accurate. Figure 6 shows a binscatter of the relationship between log-population and log-income in 1990, both in the model and in the data. Although this correlation is not targeted, the model captures it closely.[43]

Second, our choice of limiting the stock of local patents in (29) to include grants in the current and in the past period can be motivated by postulating that ideas created more than 40 years earlier become commonly known and do not contribute to explaining variation in the stock of knowledge across cities, and would, as such, be captured by the constant G_t.[44]

Third, in quantifying the model we assume that residential amenities are time-invariant. This assumption is crucial for the identification of the shape parameter ζ but comes at the cost of not matching population by city exactly outside the steady state. However, for given values of the structural parameters, allowing for time-varying residential amenities would be an immediate extension of the model.[45] As we show in Section 6, even in the absence of time-

[42] This identification of the dispersion of idiosyncratic preference draws follows the same intuition as in Peters (2019).

[43] The slope of the regression line is 0.12 both in the data and in the model.

[44] Notice also that, since the quantitative analysis starts in the period 1880-1899, allowing for multiple time periods in the aggregator (29) would induce an unavoidable truncation of the measure in the early periods.

[45] A further extension would be to allow for amenities that combine endogenous and exogenous components. A simple formulation would impose $u_{n,t} = v_{n,t} \times L^{\omega}_{n,t}$, where $v_{n,t}$ is the exogenous component, and ω is the

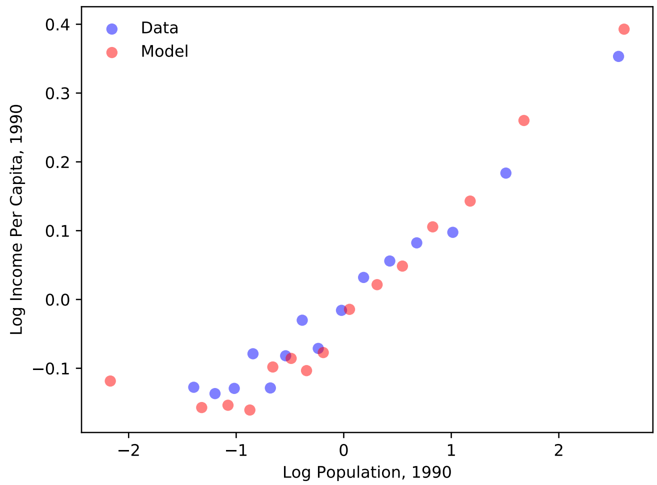


Figure 6: **Population and Income: Model vs. Data**

Notes: The binscatter plot compares the relationship between log population and log income in 1990 in the data and in the model.

varying amenities, the model goes a long way in fitting population growth by city over the last century.

5.3 Technology cycles and structural residuals

Up to this point in the calibration, we have treated $\tilde{\lambda}_{n,s,t}$ as observable objects that can be inferred directly from data on local patenting by sector. In the third step of the quantification, we use these observed values in combination with their model-based law of motion (equation 14) and network of citations (equation 15) to estimate the aggregate shocks that control the technology cycles, $\alpha_{s,t}$, and to derive the structural residual term, $\epsilon_{n,s,t}$. This step also allows to estimate separate values for the structural parameters θ and σ (recall that the steps of Section 5.2 only allowed to identify the composite parameter $\tilde{\theta} = \frac{\theta}{\sigma}$).

For a given combination of values of σ, θ, transmission costs $d^{n,s}_{m,r,t}$, and aggregate technology shocks $\alpha_{s,t}$, the equilibrium conditions allow to back up, via equation (15), the citation probabilities for each origin and destination sector-city pair. This implicitly defines a network of citations akin to the one used for the empirical results of Section 3. In particular, within this network of citations we can define the PageRank of each of the sectors in S by applying the formula in equation (1).

From the citation probabilities (15) it is evident that relative values of the technology shocks $\alpha_{s,t}$ control the centrality of each sector in the overall network of citations. Hence, we can set

elasticity of amenities to local population, that can account for congestion forces in the case $\omega < 0$.

the relative values of $\alpha_{s,t}$ within each period to exactly replicate the empirical centrality of sectors over time. When a sector gains importance in the technological landscape, its empirical PageRank will increase and, in the context of the model, this will be reflected in higher relative shocks $\alpha_{s,t}$.

A caveat of this procedure is that it can only successfully identify *relative* values for $\alpha_{s,t}$. This follows directly from the fact that a proportional shift to the value of *all* $\alpha_{s,t}$'s would induce an increase in aggregate growth but would leave the citation probabilities unaltered, and would not have any effect on the PageRank. In order to pin down the magnitude of the technology shocks $\alpha_{s,t}$, we first recognize that, given values of those shocks, structural residuals $\tilde{\epsilon}_{n,s,t}$ can be backed up uniquely by inverting the law of motion for $\lambda_{n,s,t}$ (equation 14). So, any restriction on the average of the structural residuals will directly imply a condition on the average magnitude of the technology shocks. Specifically, we impose structural residuals to have a weighted average of 1 in the aggregate:[46]

$$\mathbb{E}\left[\tilde{\epsilon}_{n,s,t}\right] = 1. \tag{30}$$

This condition requires that aggregate growth in the long-run can only be explained by the endogenous process of knowledge creation and diffusion, while idiosyncratic residuals can only affect the distribution of economic activity across space and sectors. 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, what is their spatial and temporal correlation, and whether they are systematically correlated with the other terms in the law of motion (14). We elaborate more on this point in Section 6.

The last step requires to separate the parameters θ and σ. Given an estimated value for $\tilde{\theta} = \frac{\theta}{\sigma}$, panel (c) of Figure 17 shows that σ is well identified by the share of innovators in the economy. To pin down the value of σ we target a measure of the share of innovators in the US economy in 1990. Formulating a sensible estimate of this number is not a trivial task. We set this share to be equal to 38%, as a compromise between a strict interpretation of innovators as members of the "Creative Class" as defined by Florida (2014), who estimates that the share of individuals in occupations with highly creative content is around 30% of the US labor force in 1990, and a broader definition that includes the share of the US labor force with at least some college or an associate degree, that, according to BLS estimates is around 50% of the labor force in 1992.[47] This leads to an estimate for σ of 0.14, and implies $\theta = 1.22$.

[46] Weights are given by the mass of young agents in a given sector-city pair.

[47] See the Bureau of Labor Statistics report available at https://www.bls.gov/spotlight/2017/educational-attainment-of-the-labor-force/pdf/educational-attainment-of-the-labor-force.pdf

Table 6: **Parameter values and targets**

Assigned Parameters				
Parameter	Value	Target		
β	0.60	Annual discount factor 0.98		
A_y	0.80	Experience premium 1.24		
Calibrated Parameters				
Parameter	Value	Target	Model	Data
ζ	6.3	Gross migr. to pop ratio, 1990	0.82	0.82
θ	1.22	Std. log-income across cities, 1990	0.18	0.18
σ	0.14	Share of innovators, 1990	0.38	0.38

5.4 Taking stock

Table 6 presents a summary of the estimated parameters with the corresponding targets. In the following section, we will assess the ability of the model to account for the variation in population growth across cities over the last century. There are two key aspects that are necessary for the model to generate an accurate description of the data. First, the growth in local productivity $\lambda_{n,s,t}$ induced by the law of motion in (14) should correlate well with the one obtained through the empirical patenting-based measure in (29). Second, the average growth rate of $\lambda_{n,s,t}$ for each sector must correlate well with changes in its centrality.

