Immigration, Innovation, and the Geography of Growth*

Costas Arkolakis† Sun Kyoung Lee‡ Michael Peters§

June, 2026

Abstract

Between 1880 and 1920, more than 20 million immigrants settled in the United States. We study how this migration wave affected innovation and growth. Using a newly constructed dataset linking individual census records to historical immigration records and the universe of US patents, we highlight a new channel through which immigrants contributed to growth: they disproportionately settled in urban innovation hubs. To quantify the aggregate and regional effects of this mass migration episode, we develop a new spatial growth model in which skilled workers have a comparative advantage in innovation and sort endogenously across space. We find that international arrivals after 1880 raised US income per capita by 8.2% by 1940. Removing the subsequent immigration restrictions of the 1920s would have raised income per capita by a further 1.7% by 2000. Immigrants' skill composition and their concentration in urban hubs are key drivers of these effects.

JEL Codes: J61, O31, O41, N11, R12, J24

Keywords: Immigration, Innovation, Urban Agglomeration, Knowledge Spillovers, Semi-Endogenous Growth, Age of Mass Migration

*We thank Ariel Burstein, Don Davis, Jonathan Dingel, Jose-Antonio Espin-Sanchez, Chad Jones, Gaurav Khanna, Pete Klenow, Sam Kortum, Naomi Lamoreaux, Ellen McGrattan, Ferdinando Monte, Dan OFlaherty, Tommaso Porzio, Paul Rhode, Esteban Rossi-Hansberg, José Scheinkman, Chris Tonetti, Jon Vogel, and Alwyn Young for comments and suggestions. We also thank many seminar participants. We are grateful to Ariel Boyarsky, Arthur Erlendsson, Shizuka Inoue, Adrian Kulesza, Jack Liang, Roxane Spitznagel, Meili Wang, and, especially, Yongjie Xiong for outstanding research assistance. The authors would like to thank the NSF for support under RIDIR grant #1831524 as well as the NBER and Kauffman Foundation for the Entrepreneurship Grant. We thank the Opportunity & Inclusive Growth Institute at the Federal Reserve Bank of Minneapolis for their hospitality. Sun Kyoung Lee acknowledges the support from the US Census Bureau under cooperative agreement No. CB16ADR0160001. The views expressed in this paper are those of the authors and not those of the US Census Bureau or any other sponsor.

†Yale University and NBER; costas.arkolakis@yale.edu

‡University of Michigan; sunkylee@umich.edu

§Yale University and NBER; m.peters@yale.edu

1. INTRODUCTION

Between 1880 and 1920, more than 20 million immigrants made the US their home. Did their arrival contribute to US productivity growth? Two mechanisms come immediately to mind. First, by expanding the *scale* of the workforce and the size of the market, migrants could have increased the demand for innovation. Second, migrants could have increased the stock of available *skills* with Italian textile workers, German craftsmen, and British engineers bringing their existing knowledge to the American shore. In this paper, we highlight a novel third mechanism: *space*. Migrants could have contributed to US growth by settling in the urban innovation hubs of New York, Chicago, and Philadelphia, where most innovative activity took place.

We quantify the importance of these three channels and evaluate the aggregate and regional effects of immigration in the United States, by using new data and new theory. On the empirical side, we create a dataset on immigrants' patent behavior by linking, at the individual level, the complete-count Population Census, the universe of US patents, and millions of historical immigration records, which provide direct microdata on immigrants' pre-migration occupations and hence the skill content of international migrants. We then develop a spatial semi-endogenous growth model that allows us to quantify the impact of immigration on growth. We find that without the inflow of international migrants between 1880 and 1920, US income per capita would have been 8.2 percent lower by 1940, and that both immigrants' innate skills and their concentration in urban innovation hubs drive most of this effect.

Our data reveal three key empirical patterns. First, immigrants played an important role in US innovation: they created patents at roughly the same rate as natives. Second, spatial sorting played a key role because immigrants' patenting rates are substantially lower within local labor markets. Hence, immigrants' relative innovativeness reflected their disproportionate concentration in cities, where patent intensity is highest. Third, our data highlight the importance of innate innovative skill. For example, German and British immigrants were substantially more innovative than Italian and Irish immigrants, and we document that they were positively selected based on their pre-migration occupations.

On the theoretical side, we propose a spatial semi-endogenous growth model that speaks directly to this evidence. The key feature of our theory is that the local supply of innovative human capital matters. We assume that skilled workers of each nationality are heterogeneous and can work as production workers or, as in Romer (1990) or Krugman (1980), become innovators by inventing new products. Goods are tradable and the demand for new varieties stems from local forward-looking capitalists, who hire innovators to come up with new blueprints. As a result, the costs of product creation differ across people *within* a location (because skilled individuals have a comparative advantage in innovation) and *across* space because they depend on the endogenous stock of accumulated knowledge

("building on the shoulders of *local* giants") and because some locations might have higher innovation TFP than others.

Our model captures the three channels of how immigrants can contribute to growth discussed above. First, the *size* of the local workforce increases the return to innovation so that immigrant inflows induce innovation by native skilled workers. Second, the *skill* composition of the local population determines the local innovation rate: an inflow of skilled immigrants thus contributes to the creation of ideas. Finally, *spatial* differences in the quality of the local knowledge pool imply that the productivity of innovators differs across labor markets. Immigrants' bias to move to innovative urban centers thus has a direct effect on innovative activity.

We analytically characterize the spatial balanced growth path (SBGP) of the model. Since it is a semi-endogenous growth model in the tradition of Jones (1995), the long-run growth rate is tied to the rate of population growth, and the cross-sectional distribution of wages, patent activity, and population is stationary in this growth path. Hence, more populous locations are richer and have a higher stock of patents but they do not grow at a faster rate in the long run. We also show that the share of people engaged in innovation as opposed to production work is equalized across space. Patents per innovator, by contrast, are larger in locations with high innovation TFP and a large knowledge stock — that is, in cities, which in equilibrium are richer and more populous.

A key advantage of our theory is that it can be readily taken to our data. Three aspects govern the innovation side of our model. First, individuals' innate human capital determines their innovative potential. We assume that the share of skilled individuals differs across nationalities and that the innovative talent of skilled individuals is drawn from a distribution whose mean depends on a location's endogenous knowledge stock and exogenous innovation TFP. We show that we can estimate the parameters of this distribution and the share of skilled individuals by nationality directly using our linked data on the extensive and intensive margin of patenting. Second, we estimate the spatial distribution of innovation TFP, which is akin to a location fixed effect. We do so by using data on the existing stock of patents, the flow of patents per capita, and the share of people engaged in innovation. Intuitively, if a location has a large flow of patents given its population and stock of patents, but only a small share of individuals contributes to new inventions, innovation TFP must be large. Finally, as in every model of (semi-)endogenous growth, the extent to which the existing stock of knowledge makes future innovation easier is crucial. We estimate this parameter directly from the panel data on patent activity at the regional level, leveraging the autocorrelation between the local patent stock and subsequent patent flows. We estimate this intertemporal scale elasticity to be substantially below unity, consistent with growth being semi-endogenous and ideas getting harder to find (Bloom et al., 2020).

With the calibrated model in hand, we quantify the importance of international migration

for US growth, more generally, and the role of immigrants' skills and the "urban innovation hubs" hypothesis, in particular. First, to capture the overall importance of international migration, we compute the economic consequences of a complete migration ban between 1880 and 1920. Such a policy will have a long-lasting negative effect on the US population, but in the long run, the economy will return to its old balanced growth path. We find that the inflow of immigrants during the Age of Mass Migration played an important role in US innovation and growth: without the inflow of international migrants between 1880 and 1920, aggregate GDP per capita would have been 8.2% lower by 1940.

To isolate the role of migrants fueling the population of urban innovation hubs and the role of skilled migration, we compare two alternative scenarios: one where the overall number of migrants remains unchanged but their initial settlement is in line with the settlement of natives — that is, relatively less urban — and one where we keep the number of immigrants constant but assume that all such immigrants are unskilled. In the first counterfactual, both the number and the skill-nationality composition of international migrants do not change, but they arrive in rural America rather than New York. This change would reduce GDP per capita by almost 2% and thus would account for about a quarter of the overall GDP impact of a full migration stop. The second counterfactual highlights the importance of innate innovative skills — the aforementioned British engineers and German craftsmen. In this case, overall GDP per capita would be reduced by 4%. Hence, migrants' bias to settle in urban innovation hubs was half as important for US growth as the entire difference in innate skills. This highlights a new and important aspect of successful immigration policy: not only does it matter who enters the country, but also where migrants settle.

Finally, we use our model to evaluate the growth consequences of the 1921 Emergency Quota Act and the 1924 Johnson–Reed Act, the first large-scale immigration restrictions in the US, which substantially reduced new arrivals, especially from Eastern and Southern Europe and Asia. We find that the removal of these bans leads to an increase in US GDP per capita of about 1.7% by 2000. There are two main reasons why this policy had relatively modest effects. First, given the size of the US population, the implied change in population growth is relatively small. Second, the policies mainly affected countries with a large share of unskilled immigrants, who, according to our estimates, contributed relatively little to overall innovation.

Related Literature The rising prominence of immigration flows, particularly high-skill immigration (see, e.g., Kerr et al. (2016), Kerr (2018), and Bernstein et al. (2022)), has led to increased interest in estimating the economic effects of immigration. Traditionally, the literature has focused on the short-term consequences on labor outcomes of natives (see, e.g., Card (1990), Burstein et al. (2020), Dustmann et al. (2017), and Peri (2016) for a survey). More recent papers, including ours, focus on the long-term impact. Sequeira et al. (2020) and Akcigit et al. (2017) document a large positive effect of 19th-century immigration

inflows on contemporary economic activity and patent creation. Burchardi et al. (2026) find positive effects of immigration on local innovation and wages in the US since 1965. Our paper complements this empirically oriented literature. By proposing a spatial model of migration and growth, we quantify the aggregate and regional consequences of immigration and highlight the importance of urban innovation hubs, isolating the role of where migrants settle, not just how many arrive.

Our paper adds to the growing literature in spatial economics. Early contributions in this area focus on static models, taking regional productivity as given — see, for example, Allen and Arkolakis (2014) or the recent survey by Redding and Rossi-Hansberg (2017). A growing number of papers model the evolution of regional productivity explicitly (e.g., Desmet et al., 2018; Nagy, 2023; Peters, 2022; Eckert and Peters, 2022). We build on the idea-based, semi-endogenous growth models in the spirit of Jones (1995) or Kortum (1997) (see, for example, Akcigit (2017) and Jones (2022) for recent surveys) and propose a framework where innovators build on local ideas and sort endogenously across space. To the best of our knowledge, our paper is the first to connect a micro-founded model of spatial growth with direct evidence on regional patenting at the individual level.

On the methodological side, we relate to the growing literature in spatial economics that explicitly models forward-looking investment behavior. Kleinman et al. (2023) introduce the accumulation of physical capital in a spatial model; Walsh (2025) considers a framework of firms with free entry. We directly model the process of innovation, whereby long-lived capitalists hire local researchers to generate new product varieties.

Finally, we speak to the literature on the "Era of Mass Migration" that has used record-linking techniques to study the experience of immigrants in the late 19th and early 20th century (see, e.g., Abramitzky et al. (2014), Abramitzky et al. (2012), or Abramitzky and Boustan (2022)). Complementing this line of research, we link population census data to patents and use our microdata to estimate a structural model and quantify the aggregate impacts of immigration inflows.

The rest of the paper is structured as follows. Section 2 documents the main empirical facts about immigrants' patenting behavior and the urban innovation premium. In Section 3 we present our theory and structurally estimate it in Section 4. Section 5 quantifies the impact of immigration on US growth. A number of theoretical and empirical details are contained in the Appendix and in the Supplementary Material.

2. Immigrants and Innovation: The Role of Cities

In this section, we present the main motivating facts for our study. First, immigrants played an active part in aggregate innovation activity in the United States between 1880 and 1920. Second, this is partly due to the fact that they settled predominantly in urban innovation hubs.

Data: A New Dataset on Immigrants' Patent Activity To document these facts, we create a novel database that captures individual-level patent activity. We create these data by linking the universe of patents to tens of millions of individuals in the complete-count Population Census between 1880 and 1920. Because the Census reports detailed information on individuals' immigration status and places of residence, our matched data allow us to directly measure whether immigrants are engaged in innovation and where innovation takes place.

In order to link patents to individuals in the Census, we employ supervised, machine-learning-based record-linking techniques. Specifically, we opt for a *random forest classifier* (RFC). Our study concentrates exclusively on males, given that the majority of patent holders during the period from 1880 to 1920 were male.[1] We further restrict the sample to working-age individuals aged between 16 and 60 years. To match patents and individuals, we rely on information regarding the names and the residential locations of individuals. We match patent records within a fifteen-year window around the Census, from five years prior to ten years after. We train our RFC using a training dataset of IPUMS-linked records and apply the trained classifier to pairs of records that meet feasibility constraints, such as matching the county of residence. Subsequently, we select unique matches through a "pruning" procedure. We first retain candidate pairs whose predicted match probability is at least 0.85; these constitute the set of potential matches. Among these, a pair (r_A, r_B) is selected as a unique match if either (i) r_A and r_B are matched only to each other, or (ii) the pair's predicted match probability exceeds that of every competing pair involving r_A or r_B by at least a 15% margin.

A detailed description of our methodology is contained in Appendix Section A.2.2. In the Appendix we also provide more detailed insights into the quality of our patent-matching procedure (see Section A.2.3) and compare our data with those generated using the alternative linking methods discussed in Abramitzky et al. (2021) and Abramitzky et al. (2025); see Section A.2.4.

In Table 1 we report the matching rates for each period. On average, we achieve a matching rate ranging from one-third to one-quarter of all patents, resulting in a dataset of approximately 430,000 matched patents.

Results: Immigrants and Innovation Figure 1 summarizes two key findings of these data. First, as shown in the left panel, immigrants played a pivotal role in US innovation and were almost as likely to engage in patenting as natives. In 1910, among natives, slightly more than two out of 1,000 individuals were assigned any patent. Among immigrants, the rate is only moderately lower, even though many migrants did not speak the language and arguably faced other hurdles that might have discouraged them from engaging with the formal patent system.

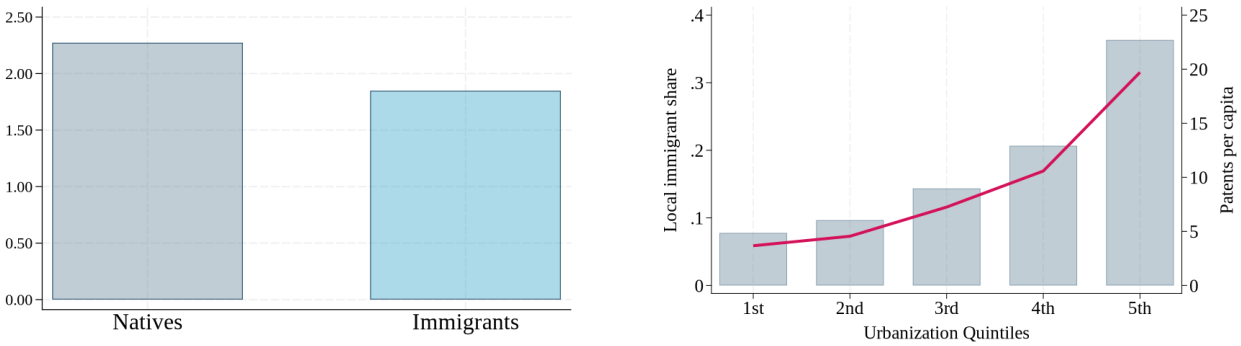
[1] We impute the gender of patent inventors based on the given names of patent holders and find that

TABLE 1: MATCHING PATENTS TO THE POPULATION CENSUS

	1880	1900	1910	1920	Total
Total number of patents issued by period	222,974	307,324	382,193	456,800	1,369,291
Number of patents matched to Census	63,975	108,600	122,861	138,363	433,799
Share of patents matched*	0.29	0.35	0.32	0.30	0.32

Notes: This table reports the total number of issued patents and the subset that we were able to match to the US Census.

FIGURE 1: IMMIGRANTS IN 1910: INNOVATORS IN CITIES



Notes: The left panel shows the number of natives and immigrants with at least one patent out of 1,000 individuals. The right panel shows the share of immigrants (bars) and patents per capita (red line) across counties in 1910 by quintiles of urbanization. We compute patents per capita using the overall population of geolocalized patents, not the patents matched to the US Census. The higher patent intensity in urban areas thus does not reflect differences in match rates.

Second, as shown in the right panel, the fact that immigrants were much more likely to live in cities played an important role in overcoming such disadvantages. Specifically, we compute patent intensity — the number of patents per capita — for each county and then aggregate counties based on their urbanization rate. It is evident that historical patent activity exhibited a significant urban bias: urban locations had about four to five times as many patents per capita as their rural counterparts. Therefore, much like today, cities served as *innovative hubs*. In addition, these regional differences in patent activity strongly correlate with migrants' location choices: migrants were significantly more likely to reside in cities. While counties in the highest urbanization quintile had an immigrant share of almost 40%, less than 10% of individuals living in rural counties were born outside of the US.

The patterns illustrated in Figure 1 suggest that immigrants derived benefits from the local knowledge pool available in cities. To see this more directly, we run a set of regressions of the form

$$\mathbb{1}\{\mathcal{P}_{irt} > 0\} = \alpha\,\text{Urban}_{irt} + \beta\,\text{Immigrant}_{irt} + x'_{irt}\gamma + \delta_r + \delta_t + u_{irt}, \tag{1}$$

where $\mathcal{P}_{irt}$ is the number of patents by individual i living in region r at time t. The

approximately 98% of patent holders were male.

TABLE 2: PATENT ACTIVITY BY IMMIGRANTS

	(1)	(2)	(3)	(4)
Immigrant	-0.113***	-1.168***		
	(0.036)	(0.043)		
Urban		0.682***	0.838***	0.694***
		(0.035)	(0.179)	(0.035)
British				0.616***
				(0.078)
German				-1.059***
				(0.070)
Irish				-2.324***
				(0.127)
Italian				-2.407***
				(0.101)
Others				-1.231***
				(0.047)
Year FE	yes	yes	yes	yes
County FE	no	yes	yes	yes
Age + Educ FE	no	yes	yes	yes
Name Frequency	no	yes	yes	yes
Individual FE	no	no	yes	no
R^2	.000157	.00467	.456	.00471
N	97,515,818	87,410,901	3,996,904	87,410,901

Notes: The table reports the results of regression (1). The coefficients are scaled to reflect the change in the number of patent holders per 1,000 individuals. All columns control for year FE. Columns 2-4 control for county FE, age and education score FE, and for 100 percentiles of name frequency. Column 3 controls for individual FE. Robust standard errors are clustered at the county-year level.

dummies Urban_{irt} and Immigrant_{irt} indicate whether the individual lived in an urban area and was an immigrant. The vector x_{irt} collects additional controls, δ_r and δ_t denote region and time fixed effects, and u_{irt} is an idiosyncratic error term.

Column 1 of Table 2 documents the unconditional comparison between immigrants and natives. Consistent with Figure 1, immigrants are slightly less likely to hold a patent, though the difference is economically small. Column 2 shows that differences in spatial sorting play a crucial role in explaining why immigrants and natives patented at similar rates overall. Once we control for county fixed effects and an urban dummy, immigrants are substantially less likely to engage in patenting: within locations immigrants have, on average, one fewer patent holder per 1,000 individuals; i.e., only half as many as natives. Note that this specification also controls for age and educational-score fixed effects, and county-specific name frequency, to highlight the importance of spatial differences and to address the concern that immigrants may have distinctive names with match rates that differ systematically from those of natives.[2] At the same time, individuals living in cities have a substantially higher probability of patenting. This coefficient captures both the

[2]We use the information on educational scores contained in the US Census, which does not directly report educational attainment in our time period. We compute name frequency as the number of times we observe a given name in each county and then control for 100 percentiles of this distribution.

direct effect of urban knowledge on individual patent intensity and selection, whereby workers with innovative potential sort toward urban locations.

Column 3 provides additional evidence for a direct effect of urban hubs on innovativeness. By matching census records over time, we construct a panel of roughly four million individuals that we observe at least twice. We can therefore estimate (1) including individual fixed effects, controlling for all time-invariant heterogeneity across individuals. The urban coefficient remains sizable and similar in magnitude to that in column 2.

Finally, column 4 exploits our microdata to document that immigrants of different nationalities were highly heterogeneous in their propensity to patent. British immigrants were more likely than natives to innovate, while Italian and Irish migrants had substantially lower rates of patenting.

Taking stock, the patterns in Figure 1 and Table 2 suggest important roles for both the aggregate skill composition of immigrants and their spatial allocation in shaping US growth. To quantify this link, we now turn to a model in which these factors play a central role.

3. THEORY

In this section, we present a novel model of spatial growth that embeds an idea-based growth model (Romer, 1990; Jones, 2005) into a canonical quantitative spatial framework (Allen and Arkolakis, 2014; Redding and Rossi-Hansberg, 2017). The key ingredient of our theory is the local nature of innovative human capital: ideas depend on local innovation labor whose human capital has both an innate component reflecting skill differences and a regional component reflecting innovation efficiency and the endogenous accumulation of local knowledge.

3.1 Environment

We consider an economy with R regions, denoted by r. Time is indexed by t. Individuals have preferences over a unique final good, which itself is a CES bundle of differentiated varieties produced in different regions, indexed by ω:

$$C_{rt} = \left(\sum_{r=1}^{R} \int_{\omega=0}^{N_{rt}} c_{rt}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}.$$

In this expression, N_{rt} denotes the number of varieties being produced in region r at a given time period t. The aggregate number of varieties available in the economy at any given time t is $N_t = \sum_r N_{rt}$, $c_{rt}(\omega)$ is the consumption of good ω in region r, and σ the elasticity of substitution across varieties.

We assume that varieties are long-lived and produced with constant returns to scale using

only labor with region-specific productivity A_{rt}, which we take as exogenous. By contrast, the mass of varieties evolves endogenously as an outcome of local innovation. Letting I_{rt} denote the flow of variety creation in region r at time t, the law of motion of varieties is given by

$$N_{rt} = N_{rt-1} + I_{rt}.$$

Each variety ω is produced by a single firm that competes in a monopolistically competitive market. Inter-regional trade is subject to iceberg trade costs, denoted by τ_{rj}.

The economy is populated by two types of agents. First, there is a mass ℓ_t of *workers*, who are spatially mobile, live for a single period, and can either work as production workers or as innovators. To stress the importance of innate skills, we assume that workers are distinguished by their capacity to engage in innovation activities. Unskilled individuals can only work as production workers. By contrast, skilled individuals face an occupational choice between working in the production sector and innovating. Each skilled worker is endowed with a single efficiency unit in production and h units of innovation efficiency, which is distributed according to a cdf $F_{rt}(h)$ in location r at time t.

Motivated by our empirical analysis in Section 2, we assume that individuals of different nationalities ν differ in their innovativeness, i.e., in the relative abundance of skilled individuals. In terms of notation, we let ℓ_{rt}^{ν} denote the mass of workers of nationality ν in region r at time t and λ_{rt}^{ν} the share of skilled individuals with nationality ν in location r.[3] Together with our assumption that the distribution of innovation efficiency, $F_{rt}(h)$, varies across space, our theory generates differences in innovativeness both across space and across individuals within locations.

Second, each location is populated by a unit mass of *capitalists*. Capitalists are long-lived, discounting the future at rate β, but spatially immobile. They run local firms and are the residual claimants to corporate profits. They hire local production workers to produce existing varieties and local innovators to create new ones.[4]

3.2 Labor Supply and the Demand for Innovation

We now characterize the behavior of capitalists and workers. We denote the prevailing wage for production workers by w_{rt} and the wage for innovation human capital by χ_{rt}. Both are determined endogenously, but taken as given by all agents.

Capitalists To determine the demand for variety creation, let $V_{rt}(N)$ denote the value of the representative capitalist in location r at time t with a stock of N local varieties. Letting π_{rt} denote the profit from each variety that is produced in region r, the value of owning

[3] The total size of the population in region r is thus $\ell_{rt} = \sum_{\nu} \ell_{rt}^{\nu}$, and the overall size of the population satisfies $\ell_t = \sum_r \ell_{rt}$. Moreover, the total number of skilled workers of nationality ν is given by $\ell_t^{\nu H} = \sum_r \lambda_{rt}^{\nu} \ell_{rt}^{\nu}$.

[4] The distinction between capitalists and workers, also made in Kleinman et al. (2023), achieves tractability: because capitalists are long-lived and spatially immobile, their forward-looking behavior on innovation decisions does not interact with overall labor supply.

such a portfolio of varieties can be written recursively as

$$V_{rt}(N) = \max_{I} \left\{ \frac{N\pi_{rt}}{P_{rt}} + I\frac{(\pi_{rt} - \chi_{rt})}{P_{rt}} + \beta V_{rt+1}(N+I) \right\}.$$

The first term captures the overall profits of the N existing varieties in region r. The key choice variable of capitalists is the number of new varieties, I, that are created. Each new variety generates a flow payoff of $\pi_{rt} - \chi_{rt}$, after the payments to innovators are taken into account. In addition, because varieties are long-lived, in the next period, the stock of varieties is given by $N + I$, which has a value of $V_{rt+1}(N + I)$.

As we show in Appendix Section A.1.1, $V_{rt}(N)$ has a tractable solution. First, as in Walsh (2025), it is linear in the number of varieties, so that $V_{rt}(N) = v_{rt}N$, where v_{rt}, the value of a single variety, is given by the present discount value of profits

$$v_{rt} = \frac{\pi_{rt}}{P_{rt}} + \beta v_{rt+1}. \tag{2}$$

Second, competition for local innovators implies that innovators get paid the discounted value, in nominal terms, of the new variety; $\chi_{rt} = P_{rt}v_{rt}$.

Workers While unskilled workers always work in the production sector, a skilled worker with h efficiency units of innovative human capital selects into the innovation process if and only if $h\chi_{rt} \geq w_{rt}$. Thus, there is a region-specific human capital cutoff, $\overline{h}_{rt}$, so that all skilled workers with $h \geq \overline{h}_{rt}$ become innovators:

$$\overline{h}_{rt} = \frac{w_{rt}}{\chi_{rt}}.$$

We assume that h is drawn from a Pareto distribution:

$$F_{rt}(h) = 1 - (\psi_{rt}/h)^{\theta},$$

where $\theta > 1$ and ψ_{rt} is a region-r and time-t specific shifter that determines the distribution of innovative abilities in a particular location at a given point in time. We put more structure on ψ_{rt}^{ν} below. At this point, we simply note that some regions might be better suited to generate ideas ($\psi_{rt} > \psi_{r't}$), and that innovation human capital can grow over time ($\psi_{rt} > \psi_{rt'}$).

Given the cutoff $\overline{h}_{rt}$ and the Pareto assumption on individuals' innovative human capital, the total mass of new varieties created in region r in time t is given by[5]

$$I_{rt} = \sum_{\nu} \lambda_{rt}^{\nu} \ell_{rt}^{\nu} \left(\int_{\overline{h}_{rt}}^{\infty} h dF_{rt}(h) \right) = \frac{\theta}{\theta - 1} \overline{h}_{rt} \sum_{\nu} \lambda_{rt}^{\nu} \ell_{rt}^{\nu} \left(\frac{\psi_{rt}}{\overline{h}_{rt}} \right)^{\theta}.$$

[5]For our theoretical exposition we assume $\psi_{rt} < \overline{h}_{rt}$. In our quantitative analysis we account for $\psi_{rt} > \overline{h}_{rt}$.

Hence, total variety creation is increasing with the number of skilled individuals, $\lambda_{rt}^{\nu}\ell_{rt}^{\nu}$, and local human capital ψ_{rt}, while it is decreasing in the innovation cutoff $\overline{h}_{rt}$.

To characterize the equilibrium, it is useful to let $\varpi_{rt}^{\nu} := \ell_{rt}^{\nu}/\ell_{rt}$ denote the share of people of nationality ν in location r and define the aggregate statistic

$$\mathcal{H}_{rt} := \psi_{rt}\left(\sum_{\nu}\lambda_{rt}^{\nu}\varpi_{rt}^{\nu}\right)^{1/\theta}.$$

We refer to $\mathcal{H}_{rt}$ as the region r's *innovation capacity* because it depends on the location's innovation efficiency ψ_{rt} and its overall share of skilled individuals. Given $\mathcal{H}_{rt}$, the overall share of innovators in region r, e_{rt}, is given by

$$e_{rt} = \frac{\sum_{\nu}\lambda_{rt}^{\nu}\ell_{rt}^{\nu}\left(\psi_{rt}/\overline{h}_{rt}\right)^{\theta}}{\ell_{rt}} = \left(\frac{\mathcal{H}_{rt}}{\overline{h}_{rt}}\right)^{\theta}. \tag{3}$$

Hence, at the regional level, the overall innovator share again resembles a Pareto distribution, with an *endogenous* scale parameter $\mathcal{H}_{rt}$.

Our theory has strong implications for the distribution of innovative activity across nationalities and space. As such it speaks directly to the empirical evidence in Section 2. The share of people of nationality ν that innovate in location r at time t is given by $e_{rt}^{\nu} = \lambda_{rt}^{\nu}\left(\psi_{rt}/\overline{h}_{rt}\right)^{\theta}$. *Within* locations, differences in innovation activity across nationalities, $e_{rt}^{\nu}/e_{rt}^{\nu'}$, reflect the relative abundance of skill supplies, $\lambda_{rt}^{\nu}/\lambda_{rt}^{\nu'}$. Hence, the fact that within locations immigrants have fewer patents than natives reflects the fact that the migrant population is, on average, less skilled. At the same time, innovative activity is concentrated in cities, if their innovative capacity $\mathcal{H}_{rt}$, relative to the cutoff $\overline{h}_{rt}$, is large. As a consequence, the *average* degree of patenting might not vary much between natives and immigrants, if migrants are more likely to live in urban locations.

3.3 Local Innovation and Growth

Using the definition of $\mathcal{H}_{rt}$, the total flow of local innovation can be written as

$$I_{rt} = \frac{\theta}{\theta-1}\overline{h}_{rt}^{1-\theta}\mathcal{H}_{rt}^{\theta}\ell_{rt} = \frac{\theta}{\theta-1}\overline{h}_{rt}e_{rt}\ell_{rt}. \tag{4}$$

Local innovation depends on the total size of the local population ℓ_{rt}, overall innovation capacity $\mathcal{H}_{rt}$, and the endogenous human capital cutoff $\overline{h}_{rt}$. In particular, given the population distribution $[\varpi_{rt}^{\nu}, \lambda_{rt}^{\nu}]$ and innovative ability ψ_{rt}, the human capital cutoff is the sole endogenous variable that fully determines innovative investments and innovation per researcher $I_{rt}/(e_{rt}\ell_{rt})$.

So far, we have taken individuals' innovative human capital in location r, ψ_{rt}, as exogenous.

We now put more structure on its determinants and incorporate dynamics. In the tradition of idea-based models of economic growth, we link individuals' innovative efficiency to the available stock of knowledge embodied in the existing mass of local varieties N_{rt}. Specifically, we assume ψ_{rt} is given by

$$\psi_{rt} = \zeta_r N_{rt-1}^{\vartheta}, \tag{5}$$

that is, ψ_{rt} reflects both an exogenous location-specific effect ζ_r and the endogenous amount of local knowledge N_{rt-1}. The former allows for systematic, exogenous differences in innovative ideas across space. The latter incorporates the usual mechanism of building on the shoulders of giants, whereby existing local knowledge is an input into the production of new entrepreneurial human capital and the source of long-run growth. The intertemporal scale elasticity $\vartheta \leq 1$ governs the strength of such spillovers.

Equation (5) stresses that our theory captures both the role of *places* and *people*. In particular, the local human capital supply $\mathcal{H}_{rt}$ is given by

$$\mathcal{H}_{rt} = \zeta_r N_{rt-1}^{\vartheta} \left(\sum_{\nu} \lambda_{rt}^{\nu} \varpi_{rt}^{\nu} \right)^{1/\theta},$$

and thus depends on the local skill supply, the region's innovation TFP ζ_r, and the local pool of ideas N_{rt-1}. Together with the law of motion of varieties and the definition of the entrepreneurial employment share e_{rt} in (3), we can express local variety *growth* as

$$g_{Nt} \equiv \frac{N_{rt} - N_{rt-1}}{N_{rt-1}} = \frac{\theta}{\theta - 1} \zeta_r N_{rt-1}^{\vartheta - 1} \left(\sum_{\nu} \lambda_{rt}^{\nu} \varpi_{rt}^{\nu} \right)^{1/\theta} \ell_{rt} e_{rt}^{\frac{\theta-1}{\theta}}.$$

Local growth is increasing in permanent research TFP ζ_r, population size ℓ_{rt}, the share of skilled individuals $\sum_{\nu} \lambda_{rt}^{\nu} \varpi_{rt}^{\nu}$, and the share of the population engaged in innovative activities, e_{rt}. Note that this intensive margin of innovation has decreasing returns: $(\theta - 1)/\theta < 1$. This reflects the scarcity of entrepreneurial talent: the more people are engaged in innovation, the lower the idiosyncratic talent draw of the marginal innovator. By contrast, growth is (weakly) decreasing in the mass of existing varieties, N_{rt-1}. If $\vartheta = 1$, our model is the spatial analogue of a fully endogenous growth model of Romer (1990): local growth is independent of the mass of varieties and there are strong scale effects whereby larger or more innovative locations grow at a faster rate. If $\vartheta < 1$, ideas are getting harder to find locally: a higher mass of varieties reduces future growth. This specification turns our model into a spatial version of the semi-endogenous growth model of Jones (2005) with long-run growth being proportional to population growth and a stationary population distribution — see Section 3.5 below.[6]

[6] As such, our theory captures the notion that "ideas are getting harder to find" (Bloom et al., 2020).

We also want to highlight what our specification in (5) does not capture. First, we abstract from contemporaneous complementarities, where each type's innovative human capital depends, e.g., on the current composition of the local population. Instead, we summarize the history of past innovative efforts in the scalar-valued dynamic state variable N_{rt}. Second, individuals only benefit from local knowledge; ideas developed in other locations, N_{jt-1}, do not impact the innovative opportunities of a given individual in r. This assumption, while consistent with the empirical evidence that, for example, citations are very localized, is obviously stark and meant to highlight the importance of the local human capital pool. In Section 6 below we extend our analysis by allowing for a richer form of spillovers, where ψ_{rt} is a function of the knowledge stock in all locations. Third, (5) imposes a strict separation between portable and non-portable human capital. In particular, the benefits from an innovative location only accrue as long as the individual is actually present in that location. This precludes the possibility of learning, whereby innovative human capital is a function of the migration history of the individual (see, e.g., Crews (2023) and Bilal and Rossi-Hansberg (2021)). We impose this assumption for tractability as otherwise we would need to keep track of individuals' migration histories as additional state variables.

3.4 Equilibrium

We can now characterize the equilibrium of our model. Given our results from above, we have to determine three endogenous variables: the local wage w_{rt}, the innovation cutoff $\overline{h}_{rt}$ (or alternatively the research wage χ_{rt}), and the mass and composition of people of nationality ν, ℓ^{ν}_{rt} and λ^{ν}_{rt}.

Demand and Supply for Innovation To solve for the equilibrium innovation cutoff $\overline{h}_{rt}$, note first that due to CES demand, aggregate profits $N_{rt}\pi_{rt}$ are a constant fraction of production workers' labor income:

$$\pi_{rt} = \frac{1}{\sigma - 1}\frac{w_{rt}\ell_{rt}\left(1 - e_{rt}\right)}{N_{rt}} = \frac{1}{\sigma - 1}\frac{w_{rt}\ell_{rt}}{N_{rt}}\left(1 - \left(\frac{\mathcal{H}_{rt}}{\overline{h}_{rt}}\right)^{\theta}\right).$$

Combining this equation with (2) allows us to express the innovation cutoff $\overline{h}_{rt}$ as

$$\overline{h}_{rt} = \frac{(\sigma - 1)}{\frac{\ell_{rt}}{N_{rt}}\left(1 - \left(\frac{\mathcal{H}_{rt}}{\overline{h}_{rt}}\right)^{\theta}\right) + \beta\left(\sigma - 1\right)\frac{P_{rt}v_{rt+1}}{w_{rt}}}. \tag{6}$$

Equation (6) summarizes the *demand* for innovation. Given the population size, ℓ_{rt}, innovation capacity $\mathcal{H}_{rt}$, and the future discounted value of profits relative to the prevailing

Research productivity, i.e., variety growth per researcher, is given by $\frac{g_{rN}}{\ell_{rt}e_{rt}}$, and hence high in locations with high research TFP ζ_r and a high share of high-type individuals, and low in locations with a larger knowledge stock N_{rt-1} and a larger innovation share, e_{rt}.

real wage, $\beta \frac{v_{rt+1}}{w_{rt}/P_{rt}}$, (6) represents an increasing relationship between the mass of products produced, N_{rt}, and the innovation cutoff $\overline{h}_{rt}$. If the number of products in region r is large, the market for innovation becomes more selective: the marginal entrepreneur has to be more innovative to break even and the cutoff $\overline{h}_{rt}$ rises. Moreover, $\mathcal{H}_{rt}$ acts like a supply shifter: a higher innovative capacity raises $\overline{h}_{rt}$ for a given N_{rt} because fewer people are required to produce a given number of goods N_{rt}.

The equilibrium law of motion for local ideas,

$$N_{rt} = N_{rt-1} + \frac{\theta}{\theta - 1} \ell_{rt} \left(\overline{h}_{rt} \right)^{1-\theta} \mathcal{H}_{rt}^{\theta},$$

is the second equation that links N_{rt} and $\overline{h}_r$. It represents the *supply* of local innovation: N_{rt} and $\overline{h}_{rt}$ are negatively related because a lower productivity cutoff increases the supply of human capital in innovation, and hence the mass of available goods N_{rt} being sold in region r.

These two equations, therefore, allow us to compute $\overline{h}_{rt}$ as a function of the pre-determined number of varieties, the overall supply of human capital, the total population, and the future value of a new variety:

$$\overline{h}_{rt} = h \left(N_{rt-1}, \ell_{rt}, \mathcal{H}_{rt}, \frac{P_{rt} v_{rt+1}}{w_{rt}} \right).$$

Three properties are instructive. First, $\overline{h}_{rt}$ is increasing in N_{rt-1} and decreasing in ℓ_{rt}: if goods are abundant relative to the population, the demand for new varieties falls, raising the selection cutoff. Second, $\overline{h}_{rt}$ is increasing in $\mathcal{H}_{rt}$: a higher supply of human capital raises the selection cutoff because fewer entrepreneurs are needed to produce a given number of ideas. Finally, $\overline{h}_{rt}$ is decreasing in the future value of variety creation because higher future profits incentivize more innovation today.

Goods Market Clearing Equilibrium wages are determined from the usual labor market clearing condition that total spending on goods produced in region r equals total income received in r, Y_{rt}:

$$Y_{rt} = N_{rt} \left(\frac{\sigma}{\sigma - 1} \frac{w_{rt}}{A_{rt}} \right)^{1-\sigma} \mathcal{D}_{rt},$$

where $\mathcal{D}_{rt} = \sum_{j=1}^{R} \left(\tau_{rj} / P_{jt} \right)^{1-\sigma} Y_{jt}$ is a measure of aggregate demand and P_{jt} is the local price index. Using that total income Y_{rt} is proportional to the payments to production workers, $Y_{rt} = \frac{\sigma}{\sigma - 1} w_{rt} \ell_{rt} (1 - e_{rt})$, yields a simple expression for local wages:

$$w_{rt}^{\sigma} = \left(\frac{\sigma - 1}{\sigma} \right)^{\sigma} A_{rt}^{\sigma-1} \mathcal{D}_{rt} \frac{N_{rt}}{\ell_{rt} (1 - e_{rt})}. \tag{7}$$

Local wages are determined by exogenous TFP A, local market size $\mathcal{D}$, and the number of ideas per production worker $N/(\ell(1-e))$.

Spatial Mobility We model agents' spatial labor supply using the usual discrete choice methodology and assume that the utility of individual i of nationality ν of skill group s to reside in location r given that she currently resides in location o is given by

$$U_{rt}^{i(\nu,s)} = \mu_{ro} B_{rt}^{\nu s} \frac{y_{rt}^{i(s)}}{P_{rt}} v_{rt}^{i},$$

where $B_{rt}^{\nu s}$ is a regional amenity, μ_{ro} parameterizes the moving costs to migrate from o to r, y_{rt}^{i} is the expected income of individual i (which depends on the individual's skill), and v_{rt}^{i} is an individual preference shock that we assume is Fréchet distributed with shape ε. Below, we impose additional assumptions on $B_{rt}^{\nu s}$.

We assume that individuals move before they know the realization of h. Hence, they base their moving decisions on the expected income. Flow income for unskilled individuals is given by the local wage, that is, $y_{rt}^{L} = w_{rt}$. By contrast, expected income for skilled individuals is given by

$$y_{rt}^{H} = w_{rt} + \int_{\bar{h}_{rt}}^{\infty} (\chi_{rt} h - w_{rt})\, dF_{rt}(h) = \left(1 + \frac{1}{\theta - 1}\left(\frac{\psi_{rt}}{\bar{h}_{rt}}\right)^{\theta}\right) w_{rt}. \tag{8}$$

Intuitively, skilled individuals' expected income reflects the local wage and the option value of innovation. This option value exactly reflects the inframarginal returns of innovation, which are given by $\frac{1}{\theta-1}\left(\psi_{rt}/\bar{h}_{rt}\right)^{\theta} = \frac{1}{\theta-1} e_{rt}^{H}$, where e_{rt}^{H} is the probability that a skilled individual innovates in location r.

In addition to the bilateral moving costs μ_{ro}, we follow Bilal (2023) and assume that individuals face a "Calvo"-type friction, where, at each point in time, only a share χ of individuals is allowed to move. Doing so allows us to capture an extensive margin of moving in a parsimonious way. As a consequence, the share of individuals of nationality ν and skill group s migrating from o to r is given by $\chi m_{rot}^{\nu,s}$, where

$$m_{rot}^{\nu s} = \frac{\left(\mu_{ro} B_{rt}^{\nu s} y_{rt}^{s} / P_{rt}\right)^{\varepsilon}}{\sum_{k} \left(\mu_{ko} B_{kt}^{\nu} y_{kt}^{s} / P_{kt}\right)^{\varepsilon}} \tag{9}$$

denotes the moving probability conditional on receiving the moving opportunity.

Demographics To capture the changing demographics, whereby migrants' children are US-born and the immigrant population grows because of new in-migration, we assume a simple stochastic law of motion. Each period, each individual faces a death rate d and a mass η^{B} of children is born. All newly born children are natives, even though their parents are immigrants. For simplicity, we assume that children keep the skills of their parents,

i.e., the children of high-skill British immigrants will be high-skill US natives. However, it would be straightforward to allow for only a partial inheritability of skills at the expense of additional notation. The immigrant population is replenished through new migration from abroad.

Specifically, let $M_{rt}^{\nu s}$ denote the mass of new international migrants of nationality ν with skill s arriving in r at time t. The mass of immigrants of nationality ν and skill s in region r evolves according to

$$\ell_{rt}^{\nu s} = (1-\chi)\left((1-d)\,\ell_{rt-1}^{\nu s} + M_{rt-1}^{\nu s}\right) + \chi \sum_{o} m_{rot}^{\nu,s}\left((1-d)\,\ell_{ot-1}^{\nu s} + M_{ot-1}^{\nu s}\right). \tag{10}$$

Similarly, the stock of natives with skill s evolves according to

$$\begin{aligned} \ell_{rt}^{USs} &= (1-\chi)\left(\left(1+\eta^B-d\right)\ell_{rt-1}^{USs} + \eta^B \sum_{\nu\in \text{Imm}} \ell_{rt-1}^{\nu s}\right) \\ &\quad + \chi \sum_{o} m_{rot}^{USs}\left(\left(1+\eta^B-d\right)\ell_{ot-1}^{USs} + \eta^B \sum_{\nu\in \text{Imm}} \ell_{ot-1}^{\nu s}\right). \end{aligned} \tag{11}$$

Note that $\eta^B \sum_{\nu\in \text{Imm}} \ell_{rt}^{\nu s}$ are the newly born children among immigrants of skill group s in region r, which add to the stock of natives. Also note that even though we take the inflow of immigrants, M_{rt}^{ν}, as exogenous, migrants' spatial labor supply still responds to local wages because, like everyone else, they have the option to move.

This system of equations determines the spatial allocation of individuals given the current population distribution and the exogenous inflow of international migrants and the vectors of wages, prices, and innovation cutoffs.

Spatial Sorting

Our theory highlights that the skill composition of the local workforce is a key determinant of local innovation capacity: everything else equal, $\mathcal{H}_{rt}$ is increasing in λ_{rt}^{ν}. This distribution of skills is, of course, endogenous, reflecting workers' optimal location decisions. To see this specifically, consider the relative likelihood of skilled individuals and unskilled individuals of the same nationality ν in location o to move to location r relative to location j. Using (8) and (9), this is given by

$$\frac{m_{rot}^{\nu H}/m_{jot}^{\nu H}}{m_{rot}^{\nu L}/m_{jot}^{\nu L}} = \left(\frac{B_{rt}^{\nu H}/B_{jt}^{\nu H}}{B_{rt}^{\nu L}/B_{jt}^{\nu L}}\right)^{\varepsilon}\left(\frac{y_{rt}^{H}/y_{jt}^{H}}{y_{rt}^{L}/y_{jt}^{L}}\right)^{\varepsilon} = \left(\frac{B_{rt}^{\nu H}/B_{jt}^{\nu H}}{B_{rt}^{\nu L}/B_{jt}^{\nu L}}\right)^{\varepsilon}\left(\frac{1+\frac{1}{\theta-1}e_{rt}^{H}}{1+\frac{1}{\theta-1}e_{jt}^{H}}\right)^{\varepsilon}, \tag{12}$$

where e_{rt}^{H} is the probability of a successful innovation in location r for skilled individuals of *any* nationality.

This expression highlights three properties about the endogenous sorting in our model. First, sorting can be driven by labor supply: if location r's amenities are particularly sought after by skilled individuals, the share of skilled workers will be higher. Second, the incentives to sort due to labor demand are fully summarized by the local innovation probabilities (which, of course, are endogenous). Skilled individuals are drawn to locations where it is likely that they can actually participate in the innovation process. These are locations that are intrinsically productive in innovation (high ζ_r), that have a large local idea pool (high N_{rt-1}), and where the competition for innovators is weak (low $\overline{h}_{rt}$). Third, sorting is independent of the location of origin o, reflecting our assumption that the bilateral moving costs, μ_{ro}, are independent of individuals' skill level.

3.5 Spatial Balanced Growth Path

Consider now a stationary equilibrium, or *spatial balanced growth path* (SBGP). Along a SBGP, the aggregate population $\ell_t \equiv \sum_r \ell_{rt}$ grows at a constant rate $\eta > 0$, exogenous productivity A_{rt} grows at rate g_A, and the distribution of economic activity is stationary. In Appendix Section A.1.2, we provide a formal characterization of the SBGP; here, we highlight three key properties.