As for the first point, Figure 18 plots the correlation between the patenting-based measures and the model-based prediction (without structural residuals) of the growth rate of $\lambda_{n,s,t}$ for all the 20 sectors in the model between 1990 and 2010 (the picture for the other decades is similar). The law of motion in (14) has a strong predictive power with respect to changes in $\lambda_{n,s,t}$ inferred by observed local patenting.[48] As for the second point, Figure 19 plots the change in each sector's PageRank and the corresponding average growth rate of $\lambda_{n,s,t}$ for each period between 1910 and 2010. The two measures are strongly correlated in all time periods, possibly with the exception of $t = 1950$, where the correlation appears to be slightly negative mostly for the presence of an outlier.

To summarize, the endogenous mechanism of innovation and knowledge diffusion closely tracks the evolution of local productivity as inferred from local patenting by sector, and, on average, the change in the centrality of each sector is a strong predictor of that sector's productivity growth. These two observations will be key ingredients for the model to generate the results of the following section.

[48] By construction, introducing the structural residual would induce a perfect fit.

6 Quantitative results

In this section, we explore the ability of the model to account for the variation in population growth across US cities in the period 1890-2010 in response to technology cycles. We start by showing that the interaction between changes in the technological environment and the endogenous mechanism of knowledge creation and diffusion closely predicts the long-run growth rate of cities. We then examine the importance of this mechanism in accounting for the two most striking episodes of technological and geographical transformation in the last century: The extraordinary rise of manufacturing-intensive cities in the early decades of the 20th century and their later decline as cities specialized in knowledge-intensive sectors, such as IT and pharmaceuticals, gained prominence. These two episodes illustrate that sectoral shocks, interacting with frictions to knowledge diffusion, can explain at the same time path dependence and reversal of fortune in the growth trajectory of cities. The model allows to separate the contribution of endogenous innovation and knowledge diffusion to the overall effect from the one of structural residuals. This exercise highlights in which cases residual factors amplify or dampen the effect of technology cycles on city growth. Finally, we use our model to predict how the economic geography of the US is likely to evolve in the coming decades in response to three plausible scenarios on changes in the technological environment.

6.1 Model Fit

Figure 7 illustrates the performance of the model in explaining the variation in the growth of cities between the initial steady state (1890) and the last period of the analysis (2010). The graph plots the 1890-2010 difference in the log-population of our sample of commuting zones in the data (horizontal axis) and in various versions of our model (vertical axis). The black lines correspond to the economy in the absence of any shocks (which will remain in the steady state by construction), and the 45-degree line (perfect fit).[49]

The blue line and circles correspond to the complete model with both aggregate technology cycles, $\alpha_{s,t}$, and structural residuals $\epsilon_{n,s,t}$. The size of each circle is proportional to initial population, that we also use to weight the regression line. The slope of the line is 0.50 and the weighted correlation is 64%. This correlation can be interpreted as the predictive power of the path of local patenting by sector (encapsulated in the values of $\lambda_{n,s,t}$ via equation (29)) on population growth across cities.

The red line and circles correspond to the case where we feed the model with the full set of shocks in the first period after the steady state (1910), after which we only provide aggregate technology cycles $\alpha_{s,t}$. In this case, we fix the structural residuals to $\epsilon_{n,s,t} = 1$ for all sector-city

[49]Recall that the model does not replicate the data exactly because residential amenities are fixed at the initial steady state.

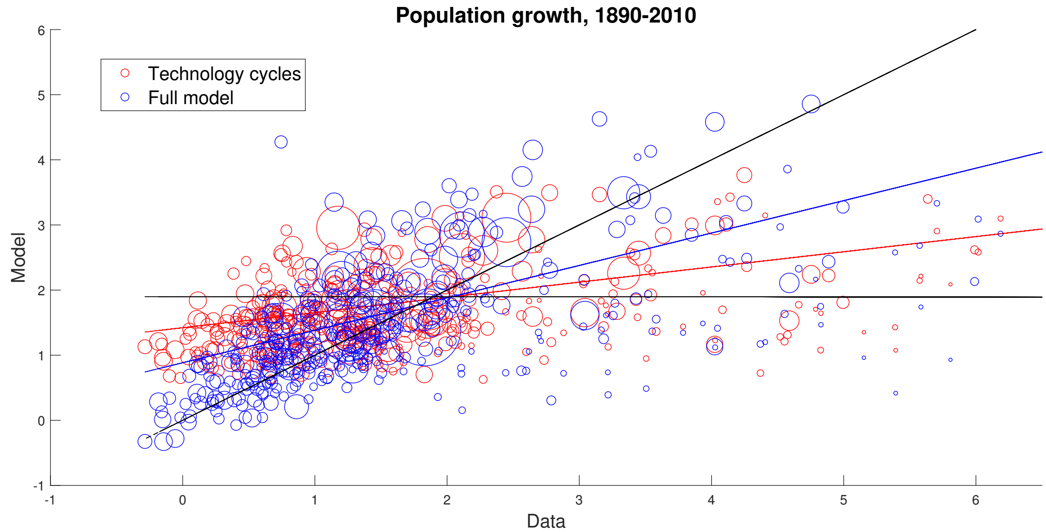


Figure 7: **Population: Model vs. Data**

Notes: The graph plots the 1890-2010 difference in the log-population of our sample of commuting zones in the data (horizontal axis) and and the model (vertical axis). The horizontal black line corresponds to the economy in the absence of any shocks (which will remain in the steady state by construction), while the dotted line corresponds to the 45-degree line (perfect fit). The blue line shows the population obtained in the full model (with structural residuals and technology cycles), whereas the red line in a model in which the structural shocks are kept constant.

pairs, and we let the path for $\lambda_{n,s,t}$ to be determined by the endogenous law of motion in (14), which reflects the interaction between the state variables in 1910 and the gradual unfolding of the technology cycles over time. The predictive power of the model declines but remains significant: The slope of the line is 0.23, while the weighted correlation is 45%. This correlation can be interpreted as the contribution of the technology cycles, interacting with the endogenous process of innovation and knowledge diffusion, in explaining the variation in population growth over the last century.

Figure 20 shows the predictive power of the model over single periods between 1950 and 2010. Specifically, for each of the displayed model periods T, we first compute the model with the full set of shocks until the previous period $T-1$, and then look at city population growth over the specified period by providing only fertility shocks (black line), by adding technology cycles shocks $\alpha_{s,t}$ (red line), and by feeding the full set of shocks (blue line). Contrary to the case of Figure 7, since the model is now not in steady state, cities preserve some persistence in their dynamics, which explains the positive slope of the black line. However, the red line is consistently above the black one, and, depending on the period considered, it closes the gap between the slopes of the black and the blue lines by 18% to 54%.

We now look specifically at how the calibrated model can account for the two most striking episodes of transformation of the economic geography of the US in the last century: The extraordinary rise of manufacturing-intensive cities in the early 20th century, and followed by their decline and contemporaneous rise of knowledge-intensive urban areas. We focus on the exemplary cases of cities that have been notoriously exposed to those transformations (some of which introduced in Section 3) and, for each of those cases, we analyze the role of the endogenous mechanism of knowledge creation and transmission in driving their growth trajectory, and the amplifying or dampening effect of the residual factors.

6.2 The rise of manufacturing-intensive cities

In Figure 8 we plot the evolution of population in the first half of the 20th century for an illustrative subset of cities in the United States. The experiment is as follows. For the two initial periods, we simulate the model with the full set of shocks. Starting from $t = 1930$, we follow the evolution of each city under three different scenarios. First, we only provide the model with fertility shocks. In this case, since the technological frontier is constant, the economy is only driven to converge towards a new steady state. We take this case as our baseline and normalize it to zero in every period. Second, we compute the model by feeding both shocks to fertility and to technology cycles $\alpha_{s,t}$, and plot in red population in log-deviations from the baseline. Third, we compute the model by feeding the full set of shocks (fertility, technology cycles, and structural residuals), and plot in blue the resulting log-deviations from the baseline.

The plots show that both Detroit and Cleveland, the largest urban centers of what would later be known as the Rust Belt not only were favorably exposed to the technology cycle in the early part of the century, but were also subject to forces outside of our model that significantly contributed to drive their growth. Consider the red line first. According to the model, the interaction between their exposure in 1910 with the technology shocks $\alpha_{s,t}$ in 1930 and 1950 induced an increase in population of 4% in Detroit and 8.5% in Cleveland, compared to a scenario with no technology shocks. Consider now the blue line, that plots the evolution of population in the model with the full set of shocks. These shocks include the structural residuals, that capture the combination between factors that evolve endogenously to the cities responses to the technology cycle (such as the opening of new factories and the induced local externalities) and exogenous forces that affect population growth by boosting local innovation and productivity (such as investment in infrastructure uncorrelated with other local disturbances). Although a systematic exploration of those factors goes beyond the scope of this paper, our model allows us to infer, for each episode, whether those forces amplified or dampened on net the effect of the technology cycle. In this scenario, Detroit and Cleveland are 38% and 29% larger in 1950 compared to the baseline.

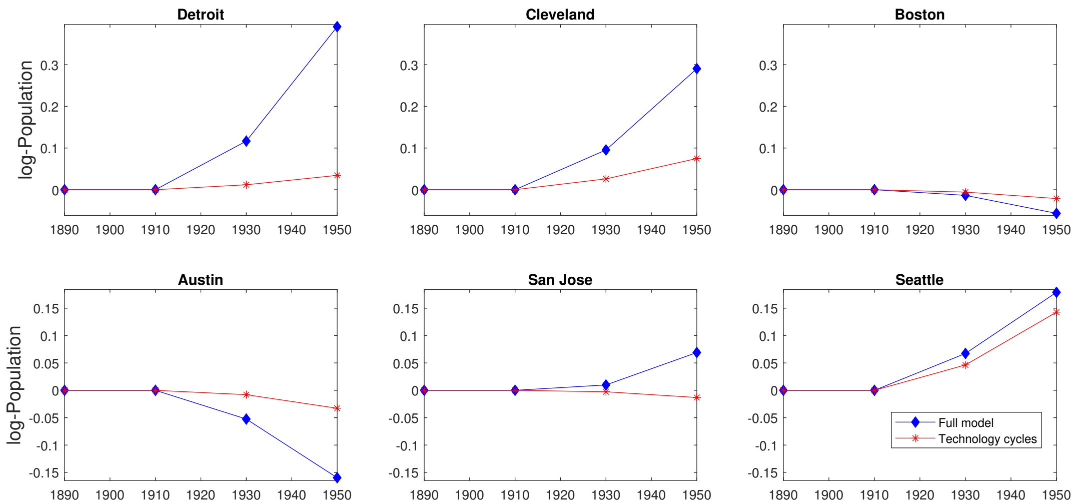


Figure 8: **Growth Decomposition: 1890-1950**

Notes: The plots show the growth trajectories of Detroit, Cleveland, Boston, Austin, Silicon Valley, and Seattle as deviations from a model without shocks. The blue line shows the growth trajectory of these six cities in the full model (with structural residuals and technology shocks). The red line represents the growth obtained in a model in which the structural shocks are kept constant.

During the same period, cities did not uniformly benefit from this transformation in the technological landscape. The commuting zone of Austin, as we already saw in Section 3, lost population. The red line shows that part of this decline was due to an unfavorable exposure to the technology cycle. However, the blue line shows that external forces were even more penalizing: The Texas Oil Boom created opportunities in the other areas of Texas, further depressing population growth in Austin. In this period, Boston and San Jose are both negatively (albeit weakly) exposed to the technology cycle, while residual factors contribute negatively to the growth of Boston and positively to the one of San Jose. Part of these factors can be explained as a general expansion of the West at the expense of the North-East.

Finally, the commuting zone of Seattle is positively affected by the technology cycle, as reflected by the rising importance of technologies related to shipbuilding first and aviation later between WWI and WWII. However, in this case the model records a smaller amplification from residual factors than in the cases of Detroit and Cleveland. The history of Seattle in the interwar period can be informative of the factors that prevented a larger amplification: The Seattle economy was severely hit by the Great Depression, and the events following the Maritime Strike of 1934 led to the relocation of major shipping companies to the port of Los Angeles.

Panel (a) of Figure 21 displays the same experiment for an extended set of cities with comparable characteristics.

6.3 The emergence of the US knowledge hubs

The experiences of Detroit and Cleveland in the first half of the 20th century were not isolated cases. Several other cities that specialized in heavy manufacturing and were mostly concentrated in what is now known as the Rust Belt witnessed exceptional growth in population. Our model suggests that part of this growth can be explained by a pre-existing availability of local ideas in fields that were complementary to the prevailing technology cycles. We now explore whether the same factors that led to the remarkable growth of manufacturing-intensive cities contributed to their later decline to the benefit of newly emerging knowledge hubs specialized in fields such as information technology and pharmaceuticals.

Figure 9 shows the results of an experiment analogous to the one in Figure 8, in which we provide the model with the full set of shocks until 1970, and then analyze the evolution of population until 2010 with only fertility shocks (baseline), adding technology cycles $\alpha_{s,t}$ (red line), and including also the structural residuals (blue line). To provide a sense of the magnitude of the technological transformation over this period, the centrality of field B4 ("Transporting") fell from 0.11 in 1950 to 0.06 in 2010, and the one of field C2 ("Metallurgy") declined from its 1970 peak of 0.032 to 0.021 in 2010. At the same time, the centrality of fields G1 ("Physics") and H1 ("Electricity") increased from 0.10 and 0.09, respectively, to 0.19 and 0.16.

As a result of this transformation, cities that were heavily exposed to knowledge in declining fields experienced population declines. The model suggests that, as a direct effect of the technology cycle (red line), population in Detroit and Cleveland declined by 3% and 5%, respectively, compared to the baseline. At the same time, the residual factors in the evolution of productivity led to an additional decline of 3% in Detroit and, more dramatically, almost 20% in Cleveland (blue line). The reason why the structural residual imposes a more severe loss in the commuting zone of Cleveland compared to Detroit is interesting and worth further investigation. One candidate explanation is that the policy response to the decline of the automotive industry compressed the amplification mechanisms in Detroit but not in Cleveland.

Throughout the same decade, a handful of cities emerged as modern leading technological hubs. The commuting zones of Austin and San Jose are archetypal examples of this expansion. Our model suggests that population in Austin and San Jose increased, relative to the baseline, by 15% and 22%, respectively, as a direct effect of the technology cycle interacted with their local characteristics in 1970. However, the amplification effect coming from the structural residuals is significantly larger for Austin than it is for San Jose. Even more dramatically, in the commuting zone of Boston, that our model also predicts having a positive exposure to the technology cycle over this period, structural residuals work in the opposite direction and prompt a *decline* (albeit small) compared to the baseline. Why does the contribution of structural residuals vary so much among these cases? Again, while an exact answer to this