First, local knowledge N_{rt} and wages w_{rt} grow at rates g_N and g_w, which are given by

$$1 + g_N = (1+\eta)^{\frac{1}{1-\vartheta}} \quad \text{and} \quad 1 + g_w = (1+g_A)(1+g_N)^{\frac{1}{\sigma-1}}.$$

As in Jones (1995), population growth η is the ultimate driver of long-run growth, with the scale elasticity ϑ amplifying the link between population and income growth. Intuitively, as stressed in (4), along a SBGP, the stock of knowledge is proportional to $\overline{h}_{rt}\ell_{rt}$. Since human capital $\overline{h}_{rt}$ is proportional to N_{rt}^{ϑ} by (5), we have $N_{rt} \propto \ell_{rt}^{\frac{1}{1-\vartheta}}$, so that the scale elasticity ϑ shapes the link between population and income growth. The mapping from variety growth to income growth is then determined by the elasticity of substitution. Without any inter-temporal elasticity $\vartheta = 0$, the effect of variety growth is simply $g_w \approx \frac{1}{\sigma-1}\eta$ as in the static model of Krugman (1980). With the possibility of building on the shoulders of local giants ($\vartheta > 0$), population growth has a multiplier $\frac{1}{\sigma-1}\frac{1}{1-\vartheta}\eta$ as in Jones (1995).

Our theory also implies that the share of innovators, e_{rt}, is common across locations and, as we show in Appendix Section A.1.2, given by

$$e_{rt} = e^* = \frac{1}{1 + \frac{\theta}{\theta-1}\frac{1+g_N}{g_N}(\sigma-1)\left(1 - \beta(1+g_A)(1+g_N)^{\frac{1}{\sigma-1}-\vartheta}\right)}.$$

Hence, locations with high innovation efficiency ζ_r or a large population ℓ_{rt} generate more ideas per person, *not* by having more innovators but rather by generating more ideas per innovator. The share of innovators e_r is (i) increasing in the growth rate g_N (as a higher

growth requires more innovation inputs), (ii) decreasing in the elasticity of substitution σ (as more substitutability lowers markups and hence the returns to innovation), and (iii) increasing in the tail of the Pareto distribution θ (as a thicker tail reduces the mass of "superstar" inventors, so that more people have to work in the idea-creation sector).

Finally, the human capital cutoff and the local stock of local knowledge are given by

$$\overline{h}_{rt} \propto \zeta_r^{\frac{1}{1-\vartheta}} \ell_{rt}^{\frac{\vartheta}{1-\vartheta}} \left(\sum_{\nu} \lambda_r^{\nu} \omega_r^{\nu} \right)^{\frac{1}{\theta}\frac{1}{1-\vartheta}} \quad \text{and} \quad N_{rt} \propto \overline{h}_{rt} \ell_{rt}$$

These expressions highlight the mechanics of local agglomeration. The local innovation cutoff and the stock of local knowledge are determined by three forces: (i) innovation efficiency ζ_r, (ii) the size of the local population ℓ_{rt}, and (iii) the share of skilled residents. The scale elasticity ϑ again emerges as the crucial elasticity and acts as an amplifying force. Immigrants, by shaping labor supply both across space and time, therefore affect the geography of economic growth.

4. Structural Estimation

We now structurally estimate our model. Specifically, we estimate the structural parameters $\mathscr{P}$, the set of region-specific local fundamentals $\mathscr{F} = \{A_r, \zeta_r, \mathcal{B}_r\}$ (local TFP, innovation efficiency, and amenities), and the initial conditions $\mathscr{I} = \{N_{r0}, \ell_{r0}^{\nu}, \lambda_{r0}^{\nu}\}$, that is, the prevailing stock of knowledge, the number of people of nationality ν, and their skill composition.

We take a location in our theory to be a State Economic Area (SEA), which ensures that boundary definitions are consistent throughout our sample period and that we observe patents in all locations. There are 452 SEAs in the US, and we use a crosswalk provided by IPUMS to aggregate county-level data to the SEA level. We treat 1880 as the first period of our analysis.

4.1 Data

We estimate our model using a combination of micro, macro, and spatial data. In addition to the matched data on individual patenting described in Section 2, we also employ (i) newly acquired microdata on immigration records from the Castle Garden Immigration Database (CG) and the Hamburg Passenger Lists (HPL), (ii) data on average wages at the county level from the Census of Manufactures, (iii) data on the population of patents from PatentCity, and (iv) time-series data on aggregate GDP per capita from Historical Statistics of the United States.

Immigration Records The CG data, compiled by the Battery Conservancy, encompasses a comprehensive list of all immigrants entering the United States via the port of New York between 1820 and 1914, totaling approximately 11 million individual records. The HPL

records all passengers departing from the port of Hamburg to the United States between 1850 and 1914, comprising approximately 6 million records.[7]

Crucially for our purposes, both datasets report immigrants' *pre*-migration occupations, information that we use to measure the skill content of international arrivals.[8] As with patents, we match these immigration records to the Population Census. We link all immigration records up until the respective Census year. For example, when linking the 1900 Census, we attempt to link all immigration records up until 1900. To do so, we insist that individuals in the Census and the immigration records have a matching foreign birthplace (at the country level) and race, and that the immigration year reported in the CG/HPL is within two years of the Census-recorded immigration year. For the entire four decades from 1880 to 1920, we match more than 600,000 immigrants between the ages of 16 and 60; see Appendix Table A.6 for a summary of the match rates.

Table 3 summarizes the distribution of pre-migration occupations by immigrants' nationality. The most striking pattern is the heterogeneity in the importance of skilled occupations such as craftsmen, managers, operatives, and technical or sales workers. British and German immigrants were much more likely to work in such occupations prior to migrating compared to Irish and Italian immigrants.[9] This pattern is consistent with the higher patenting rates of British and German immigrants documented in Table 2, suggesting that pre-migration skills shaped post-arrival innovative activity.

Wages and Aggregate GDP Because the Census microdata does not contain information on individual earnings, we measure average local earnings from the county-level tabulations of the Census of Manufactures for 1880, 1900, and 1920 from the NHGIS. We compute average wages by dividing total manufacturing payroll by total manufacturing employment and aggregate the county-level data to the SEA level. To measure aggregate GDP, we use the canonical Historical Statistics of the United States, which reports a time-series for real GDP growth.

Regional Data on Patenting We measure aggregate patenting at the regional level from the PatentCity Database (Bergeaud and Verluise, 2024). From these data, we compute the decadal flow of patents in region r between time t and $t+9$, $\mathcal{P}_{rt}$, and the stock of patents filed in location r up to time t, $\mathcal{S}_{rt}$, by cumulating the flow of patents, i.e., $\mathcal{S}_{rt} := \sum_{k=0}^{t-1} \mathcal{P}_{rk}$.

[7] We have complete access to the records of the Hamburg Passenger Lists through a collaborative arrangement with the Hamburg State Archive.

[8] Since the original databases presented information as unorganized string variables, we created a crosswalk to the official occupational classification used by the US Census. This process utilized the lexical database "WordNet" from the NLTK library in Python. WordNet, with its extensive network of synonyms arranged hierarchically, allowed us to classify immigrants' pre-migration occupations based on shared synonyms with the US Census system. For a comprehensive explanation of our occupational classification methodology, please refer to Section OA.1.3 in the Online Appendix.

[9] Note that we classify individuals with missing information on pre-migration occupations as unskilled. We do so based on the assumption that skilled individuals are more likely (and able) to document this information in their arrival records. For our estimation, we use this information only for the initial sorting upon arrival (see (18)), not to estimate migrants' skill content.

TABLE 3: PRE-MIGRATION SKILLS AND THE COMPOSITION OF IMMIGRATION

Occupation	Immigrants	Distribution of Occupations (%)				
		Immigrants	British	German	Italian	Irish
Craftsmen	67,488	20.4	20.9	23.3	14.8	9.5
Managers, Officials, Proprietors, and Clerks	6,629	2.0	3.3	2.1	0.5	1.3
Operatives	5,321	1.6	2.3	1.5	2.2	0.8
Professional and Technical	7,029	2.1	2.6	2.4	1.2	0.7
Sales Workers	26,399	8.0	5.1	10.5	1.7	2.1
Farmers and Farm laborers	79,353	24.0	8.6	29.9	28.1	8.9
Laborers	2,005	0.6	0.6	0.7	0.7	0.2
Service Workers	6,067	1.8	1.1	1.8	3.7	1.9
Other	130,951	39.5	55.5	27.7	47.0	74.6
Number of People	.	331,242	42,188	191,779	29,701	26,407
Immigrant Group Share	.	.	12.7	57.9	9.0	8.0

Notes: The table reports the distribution of immigrants according to their pre-migration occupations. The "Other" category captures observations with missing or unreported occupation data.

4.2 Measurement: Patents, Ideas, and Innovative Skills

We relate patenting in the data to the creation of varieties in our theory. To operationalize this mapping, we assume that the flow of patents of individual i in region r at time t, $\mathcal{P}_{rt}^i$, is proportional to the number of created varieties conditional on patenting, i.e.,

$$\mathcal{P}_{rt}^i = \begin{cases} 0 & \text{if } h^i < \bar{h}_{rt} \\ \varsigma h^i & \text{if } h^i \geq \bar{h}_{rt} \end{cases}, \tag{13}$$

where $\varsigma > 0$ is a constant of proportionality. If $\varsigma < 1$, not every productive idea is patented; if $\varsigma > 1$, it takes multiple patents to implement a marketable idea. As we show below, we do not have to take a stand on ς to estimate our model.

Furthermore, we need to take into account that we only successfully match a subset of patents to individuals. We thus assume that individual i is matched to their patents with probability q_t^M.[10] We calibrate q_t^M to ensure that our model is consistent with both the micro and macro data on patenting and hence set q_t^M to the match rates reported in Table 1. The share of people with a patent in the model, e_{rt}, is then related to the measured share of patent holders, $\mathcal{M}_{rt}$, according to $\mathcal{M}_{rt} = q_t^M e_{rt}$. To discipline the spatial distribution of skilled individuals, λ_{rt}^ν, we utilize occupations as reported in the US Census as a proxy for innovative skills. Specifically, as in Table 3, we define individuals as working in high-skilled occupations if they work in professional, technical, and managerial occupations or as craftsmen.

[10] Conditional on being matched, we assume that the individual is matched to all patents they issued, because we match on the individual's name. Formally, letting M_i denote an indicator variable that takes the value one if individual i is matched to a patent, we assume that $M_{it} = 1$ with probability $q_t^M P\left(\mathcal{P}_{rt}^i > 0\right)$ and $M_{it} = 0$ with probability $1 - q_t^M P\left(\mathcal{P}_{rt}^i > 0\right)$.

4.3 Estimation of Structural Parameters $\mathscr{P}$

Even though many of our parameters are estimated jointly, our theory allows for a tight mapping between specific structural parameters and empirical moments. We describe this mapping here.

We calibrate three parameters exogenously from the literature. We set $\sigma = 5$, which lies between the median and average estimates reported by Broda and Weinstein (2006) and within the range reported by Oberfield and Raval (2021). For the migration elasticity, we target $\varepsilon = 2$, in line with estimates reported in Fajgelbaum et al. (2019) and Peters (2022). For investors' rate of time preference β we assume an annual discount rate of 10%. Given that we take each period in the model to be a decade, this implies that $\beta = 0.9^{10} \approx 0.35$.

The dispersion of human capital: θ To estimate θ, we exploit information on the distribution of the number of patents. Let $Q_{rt}^{p|\nu}$ be the share of innovators of nationality ν in region r at time t with at least p patents. Our theory implies that

$$Q_{rt}^{p|\nu} \equiv P_{rt}^{\nu}\left(\mathcal{P}_i \geq p | \mathcal{P}_i > 0\right) = P_{rt}\left(h \geq \varsigma^{-1} p | h \geq \overline{h}_{rt}\right) = \left(\varsigma \frac{\overline{h}_{rt}}{p}\right)^{\theta},$$

where ς is the constant of proportionality defined in the measurement equation (13). Note that $Q_{rt}^{p|\nu}$ does not vary across types ν within a location r once we condition on patent holders. Hence, we estimate θ from the regression

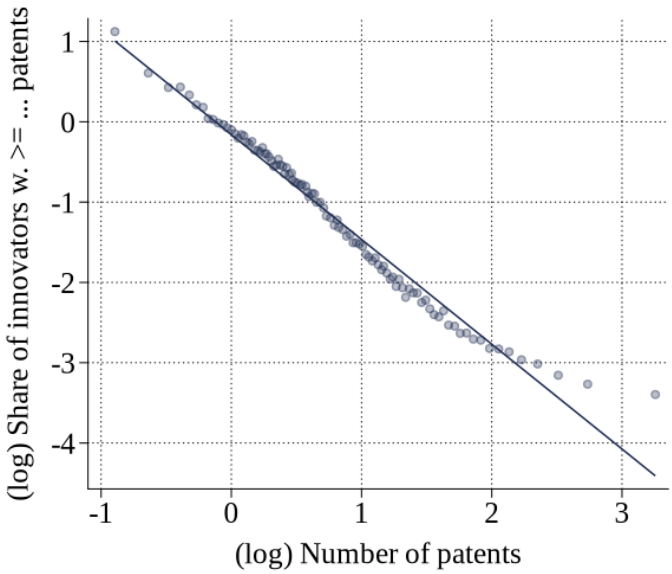
$$\ln Q_{rt}^{p|\nu} = \delta_{rt} - \theta \ln p + u_{rt}^{p\nu}, \tag{14}$$

where δ_{rt} is a region-year fixed effect. The relationship between $\ln Q_{rt}^{p|\nu}$ and $\ln p$ should be linear, and θ is exactly the slope coefficient. To implement (14), we compute $Q_{rt}^{p|\nu}$ for all ν-groups within all localities for $p = 1, 2, 3, \ldots$. In Figure 2 we depict a binscatter of the relationship in (14) after controlling for a set of region-year fixed effects. Consistent with the theory, the estimates indicate a tightly estimated and approximately linear relationship, except at the upper tail where the slope flattens. Quantitatively, the slope in Figure 2 implies that $\theta = 1.3$. In Section A.3.1 in the Appendix, we report the regression results of (14) for a variety of specifications.

Scale Effects: ϑ To estimate the strength of scale effects ϑ, we exploit the autocorrelation between the local patent stock and the flow of new patents. Our model implies that (see equations (3) and (4)) the number of new ideas per capita is given by

$$\frac{I_{rt}}{\ell_{rt}} = \frac{\theta}{\theta - 1} e_{rt}^{\frac{\theta-1}{\theta}} \mathcal{H}_{rt} = \frac{\theta}{\theta - 1} e_{rt}^{\frac{\theta-1}{\theta}} \zeta_r N_{rt-1}^{\vartheta} \left(\sum_{\nu} \lambda_{rt}^{\nu} \omega_{rt}^{\nu}\right)^{1/\theta}.$$

FIGURE 2: ESTIMATION OF PARETO TAIL: θ



Notes: The figure contains a binscatter plot of (14) after controlling for a full set of SEA-Decade fixed effects.

Using the measurement equation (13), we can thus relate the flow of patents per capita in region r, $\mathcal{P}_{rt}/\ell_{rt}$, to the stock of patents, $\mathcal{S}_{rt}$, to arrive at the empirical relationship,

$$\ln\left(\mathcal{P}_{rt}/\ell_{rt}\right) = \alpha + \ln\zeta_r + \vartheta \ln \mathcal{S}_{rt} + \left(\frac{\theta-1}{\theta}\right)\ln e_{rt} + \frac{1}{\theta}\ln\left(\sum_{\nu}\lambda_{rt}^{\nu}\varpi_{rt}^{\nu}\right).$$

The variation in patents per capita is due to four sources: (i) systematic differences in regional patent productivity (ζ_r), (ii) local learning opportunities related to the existing stock of ideas $(\mathcal{S}_{rt})$, (iii) the profitability of variety creation captured by the share of innovators e_{rt}, and (iv) the supply of human capital as summarized by the composition of the local population $(\lambda_{rt}^{\nu}\varpi_{rt}^{\nu})$.

To turn this relationship into a regression, we estimate it at the region-*industry* level to control for differences in the local industrial composition. Specifically, our estimation equation is

$$\ln\left(\mathcal{P}_{rjt}\right) = \delta_{jt} + \delta_r + \vartheta \ln \mathcal{S}_{rjt} + \phi \ln e_{rt} + \xi \ln \ell_{rt} + \rho \ln\left(\sum_{\nu}\lambda_{rt}^{\nu}\varpi_{rt}^{\nu}\right) + u_{rjt}, \tag{15}$$

where $\mathcal{P}_{rjt}$ and $\mathcal{S}_{rjt-1}$ are the patent flows and stocks in industry j in region r at time t, δ_{jt} is an industry-time fixed effect to control for industry-specific trends and δ_r is a region fixed effect. Hence, we identify ϑ from the within-industry relationship between changes in local patent flows and patent stocks, controlling for the local population composition and the local innovator share, which vary over time. In Section 4.4 below we discuss in detail how we measure the skill supply $\sum_{\nu}\lambda_{rt}^{\nu}\varpi_{rt}^{\nu}$.

TABLE 4: ESTIMATION OF SCALE EFFECTS: ϑ

	(1)	(2)	(3)	(4)	(5)	(6)
log Patent stock	0.445***	0.442***	0.401***	0.490***	0.509***	0.470***
	(0.029)	(0.029)	(0.044)	(0.025)	(0.058)	(0.041)
SEA FE	yes	yes	yes	yes	yes	yes
Industry-Year FE	yes	yes	yes	yes	yes	yes
Skill composition	no	yes	yes	yes	yes	yes
Share of innovators	no	yes	yes	yes	yes	yes
Unit elasticity of ℓ	no	no	yes	no	no	no
Estimator	OLS	OLS	OLS	IV	OLS	OLS
Industry Def.	3 dig	3 dig	3 dig	3 dig	1 dig	2 dig
R^2	.773	.774	.751	.214	.902	.825
N	52,682	52,516	52,516	36,676	8,861	33,181

Notes: This table contains the estimates of ϑ in regression (15). All specifications control for SEA and Industry-Year fixed effects. Industries are defined at the three-digit level. From columns 2 to 6, we also control for the skill supply, $\sum_\nu \lambda_{rt}^\nu \varpi_{rt}^\nu$. In column 3, we impose the unit elasticity of ℓ. In column 4 we report the IV estimates. In columns 5 and 6, we classify industries at the one- and two-digit level. Standard errors are clustered at the SEA-Year level.

We estimate (15) using both OLS and a Bartik IV strategy, where we construct the predicted patent stock in industry j in region r, $\mathcal{S}_{rjt}$, from the 1880 cross-sectional distribution of patents across region-industry cells and the aggregate number of patents in each industry. Identification requires the initial 1880 distribution to be orthogonal to subsequent region-industry-specific shocks to patent creation.

In Table 4 we report the results. Column 1 includes SEA and industry-year effects plus $\ln \ell_{rt}$ and yields a scale elasticity of 0.45. Column 2 adds the overall skill supply $\ln\left(\sum_\nu \varpi_{rt}^\nu \lambda_{rt}^\nu\right)$ and the local innovator share with essentially no change in the estimated scale elasticity. Column 3 imposes $\xi = 1$ (unit elasticity of population). Column 4 reports the Bartik IV estimate, which is somewhat larger but qualitatively very similar. Columns 5 and 6 show that our results are robust to defining industries at the 1- or 2-digit level. For our quantitative analysis, we set $\vartheta = 0.45$.

Local Amenities: Congestion, Homophily, and Urban Bias We posit that local amenities, $B_{rt}^{\nu s}$, have four underlying determinants and parametrize them as follows:

$$B_{rt}^{\nu s} = \mathcal{B}_r \ell_{rt}^{-\varrho} \left(\varpi_{rt}^\nu\right)^\iota \left(1 + \text{urb}_{rt}\right)^{\mathbf{1}\{\nu=\text{Imm}\}\gamma^I + \mathbf{1}\{\nu=\text{Skilled}\}\gamma^S}. \tag{16}$$

Here, $\mathcal{B}_r$ is an exogenous local fundamental, $\ell_{rt}^{-\varrho}$ summarizes local congestion due to a larger population, ϖ_{rt}^ν is the population share of type ν in location r and captures nationality-specific homophily as in Altonji and Card (1991) and Burchardi et al. (2026), and urb_{rt} is the measured urbanization rate, allowing for the fact that urban locations can be particularly attractive for immigrants ($\gamma^I > 0$) or skilled individuals ($\gamma^S > 0$).

To derive an estimation equation for the homophily elasticity ι, we take the logs of

TABLE 5: ESTIMATION OF HOMOPHILY AND MIGRATION ELASTICITY

	(1)	(2)	(3)	(4)
Log Nationality Share	0.216***	0.219***	0.263***	0.419***
	(0.019)	(0.019)	(0.018)	(0.086)
Log Trade Cost	-3.050***	-3.051***	-3.152***	-3.145***
	(0.418)	(0.418)	(0.401)	(0.403)
Innovation Share		3.870**	6.018**	17.610*
		(1.322)	(1.713)	(7.227)
Origin-Type-Year FE	Yes	Yes	Yes	Yes
Destination–Year FE	Yes	Yes	Yes	Yes
IV	No	No	Card	Burchardi et al.
R^2	.847	.847	.168	.154
N	184,158	184,158	174,461	171,387

Notes: The table reports estimates of equation (17). Columns (1) and (2) present OLS estimates. Column (3) presents estimates using the instrument from Card (2001). Column (4) presents estimates using the extension of the Card instrument proposed by Burchardi et al. (2026). Details for the construction of both instruments are reported in Appendix A.3.1.

expression (9) and consider a fixed-effects regression of the form:

$$\ln m^{\nu}_{rot} = \delta_{rt} + \delta^{\nu}_{ot} + \beta_1 e^{\nu}_{rt} + \iota\varepsilon \times \ln \varpi^{\nu}_{rt} + \varepsilon \ln \mu_{rot} + u^{\nu}_{rot}, \tag{17}$$

where m^{ν}_{rot} is the share of migrants of nationality ν from origin o choosing to migrate to region r and δ's are fixed effects, which absorb a number of endogenous variables, including total population and local wage.[11] Since we lack specific information on the ratio $\psi_{rt}/\bar{h}_{rt}$ in equation (9), we approximate it with the share of innovators among each nationality in the region, e^{ν}_{rt}. The coefficient on ϖ^{n}_{rt}, the population share of nationality n in location r, allows us to identify the product of the migration elasticity ε and the homophily elasticity ι.

To implement this empirical specification, we link the individual-level census data from 1880 to 1920 over time and construct nationality-specific migration flows across SEAs (see Appendix Section A.2). In column 1 of Table 5, we report the baseline OLS regression, including the full set of fixed effects. The coefficient on $\ln \varpi^{n}_{rt}$ is 0.216. To address endogeneity, we instrument ϖ^{n}_{rt} using (i) the shift-share instrument of Card (2001), which interacts total immigration from each nationality with its historical spatial distribution in 1880, 1900, and 1910, and (ii) the predicted ancestry distribution of Burchardi et al. (2026). The Card instrument is our preferred specification, since the pre-1920 sample is relatively short for a reliable first stage of the Burchardi et al. (2026) instrument. These IV strategies raise the coefficient to 0.263 (column 3) and 0.420 (column 4). Given our calibration of $\varepsilon = 2$, we choose $\iota = 0.13$ for our quantitative analysis.

[11] Note that, strictly speaking, (16) requires that δ_{rt} would be skill specific. Given that we observe migrants' pre-migration occupations only for a limited number of individuals, we estimate (17) for the entire sample and do not control for an interaction of the destination fixed effect and individual skills.

We estimate the parameters determining the urban bias for immigrants (γ^I) and skilled individuals (γ^S) within our model. Specifically, we choose these parameters to match the share of immigrants and the share of skilled individuals in "urban hubs" (locations with an urbanization rate above 80%) relative to rural locations (regions with an urbanization rate below 20%) in 1920. That way we guarantee that our model replicates the salient features of how immigrants and skilled individuals sort across space. Doing so yields $\gamma^I = 0.98$ and $\gamma^S = 0.76$, indicating that urban amenities are particularly valuable for skilled individuals and immigrants. For the congestion elasticity, we set $\varrho = 0.3$, corresponding to the typical expenditure share on housing, which we take to capture a key channel of local congestion.

Spatial Frictions: Trade, Mobility Costs, and Moving Opportunities In our model, both the flow of goods and people are subject to frictions. For the bilateral trade costs τ_{ij}, we draw on the measured trade costs constructed by Hornbeck and Rotemberg (2024) for the period 1880–1920, which are based on waterway and railroad distances between counties. We aggregate their county-level estimates to the SEA level using county population as weights. We then parameterize migration costs as a power function of trade costs, $\mu_{ij} = \tau_{ij}^{\alpha_\mu}$, and leverage the relationship implied by equation (17). As seen in Table 5, across specifications, the point estimates of $\varepsilon\alpha_\mu$ range from -3.05 to -3.15. Given that we target $\varepsilon = 2$, we adopt $\alpha_\mu = -1.5$ as our baseline value.

The remaining parameter governing spatial mobility is the probability of receiving a moving opportunity, χ. In our data, approximately 30 percent of the population moved across SEA boundaries over this period. Conditional on the other parameters, we calibrate χ to ensure that in our model 30 percent of individuals relocate across SEAs.

International Arrivals: $M_{rt}^{\nu s}$ Our theory takes the inflow of migrants, $M_{rt}^{\nu s}$, as exogenous. We model this initial allocation of migrants upon arrival parallel to our treatment of amenities in (16) and assume that

$$M_{rt}^{\nu s} = M_t^\nu \lambda_t^{\nu s} \frac{(1+\text{urb}_{rt})^{\alpha^s} \ell_{rt}^{\xi} (\omega_{rt}^\nu)^\varsigma}{\sum_k (1+\text{urb}_{kt})^{\alpha^s} \ell_{kt}^{\xi} (\omega_{kt}^\nu)^\varsigma}, \tag{18}$$

where M_t^ν is the aggregate mass of migrants of nationality ν arriving in year t, and $\lambda_t^{\nu s}$ is the share with skill s. The parameters α^s, ξ, and ς thus govern the extent to which individuals that arrive in the US sort based on the local urbanization rate, overall size, and the share of individuals of their own nationality ν. In line with (16) we allow the effect of the urbanization rate to differ between skilled and unskilled immigrants. Equation (18) is purely a measurement equation; after arrival, immigrants face the same migration opportunities as natives. Using (18), we can estimate $(\alpha^s, \xi, \varsigma)$ from the regression

$$\ln M_{rt}^{\nu s} = \delta_t^{\nu s} + \alpha^s \ln(1+\text{urb}_{rt}) + \xi \ln \ell_{rt} + \varsigma \ln \omega_{rt}^\nu + u_{rt}^{\nu s}. \tag{19}$$

We construct $M_{rt}^{\nu s}$ from our matched data between the Census and immigration records. As in Table 3, we classify migrants with professional, technical, managerial, or craftsmen pre-migration occupations as skilled, and measure $M_{rt}^{\nu H}$ ($M_{rt}^{\nu L}$) as the mass of high-skilled (low-skilled) nationality-ν immigrants in region r arriving within the last 5 years. As we show in Appendix Section A.3.2, we find that $0 < \alpha^L < \alpha^H$, $\xi > 0$, and $\varsigma > 0$, that is, international arrivals feature a bias towards heavily populated and urban locations (especially among the skilled), and nation-specific homophily plays an important role, highlighting the importance of existing immigration networks for new arrivals.

With the estimates $(\alpha^s, \xi, \varsigma)$ at hand, we can compute $M_{rt}^{\nu s}$ from (18) for our entire time series, given measures of $\{\text{urb}_{rt}, \ell_{rt}, \varpi_{rt}^{\nu}, M_t^{\nu}, \lambda_t^{\nu s}\}_t$. From the Census data for the years 1880–2010, we can directly compute $\{\text{urb}_{rt}, \ell_{rt}, \varpi_{rt}^{\nu}\}_t$. For most years, we can also directly measure the flow of new immigrant arrivals, M_t^{ν}. However, in the years 1940, 1950, and 1960 the Census does not contain information on immigrants' year of arrival. As we explain in detail in Appendix Section A.3.1, we therefore rely on information from the Department of Homeland Security for these decades. Finally, we assume that the aggregate skill composition $\lambda^{\nu s}$ is time-invariant and describe in detail below how we estimate it.

Demographics: η, η^B, and d Given the inflows $M_{rt}^{\nu s}$ and the observed initial population in 1880, we calibrate the birth and death rates η^B and d to (i) match population growth between 1880 and 2000 and (ii) the decadal time-series of the aggregate share of immigrants between 1880 and 2000. We also need to take a stand on the long-run growth rate of immigrant inflows η. While η does not affect any allocations between 1880 and 2000 directly, it determines the long-run BGP, because η is also the rate of population growth in the long run. We target an annual growth rate of 1.2%, which is in line with the observed population growth between 1880 and 2000, and hence set $M_{2000+10t}^{\nu s} = (1+\eta)^t M_{2000}^{\nu s}$. Furthermore, we assume that the initial spatial allocation of arrivals remains constant at its 2000 value: $M_{rt}^{\nu s} / M_t^{\nu} = M_{r2000}^{\nu s} / M_{2000}^{\nu}$.

Exogenous technological progress g_A and the cardinality of knowledge Our model features two sources of growth: the endogenous accumulation of ideas and exogenous growth in overall productivity g_A. Because our model features transitional dynamics, the level of the initial knowledge stock, N_{r0}, is consequential: holding everything else constant, starting from a lower level of N_{r1880} increases overall growth because the economy is further away from a balanced growth path. We therefore set the initial stock of knowledge, N_{r0}, as proportional to the observed local patent stocks, $N_{r1880} = \overline{N} \mathcal{S}_{r1880}$, and treat $\overline{N}$ as a structural parameter that we estimate. Specifically, we estimate $\overline{N}$ and g_A by targeting two moments: GDP pc in 2000 (relative to 1880), y_{2000}/y_{1880}, and average patent growth between 1880 and 1910. While we calibrate these parameters jointly, g_A is mostly informed by the long-run growth of GDP pc and $\overline{N}$ is mostly informed by the growth of patents.

4.4 Local fundamentals $\mathscr{F}$ and Initial Conditions $\mathscr{I}$

Next, we explain how, given the values of the structural parameters, we identify the fundamentals $\mathscr{F} = \{A_r, \zeta_r, \mathcal{B}_r\}$ and the initial conditions $\mathscr{I} = \left\{N_{r1880}, \ell^{\nu}_{r1880}, \lambda^{\nu}_{r1880}\right\}$.

Initial conditions $\mathscr{I} = \left\{N_{r1880}, \ell^{\nu}_{r1880}, \lambda^{\nu}_{r1880}\right\}$ As explained above, we tie the initial stock of knowledge, N_{r1880}, directly to the observed stock of patents in 1880 and the estimated scaling parameter $\overline{N}$. Similarly, the population distribution, ℓ^{ν}_{r1880}, is directly observed in the 1880 US Census.

To estimate the relative skill supply we rely on observed patent outcomes. Our theory highlights that relative innovation propensities *within* labor markets are informative about relative skill supplies, i.e., $e^{\nu}_{rt}/e^{US}_{rt} = \lambda^{\nu}_{rt}/\lambda^{US}_{rt}$. To implement this idea empirically, we assume that the share of high-skilled individuals of nationality ν in location r is proportional to that of natives, i.e.[12]

$$\lambda^{\nu}_{rt} = \kappa^{\nu}\lambda^{US}_{rt}. \tag{20}$$

Hence, Italian immigrants might have fewer skilled workers overall, and they might live in different locations; however, their sorting across space is the same as that of natives. As a consequence, the share of innovators among nationality ν in location r at time t satisfies

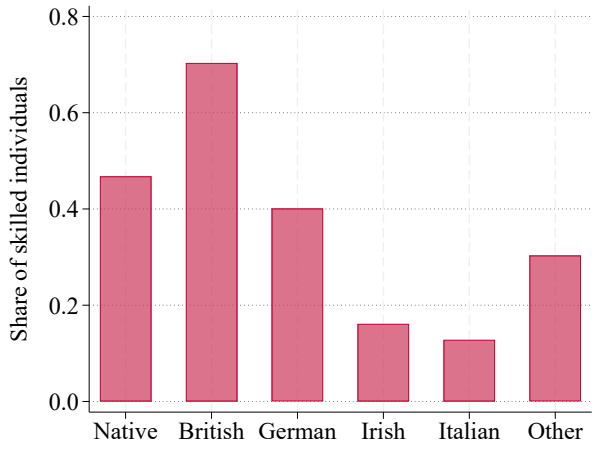
$$\ln e^{\nu}_{rt} = \ln\left(\lambda^{\nu}_{rt}\left(\frac{\psi_{rt}}{\bar{h}_{rt}}\right)^{\theta}\right) = \ln\kappa^{\nu} + \ln\left(\lambda^{US}_{rt}\left(\frac{\psi_{rt}}{\bar{h}_{rt}}\right)^{\theta}\right) = \ln\kappa^{\nu} + \delta_{rt}. \tag{21}$$

Hence, we can identify the relative skill supply of all nationalities, κ^{ν}, from a regression of (the log of) the share of patent holders on a nationality fixed effect and a location-time fixed effect. Note that this is exactly the variation we exploited in Section 2 to estimate migrants' innovativeness within local labor markets. In particular, the region-time fixed effect δ_{rt} controls for exogenous innovation TFP ζ_r, the endogenous variables N_{rt-1} and $\bar{h}_{rt}$, and the extent of spatial sorting of natives λ^{US}_{rt}. Moreover, (21) identifies the share of skilled individuals entirely from the extensive margin of patenting and does not rely on a particular cardinality of patents.

In Appendix Table A.9, we report the estimates of equation (21) for a variety of specifications. They mirror the results in Table 2: relative to natives, British immigrants are more likely to patent, German immigrants slightly less so, and Italian and Irish immigrants substantially less. Our theory thus provides a clean structural interpretation of these estimates: the estimated coefficients identify the relative skill supply κ^{ν}.

[12] Note that this assumption is in line with our migration module. Equation (16) implies that $\frac{B^{\nu H}_{rt}/B^{\nu H}_{jt}}{B^{\nu L}_{rt}/B^{\nu L}_{jt}} = \left(\frac{1+\text{urb}_{rt}}{1+\text{urb}_{jt}}\right)^{\gamma^S}$. Equation (12) therefore implies the relative moving probabilities by skill are equalized across nationalities. In the absence of new migration inflows, (20) would therefore hold along an SBGP.

FIGURE 3: SHARE OF SKILLED INDIVIDUALS IN 1880



Notes: In this figure we depict the share of skilled individuals by nationality in 1880.

Given the estimates for κ^ν we can use (20) to compute the local skill supply as a function of the skill supply of natives. To measure λ^{US}_{r1880} we use occupational data for US-born individuals in the Census and set it equal to the share of US-born individuals in region r working in high-skilled occupations in the non-agricultural sector. By using only natives' occupations we avoid concerns that immigrants' occupations may be distorted by discrimination or occupational downgrading. Furthermore, by focusing on the occupational distribution outside of agriculture, we avoid conflating spatial differences in skill supplies with differences in the sectoral composition.

In Figure 3 we report the implied aggregate share of skilled individuals by nationality, $\lambda^\nu_{1880} = (\sum_r \kappa^\nu \lambda^{US}_{r1880} \ell^\nu_{r1880}) / \sum_r \ell^\nu_{r1880}$. Among natives, our procedure leads us to infer that around 40% of individuals are skilled and have the ability to innovate (of course, in equilibrium only a small fraction actually engages in innovation). Among British immigrants, the share of skilled individuals is larger and equal to about 70%. By contrast, Italian and Irish immigrants are mostly unskilled.

Local Fundamentals $\mathscr{F} = \{\zeta_r, A_r, \mathcal{B}_r\}$ To identify each region's innovation TFP ζ_r, note that the endogenous innovation cutoff $\overline{h}_{rt}$ is proportional to the average number of patents per innovator

$$\overline{h}_{rt} = \frac{\theta - 1}{\theta} \frac{I_{rt}}{e_{rt} \ell_{rt}} = \frac{q^M}{\varsigma} \frac{\theta - 1}{\theta} \frac{\mathcal{P}_{rt}}{\mathcal{M}_{rt} \ell_{rt}},$$

where the second equality exploits the measurement equations $\mathcal{M}_{rt} = q^M e_{rt}$ and $\mathcal{P}_{rt} = \varsigma I_{rt}$. Using the expression for $\overline{h}_{rt}$ and exploiting equation (20), we can express the share of innovators in location r as

$$e_{rt} = \sum_\nu \omega^\nu_{rt} \lambda^\nu_{rt} \left(\frac{\psi_{rt}}{\overline{h}_{rt}} \right)^\theta = \psi^\theta_{rt} \left(\frac{\theta \varsigma}{q^M (\theta - 1)} \right)^\theta \lambda^{US}_{rt} \left(\frac{\mathcal{M}_{rt} \ell_{rt}}{\mathcal{P}_{rt}} \right)^\theta \left(\sum_\nu \omega^\nu_{rt} \kappa^\nu \right). \tag{22}$$

Because we already estimated ϑ, θ, and κ^{ν}, equation (22) expresses each location's innovation productivity ψ_{rt} directly as a function of observables. Note, in particular, that identification uses both the observed number of patents *per innovator* and the share of people who patent — two moments that require our matched data between patents and the US Census.[13] In a second step, we can then use the identity $\psi_{rt} = \zeta_r N_{rt-1}^{\vartheta}$ to decompose ψ_{rt} into the variation explained by the local knowledge stock, N_{rt-1}^{ϑ}, and the variation that is, residually, attributed to exogenous research TFP ζ_r. Note that the relative importance of local knowledge is modulated by the scale elasticity ϑ. Without intertemporal spillovers, $\vartheta = 0$, ζ_r has to do all the work. If $\vartheta > 0$, the measured patent stock has the potential to explain the estimated variation in innovative human capital ψ_{rt}.

To identify the distribution of local productivity A_r we rely on observed wages w_{rt}. Using the observed distribution of patent stocks, N_{rt}, and production workers, $\ell_{rt}(1 - e_{rt})$, we can compute $\{A_r\}_r$ directly from (7). Note that this computation does not involve any measures of skill shares or innovation human capital.

Finally, we identify the exogenous level of local amenities, $\mathcal{B}_r$, from the observed population distribution of unskilled natives. Using (9) and (16), the moving probabilities can be computed as

$$m_{rot}^{USL} = \frac{\left(\mu_{ro}\mathcal{B}_r L_{rt}^{-\varrho}\left(\varpi_{rt}^{US}\right)^{\iota} w_{rt}/P_{rt}\right)^{\varepsilon}}{\sum_k \left(\mu_{ko}\mathcal{B}_k L_{kt}^{-\varrho}\left(\varpi_{kt}^{US}\right)^{\iota} w_{kt}/P_{kt}\right)^{\varepsilon}}. \tag{23}$$

Given the structure of the model, all components of the migration probabilities are observed except for $\mathcal{B}_r$. Hence, for any candidate vector $\{\mathcal{B}_r\}$, the implied migration probabilities $m_{rot}^{US,L}(\{\mathcal{B}_r\})$, together with the laws of motion in equations (10) and (11), determine the mass of low-skilled natives in each region in the next period, $\ell_{r,t+1}^{US,L}$, given the current distribution $\ell_{rt}^{US,L}$.[14] We thus recover $\{\mathcal{B}_r\}$ by requiring that the observed distribution of low-skilled natives across regions, $\ell_{rt}^{US,L}$, is consistent with equation (23).

Note that (7), (22), and (23) should, in principle, hold at every point in time. Given the time-invariance of the local fundamentals $\{\zeta_r, A_r, \mathcal{B}_r\}$, this will generically not be the case. We therefore solve these equations independently for all decades between 1880 and 1920, and then recover $\{\zeta_r, A_r, \mathcal{B}_r\}$ as the regional fixed effect from these estimates — see Appendix Section A.3.1 for details. The resulting calibrated model therefore does not replicate the cross-sectional data exactly. However, we show below that it still provides

[13] Note also that the cardinality of the skill supply, λ_{rt}^{US}, is not essential for identification. If we had assumed that λ_{rt}^{US} was proportional to the share of skilled natives in the manufacturing sectors, we would have simply changed the level of ψ_{rt}: twice as many potential innovators, or each twice as likely to innovate, would yield observationally equivalent outcomes.

[14] Here $\ell_{r,t+1}^{US,L}$ denotes the population after migration but before natural population changes from births and deaths.

an excellent fit while reducing sensitivity to measurement error in regional wages or innovation shares.

4.5 Results and Model Fit

In Table 6 we summarize the estimated structural parameters. As discussed in the main text, we estimate a scale elasticity of around 0.45, which implies that our model features semi-endogenous growth and is consistent with a stationary distribution of economic activity across space. It also implies a direct decomposition of aggregate growth along the BGP:

$$\underbrace{g}_{1.38\%} = \underbrace{g_A}_{0.83\%} + \underbrace{\frac{1}{1-\vartheta}\frac{1}{\sigma-1}\eta}_{0.55\%}$$

Hence, in the long run, roughly 40% of long-run growth is driven by the accumulation of ideas, with the rest attributed to sources that we take to be exogenous and that do not respond to changes in immigration policy. The estimate $g_A = 0.83\%$ is required for our model to match income per capita growth between 1880 and 2000.

In terms of the underlying supply of human capital, our estimate of $\theta \approx 1.3$ implies that the distribution of innovative talent is very thick-tailed. Given that, on average, 7.7% of all high-skilled individuals become innovators, they earn, on average between 1880 and 1920, a premium of $1 + \frac{1}{\theta-1} e_{rt}^H \approx 1.26$ relative to unskilled individuals. The observed differences in patent probabilities across nationalities within locations imply substantial differences in relative skill supplies κ^ν.

Regarding the spatial labor supply function, we estimate a strong role of homophily towards one's own nationality, especially among recent arrivals ($\varsigma > \iota$). We also estimate an urban amenity premium for immigrants and skilled individuals ($\gamma^I > 0$ and $\gamma^S > 0$), which is required to match the empirically observed degree of urban sorting by skilled individuals and by immigrants.

In Figure 4 we display the cross-sectional fit of our model. Our model replicates the spatial patterns of economic activity in 1900 very well: urban locations are more populous and richer, generate more patents per capita, and contain a higher share of skilled individuals among natives.

In Figure 5 we report the cross-sectional variation of the estimated time-invariant fundamentals (A_r, $\mathcal{B}_r$, ζ_r) and the initial knowledge stock $\ln N_{r1880}^{\vartheta}$. To highlight the systematic differences between rural and urban areas, we construct population-weighted deciles by urbanization (so that each decile accounts for 10% of the population) and then compute averages within each decile (again, weighted by population).[15] For each outcome, we normalize the lowest urban decile to unity and hence plot the premium relative to the most rural locations.

[15] In the Appendix, we report the estimates for the fundamentals $(\zeta_r, A_r, \mathcal{B}_r)$ for each location r.

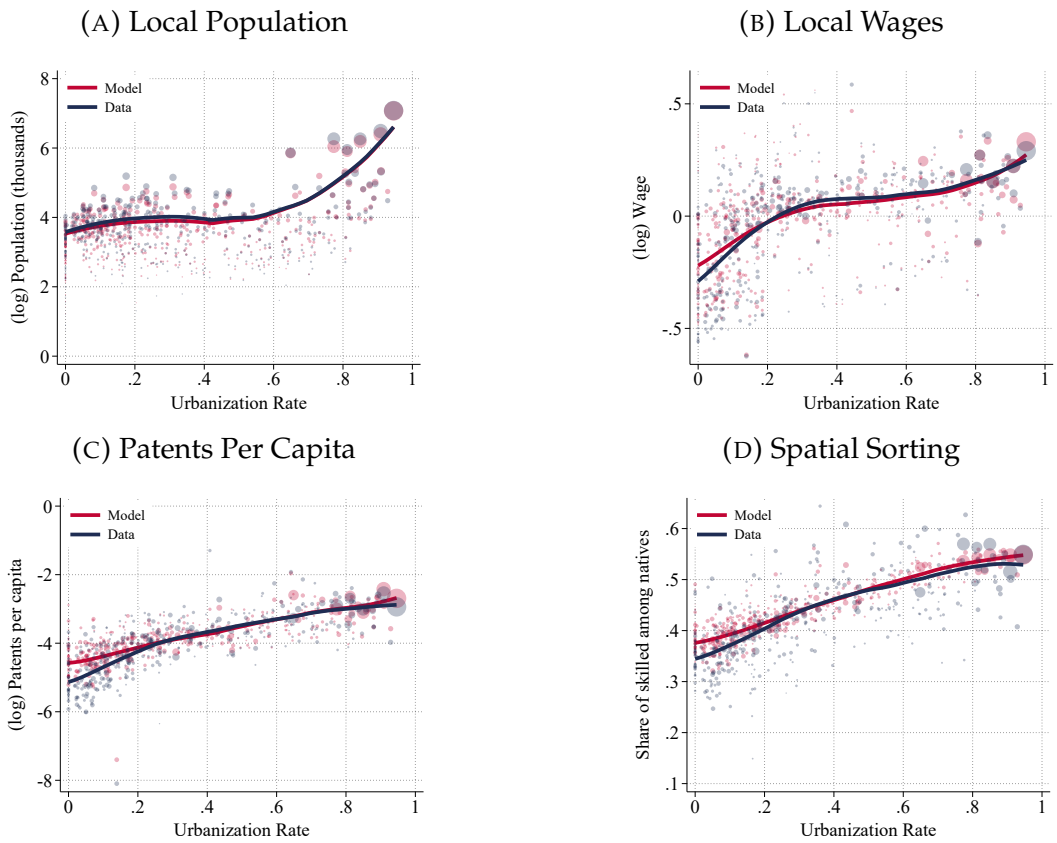
TABLE 6: STRUCTURAL PARAMETERS

Structural Parameters			Source/Target
ϑ	Scale elasticity	0.45	Table 4
θ	Human capital distribution	1.3	Figure 2
κ^{UK}	Share of skilled British (rel. to natives)	1.30	Table A.9
κ^{GER}	Share of skilled Germans (rel. to natives)	0.76	Table A.9
κ^{ITA}	Share of skilled Italians (rel. to natives)	0.23	Table A.9
κ^{IRL}	Share of skilled Irish (rel. to natives)	0.28	Table A.9
κ^{OTH}	Share of skilled other immig. (rel. to natives)	0.59	Table A.9
ι	Own-nationality homophily	0.13	Table 5
χ	Mobility hazard	0.304	Cross-SEA moving rate
γ^{I}	Urban amenity for immig.	0.976	Immig.-urbanization gradient (Fig. 6)
γ^{S}	Urban amenity for skilled	0.759	Skill-urbanization gradient (Fig. 4)
α^{L}	Urban amenity for unskilled arrivals	0.228	Table A.8
α^{H}	Urban amenity for skilled arrivals	0.447	Table A.8
ξ	Population elasticity for arrivals	0.734	Table A.8
ς	Homophily for arrivals	0.530	Table A.8
η	Growth rate of intl. migration	0.012	Long-run pop. growth = 1.2%
η^{B}	Birth rate	0.029	Pop. growth 1880-2000 (Fig. A.5)
d	Death rate	0.019	Time series of immig. share (Fig. A.5)
g_A	Exogenous TFP growth	0.008	Long-run growth rate (Fig. A.5)
σ	Elasticity of substitution	5	Set exogenously
ϱ	Local congestion elasticity	0.3	Set exogenously
ε	Migration elasticity	2	Set exogenously
β	Discount rate for capitalists	0.35	Set exogenously

In Panel (A) we focus on the innovation side and plot overall innovation human capital (ψ_{r1890}) and its components, i.e., local knowledge (N^{ϑ}_{r1880}) and exogenous research TFP (ζ_r). Overall innovative human capital is substantially higher in urban labor markets: between the highest and lowest urbanized locations, it differed by a factor of five. Strikingly, this "innovative premium" of cities is entirely explained by their measured local knowledge stock. For the first eight deciles, our model infers exogenous research TFP ζ_r to be basically flat. At the very top, it is, if anything, lower because we measure the stock of patents to be much higher (recall that we infer $\zeta_r = \psi_{rt}/N^{\vartheta}_{rt-1}$ residually).