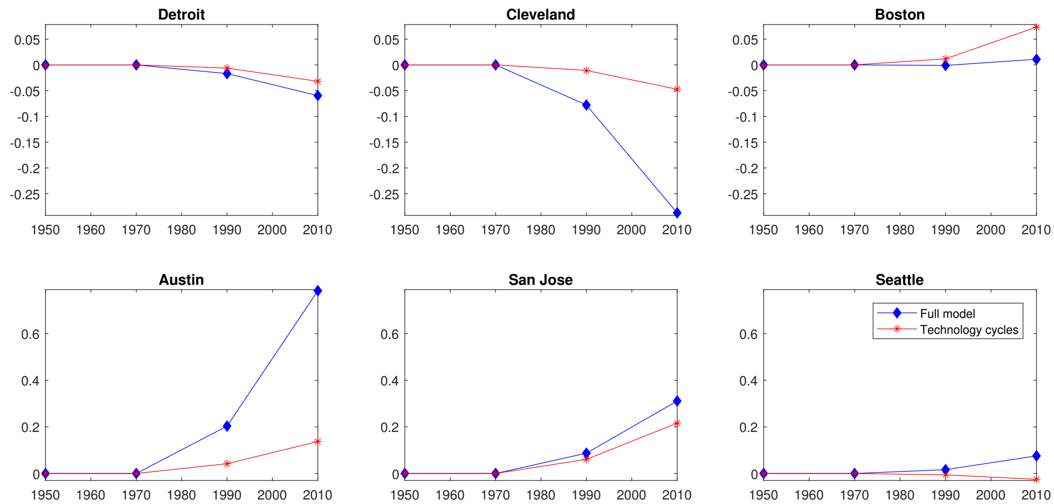


Figure 9: **Growth Decomposition: 1950-2010**

Notes: The plots show the growth trajectories of Detroit, Cleveland, Boston, Austin, Silicon Valley, and Seattle as deviations from a model without shocks. The blue line shows the growth trajectory of these six cities in the full model (with structural and technology shocks). The red line represents the growth obtained in a model in which the structural shocks are kept constant.

question is beyond our scope, a candidate explanation can be found in the different constraints imposed by local taxation and land-use regulation that characterize those commuting zones. This hypothesis is in line with the evidence in recent studies, such as Glaeser and Ward (2009) and Hsieh and Moretti (2019), that document the consequences of land-use restrictions on the misallocation of people across US cities.

Finally, the commuting zone of Seattle appears to be weakly but negatively affected by the technology cycle, and to receive instead a positive contribution from structural residuals. This finding is in line with the fact that most of the recent extraordinary growth in the IT sector in Seattle is a consequence of local events that happened after 1970 (such as the relocation of Microsoft in to Bellevue in 1979 and the establishment of Amazon in 1994). In fact, it is worth mentioning that we find a positive direct effect of the technology cycle if we consider the model with the full set of shocks until 1990, and only provide the technology shocks $\alpha_{s,t}$ in 2010.

It is interesting to note that the effect of technology cycles on Boston's and Seattle's population appears to be smaller in magnitude compared to the effect on Austin and San Jose. The key for this finding lies in the different degrees of diversification and specialization found in those cities. Boston and Seattle are significantly more diversified, with patents in classes G1 and H1 ("Physics" and "Electricity") making up 46% and 44% of their overall innovation portfolio in 1990, while the corresponding figures for Austin and San Jose are 78% and 74%, respectively.

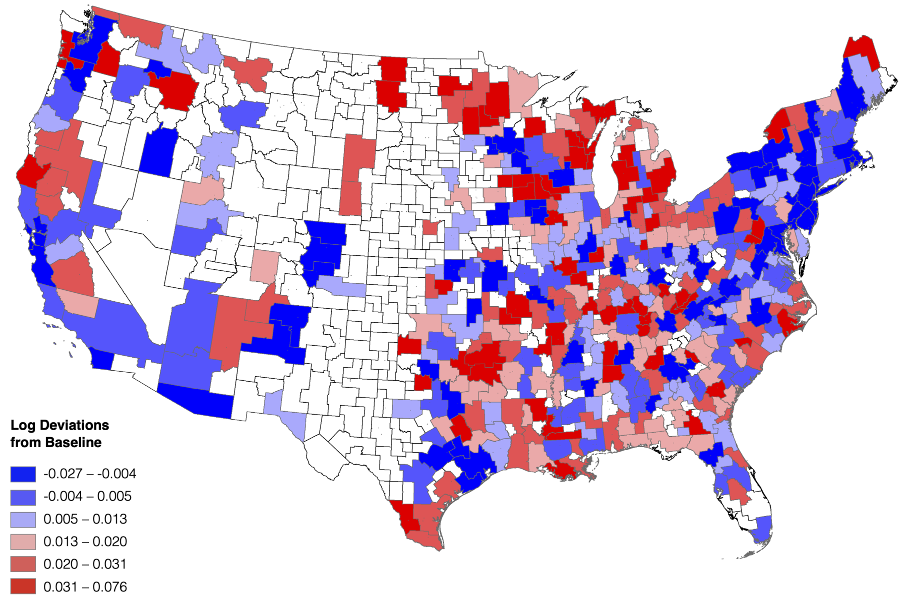


Figure 10: **Counterfactual: Autonomous Vehicles**

Notes: The map shows the changes in population over the next four decades if Transportation went back to its 1950 levels in terms of importance relative to a status-quo scenario. The log deviations are divided into six quantiles.

Panel (b) of Figure 21 proposes the same experiment for an extended set of comparable cities.

6.4 Three scenarios for the future

The quantitative model can also be used to predict the evolution of the US economic geography in the coming decades in response to transformations in the technological landscape. In this Section, we propose three plausible scenarios for future technology cycles and look at which cities and regions will be most positively and negatively affected by these changes. Specifically, we project population flows across cities until 2050 under different assumptions on how the centrality of fields of knowledge will evolve in the near future, and compare the outcome with a baseline in which centrality for all sectors is constant to its 2010 values.

In the first scenario, we assume that the centrality of sector B4 ("Transporting") experiences a comeback to its 1950 peak, as new advances in transit technologies and autonomous vehicles induce innovation in transportation to return to a pivotal role.[50] The map in Figure 10 visually illustrates the results. Commuting zones in red (blue) experience a net gain (loss) of population compared to the baseline. The results indicate that, given the state variables observed in the

[50] We rescale the centrality of the other sectors uniformly so that the total PageRank adds up to one.

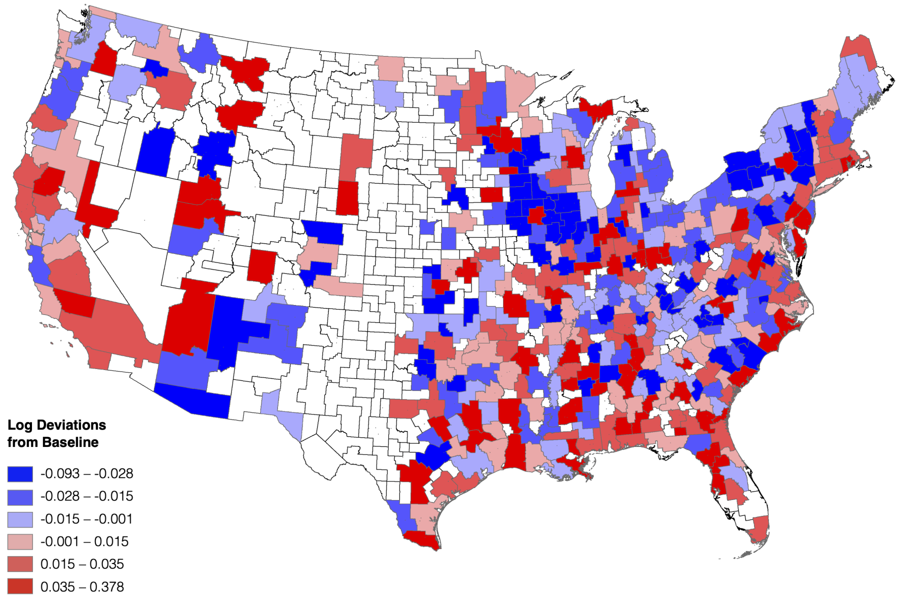


Figure 11: **Counterfactual: Pharmaceuticals and Biotech**

Notes: The map shows the changes in population over the next four decades if innovations related to Pharmaceuticals and Biotech experienced a surge in importance relative to a status-quo scenario. The log deviations are divided into six quantiles.

last period of our sample ($t = 2010$), cities in the Rust Belt are still the areas that are best prepared to take advantage from this transformation. Detroit would experience an increase in population of 3.5% compared to the baseline. Cities in the North-East and in California, as well as other technology hubs such as Austin, Denver-Boulder and the North Carolina Research Triangle would instead experience a relative loss in population.

In the second scenario, we simulate a rise of class A4, that is focused on technologies related to pharmaceuticals and biotech, as the most central knowledge field,[51] possibly in response to new challenges in global health such as the COVID-19 pandemic. The results are depicted in the map of Figure 11. The counterfactual suggests that major commuting zones in the North-East (including Boston, Providence, and New York City) and in California (including Los Angeles, San Diego, and San Francisco) would experience a net inflow of population, at the expense of specialized IT clusters such as San Jose (-2.5%) and Austin (-6.1%).

In the third scenario, we assume that class A1, that includes technologies related to agriculture and animal husbandry, experiences a comeback in its centrality to its peak in 1890. 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

[51] Specifically, we assign to class A4 the PageRank of the most central class in 2010 (G1), and rescale all the other values so that the PageRank adds up to one.

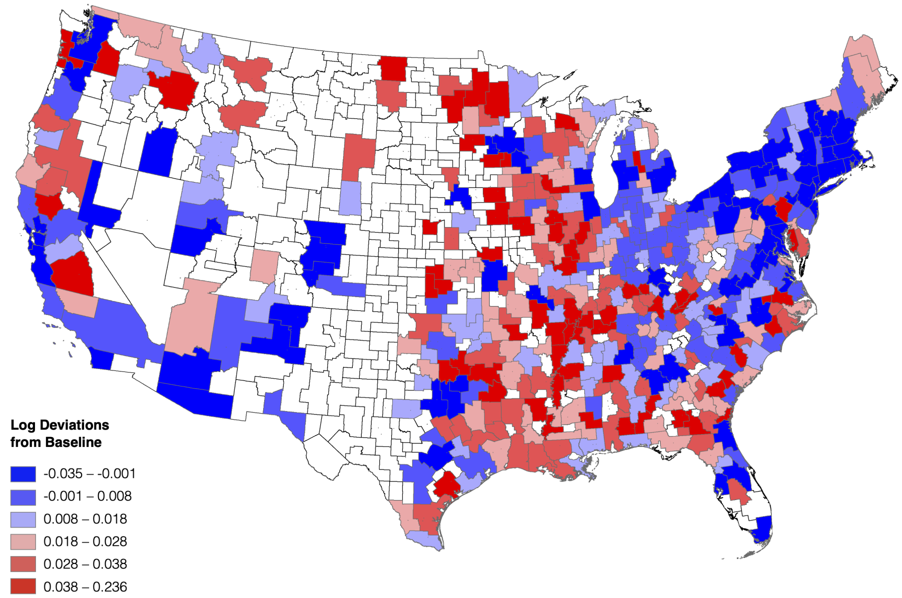


Figure 12: **Counterfactual: Agriculture**

Notes: The map shows the changes in population over the next four decades if Agriculture experienced a surge in importance relative to a status-quo scenario. The log deviations are divided into six quantiles.

as climate change. Results are in the map of Figure 12. Under this scenario, the economic geography of the US experiences a pronounced shift away from the coasts and the Rust Belt, towards the Central States. Among the major commuting zones, Des Moines (IA) receives the highest net gain (+12.8%). This scenario would represent a significant convergence force in relative population across commuting zones: A regression of log-population in 2010 with the log-deviation from the baseline in 2050 delivers a coefficient of -0.9%, implying that population would mostly relocate away from larger commuting zones and towards low-population ones.

7 Conclusion

The economic geography of countries is in constant evolution. Some cities remain large and important throughout long time spans, while some others experience sharp episodes of growth and decline. We explore and quantify the hypothesis that the pattern of specialization of cities across different technologies, coupled with the constantly evolving technological landscape, affects the trajectory of city growth. Our theory rests on the idea that there are geographical and sectoral frictions in the diffusion of ideas. This implies that upon a change in the technological landscape, some cities are better positioned to reap the benefits of new innovation possibilities.

Thus, these frictions in idea diffusion make city growth trajectories sensitive to the rise and fall of the their fields of specialization. We also argue that diversification can make cities more resilient by insulating them from this source of volatility.

For the empirical analysis, we exploit a novel dataset of historical US patents that covers the period 1836-2010. The information contained in these data allows us to construct a proxy of the network of knowledge using different patent technology fields, and document the changes in the technology landscape over the last 160 years. Moreover, we use the geocoded patents to infer the pattern of city specialization across different fields of knowledge over the same time span.

We provide evidence supportive of our hypotheses using both a reduced-form and a structural approach. We first show in a reduced-form exercise that the technology cycles have an effect on city growth. This effects depends on the pre-existing city pattern of specialization across fields of knowledge. We find that the effect is substantial in magnitude: up to 24.3% of the variation in city growth in our sample can be attributed to differential exposure to the technology cycle across cities. We also document that cities with a more diversified portfolio, tend to experience less volatile growth.

To interpret these reduced-form findings, we develop a parsimonious dynamic spatial-equilibrium model with endogenous innovation and frictional knowledge diffusion across cities and fields of knowledge. The model can be solved in closed form and it implies gravity equations for migration flows and idea diffusion, which facilitate its quantification. We incorporate sectoral technology shocks to match the changes in the network of knowledge that we observed in the data. We find that our proposed mechanism accounts for 45% of the overall variation in city growth in the last century. Moreover, the model features a structural error term in city-sector innovation that encompasses unmodeled city-sector specific shocks (e.g., placed-based policies, endogenous R&D effort by innovators, etc.). Including the structural error the model increases its predictive power substantially, up to 64% of the overall city growth over the same time span.

We then use the calibrated model to analyze some important episodes of transformation of the US economic geography observed in the 20th century. Our calibrated model can account for the rise in manufacturing-intensive cities in the Rust Belt, driven by the increase in centrality of transportation technologies, and the recent emergence of modern knowledge hubs, driven by the increase in the centrality of fields related to physics and electricity. Finally, we use our model to speculate on how the US economic geography will evolve under different technological scenarios, such as a come back of transportation and agriculture and a further rise in the centrality of medical sciences.

In our model, the structural residual contributes significantly to the dynamics of local innovation and to the variation in population growth. Our framework allows us isolate the direct effect of the technology cycle on city growth via innovation and knowledge diffusion, and does not require to make specific assumptions on the nature of this residual. In a current model extension, we partly endogenize this error term by allowing innovators to exert effort to improve their ideas in the spirit of an endogenous growth theory with expanding varieties as in Jones (2005). Other endogenous forces that can enter the residual include congestion and 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 technology cycles is the next step of our agenda.

References

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2019): "Democracy Does Cause Growth," *Journal of Political Economy*, 127, 47–100.

Adão, R., M. Beraja, and N. Pandalai-Nayar (2020): "Technological Transitions with Skill Heterogeneity Across Generations," Tech. rep., National Bureau of Economic Research.

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.

Arcand, J.-L., E. Berkes, and U. Panizza (2015): "Too much finance?" *Journal of Economic Growth*, 20, 105–148.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 58, 277–297.

Arellano, M. and O. Bover (1995): "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68, 29 – 51.

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: https://ssrn.com/abstract=3567425.

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.

Beck, T. and R. Levine (2004): "Stock markets, banks, and growth: Panel evidence," *Journal of Banking & Finance*, 28, 423 – 442.

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.

Blundell, R. and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115 – 143.

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.

Brin, S. and L. Page (1998): "The Anatomy of a Large-Scale Hypertextual Web Search Engine," in *Seventh International World-Wide Web Conference (WWW 1998)*.

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 (2016): "The Global Diffusion of Ideas," Working Paper 21844, National Bureau of Economic Research.

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," Available at: https://econweb.ucsd.edu/~j1clemens/personalResearch.html.