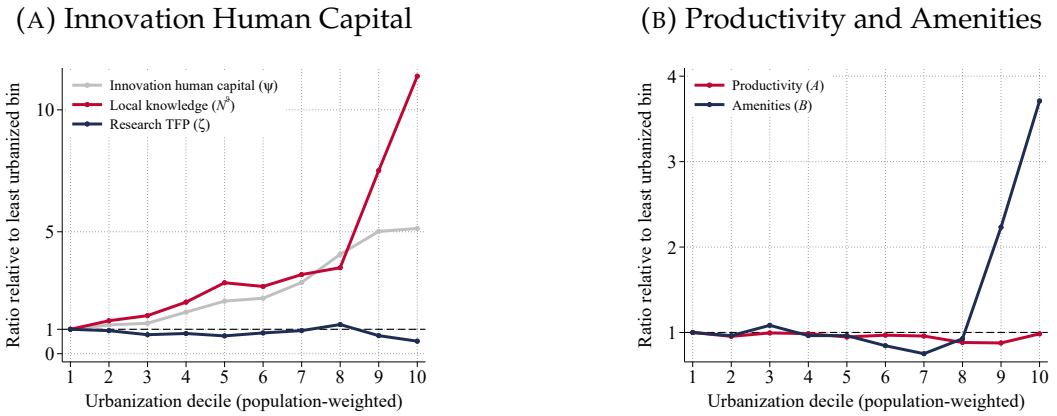
In Panel (B) we display the same results for amenities $\mathcal{B}_r$ and exogenous productivity A_r. Similar to research TFP ζ_r, productivity is not systematically correlated with urbanization. Hence, measured differences in local knowledge, together with the structure of our theory and the estimated scale elasticity ϑ, are sufficient to explain the fact that cities have higher wages and more patents per capita. As for amenities, we find that exogenous amenities are higher in the most urban locations to rationalize the very large population differences between cities and rural areas. We find this not surprising given that amenities also implicitly control for differences in the size of different SEAs and that cities, in our theory, suffer from congestion so that overall amenities are given by $\mathcal{B}_r \ell_{rt}^{-\varrho}$.

FIGURE 4: MODEL FIT: CROSS-SECTIONAL MOMENTS



Notes: Panels (a)–(d) report the correlation between the 1880 urbanization rate and four targeted moments — local population, local wages, patents per capita, and the share of high-skilled individuals among natives — read off in 1900; the model (data) is shown in red (blue), and each dot is a State Economic Area (SEA) with size proportional to its population.

FIGURE 5: URBAN PREMIA: FUNDAMENTALS AND LOCAL KNOWLEDGE



Notes: The figure summarizes the local fundamentals and initial knowledge stock across population-weighted urbanization deciles, expressed relative to the least-urban decile (dashed line at 1). We depict innovation human capital $\psi_r = \zeta_r N_{r1880}^{\theta}$, the knowledge stock N_{r1880}^{θ}, and research TFP ζ_r in panel (A) and location productivity A_r and the exogenous component of amenities $\mathcal{B}_r$ in Panel (B).

4.6 Model Validation

In Figure 6 we report several additional moments to validate our model. Panel (A) shows that our simple demographic model with constant birth and death rates, in combination with the observed immigration inflows, replicates the well-known *U*-shaped immigration shares of the US. Starting with the immigration restrictions of the 1920s, the immigrant share fell from about 20% to 5% and recovered in the 1960s owing to more liberal immigration policies. Panel (B) shows that our model is also successful in generating the empirically observed urban gradient of immigrants, even though we parametrize immigrants' taste for urban amenities only with a single elasticity. In panel (C) we report the overall share of innovators in the US economy over time. Our model predicts that around 2% to 3% of individuals actively engage in the creation of new ideas. In the data, less than 1% of individuals get assigned patents, presumably because many innovations might not be patented but still generate economic value. Interestingly, both in the data and in the model, the share of innovators is declining over time. In our theory this is due to transitional dynamics, whereby ideas are relatively scarce in the late 19th century and accumulate over time. This pattern of "catch-up growth" through innovation is also visible in the cross-section. As seen in panel D, the growth rate of patents is slightly higher in rural areas. Hence, even though cities are innovation hubs in that they generate many more patents per capita, *proportional* patent growth is slightly lower because "ideas are getting harder to find" (Bloom et al., 2020).[16]

Figures 4 and 6 show that our model rationalizes the key cross-sectional patterns between urbanization, population density, local wages, immigration and innovative activity. We now turn to the model's predictive power for *local growth*, which is not targeted in our estimation. More specifically, we consider a regression of the form

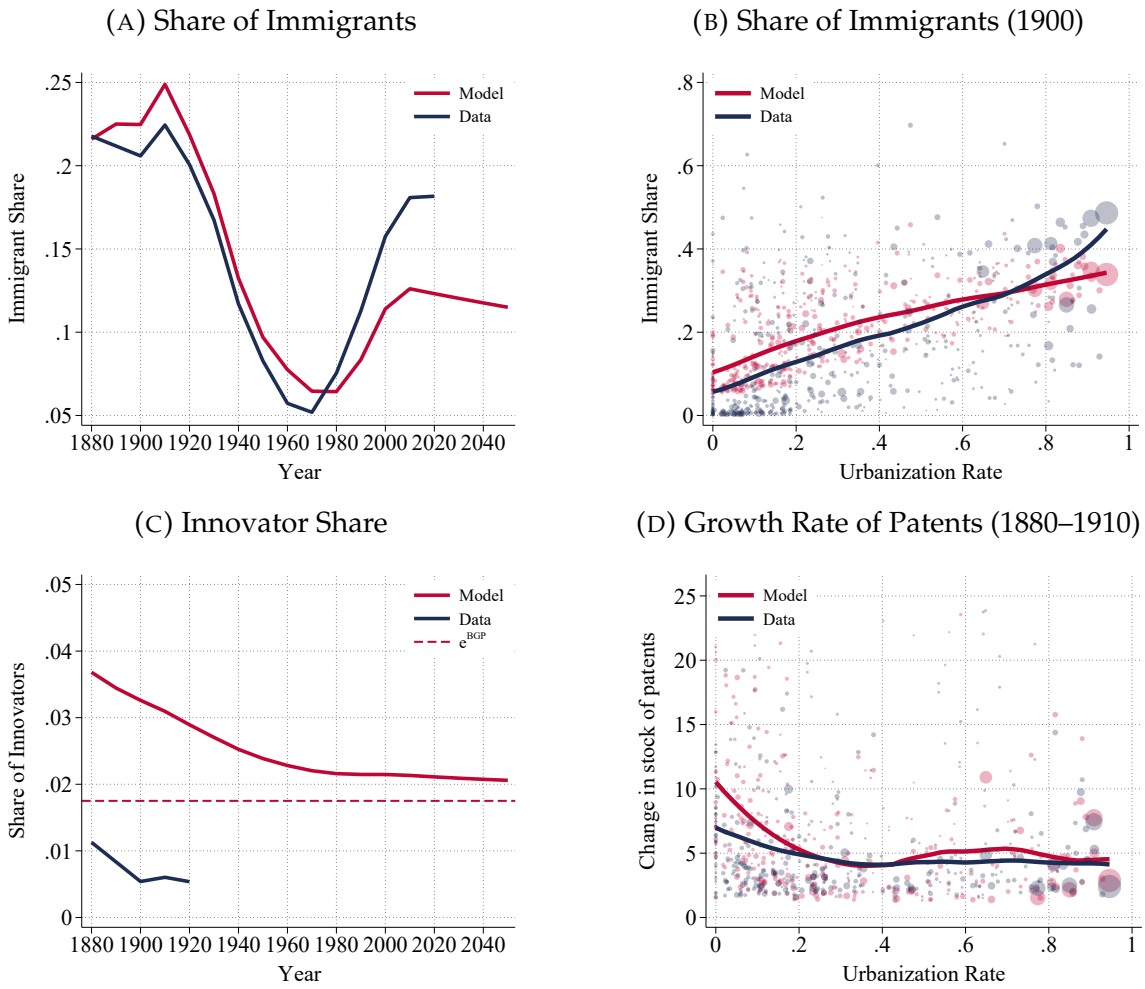
$$x_{rt}^{DATA} = \delta_r + \delta_t + \beta_x x_{rt}^{MODEL} + v_{rt}, \tag{24}$$

where x_{rt}^{DATA} and x_{rt}^{MODEL} denote local outcomes in the data and model, respectively. Given the presence of a location fixed effect, δ_r, the coefficient β_x measures the correlation between local *growth* of outcome x in the model and the data. In the model, the cross-sectional variation in local growth is due to (i) transitional dynamics in the accumulation of knowledge, (ii) transitional dynamics for the local population, and (iii) immigration shocks. We consider the regression in (24) for the main outcomes of our theory: the local population, the share of immigrants, local wages, patents per capita, and the share of innovators.

As seen in Table 7, the model's prediction for population growth, the change in the share

[16]This mechanism is also present in the time-series. As we show in Figure A.5 in the Appendix, our model implies that research productivity has been falling by a factor of 16 during the 20th century. This is qualitatively in line with the findings in Bloom et al. (2020), even though they report even larger declines, possibly because they focus on growth rates within narrowly defined goods or industries.

FIGURE 6: ADDITIONAL IMPLICATIONS



Notes: The top row displays the share of immigrants in the time series (Panel A) and as a function of the urbanization rate in 1900 (Panel B). The bottom row depicts the aggregate share of innovators (Panel C) and the correlation between the growth rates of patents and the urbanization rate (Panel D). In panel C we also put a horizontal dashed line for the innovation share along the BGP, e^{BGP} (which, recall, is equalized across space).

of immigrants, local wage growth, and the growth in patents per capita are positively correlated with the data, highlighting that the key forces of our theory are also empirically important. By contrast, as seen in column 5, the correlation for the change in the share of innovators is negative, owing to the fact that the model predicts too high a level (and hence a decline in) innovation shares for rural locations, where the stock of knowledge is very low and hence innovation incentives are high.

5. THE ROLE OF IMMIGRATION FOR US GROWTH

We now use our model to quantify the impact of international migration on US growth. To do so, we consider two counterfactuals. In our first exercise, we focus directly on the Era of Mass Migration and ask how much growth the US would have lost if it had closed its doors to the roughly 23 million international migrants that arrived between 1880 and 1920. In our second exercise, we evaluate the effects of two specific US immigration policies:

TABLE 7: MODEL VS DATA

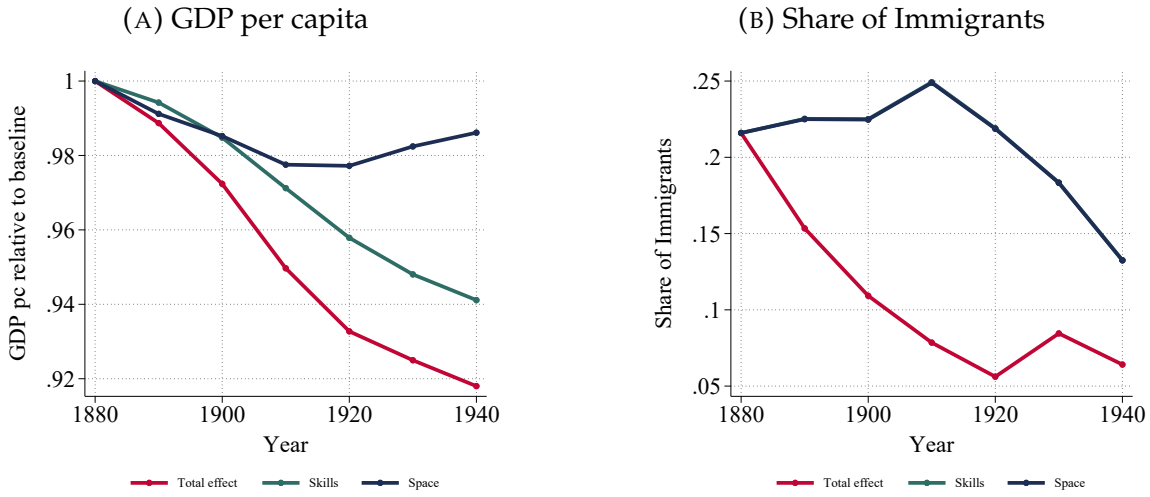
	(1) $\ln \ell_{rt}$	(2) ϖ^{IMM}	(3) $\ln w_{rt}$	(4) $\ln(I_{rt}/\ell_{rt})$	(5) $\ln e_{rt}$
Model Prediction	1.769***	0.856***	0.125***	0.366***	-0.270***
	(0.032)	(0.026)	(0.017)	(0.034)	(0.010)
Year FE	yes	yes	yes	yes	yes
SEA FE	yes	yes	yes	yes	yes
R^2	.993	.974	.999	.965	.866
N	1,704	1,704	1,704	1,704	1,704

Notes: The table contains the results of regression (24) for different outcomes. All specifications contain time and SEA fixed effects. We weight all local outcomes by the size of the local population.

the 1921 Emergency Quota Act and the 1924 Johnson–Reed Act, which severely limited international migration, especially from Southern and Eastern Europe and Asia.

5.1 The Era of Mass Migration: 1880–1920

FIGURE 7: THE AGGREGATE IMPACT OF INTERNATIONAL MIGRATION



Notes: The figure shows the change in aggregate income per capita, relative to the baseline economy (Panel A) and the aggregate share of immigrants (Panel B). We depict the baseline calibration in grey, the counterfactual with no immigrant inflows between 1880 and 1920 in red, the "no skilled immigrants" counterfactual in green, and the "no urban bias" counterfactual, where immigrants initially settle in proportion to the spatial allocation of natives in blue.

What would US growth have looked like if all immigration inflows between 1880 and 1920 were set to zero? The answer is shown in the red lines of Figure 7. The left panel depicts the impact on aggregate GDP per capita, relative to the baseline economy; the right panel shows the aggregate share of immigrants. In the absence of new migration flows, the share of immigrants would fall from 21.6% in 1880 to 6.4% by 1940, and the overall population would be 27.3% lower than the baseline by 1940 (see Appendix Section A.3.2). As a consequence, we find that GDP per capita by 1940 would have been 8.2% lower. Interestingly, these missing cohorts of immigrants cast a long-lasting shadow due

to the cumulative loss of ideas. By 2000, income per capita would have been 9.6% lower.

Our theory highlights that migration affects US growth through three channels: *scale*, whereby immigrants increase the size of the population; *skills*, whereby some immigrants bring innovative skills; and *space*, whereby immigrants increase population growth in urban innovation hubs. To isolate the role of these channels, we compute two alternative counterfactuals.

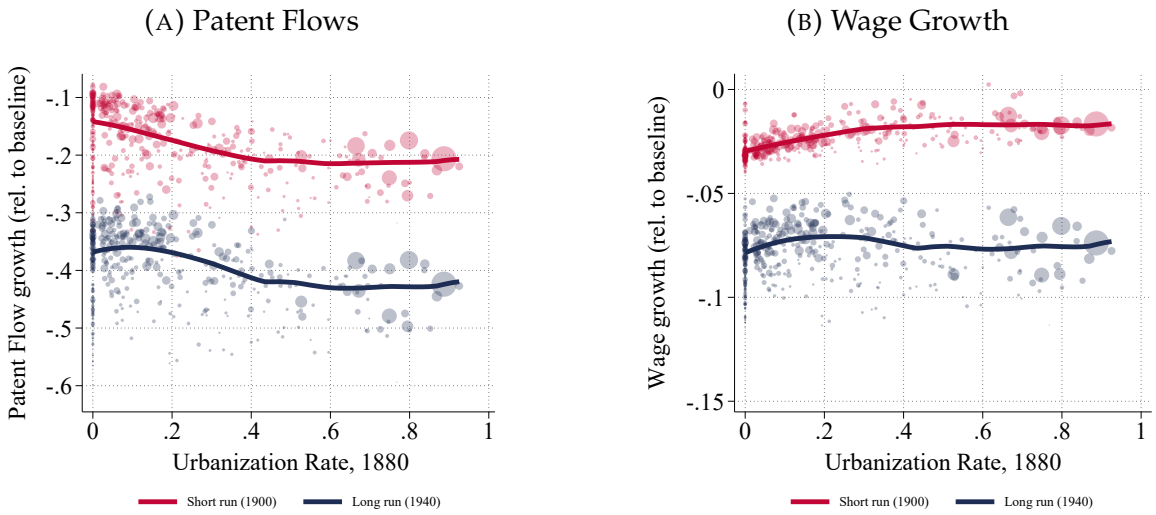
First, to quantify the importance of *skills*, we compute the equilibrium path assuming that all international migrants between 1880 and 1920 were unskilled. Hence, the total amount of "bodies" (and hence the share of immigrants) is exactly the same as in the baseline economy, but the supply of "ideas" is reduced. The impact of this counterfactual is shown in green. Even though the share of immigrants actually patenting is quite low, they account for a large part of the adverse effects of immigration restrictions: missing out on this small number of skilled immigrants would have reduced GDP per capita by 5.9% by 1940, accounting for about 75% of the overall effect of migration restrictions.

Second, to highlight the role of *space*, in particular immigrants' sorting toward cities, we consider a counterfactual where we keep both the size and the skill content of international arrivals as in the baseline calibration, but assume that migrants do not have a specific urban bias, that is, the initial spatial distribution of incoming immigrants is the same as that of natives, and migrants do not specifically value urban amenities, $\gamma^I = 0$. Hence, this scenario quantifies the importance of immigration-led urbanization for overall innovation and growth. As seen in the blue line, without migrants' "taste" for cities, overall economic activity would have been lower. Interestingly, for the first three decades, migrants' spatial sorting was as important for US growth as their innate skills. In the longer run, the effects subside because of migrants' spatial mobility within the US.

In Appendix Section A.3.2 we also document the effects on overall innovation activity and urbanization. In the absence of migration, innovation would be lower, reflecting the drop in market size. Recall that immigrants' innovation share is *lower* than that of natives. Hence, one might have thought this composition effect would cause the share of innovators to be *higher* in the absence of immigrants. However, natives' incentives to innovate are lower, given the non-rival nature of idea-creation. As for urbanization, between 1880 and 1920, the share of people living in the most urbanized SEAs increased from 35% to 45%. Without international migration, this increase would have been substantially lower.

Spatial Impacts Because immigrants arrive in a spatially unbalanced way, locations are differentially exposed to the shock. In Figure 8, we depict the change in patent flows (left panel) and local wages (right panel) in the absence of migration inflows, relative to the baseline economy, as a function of the urbanization rate. To highlight the differential dynamic responses, we depict both the short-run effects (in 1890) and the long-run effect (in 1940).

FIGURE 8: SPATIAL IMPACTS



Notes: The left panel shows the change in the flow of patents (I_{rt}) in the absence of migration relative to the baseline economy, $\ln I_{rt} - \ln I_{rt}^{\text{Baseline}}$ as a function of the urbanization rate. The right panel shows $\ln w_{rt} - \ln w_{rt}^{\text{Baseline}}$. We show both the short-run effect in 1890 and the long-run effects in 1940.

The left panel of Figure 8 shows the importance of international migration for innovation, especially in cities. In the absence of international migrants, patent creation would have declined by 15% in the short run and by more than 40% in the long run. The right panel depicts the impact on local wages. Consistent with the overall impact on GDP per capita reported in Figure 7, wages fall by around 3% in the short run and by about 8% in the long run. Interestingly, the wage impact is, if anything, slightly larger in rural areas. The reason is that rural areas are knowledge-scarce. As a consequence, a given decline in innovation has a larger impact on local wages.[17]

5.2 The Immigration Quotas of the Early 20th Century

In our second exercise, we evaluate the effects of the 1921 Emergency Quota Act and the 1924 Johnson–Reed Act, which together imposed the first major restrictions on US immigration. These policies disproportionately targeted Southern and Eastern Europeans and effectively barred all Asian immigration. The 1921 Act capped each nationality's inflow at three percent of its foreign-born population recorded in the 1910 Census. The 1924 Act tightened these quotas to two percent of the observed population in the 1890 Census. Both policies, explicitly chosen to favor Western European source countries that already had a large stock of immigrants in the 1890s, remained in place until the Hart–Celler Act of 1965 dismantled the national-origins framework.

To quantify the impact of these restrictions, we construct a counterfactual immigrant inflow series for the quota period 1930–1960. As we describe in more detail in Appendix

[17] As we show in Figure A.8 in the Appendix, the relative decline in population growth is larger in cities, given immigrants' urban bias.

Section A.3.2, we hold the total number of immigrants constant at the 1910–1920 average and apportion the additional inflow equally between the two restricted groups: Southern and Eastern Europeans and Asians. While Europeans dominated pre-WWII immigration, given the rapid growth of Asian immigration at the time and the policies' explicit intent to block Asian inflows, we treat this as a reasonable baseline. We also estimate the skill levels of Southern and Eastern Europeans and Asian immigrants, finding that both groups had substantially lower skills than natives in 1920 (see Table A.10 in the Appendix).

Figure 9 summarizes the economic impact of the immigration quotas of the 1920s. Compared to a situation where immigrant inflows had continued at the same pace as in the early 20th century, GDP per capita could have been 1.7% higher by 2000. The main reason why this effect is relatively modest is that the additional immigrant inflow is large in absolute numbers, yet its effect on the overall population and the immigrant share is limited given the size of the US population; see panel B of Figure 9, which shows that the overall immigrant share does not dramatically differ from the actual data.

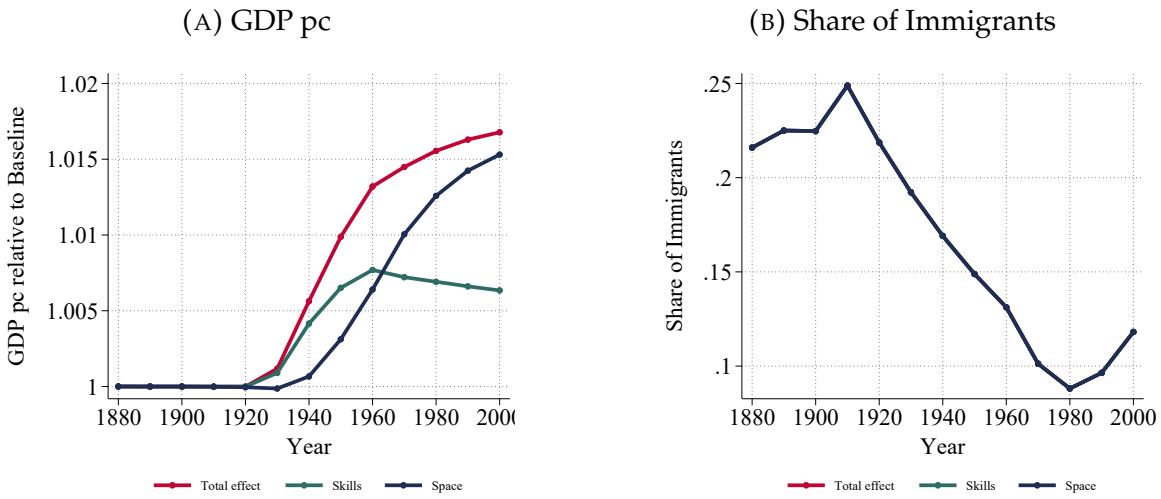
As before, we also decompose the overall effect into the role of skills (where we assume that all additional immigrants are unskilled) and the role of space (where we assume that all additional immigrants feature no specific urban bias). Figure 9 shows again that migrants' innovative skills played an important role: if all immigrants had been unskilled, the effect on GDP per capita in 1960 would only have been half as large and, if anything, slightly declining after 1960 once the quota was lifted. Interestingly, migrants' spatial allocation plays an important role in the short run. If the new immigrants had arrived without an urban bias, GDP per capita would not have risen until the 1940s. Over time, people reallocate spatially, and the gap relative to the full counterfactual becomes smaller.

6. Robustness

While most of the structural parameters of our model are directly estimated, we calibrate some externally. In this section, we discuss the robustness of our main results with respect to these choices; the results are summarized in Table 8. For each alternative specification, we recalibrate our model and conduct both counterfactuals reported in Section 5. We always report the estimated rate of exogenous technological progress (g_A), the effects of international arrivals between 1880 and 1920 on GDP per capita by 1940, both the overall effect and the respective roles of *space* and *skills*, and the overall GDP impact of removing the 1920s immigration quotas. For comparison, we report our baseline results in the first row.

In rows two and three, we analyze the importance of aggregate knowledge spillovers. In our baseline model, we assumed that innovators only build on local knowledge, N_{rt-1}. Now we assume that $\psi_{rt} = \zeta_r N_{rt-1}^{\vartheta} (\sum_j N_{jt-1})^{\gamma}$. The higher γ, the less exogenous growth is "required" to match the overall growth experience of the US, and the larger the impact

FIGURE 9: THE AGGREGATE IMPACT OF THE 1920S MIGRATION RESTRICTIONS



Notes: The figure shows the change in aggregate income per capita relative to the baseline economy (Panel A) and the aggregate share of immigrants (Panel B). We depict the counterfactual of removing the 1920s immigration restrictions in red, the counterfactual in which the additional immigrants generated by removing the restrictions are assumed to be unskilled in green, and the counterfactual in which the additional immigrants have no urban bias in blue.

TABLE 8: ROBUSTNESS

Specification	g_A	Mass Migration 1880-1920			1920s
		Full effect	Space	Skills	Quotas
Baseline	0.086	0.918	0.990	0.941	1.007
Agg. spillover ($\gamma = 0.15$)	0.071	0.904	0.988	0.929	1.009
Agg. spillover ($\gamma = 0.30$)	0.023	0.882	0.985	0.913	1.015
EoS ($\sigma = 4$)	0.053	0.893	0.988	0.922	1.010
EoS ($\sigma = 7$)	0.121	0.944	0.993	0.961	1.005
Myopic capitalists ($\beta = 0$)	0.087	0.917	0.990	0.939	1.007
Dynamic migration	0.088	0.918	0.991	0.941	1.007

Notes: This table reports the estimated g_A (row 2), the aggregate impact of international arrivals between 1880 and 1920 on GDP per capita in 1940 (rows 3–5), and the impact of the 1920s immigration quotas (row 6). See text for the details on these respective counterfactuals.

of immigrants on growth. If $\gamma = 0.3$, the impact would be roughly 30% higher than in our baseline estimation (a decline of 11.8% as opposed to 8.2%). In rows four and five, we explore different values of the elasticity of substitution σ, which determines the strength of variety gains. The lower σ, the stronger the link between immigration and growth. Quantitatively, if σ were four as opposed to five, the impact on GDP would increase from 8.2% to 10.7%. In row six, we present the results from a calibration where capitalists are myopic and only pay researchers one-period profits. Doing so leaves our results essentially unchanged. Finally, we also estimate a version of our model where workers solve a dynamic migration problem (see Appendix Section OA.4 for details). Again, this does not change the impact of immigrants on growth once the model is recalibrated.

7. Conclusion

Between 1880 and 1920, more than 20 million immigrants made the United States their home. In this paper, we evaluate the impact of this large immigration wave on US innovation and growth.

Using our dataset on individual patent behavior, we document that immigrants played an important role in US innovation and show that spatial sorting played a key role in immigrants' innovativeness. Much like today, immigrants were overrepresented in urban innovation hubs, where patent intensity was five times larger than in rural areas.

Motivated by this evidence, we propose a model of spatial growth by incorporating the insights of idea-based, semi-endogenous growth theory into a canonical spatial model. Our theory emphasizes the local supply of innovative human capital as the main determinant of local productivity, admits a balanced growth path with a stationary distribution of economic activity, yet is still sufficiently tractable to analyze the transitional dynamics. As such, our framework is useful to quantify the importance of immigrants in a realistic setting featuring many locations, trade costs, and frictional population mobility.

For the historical episode, we find that without the inflow of international migrants between 1880 and 1920, income per capita by 1940 would have been 8.2% lower. Turning from history to policy, we apply the same framework to the immigration quotas of the 1920s and find that lifting these restrictions would have raised US income per capita by a further 1.7% by 2000—a modest effect, because the excluded migrants were both demographically small relative to the US population and predominantly low-skilled. In both exercises, immigrants' innate innovative skills and their bias to arrive in and move to urban locations are quantitatively important determinants of the productivity gains from migration.

We view our paper as a natural starting point for further research at the intersection of mobility, migration, innovation, and long-run growth. We hope the tools introduced here, together with the growing availability of historical microdata, will enable a deeper understanding of how immigration shapes innovation, regional development, and long-run growth in both historical and contemporary settings.

References

Abramitzky, R. and L. Boustan (2022): *Streets of gold: America's untold story of immigrant success*, Hachette UK.

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum, and S. Pérez (2021): "Automated Linking of Historical Data," *Journal of Economic Literature*, 59, 865–918.

Abramitzky, R., L. Boustan, K. Eriksson, S. Pérez, and M. Rashid (2025): "Census Linking Project: Version 3.0 [dataset]," https://censuslinkingproject.org.

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102, 1832–56.

——— (2014): "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 122, 467–506.

ABRAMITZKY, R., R. MILL, AND S. PÉREZ (2020): "Linking individuals across historical sources: A fully automated approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53, 92–111.

AKCIGIT, U. (2017): "Economic Growth: The Past, the Present, and the Future," *Journal of Political Economy*, 125, 1736–1747.

AKCIGIT, U., J. GRIGSBY, AND T. NICHOLAS (2017): "Immigration and the Rise of American Ingenuity," *American Economic Review*, 107, 327–31.

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

ALTONJI, J. G. AND D. CARD (1991): "The effects of immigration on the labor market outcomes of less-skilled natives," in *Immigration, trade, and the labor market*, University of Chicago Press, 201–234.

BERGEAUD, A. AND C. VERLUISE (2022): "PatentCity: a dataset to study the location of patents since the 19th century," .

——— (2024): "A New Dataset to Study a Century of Innovation in Europe and in the US," *Research Policy*, 53, 104903.

BERNSTEIN, S., R. DIAMOND, A. JIRANAPHAWIBOON, T. MCQUADE, AND B. POUSADA (2022): "The contribution of high-skilled immigrants to innovation in the United States," Tech. rep., National Bureau of Economic Research.

BILAL, A. (2023): "The geography of unemployment," *The Quarterly Journal of Economics*, 138, 1507–1576.

BILAL, A. AND E. ROSSI-HANSBERG (2021): "Location as an Asset," *Econometrica*, 89, 2459–2495.

BLOOM, N., C. I. JONES, J. VAN REENEN, AND M. WEBB (2020): "Are Ideas Getting Harder to Find?" *American Economic Review*, 110, 1104–1144.

BRODA, C. AND D. WEINSTEIN (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121, 541–585.

BURCHARDI, K. B., T. CHANEY, T. A. HASSAN, L. TARQUINIO, AND S. J. TERRY (2026): "Immigration, Innovation, and Growth," *American Economic Review*, 116, 828–861.

BURSTEIN, A., G. HANSON, L. TIAN, AND J. VOGEL (2020): "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence From the United States," *Econometrica*, 88, 1071–1112.

CARD, D. (1990): "The Impact of the Marial Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, 43, 245–257.

——— (2001): "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 19, 22–64.

CREWS, L. G. (2023): "A Dynamic Spatial Knowledge Economy," Working Paper.

DESMET, K., D. K. NAGY, AND E. ROSSI-HANSBERG (2018): "The Geography of Development," *Journal of Political Economy*, 126, 903–983.

DIAMOND, R., T. MCQUADE, AND F. QIAN (2019): "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review*, 109, 3365–94.

DUSTMANN, C., U. SCHÖNBERG, AND J. STUHLER (2017): "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment," *The Quarterly Journal of Economics*, 132, 435–483.

ECKERT, F. AND M. PETERS (2022): "Spatial Structural Change," Working Paper.

FAJGELBAUM, P. D., E. MORALES, J. C. SUÁREZ SERRATO, AND O. M. ZIDAR (2019): "State Taxes and Spatial Misallocation," *The Review of Economic Studies*, 86, 333–376.

FEIGENBAUM, J. J. (2016): "Machine Learning Approach to Census Record Linking," Tech. rep.

FUSHIKI, T. (2011): "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing*, 21, 137–146.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): in *The elements of statistical learning*, Berlin, Germany: Springer.

HORNBECK, R. AND M. ROTEMBERG (2024): "Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies," *Journal of Political Economy*, 132, 3547–3602.

JARO, M. A. (1989): "Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida," *Journal of the American Statistical Association*, 84, 414–420.

——— (1995): "Probabilistic linkage of large public health data files," *Statistics in medicine*, 14, 491–498.

JONES, C. (2005): "Growth and Ideas," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1, Part B, chap. 16, 1063–1111, 1 ed.

JONES, C. I. (1995): "R&D-based Models of Economic Growth," *Journal of Political Economy*, 103, 759–784.

——— (2022): "The past and future of economic growth: A semi-endogenous perspective," *Annual Review of Economics*, 14, 125–152.

KERR, S. P., W. KERR, ÇAĞLAR ÖZDEN, AND C. PARSONS (2016): "Global Talent Flows," *Journal of Economic Perspectives*, 30, 83–106.

KERR, W. (2018): *The Gift of Global Talent: How Migration Shapes Business, Economy & Society*, Stanford University Press.

KLEINMAN, B., E. LIU, AND S. J. REDDING (2023): "Dynamic spatial general equilibrium," *Econometrica*, 91, 385–424.

KONERU, K., V. S. V. PULLA, AND C. VAROL (2016): "Performance Evaluation of Phonetic Matching Algorithms on English Words and Street Names - Comparison and Correlation," in *Proceedings of the 5th International Conference on Data Management Technologies and Applications - Volume 1: DATA*, Insticc, SciTePress, 57–64.

KORTUM, S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65, 1389–1419.

KRUGMAN, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70, 950–959.

LEE, S. K. (2019): "Essays in History and Spatial Economics with Big Data," Working Paper.

MANSON, S., J. SCHROEDER, D. VAN RIPER, AND S. RUGGLES (2019): "IPUMS National Historical Geographic Information System: Version 14.0 [Database]," .

NAGY, D. K. (2023): "Hinterlands, City Formation and Growth: Evidence from the U.S. Westward Expansion," *The Review of Economic Studies*, 90, 3238–3281.

OBERFIELD, E. AND D. RAVAL (2021): "Micro Data and Macro Technology," *Econometrica*, 89, 703–732.

PEDREGOSA, F., G. VAROQUAUX, A. GRAMFORT, V. MICHEL, B. THIRION, O. GRISEL, M. BLONDEL, P. PRETTENHOFER, R. WEISS, V. DUBOURG, J. VANDERPLAS, A. PASSOS, D. COURNAPEAU, M. BRUCHER, M. PERROT, AND E. DUCHESNAY (2011): "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 12, 2825–2830.

PERI, G. (2016): "Immigrants, Productivity, and Labor Markets," *The Journal of Economic Perspectives*, 30, 3–30.

PETERS, M. (2022): "Market size and spatial growth—evidence from Germany's post-war population expulsions," *Econometrica*, 90, 2357–2396.

PETRALIA, S., P.-A. BALLAND, AND D. RIGBY (2016): "HistPat Dataset," .

RAJKOVIC, P. AND D. JANKOVIC (2007): "Adaptation and Application of Daitch-Mokotoff Soundex Algorithm on Serbian Names," Tech. rep.

REDDING, S. J. AND E. ROSSI-HANSBERG (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9, 21–58.

ROMER, P. M. (1990): "Endogenous Technological Change," *The Journal of Political Economy*, 98, S71–s102.

RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, AND M. SOBEK (2020): "Integrated Public Use Microdata Series: Version 10.0 [Machine-readable database]," .

SEQUEIRA, S., N. NUNN, AND N. QIAN (2020): "Immigrants and the Making of America," *The Review of Economic Studies*, 87, 382–419.

TRAIBERMAN, S. (2019): "Occupations and import competition: Evidence from Denmark," *American Economic Review*, 109, 4260–4301.

WALSH, C. (2025): "The Entry Multiplier," Working paper, available at SSRN: https://ssrn.com/abstract=5468409.

WINKLER, W. E. (2014): "Advanced Methods for Record Linkage," Tech. rep.

APPENDIX

"IMMIGRATION, INNOVATION, AND THE GEOGRAPHY OF GROWTH" by C. Arkolakis, S. Lee, and M. Peters

(FOR ONLINE PUBLICATION)

A.1. THEORY

A.1.1 Characterization of Equilibrium

In this section we provide the details of the equilibrium characterization.

Solving the Value Function $V_{rt}(N)$

Consider the value function $V_{rt}(N)$ given in (3.2). Conjecture that $V_{rt}(N)$ is homogeneous in N, i.e. $V_{rt}(N) = v_{rt}N$. Equation (3.2) then implies that

$$v_{rt}N = \max_{I}\left\{\frac{N\pi_{rt}}{P_{rt}} + I\frac{(\pi_{rt}-\chi_{rt})}{P_{rt}} + \beta v_{rt+1}(N+I)\right\} \tag{A.1}$$

$$= N\left(\frac{\pi_{rt}}{P_{rt}} + \beta v_{rt+1}\right) + \max_{I}\left\{I\left(\frac{\pi_{rt}-\chi_{rt}}{P_{rt}} + \beta v_{rt+1}\right)\right\}. \tag{A.2}$$

The optimality condition for I, together with the equilibrium condition that $I < \infty$, implies that $\chi_{rt} = \pi_{rt} + P_{rt}\beta v_{rt+1}$. Hence,

$$v_{rt} = \frac{\pi_{rt}}{P_{rt}} + \beta v_{rt+1}, \tag{A.3}$$

as claimed in (2). This also implies that $\chi_{rt} = P_{rt}v_{rt}$.

Dynamic Equilibrium System

Our theory has an intuitive modularized structure:

1. ***The Growth Block:*** Given the distribution of the population, the stock of knowledge N_{rt-1}, the growth rates of real wages g^{w}_{rt+1}, and the growth rates of innovation cutoffs $g^{\bar{h}}_{rt+1}$, we can solve for the equilibrium innovation cutoff $\bar{h}_{rt}$, and the new stock of knowledge N_{rt}.

2. ***The Trade Block:*** Given the population distribution ℓ^{vs}_{rt}, the stock of local varieties N_{rt} and the cutoff $\bar{h}_{rt}$, we can compute equilibrium wages w_{rt}.

3. ***The Spatial Supply Block:*** Given the initial population distribution ℓ^{vs}_{rt-1}, the distribution of wages w_{rt} and the innovation cutoffs $\overline{h}_{rt}$, we can compute the current population distribution.

The Growth Block The equilibrium cutoff $\overline{h}_{rt}$ is given in (6). Now note that

$$\frac{P_{rt}v_{rt+1}}{w_{rt}} = \frac{P_{rt}v_{rt}}{w_{rt}}\frac{v_{rt+1}}{v_{rt}} = \frac{\chi_{rt}}{w_{rt}}\frac{\chi_{rt+1}/P_{rt+1}}{\chi_{rt}/P_{rt}} = \frac{1}{\overline{h}_{rt}}\frac{w_{rt+1}/P_{rt+1}}{w_{rt}/P_{rt}}\frac{\overline{h}_{rt}}{\overline{h}_{rt+1}} \equiv \frac{1+g^{w}_{rt+1}}{\overline{h}_{rt+1}},$$

where g^{w}_{rt+1} denotes the growth rate of *real* wages. Substituting this expression and the law of motion $N_{rt} = N_{rt-1} + \frac{\theta}{\theta-1}\overline{h}_{rt}^{1-\theta}\mathcal{H}^{\theta}_{rt}\ell_{rt}$ into (6) yields an implicit equation for $\overline{h}_{rt}$:

$$1 = \frac{(\sigma-1)}{\frac{\ell_{rt}\overline{h}_{rt}}{N_{rt-1}+\frac{\theta}{\theta-1}\overline{h}_{rt}^{1-\theta}\mathcal{H}^{\theta}_{rt}\ell_{rt}}\left(1-\left(\frac{\mathcal{H}_{rt}}{\overline{h}_{rt}}\right)^{\theta}\right) + \beta(\sigma-1)\frac{1+g^{w}_{rt+1}}{1+g^{\overline{h}}_{rt+1}}}.$$

Given $\overline{h}_{rt}$, we can compute N_{rt} from the law of motion above.

The Trade Block Given $\overline{h}_{rt}$, $\mathcal{H}_{rt}$ and N_{rt}, we can solve for the distribution of wages $\{w_{rt}\}_r$. Total income in region r is given by

$$Y_{rt} = \frac{\sigma}{\sigma-1}w_{rt}\ell_{rt}\left(1-\left(\mathcal{H}_{rt}/\overline{h}_{rt}\right)^{\theta}\right). \tag{A.4}$$

Labor market clearing requires that

$$Y_{rt} = N_{rt}\left(\frac{w_{rt}}{A_r}\right)^{1-\sigma}\sum_{j=1}^{R}\frac{\tau_{rj}^{1-\sigma}}{\sum_{m=1}^{R}N_{mt}\left(\tau_{mj}\frac{w_{mt}}{A_m}\right)^{1-\sigma}}Y_{jt},$$

where τ_{rj} denote the trade costs between r and j. These equations can be solved for the set of wages w_{rt} together with a choice of numeraire.

The Migration Block Given w_{rt} and $\overline{h}_{rt}$, we can compute the migration probabilities according to (9). Given the pre-determined population distribution ℓ^{vs}_{rt-1}, we can then compute the current population distribution ℓ^{vs}_{rt} from (10) and (11).

A.1.2 Characterization of Spatial BGP

In this section, we characterize the spatial BGP in our economy.

Proposition 1. *Suppose that $\vartheta < 1$, let the aggregate population $\ell_t \equiv \sum_r \ell_{rt}$ grow at rate $\eta > 0$, and labor productivity A_{rt} grow at rate g_A. Let the population distribution be stationary, that is $\varpi^{v}_{r} = \ell^{v}_{rt}/\ell_{rt}$ and λ^{v}_{r} is constant. Then:*

1. *The growth rates for knowledge N_{rt}, human capital stock $\mathcal{H}_{rt}$, and wages w_{rt} are given by*

$$1+g_N=(1+\eta)^{\frac{1}{1-\vartheta}} \quad \textit{and} \quad 1+g_{\mathcal{H}}=(1+\eta)^{\frac{\vartheta}{1-\vartheta}} \quad \textit{and} \quad 1+g_w=(1+g_A)(1+\eta)^{\frac{1}{1-\vartheta}\frac{1}{\sigma-1}},$$

2. *The innovation cutoff $\bar{h}_{rt}$, local wages w_{rt}, and the share of innovators e_{rt} are given by*

$$\bar{h}_{rt} = \alpha\zeta_r^{\frac{1}{1-\vartheta}}\ell_{rt}^{\frac{\vartheta}{1-\vartheta}}\left(\sum_\nu \lambda_r^\nu \varpi_r^\nu\right)^{\frac{1}{\theta}\frac{1}{1-\vartheta}} \tag{A.5}$$

$$w_{rt} = \frac{(N_{rt})^{1/\sigma}A_{rt}^{\frac{\sigma-1}{\sigma}}}{((1-e_r)\ell_{rt})^{1/\sigma}}\left(\sum_{j=1}^R \frac{\tau_{rj}^{1-\sigma}w_{jt}(1-e_j)\ell_{jt}}{\sum_{m=1}^R N_{mt}\left(\tau_{mj}w_{mt}/A_{mt}\right)^{1-\sigma}}\right)^{1/\sigma} \tag{A.6}$$

$$e_{rt}=e^* = \left(\frac{\mathcal{H}_{rt}}{\bar{h}_{rt}}\right)^\theta = \frac{1}{1+\frac{\theta}{\theta-1}\frac{1+g_N}{g_N}(\sigma-1)\left(1-\beta\frac{1+g_w}{1+g_{\mathcal{H}}}\right)}. \tag{A.7}$$

where α is an inconsequential constant, which is common across space.

3. *Let $\ell_t^{s|I}=\sum_r \ell_{rt}^{s|I}$ denote the aggregate mass of immigrants with skill s and ℓ_t^s the aggregate population. Then*

$$\frac{\ell_t^{s|I}}{\ell_t^s}=1-\frac{\eta^B}{\eta+d} \quad \textit{and} \quad \ell_t^\nu=\frac{1}{\eta+d}M_t^\nu. \tag{A.8}$$

Moreover, ℓ_{rt}^ν solves the stationary version of equations (10) and (11) evaluated at the BGP prices.