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 (2017): "Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Preindustrial Economy," *The Quarterly Journal of Economics*, 133, 1–70.

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 Gibrats 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. 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.

Florida, R. (2014): *The Rise of the Creative Class–Revisited: Revised and Expanded*, Basic Books.

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.

Glaeser, E. L. and B. A. Ward (2009): "The causes and consequences of land use regulation: Evidence from Greater Boston," *Journal of Urban Economics*, 65, 265–278.

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 (2018): "The making of the modern metropolis: evidence from London," Tech. rep., National Bureau of Economic Research.

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.

Hsieh, C.-T. and E. Moretti (2019): "Housing constraints and spatial misallocation," *American Economic Journal: Macroeconomics*, 11, 1–39.

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*, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1 of *Handbook of Economic Growth*, chap. 16, 1063–1111.

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.

Lind, N. and N. Ramondo (2019): "The Economics of Innovation, Knowledge Diffusion, and Globalization," *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.

Marshall, A. (1890): "Principles of economics Macmillan," *London (8th ed. Published in 1920)*.

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

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," Tech. rep.

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

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.

Windmeijer, F. (2005): "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of Econometrics*, 126, 25 – 51.

A Additional Tables

Table 7: **Summary Statistics**

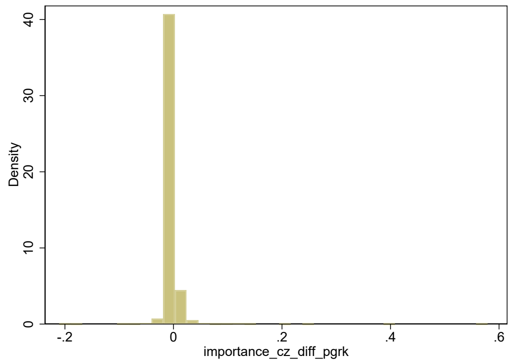
Variable	Obs.	Mean	Std. Dev.	Min	Max
Decade	5,950	N/A	N/A	1880	2010
Population	5,950	349,985.4	858,144.6	62.24	1.79e+07
Log Population	5,950	11.98	1.11	4.13	16.70
Population Growth	5,525	0.13	0.19	-0.31	2.79
Population Growth (wins.)	5,525	0.12	0.16	-0.14	0.85
Tot. Patents (CZ/decade)	5,950	822.80	3,339.41	0	92,309
Log Tot. Pats.	5,928	4.68	1.84	0	11.43
Patents per capita (CZ/decade)	5,950	0.001	0.002	0	0.039
Log Pats. per capita	5,928	-7.30	1.15	-11.95	-3.26
HHI	5,950	0.16	0.10	0	1
Human Capital (rank)	5,928	0.48	0.28	0.001	1
Tech Class Importance	5,950	0.00	0.018	-0.42	0.58
Tech Class Importance (rank)	5,950	0.50	0.29	0.002	1
Specialization	5,950	0.26	0.14	0	1.08
Specialization (rank)	5,950	0.50	0.29	0.002	1

Table 8: **IPC Classes Groups**

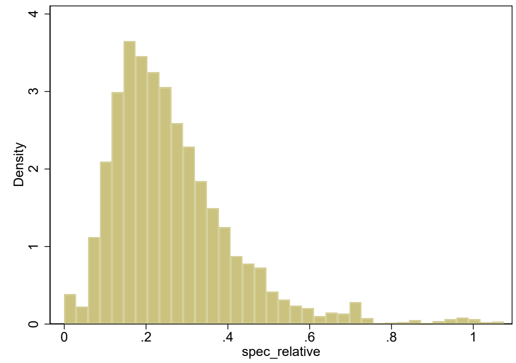
Class ID	Class Group	IPC Class Range	Label
1	A1	A01	Agriculture
2	A2	A21-A24	Foodstuffs; Tobacco
3	A3	A41-A47	Personal or Domestic Articles
4	A4	A61-A99	Health; Life-Saving; Amusement
5	B1	B01-B09	Separating; Mixing
6	B2	B21-B33	Shaping
7	B3	B41-B44	Printing
8	B4	B60-B68	Transporting
-	B5	B81-B99	Microstructural Technology; Nanotechnology
9	C1	C01-C14	Chemistry
10	C2	C21-C30	Metallurgy
-	C3	C40-C99	Combinatorial Technology
11	D1	D01-D07	Textiles or Flexible Materials Not Otherwise Provided For
12	D2	D21-D99	Paper
13	E1	E01-E06	Building
14	E2	E21-E99	Earth or Rock Drilling; Mining
15	F1	F01-F04	Engines or Pumps
16	F2	F15-F17	Engineering in General
17	F3	F21-F28	Lighting; Heating
18	F4	F41-F99	Weapons; Blasting
19	G1	G01-G16	Physics
-	G2	G21-G99	Nuclear Physics; Nuclear Engineering
20	H1	H01-H99	Electricity

Notes: This table provides label and a mapping to the original IPC classes for the class groups used for the quantitative analysis of this paper. Groups B5, C3, and G2 are excluded from our sample since they cover knowledge, such as nuclear physics, that was acquired only later in our sample.

B Additional Figures



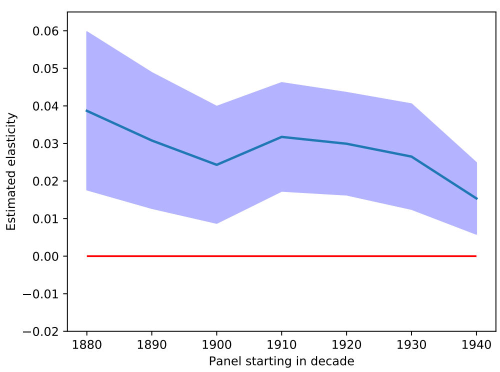
(a) Exposure



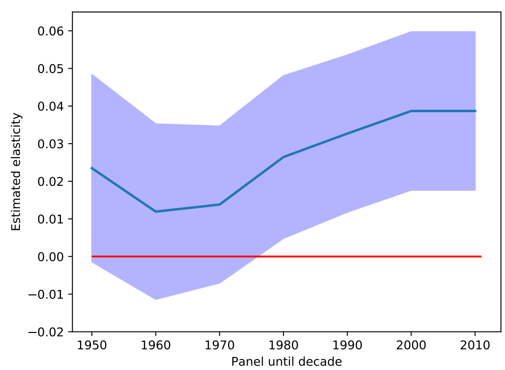
(b) Diversity

Figure 13: **Distribution of Main Variables**

Notes: The two histograms show the distribution of the exposure measure (left) and log diversity (right).



(a)



(b)

Figure 14: **Temporal Robustness**

Notes: Point estimates of the exposure parameter obtained when dropping decades to the right and to the left.