We now prove these results in turn:

The BGP Growth Rate and the Innovation Cutoff Let $e_{rt}=\left(\mathcal{H}_{rt}/\bar{h}_{rt}\right)^\theta$ denote the share of the population that is engaged in idea creations. The equilibrium cutoff is given by

$$\bar{h}_{rt}=\frac{(\sigma-1)}{\frac{\ell_{rt}}{N_{rt}}(1-e_{rt})+\beta(\sigma-1)\frac{v_{rt+1}}{w_{rt}/P_{rt}}}. \tag{A.9}$$

Along a BGP, the number of ideas N_{rt} grows at a common rate g_N. Hence, $N_{rt}=\frac{1+g_N}{g_N}I_{rt}$. Using (4), we get that

$$\frac{\ell_{rt}}{N_{rt}}=\frac{\ell_{rt}}{I_{rt}}\frac{g_N}{1+g_N}=\frac{\theta-1}{\theta}\frac{1}{\bar{h}_{rt}e_{rt}}\frac{g_N}{1+g_N}$$

Substituting this in (A.9) yields

$$1=\frac{(\sigma-1)}{\frac{g_N}{1+g_N}\frac{\theta-1}{\theta}\left(\frac{1-e_{rt}}{e_{rt}}\right)+\beta(\sigma-1)\frac{v_{rt+1}\bar{h}_{rt}}{w_{rt}/P_{rt}}} \tag{A.10}$$

Now note that $v_{rt+1} = \chi_{rt+1}/P_{rt+1} = w_{rt+1}/(P_{rt+1}\overline{h}_{rt+1})$. Hence,

$$\frac{v_{rt+1}\overline{h}_{rt}}{w_{rt}/P_{rt}} = \frac{\chi_{rt+1}\overline{h}_{rt}/P_{rt+1}}{w_{rt}/P_{rt}} = \frac{w_{rt+1}/P_{rt+1}}{w_{rt}/P_{rt}}\frac{\overline{h}_{rt}}{\overline{h}_{rt+1}} = \frac{1+g}{1+g_{\overline{h}}}, \tag{A.11}$$

where g is the rate of real wage growth and $g_{\overline{h}}$ is the growth rate of the human capital cutoff. Both these growth rates are endogenous and we will solve for them below. However, both of them are also constant along a BGP. The innovator share is therefore also constant across space and given by

$$e^{BGP}_{rt} = e^{BGP} = \frac{1}{1+\frac{\theta}{\theta-1}\frac{1+g_N}{g_N}(\sigma-1)\left(1-\beta\frac{1+g}{1+g_{\overline{h}}}\right)}. \tag{A.12}$$

The cutoff $\overline{h}_{rt}$ is then given by

$$\overline{h}_{rt} = \left(e^{BGP}\right)^{-1/\theta}\mathcal{H}_{rt} = \left(e^{BGP}\right)^{-1/\theta}\zeta_r N^{\vartheta}_{rt-1}\left(\sum_\nu \lambda^\nu_{rt}\omega^\nu_{rt}\right)^{1/\theta}. \tag{A.13}$$

Rearranging terms yields

$$1 = \left(e^{BGP}\right)^{-1/\theta}\zeta_r\left(\frac{N_{rt-1}}{\ell_{rt}\overline{h}_{rt}}\right)^{\vartheta}\frac{\ell^{\vartheta}_{rt}}{\overline{h}^{1-\vartheta}_{rt}}\left(\sum_\nu \lambda^\nu_{rt}\omega^\nu_{rt}\right)^{1/\theta}. \tag{A.14}$$

Because $\frac{N_{rt-1}}{\overline{h}_{rt}\ell_{rt}} = \frac{\theta}{\theta-1}e^{BGP}\frac{N_{rt}}{I_{rt}}$ is constant along the BGP, this implies that $\overline{h}^{\frac{1-\vartheta}{\vartheta}}_{rt} \propto \ell_{rt}$. Hence, as in the semi-endogenous growth model of Jones (1995), the growth in $\overline{h}_r$ is tied to the rate of population growth.

Let the aggregate population growth rate be equal to $\dot{\ell}_t/\ell_t = \eta$. Because $N_{rt-1}/\overline{h}_{rt}\ell_{rt}$ is constant, we have that

$$1+g_N = \frac{N_{rt}}{N_{rt-1}} = \frac{\overline{h}_{rt+1}\ell_{rt+1}}{\overline{h}_{rt}\ell_{rt}} = \left(\frac{\ell_{rt+1}}{\ell_{rt}}\right)^{1+\frac{\vartheta}{1-\vartheta}} = (1+\eta)^{\frac{1}{1-\vartheta}}. \tag{A.15}$$

The human capital cutoff $\overline{h}_{rt}$ and the human capital supply $\mathcal{H}_{rt}$ therefore grow at the rates

$$1+g_{\overline{h}} = 1+g_{\mathcal{H}} = (1+\eta)^{\frac{\vartheta}{1-\vartheta}}.$$

Given (A.15) we can solve for the human capital cutoffs $\overline{h}_{rt}$ as

$$g_N = \frac{\theta}{\theta-1}e^{BGP}\frac{\overline{h}_{rt}\ell_{rt}}{N_{rt-1}} \quad \text{and} \quad \overline{h}^{\theta}_{rt} = \frac{1}{e^{BGP}}\mathcal{H}^{\theta}_{rt} = \frac{1}{e^{BGP}}\zeta^{\theta}_r N^{\vartheta\theta}_{rt-1}\sum_\nu \lambda^\nu_{rt}\omega^\nu_{rt}.$$

Combining these equations and noting that ϖ_{rt}^{ν} and λ_{rt}^{ν} are stationary along a BGP yields

$$\bar{h}_{rt} = \left(e^{BGP}\right)^{\frac{\vartheta-\frac{1}{\theta}}{1-\vartheta}} \left(\frac{\theta}{\theta-1}\frac{1}{g_N}\right)^{\frac{\vartheta}{1-\vartheta}} \zeta_r^{\frac{1}{1-\vartheta}} \ell_{rt}^{\frac{\vartheta}{1-\vartheta}} \left(\sum_{\nu} \lambda_r^{\nu} \varpi_r^{\nu}\right)^{\frac{1}{\theta}\frac{1}{1-\vartheta}}. \tag{A.16}$$

As for wages, taking the price of output in a particular region j as the numeraire, the growth rate of wages (and hence income per capita) is given by

$$1 + g_w = (1 + g_A)(1 + \eta)^{\frac{1}{1-\vartheta}\frac{1}{\sigma-1}}. \tag{A.17}$$

Finally, we can solve for the local distribution of spatial knowledge N_{rt} along a BGP. In particular,

$$N_{rt} = \frac{1+g_N}{g_N} I_{rt} = \frac{1+g_N}{g_N}\frac{\theta}{\theta-1}\bar{h}_{rt} e^{BGP} \ell_{rt} \tag{A.18}$$

Labor Supply Along the BGP Given the BGP distribution of wages, we can use the labor supply module to compute the population distribution. Along the BGP, the aggregate population grows at a constant rate and is fully determined from the mass of international inflows. In particular, (10) and (11) imply that

$$\ell_t^{\nu s} = \frac{1}{\eta + d} M_t^{\nu s} \quad \text{and} \quad \ell_t^{USs} = \frac{\eta^B}{\eta - \eta^B + d}\frac{1}{\eta + d}\sum_{\nu \in \mathrm{I}} M_t^{\nu s}. \tag{A.19}$$

Hence, given $M_t^{\nu s}$, which is exogenous in our theory, we can directly compute the mass of migrants and natives by skill group. The distribution across space then follows from (10) and (11) under the restriction that the distribution is stationary, i.e. $\ell_{rt}^{\nu s} = (1+\eta)\ell_{rt-1}^{\nu s}$.

A.2. Additional Data Details

Our analysis combines four primary data sources: individual-level U.S. population census from 1880 to 1920, the complete universe of U.S. patents since 1790, immigration arrival records from Castle Garden, and passenger departure records from the Hamburg Passenger Lists.

Section A.2.1 describes these datasets. Section A.2.2 explains the linkage procedures, and Sections A.2.3 and A.2.4 evaluate linked-data quality. The harmonization procedure is described in the Online Appendix.

A.2.1 Main Datasets for Analysis

U.S. Population Census *(1880—1920)*

The U.S. Population Census covers the full population and reports nativity, immigration status, and employment. We use the restricted-access version from Ruggles et al. (2020), which contains individual-level identifiers such as names and places of birth. These identifiers allow us to link individuals across census waves (intercensal linkage) and to integrate census records with patent and immigration data.[18]

Historical Immigration Records *(1820–1914)*

We construct our immigration database from two sources: the Castle Garden (CG) Immigration Database, which records *inflows* to the United States, and the Hamburg Passenger Lists (HPL), which record *outflows* from Europe to the United States.

The *Castle Garden Immigration Database (1820–1914)* The CG Immigration Database contains records of individuals arriving in New York City between 1820 and 1914. During this period, most immigrants to the United States entered through the port of New York, making the CG database a comprehensive source for immigration in this era. The database includes approximately 11 million individuals and reports name, age, nationality, year of arrival, and, notably, pre-migration occupation. These records were compiled from the Customs Passenger Lists prepared by ship captains and filed upon arrival. To link CG records to the U.S. Population Census, we use name, age, nationality, and arrival year. Additional details are provided in Section A.2.2.

The *Hamburg Passenger Lists (1850–1914)* The HPL provide records of individuals departing from Hamburg, Germany, to destinations worldwide, including the United States, between 1850 and 1939. The database contains approximately six million individuals and reports name, age, nationality, departure date, final destination, and, importantly, occupation. For consistency with the CG database, we use four variables — name, age, nationality, and year of departure — to link HPL records to the U.S. Population Census, following the procedure described in Section A.2.2.

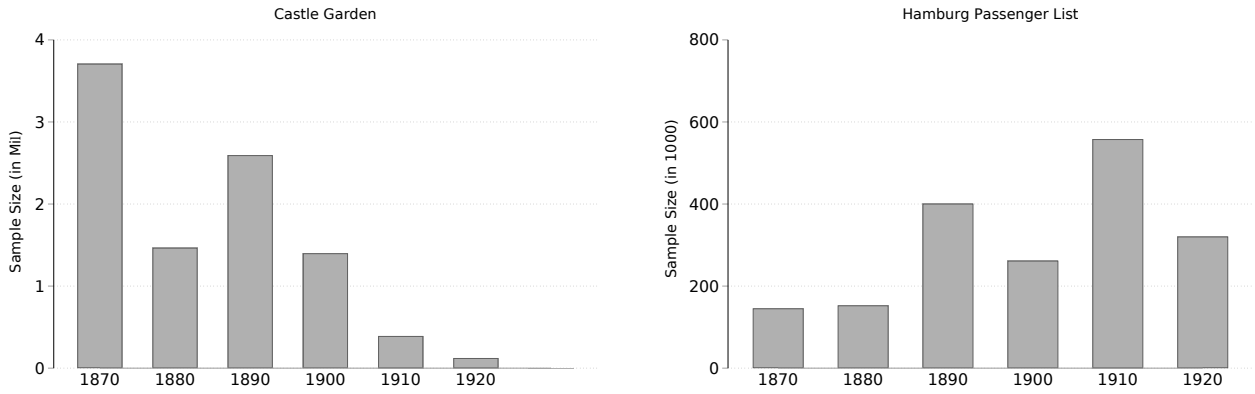
Access to the HPL is granted through a cooperation agreement with the State Archive of Hamburg. Due to privacy restrictions, we use data only through 1914, covering approximately 4.6 million individual records. Of these, approximately 77 percent list the United States as the intended destination. Additional details on data preparation are provided in the Online Appendix.[19]

[18] We additionally link the 1920 and 1930 censuses to study internal migration patterns of natives and immigrants within the United States.

[19] Together, the CG and HPL databases contain approximately 13 million individual records. Our working

In Figure A.1, we present the number of individuals in the CG Database (left panel) and the HPL (right panel) by decade, showing that the two sources provide complementary temporal coverage: the CG database contains more immigrants in earlier decades, whereas the HPL includes more immigrants after 1890.

FIGURE A.1: THE HISTORICAL IMMIGRATION RECORDS



Notes: The figure shows the number of individuals in the Castle Garden (left panel) and Hamburg Passenger Lists (right panel) databases, by decade. Bars report the working sample of male immigrants used for record linkage; for the HPL, this further restricts to records listing the United States as the intended destination. These counts are therefore smaller than the full database totals reported in the text. Records are binned by decade; the final bin (labeled 1920) covers 1910–1914, as both databases end in 1914 with the outbreak of the First World War.

Historical Data on U.S. Patents

To measure innovative activity, we use two primary sources of patent data: Google's Patent Search Engine (https://patents.google.com/) and PatentCity (Bergeaud and Verluise, 2024).

Google Patent Data. Since 1790, the United States Patent and Trademark Office (USPTO) has granted millions of patents, with digitized records available beginning in 1836. In collaboration with Reed Tech, Google applied Optical Character Recognition (OCR) to pre-1975 U.S. patent documents, converting them into machine-readable text. From Google's Patent Search Engine, we extract patent numbers and grant years, inventor and assignee names and locations, invention descriptions, and industry classifications based on the Cooperative Patent Classification (CPC) system.

PatentCity Data. Bergeaud and Verluise (2024) extend this digitization effort by collecting original patent document images, applying OCR, and using Named Entity Recognition (NER) to identify inventors, assignees, and geographic information. PatentCity systematically records patent numbers, grant years, inventor and assignee names, and detailed location information (city, county, and state).

sample is restricted to male immigrants: we exclude women because last-name changes at marriage make record linkage unreliable.

Together, these sources provide complementary strengths: Google offers detailed classification and descriptive textual information, while PatentCity provides enhanced disambiguation of inventors, assignees, and locations.

We construct a unified patent-level dataset by combining patent descriptions and industry classifications from Google with inventor and location information from PatentCity. We harmonize CPC-based industry classifications to the 1950 Census industry classification (IND1950). We then link patent records to the population census using inventor names, locations, and patent grant years. Additional details on the linkage procedure are provided in Section A.2.2.

A.2.2 Record Linking

This section describes (i) linking immigrants from the CG and HPL databases to the U.S. Population Census, (ii) linking patent inventors from U.S. patent records to the census, and (iii) linking census records across years.

Linking with Random Forest Classifiers: Supervised Machine-Learning

We implement a supervised machine-learning approach—random forest classification (henceforth, RFC)—to link records. RFC is a bagging method that trains multiple decision trees on subsamples of the training data. This approach reduces overfitting and improves predictive stability (Lee, 2019).

Before proceeding with RFC, we compared its performance to two commonly used machine-learning record-linking methods: support vector machines (SVM) and XGBoost (XGB). Relative to SVM, RFC produced fewer false positives and a larger number of unique matches. RFC and XGBoost performed similarly, but we select RFC due to its interpretability, scalability, and ease of tuning.

The linking procedure proceeds in two stages: identifying potential matches using RFC and selecting unique matches from this candidate set.

Training the Classifier We train the classifier using existing linked data. Because RFC is supervised, training requires labeled training data. We use hand-labeled and SVM-labeled matches provided by IPUMS, which contain census individuals linked across waves. The training data include a binary match indicator equal to 1 for matches and -1 for non-matches.

From the labeled data, we construct features that allow RFC to distinguish matches from non-matches. To improve model tractability, we apply recursive feature elimination (RFE) to select the final feature set.[20]

[20] RFE iteratively removes the least important variable until a user-specified number of features remains. We measure feature importance using logistic regression coefficients and select the final feature set via grid search to maximize classifier accuracy.

The selected features are: **name similarity**, the Jaro–Winkler similarity for given name and surname;[21] **Double Metaphone match**, an indicator equal to one if records share at least one phonetic code under the Double Metaphone algorithm, evaluated for given name and surname;[22] **NYSIIS surname match**, an indicator equal to one if surnames share the same NYSIIS phonetic code (Rajkovic and Jankovic, 2007; Koneru et al., 2016); **initial match**, an indicator equal to one if the first letters of given and surnames match; **initial only**, an indicator set to true if either record has a single-letter given name; **age difference**, the absolute age difference in years (CG–Census and HPL–Census only); and **percentage age difference**, the age difference divided by mean age (CG–Census and HPL–Census only).

We then tune RFC hyperparameters using k-fold cross-validation (Hastie et al., 2009). The dataset is partitioned into folds, with each fold serving as a validation set while the remaining folds are used for training. The average validation accuracy estimates the out-of-sample predictive performance (Fushiki, 2011). We conduct a grid search over RFC hyperparameter values and select the specification with the highest cross-validated accuracy (Pedregosa et al., 2011).

Identifying Potential Record Matches After training, the classifier generates predicted match probabilities for record pairs across datasets.

The predicted probability equals the mean predicted match probability for all decision trees in the random forest. We retain high-probability pairs as potential matches (See Section A.2.2).

When linking immigration records to the Census, we link all immigration records observed up to the census year. For example, when linking the 1900 Census to CG data, we consider all CG records through 1900. To reduce false positives and computational burden, we restrict comparisons to census records with matching country of birth. When the immigration year is observed, we require that the difference between the immigration year in CG or HPL and the census-reported immigration year is at most two years.[23]

When linking patents to the Census, we match patents within a 15-year window spanning the five years before to the nine years after the census year.[24] We restrict comparisons to

[21] Jaro–Winkler similarity enters the procedure in two ways: as a blocking criterion, candidate pairs with Jaro–Winkler similarity below 0.8 in either given name or surname are excluded prior to classification; and as a continuous feature supplied to the random forest. The 0.8 blocking floor follows prior literature (Jaro, 1989, 1995; Winkler, 2014); Feigenbaum (2016) reports that 95% of IPUMS-linked records exceed 0.8 similarity, and Abramitzky et al. (2021) describe a 0.9 threshold as conservative.

[22] The Double Metaphone is an algorithm that codes English words (and foreign words often heard in the United States) phonetically by reducing them to a combination of 12 consonant sounds, reducing matching problems from incorrect spelling.

[23] The 1900, 1910, and 1920 censuses report immigration year for foreign-born individuals. The 1880 Census does not; for 1880, matches are generated using the remaining identifiers—birthplace, birth year, name agreement, and middle initial—without the immigration-year filter. Because 1880 relies on a weaker constraint set, we place greater weight on the remaining identifiers and verify 1880 match quality using the departure-port validation described below.

[24] For example, we match the 1900 Census to patents from 1895 to 1909. Because the 1890 Census was

literate, English-speaking census individuals residing in the same county as the patent inventor.[25] The classifier then evaluates candidate pairs using the features described above.

Table A.1 summarizes the search constraints for each dataset pair.

Identifying Unique Record Matches After generating potential matches, we apply a pruning procedure to obtain uniquely matched pairs. A candidate pair qualifies as a potential match only if its predicted match probability is at least 0.85, as described in Section 2; the pruning procedure below therefore operates exclusively on pairs that already satisfy this threshold. Let r_A denote a record in dataset A and r_B a record in dataset B. Among potential matches, we retain the match (r_A, r_B) if either: (i) r_A matches only r_B and r_B matches only r_A; or (ii) the predicted match probability for (r_A, r_B) exceeds all alternative matches involving r_A or r_B by at least 15%.

In Section A.2.3, we evaluate the performance of our RFC-based methodology relative to alternative linking approaches.

Linking Immigrants to the U.S. Census

This subsection reports match counts and match rates for the linked immigration datasets.

Linking the CG Immigration Records to the U.S. Census The upper panel of Table OA.1 (reported in the Supplementary Material) reports match counts and match rates by nationality for the CG data linked to the Census. German immigrants account for approximately 50% of CG-Census matches and are slightly overrepresented in the final linked sample. In contrast, match rates for Irish and British immigrants are systematically lower. The table also shows a sharp increase in Italian immigration around the turn of the twentieth century.

Linking the HPL Immigration Records to the U.S. Census The lower panel of Table OA.1 reports match counts and match rates for HPL-Census links by nationality and period. Because Hamburg is a German port city, German immigrants constitute the majority of recorded passengers. Italian and Irish passengers are underrepresented, as they typically departed for the United States from other ports.

Section A.2.3 discusses potential sources of variations in match rates across nationalities and over time and evaluates implications for sample representativeness.

destroyed, we match the 1880 Census to patents from 1875-1899.

[25] Following Abramitzky et al. (2020), restricting to county may exclude individuals who moved between patenting and census enumeration. We therefore also test broader geographic constraints using State Economic Areas (SEAs).

TABLE A.1: SEARCH CONSTRAINTS FOR RECORD LINKING

Datasets Linked	Search Constraints
CG–Census	– Matching birthplace country in CG and census – $\lvert$CG arrival year $-$ census immigration year$\rvert \leq 2$ (applied when census immigration year is observed; see note) – $\lvert$CG birthyear $-$ census birthyear$\rvert \leq 2$ – Matching middle initial (if exists in both records)
HPL–Census	– Matching birthplace country in HPL and census – $\lvert$HPL departure year $-$ census immigration year$\rvert \leq 2$ (applied when census immigration year is observed; see note) – $\lvert$HPL birthyear $-$ census birthyear$\rvert \leq 2$ – Matching middle initial (if exists in both records)
Patent–Census	– Literate and English-speaking US residents only – Matching US residence county in patent and census records – Matching middle initial (if exists in both records)
Census–Census[†]	– Matching birthplace (birth country if foreign-born; birth state if US-born) – $\lvert$census birthyear difference$\rvert \leq 2$ – Close agreement in given name, surname, and parental birthplaces

Notes: This table enumerates the search constraints used to generate the RFC matches. We only perform random forest classification on record pairs that meet the search constraints. When pruning, we apply the rules in the order specified. Before applying each rule, we save the uniquely matched record pairs and ignore them when applying subsequent pruning rules. The immigration-year constraint for CG– and HPL–Census links applies only to censuses that report immigration year (1900, 1910, and 1920); the 1880 Census does not record immigration year, so for 1880 candidate pairs are screened on the remaining constraints—birthplace country, birth year (± 2), name agreement, and middle initial—without the immigration-year filter. [†] For census-to-census linkage, geographic location and occupation are deliberately excluded from the search constraints to avoid mechanically reducing match rates for individuals who migrate or change occupation or industry.

Linking Patents to the U.S. Census

This subsection reports match counts and rates between patent records and Census individuals and describes the temporal matching strategy.

Table A.2 reports the number of matches between patents and census individuals by year. The first row reports matched patent–inventor records, while subsequent rows report the number of distinct matched inventors. The mapping between patents and inventors is many-to-many: a single patent may list multiple inventors, each matched separately to the Census, and a single inventor may hold many patents. For this reason, the patent–inventor record count differs from the count of unique matched patents reported in Table 1, and the two need not coincide in any given year.

Figure OA.5, reported in the Supplementary Material, presents the distribution of patent grant years among matched inventor records. Patents are matched within a 15-year

TABLE A.2: MATCH COUNTS, PATENTS TO CENSUS

	1880	1900	1910	1920
Patent Match Count	115,100	94,867	120,298	120,652
Inventor Count: Total	57,027	53,349	68,561	67,852
Inventor Count: Native	44,145	42,296	55,093	54,167
Inventor Count: Foreign	12,882	11,053	13,468	13,685

Notes: The first row reports the number of patent–inventor match records linking patents to Census individuals; because a single patent may list multiple inventors, each matched separately to the Census, this count exceeds the number of unique matched patents reported in Table 1. Subsequent rows report the number of distinct Census individuals (inventors) matched to at least one patent; a single individual may hold multiple patents.

window spanning the five years before to the nine years after each census year; the 1880 window is extended through 1899 because the 1890 Census was destroyed. Shaded regions indicate patent years that fall within the windows of two adjacent censuses and are therefore eligible for matching to either (1895–1899, 1905–1909, and 1915–1919). When an inventor is matched in both censuses, we retain the match associated with the census year closest to the patent grant year. This rule minimizes temporal misalignment between inventive activity and observed residence and improves internal consistency, as shown in Section A.2.3.

Linking Censuses Across Years

We match male individuals across the decennial U.S. Population Censuses between 1880 and 1920. The linkage criteria require (i) identical birthplaces (i.e., birth country for foreign-born individuals; birth state for U.S.-born individuals), (ii) age differences of at most two years, and (iii) close agreement in names and parental birthplaces. We deliberately exclude geographic location (e.g., county or state of residence) and occupation from the linking criteria to avoid mechanically reducing match rates for individuals who migrate or change occupation/industry.

Other Approaches to Linking Historical Records

We compare our supervised machine-learning approach (RFC-based) to commonly used linkage procedures in the literature. Section A.2.4 compares match datasets generated by alternative approaches. Replication is facilitated by publicly available documentation and code.[26] We apply each method to the same immigration datasets and the U.S. population

[26]See https://ranabr.people.stanford.edu/historical-record-linking for code and descriptions of commonly used historical linkage procedures.

census.[27]

Abramitzky, Boustan, and Eriksson (ABE): Iterative Linking

We implement the iterative procedure of Abramitzky et al. (2012, 2014). For datasets A and B, we (i) restrict to records in A unique on the matching key (given name, surname, age, birthplace) within a five-year band (± 2 years); (ii) retain only one-to-one candidates in B, discarding multiple-candidate cases; (iii) repeat allowing age differences of one then two years for unmatched records; and (iv) run the reverse pass ($B \to A$) and keep the intersection. We use both the phonetic (NYSIIS) and Jaro–Winkler variants, denoted *ABE-NYSIIS* and *ABE-JW*; the JW variant matches on first and last initials and birth years within a five-year band and retains pairs with JW distance < 0.1.