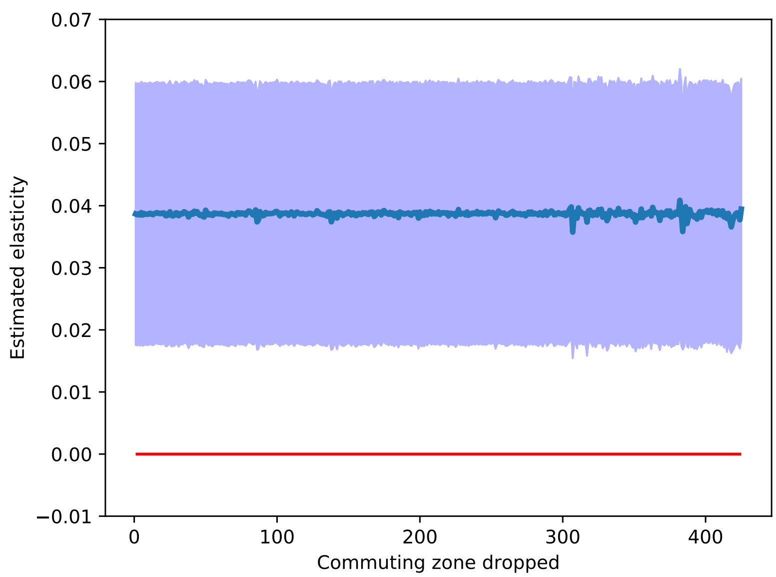
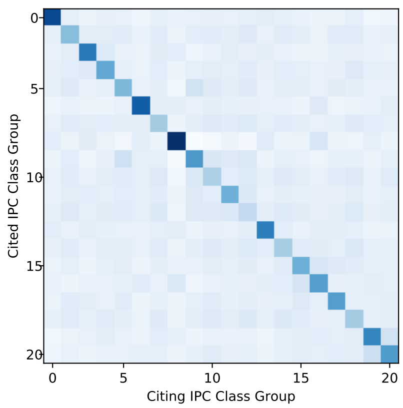
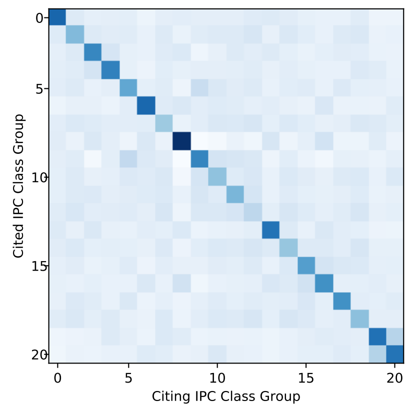


Figure 15: **Spatial Robustness**

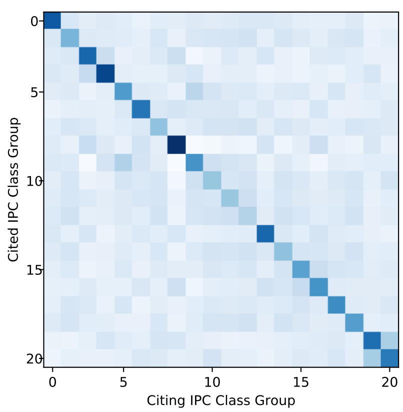
Notes: Point estimates of the exposure parameter obtained when dropping one commuting zone at a time.



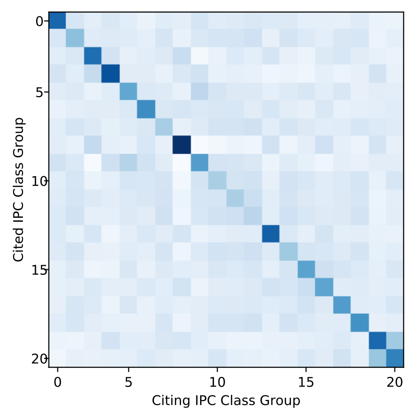
(a) 1950



(b) 1970



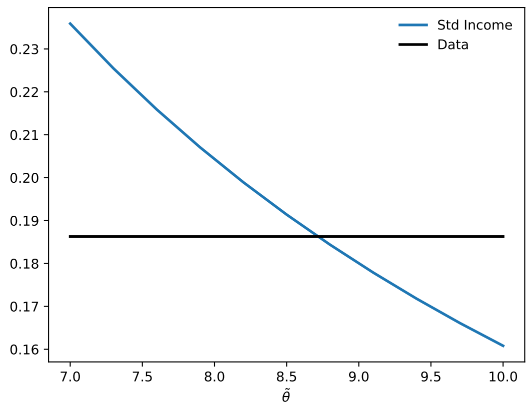
(c) 1990



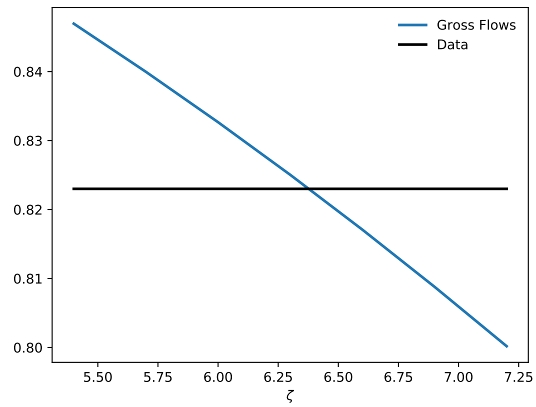
(d) 2010

Figure 16: **Citation Intensity Between Classes**

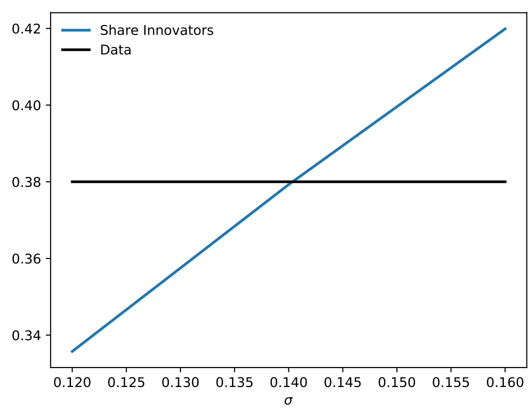
Notes: Heatmap citing/cited IPC classes by decade



(a) Theta Tilde



(b) Zeta



(c) Sigma

Figure 17: **Parameters Identification**

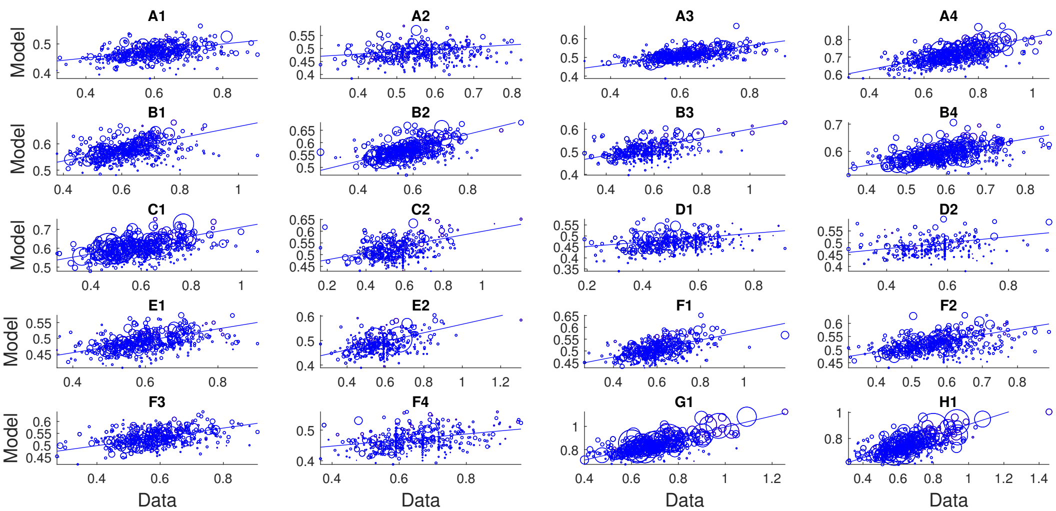


Figure 18: **Growth in $\lambda_{n,s,t}$: Data VS Model, 1990-2010**

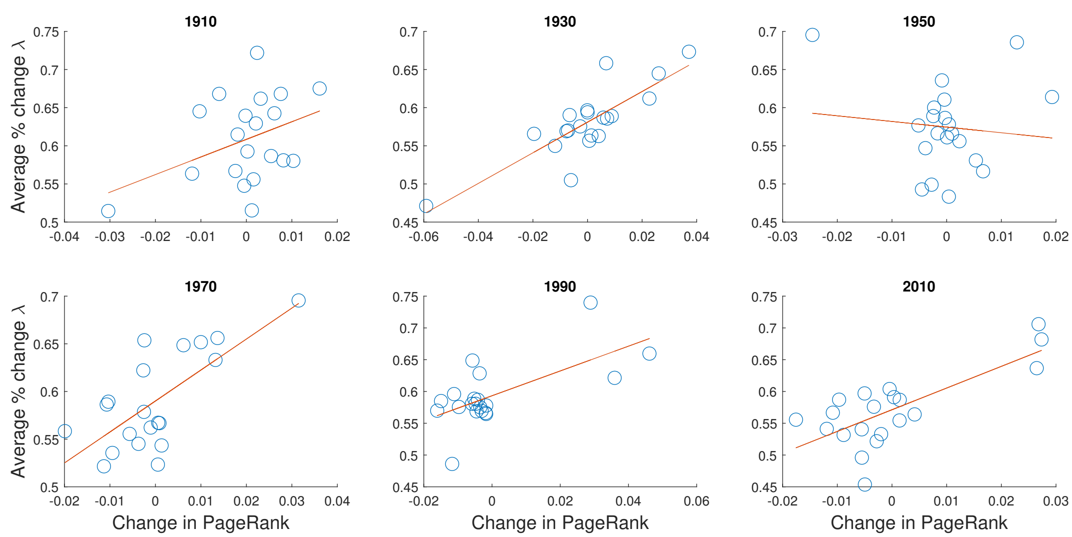


Figure 19: **Average growth growth in $\lambda_{n,s,t}$ and change in PageRank of the 20 sectors**

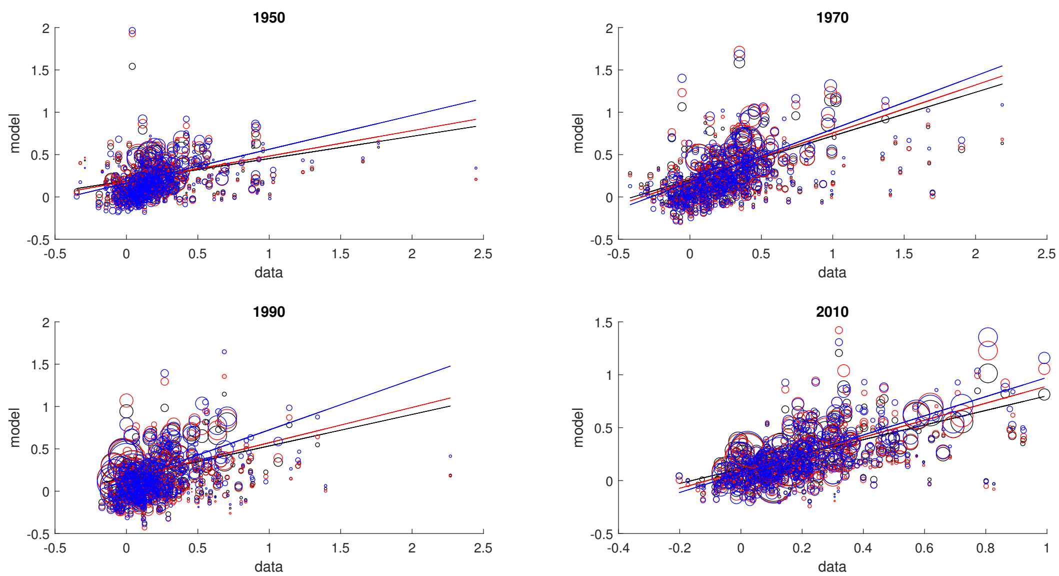
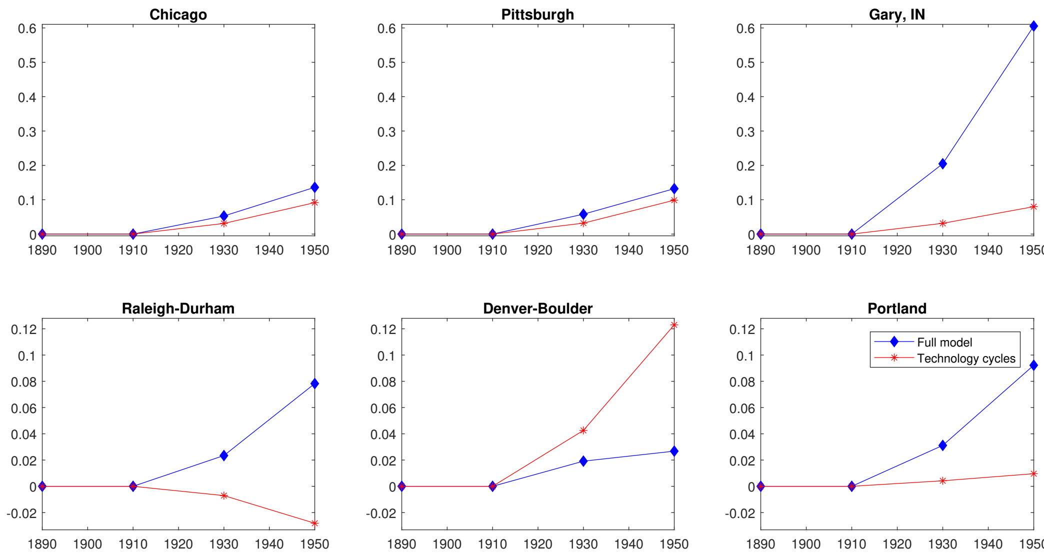
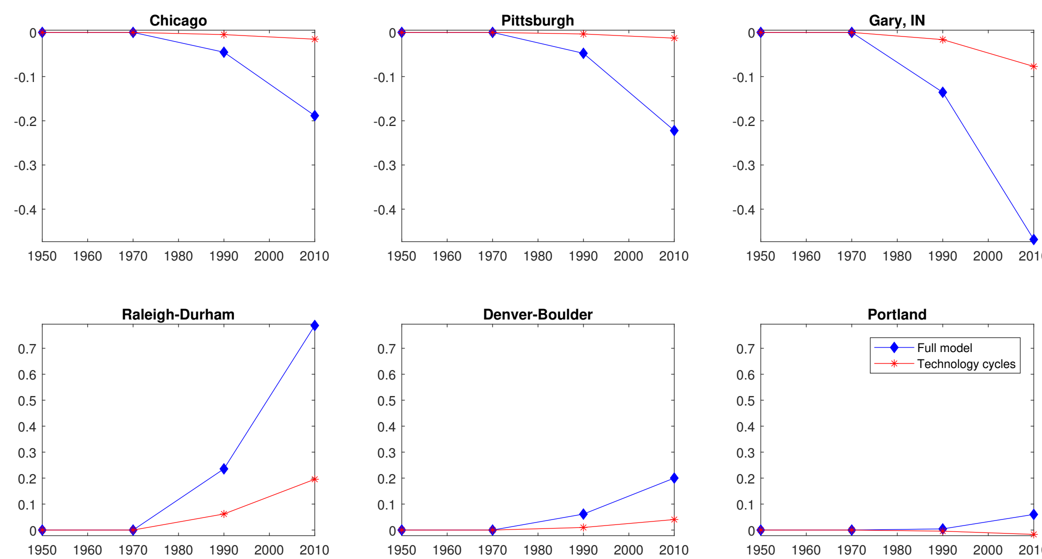


Figure 20: **Population Growth Period by Period**

Notes: The blue line corresponds to the full model, the red line corresponds to the model with only technology cycles, the black line is the model with no shocks.



(a) 1890-1950



(b) 1950-2010

Figure 21: **Growth Decomposition**

Notes: The plots show the growth trajectories of Chicago, Pittsburgh, Gary (IN), Raleigh-Durham, Denver-Boulder, and Portland as deviations from a model with only fertility shocks. The blue line shows the growth trajectory of these six cities in the full model (with structural residuals and technology shocks). The red line represents the growth obtained in a model in which the structural shocks are kept constant.