Expectation Maximization (EM)

We also implement the unsupervised EM procedure of Abramitzky, Mill, and Pérez (2020), which needs no hand-linked training data. For datasets A and B, we (i) block candidates in B on birthplace, first and last initials, and age differences ≤ 2 years; (ii) compute Jaro–Winkler distances for given names and surnames; (iii) estimate the posterior probability of a true match via EM; and (iv) retain pairs with match probability $\geq p_m$ whose next-best candidate falls below L, setting $p_m = 0.5$ and $L = 0.45$.

A.2.3 Quality of Matches

While the previous section compares algorithm outputs, this section evaluates the accuracy of our RFC-based linking procedure using independent validation information, assessing whether it introduces systematic bias or disproportionately represents particular groups relative to alternative methods.

Match Accuracy

We evaluate RFC linking accuracy along two dimensions. First, we use departure-port information in the immigration datasets to assess immigration-census linkage quality. Second, we use city-of-residence information in the patent data to evaluate patent-census linkage accuracy.

Immigration–Census Accuracy: Departure-Port Validation The Castle Garden (CG) dataset records the departure port of all arriving passengers. By construction, all Hamburg

[27] Inventors frequently file multiple patents, which implies a many-to-one structure for patent-census linkage. Many existing linkage procedures (including iterative linking and EM) are designed for one-to-one matches. We therefore restrict method comparisons to immigration-census linkage to ensure comparability across approaches.

Passenger Lists (HPL) records originate from Hamburg. Therefore, for census individuals matched to both CG and HPL, the corresponding CG record should list Hamburg as the departure port. While the ground truth of match status of individuals is unobserved, this implication provides a natural validation criterion.

We construct three validation sets of HPL-census matches: (i) matches generated solely by RFC; (ii) unique matches identified by at least two independent linking methods; and (iii) unique matches identified by at least three independent linking methods.

Treating individuals in each validation set as true Hamburg-origin individuals, we define precision as the share of CG matches whose recorded departure port includes Hamburg (either directly or via transit). A classifier with perfect precision would assign Hamburg as the departure port for all CG-HPL matched individuals.

Table A.3 reports the five most common departure ports for CG-HPL-matched individuals across the three validation sets. Across specifications, between 77 and 85 percent of matched individuals list a Hamburg-related departure port — Hamburg directly (roughly 40–43%), Hamburg–Havre (31–35%), or Hamburg–Southampton (7–8%) — consistent with our precision definition. The remaining matches list non-Hamburg ports such as Bremen (7%) and Liverpool–Queenstown (2%), which likely reflect a small share of transit passengers or false matches. Taken together, these results indicate high within-class accuracy across all three validation sets.

TABLE A.3: DEPARTURE PORT VALIDATION

	% of HPL-Matched Individuals		
CG Departure Port	RFC only	≥ 2 methods	≥ 3 methods
Hamburg-related ports			
Hamburg	40.0	39.9	43.3
Hamburg–Havre	30.7	30.9	34.5
Hamburg–Southampton	6.9	6.9	7.5
Hamburg-related total	**77.6**	**77.7**	**85.3**
Other ports			
Bremen	7.1	7.3	3.1
Liverpool–Queenstown	2.7	2.3	2.1

Notes: The above table reports the frequency of the top five departure ports for HPL-matched individuals and matched with RFC to CG records.

In Table A.4, we also compare the precision of each linking method, measured by the consistency of departure-port information. We report two RFC specifications: the baseline, and a more conservative variant that imposes the ABE uniqueness restriction, limiting matches to CG and Census records with unique names within a five-year (± 2) band. On

the consensus validation sets—matches agreed upon by at least two, or at least three, independent methods—RFC attains precision comparable to the alternative procedures (roughly 79–80% for the two-method overlap and 87% for the three-method overlap) while generating substantially more matches overall. Importantly, RFC's non-consensus matches achieve comparable precision: as shown in the "RFC only" column of Table A.3, matches identified by RFC alone exhibit a Hamburg-related departure-port share of 77.6%, essentially identical to the 77.7% share for matches agreed upon by at least two methods. This indicates that RFC's larger match volume does not come at the cost of lower precision on the additional matches it recovers.

TABLE A.4: COMPARISON OF LINKING METHOD PRECISION

	Match Volume		Precision (% Hamburg port)	
Linking Method	Total	Rel. to RFC	≥ 2 methods	≥ 3 methods
RFC	149,590	1.00	79.2	86.8
RFC (unique, ± 2 yrs)	120,726	0.81	80.3	87.7
ABE (NYSIIS)	70,108	0.47	79.9	87.3
EM	53,160	0.36	80.4	86.1
ABE (JW)	28,806	0.19	80.9	86.6

Notes: This table compares Hamburg departure frequency for each linking method. In this table, we construct the RFC dataset with and without the name uniqueness constraints imposed by ABE, where the procedure only matches CG and Census records with unique names in a five year band. We apply immigration year filters for all methods.

Patent–Census Accuracy: City-of-Residence Validation To assess the quality of patent-census matches, we compare inventor residence information reported in the census to residence information extracted from patent texts. As described in Section A.2.2, the linkage procedure restricts candidate matches to individuals residing in the same U.S. county in both patent and census records. Sub-county geographic information (e.g., city or town) is not used in the matching algorithm. For example, if a patent lists the inventor as residing in ‘‘`Ann Arbor, in the county of Washtenaw, State of Michigan`’’, the search space is restricted to census individuals in Washtenaw County, Michigan. We define a *city match* as a case in which the sub-county location (e.g., "Ann Arbor") coincides in both the patent text and the census record, even though city information was not used in the matching procedure. Figure OA.6, in the Supplementary Material, reports city match rates as a function of the difference between patent publication year and the census year. City match rates are highest when the patent publication year is closest to the census year, consistent with within-county mobility over time. Nevertheless, match rates remain high across specifications. For the 1880 Census, city match rates exceed 75 percent across year gaps, and for later census waves, they exceed 90 percent when patents are published near the

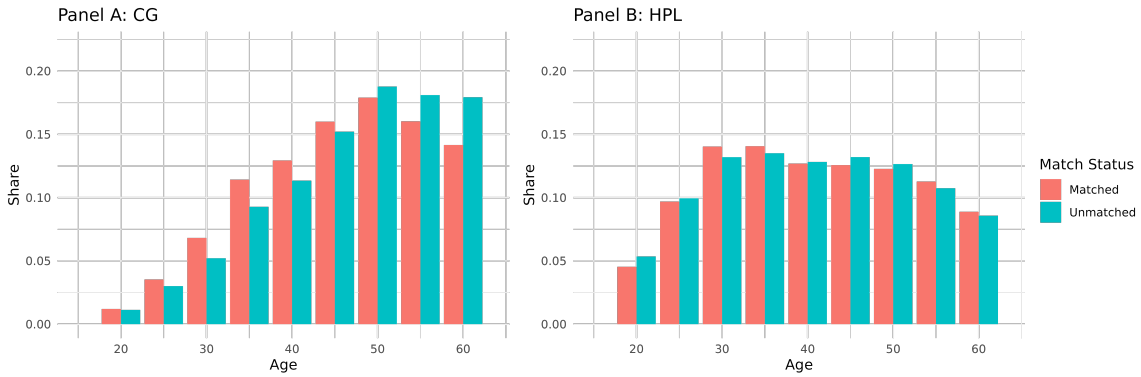
census year.

Representativeness

Beyond match accuracy, we assess representativeness by comparing observable characteristics of matched and unmatched records along two dimensions: the age distribution of immigration-census matches and the patent classification distribution of patent-census matches. Two patterns emerge: for Castle Garden, matched immigrants are somewhat younger on average than unmatched immigrants, while for the Hamburg Passenger Lists the matched and unmatched age distributions are very similar; matched patent records preserve the distribution of classifications observed in the full dataset. Taken together, these patterns indicate that the linked samples retain the central demographic and technological structure of the original populations.

Immigration–Census Representativeness: Age Distributions We begin by comparing matched and unmatched individuals in the immigration datasets. Figure A.2 reports age distributions for matched and unmatched male immigrants aged 16-60 linked to the 1910 Census. For Castle Garden (CG), matched individuals are younger on average than unmatched individuals. For the Hamburg Passenger Lists (HPL), the matched and unmatched age distributions are very similar, with no appreciable difference in mean age. The CG pattern is consistent with life-cycle bias and age misreporting among older individuals, which can reduce match probabilities; for HPL, matched and unmatched individuals are comparable in age.[28]

FIGURE A.2: AGE DISTRIBUTIONS FOR IMMIGRATION RECORD MATCHES



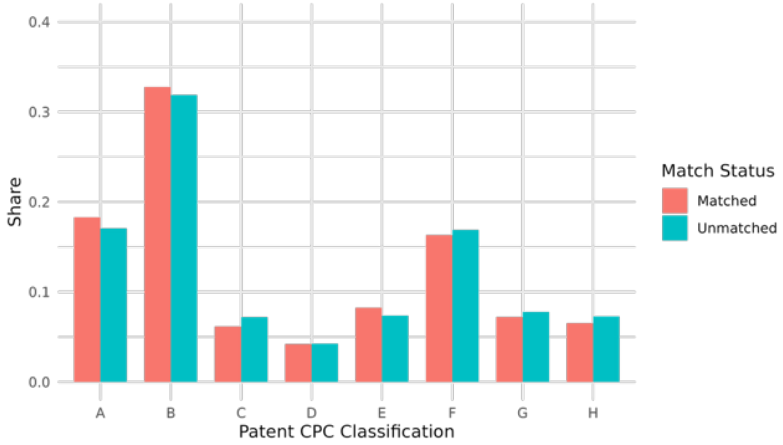
Notes: The above figures show the age distribution across matched and unmatched observations from the CG Data (left panel) and HPL (right panel) against 1910 Census. The matched group includes males aged between 16 and 60 years who are uniquely matched with 1910 Census. The unmatched group includes males aged between 16 and 60 years who are unmatched with 1910 Census.

[28] Age misreporting among older individuals is well documented in the demographic literature and can reduce linkage accuracy.

Patent–Census Representativeness: Patent-Classification Distributions We next examine whether the RFC linkage alters the distribution of patent classifications. Figure A.3 compares the distribution of patent classes in the full patent dataset to the distribution among matched patents.

The distributions are highly similar, indicating that RFC-based linkage does not disproportionately select particular technological categories. This suggests that the matching procedure preserves the composition of patent classifications.

FIGURE A.3: PATENT CLASSIFICATION DISTRIBUTIONS FOR PATENT MATCHES



Notes: This figure shows the distribution of patent classifications across matched and unmatched patent observations.

A.2.4 Comparing Match Datasets

Having described the linking procedures and alternative algorithms, we next compare the datasets generated by each method to assess robustness across specifications.

We compare the outputs of all linkage methods described in Section A.2.2 using CG-1910 Census matches.[29] For all methods, we impose identical immigration-year filters (an absolute difference of at most two years between immigration records and census reports). Table A.5 reports match counts by method and birthplace (four major immigrant-sending countries). RFC generates substantially more matches than alternative procedures.

Taken together, these comparisons indicate that RFC increases match volume without materially altering observable age or birthplace distributions: as the right panel of Figure A.4 shows, the share of matched CG records by place of birth under RFC closely tracks both the alternative algorithms and the full CG cross-section.

[29] ABE and EM produce one-to-one matches; therefore, comparisons are restricted to immigration-census linkage rather than patent-census linkage.

TABLE A.5: CG-CENSUS MATCH COUNTS BY LINKING METHOD

Birthplace	RFC	ABE (NYSIIS)	EM	ABE (JW)	RFC vs. best alt.
All matches	187,490	69,738	43,544	31,243	2.7×
British	26,682	7,676	3,168	7,249	3.5×
German	102,595	39,356	22,153	15,235	2.6×
Irish	21,377	9,869	6,750	4,134	2.2×
Italian	22,083	4,188	1,578	1,354	5.3×

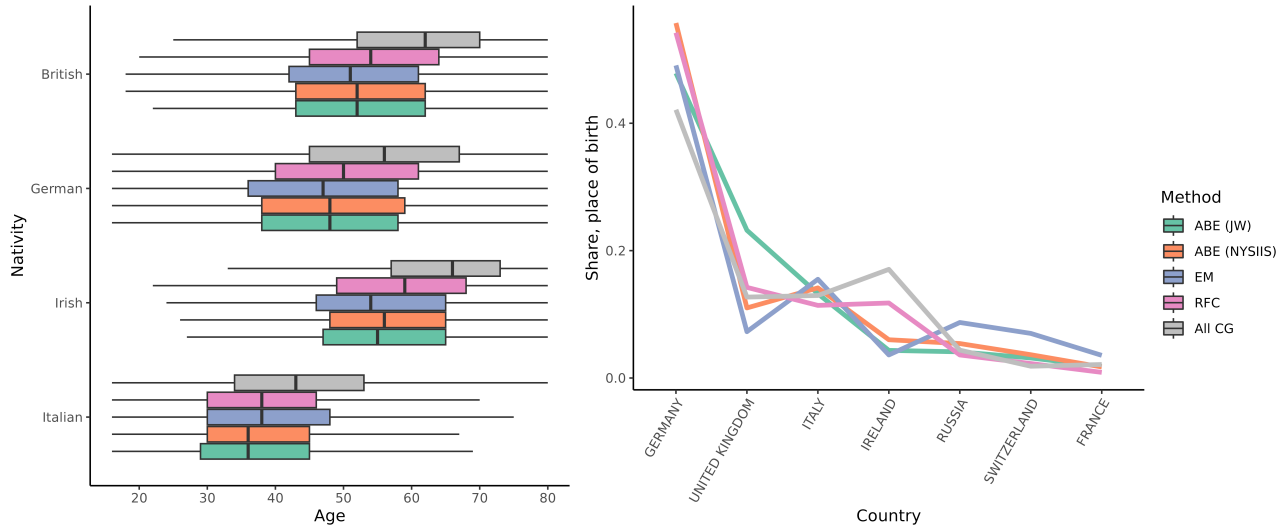
Notes: The above table shows match counts for the 1910 CG and Census, categorized by linking method and by birthplace (four major immigrant-sending countries during our study period). RFC-based linking produces a significantly larger number of unique matches than other methods, and the distribution of birthplaces falls within those of ABE and EM.

Distribution Comparisons

We next compare age and birthplace distributions across methods. The left panel of Figure A.4 plots age distributions for matches generated by ABE (JW and NYSIIS variants), EM, and RFC, along with the full 1910 CG cross-section. The RFC age distribution closely mirrors those produced by alternative methods. The original cross-sectional CG age distribution skews older, reflecting the mechanical absence of deceased immigrants in census matches.

The right panel reports birthplace distributions for each linking procedure. RFC preserves nationality shares that are similar to those generated by ABE and EM, and closely aligned with the original CG distribution.

FIGURE A.4: Age and Birthplace Distributions by Linking Method



Notes: The left panel (boxplots) shows age distributions for CG records linked by ABE-JW, ABE-NYSIIS, EM, and RFC, as well as the age distribution for all CG records. The right panel reports the share of matched records by place of birth for each linking method, alongside the distribution for all CG cross-sectional records.

Match Overlap

Figure OA.9 in the Supplementary Material illustrates the overlap of matched record pairs across methods. Each procedure identifies a substantial number of unique matches. Approximately 33 percent of RFC matches overlap with at least one alternative method. After imposing immigration-year constraints, RFC produces a larger number of unique matches without reducing accuracy, as measured by departure-port validation.

A.3. Quantitative Analysis

In this section we provide more details for our quantitative analysis. Section A.3.1 contains details for our estimation strategy. Section A.3.2 contains additional results for our counterfactual analysis of immigration restrictions. Finally, in Section OA.4 we discuss the robustness analysis reported in Section 6 in more detail.

A.3.1 Estimation Details

Matched Immigration Records As described in Section 4.1, we match immigration records from Castle Garden (CG) and the Hamburg Passenger Lists (HPL) to the US Population Census. In Table A.6 we report the number of matches and the size of the respective immigration record data. We are able to match between 85,000 and 150,000 CG records to the Census and between 25,000 and 50,000 records from the HPL.

Table A.6: Immigration Records: 1880–1920

	1880	1900	1910	1920
Number of records in CG	1105819	996906	226073	18578
Number of CG records matched to Census	136581	152276	86939	84973
Number of records in HPL	112146	205996	490550	286296
Number of HPL records matched to Census	25095	50434	35076	49034
Number of matches in both CG and HPL	7868	14922	7752	6984

Notes: The table reports the number of immigration records in the Castle Garden (CG) and Hamburg Passenger Lists (HPL) databases for each census year, along with the corresponding number of records successfully linked to the US Population Census. The final row reports the number of immigrants who are matched to the Census in both CG and HPL.

Aggregate Growth We measure the growth rate of real aggregate GDP per capita in our model using the Fisher chained index. The Fisher Index is defined by

$$g_t^F = \sqrt{\frac{P_{t-1}Y_t/L_t}{P_{t-1}Y_{t-1}/L_{t-1}} \times \frac{P_tY_t/L_t}{P_tY_{t-1}/L_{t-1}}},$$

where P_t is the price of the consumption good at time t, Y_t is the total quantity of production, and L_t is the total population.

In the context of our model with R regions $r = 1, ..., R$, we construct the following auxiliary indices of output per capita, evaluated at the respective prices:

$$Y_{t-1}(P_{t-1}) = \frac{\sum_{r=1}^{R} P_{rt-1} Y_{rt-1}}{\sum_{r=1}^{R} L_{rt-1}} \quad \text{and} \quad Y_{t-1}(P_t) = \frac{\sum_{r=1}^{R} P_{rt} Y_{rt-1}}{\sum_{r=1}^{R} L_{rt-1}}.$$

and

$$Y_t(P_{t-1}) = \frac{\sum_{r=1}^{R} P_{rt-1} Y_{rt}}{\sum_{r=1}^{R} L_{rt}} \quad \text{and} \quad Y_t(P_t) = \frac{\sum_{r=1}^{R} P_{rt} Y_{rt}}{\sum_{r=1}^{R} L_{rt}}.$$

where P_{rt} denotes the price of total output in region r at time t, and Y_{rt} denotes total physical output in region r at time t.

We then combine these expressions into the corresponding Fisher Index as follows:

$$g_t^F = \sqrt{\frac{Y_t(P_{t-1})}{Y_{t-1}(P_{t-1})} \times \frac{Y_t(P_t)}{Y_{t-1}(P_t)}}.$$

Total real GDP per capita at time t relative to 1880 is then simply given by

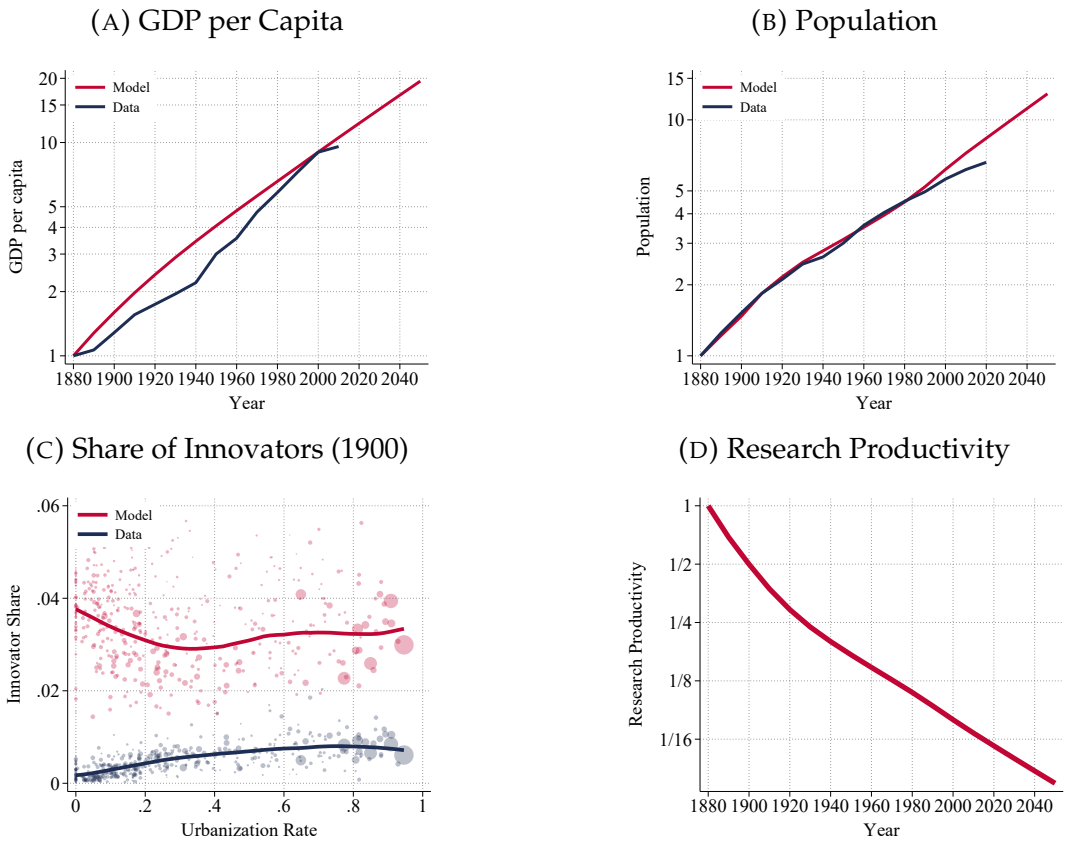
$$y_t / y_{1880} = \prod_j g_j^F \tag{A.20}$$

In Panel A of Figure A.5, we plot GDP per capita as observed in the data (blue line) and in our model as computed according to (A.20).

Estimation of θ In Figure 2 we displayed the regression in (14) graphically. In Table A.7 we report the actual regression results. In the first 5 columns, we (as suggested by our theory) only control for region-year fixed effects, where a region is defined as an "SEA-urban cell". In the different specifications, we impose different thresholds for the maximum number of patents we consider. In all specifications, we estimate θ to be around 1.3. In the last two columns, we explicitly control for a full set of region-type-year interactions, where a type is a nationality-skill combination. Hence, this specification includes fixed effects for say skilled British immigrants living in an urban location in a particular SEA in a given decade. Doing so does not affect our estimate of the slope coefficient.

Estimation of Demographics: η_b and d We discipline the birth rate η_b and the death rate d from the rate of population growth and the share of immigrants between 1880 and 2000 given the observed inflow of immigrants (M_t). Our primary source for immigrant inflows is census microdata, where we identify decadal arrivals using the information on year of immigration. Since this information is not contained in the 1940, 1950, and 1960 censuses, we fill these gaps using administrative records on legal immigrant arrivals from the Department of Homeland Security (DHS). The DHS records differ conceptually from

FIGURE A.5: MODEL FIT: ADDITIONAL IMPLICATIONS



Notes: In this figure we display GDP per capita (Panel A), the overall population (Panel B), the correlation between the urbanization rate (as measured in the data) and the share of innovators in 1900 (Panel C), and aggregate research productivity (Panel D).

TABLE A.7: ESTIMATION OF PARETO TAIL: θ

	(1)	(2)	(3)	(4)	(5)
log Number of patents	-1.305***	-1.306***	-1.309***	-1.317***	-1.356***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Region-Year FE	yes	yes	yes	yes	yes
Max num of patents	-	200	100	50	20
R^2	.785	.785	.785	.784	.778
N	40,422	40,418	40,407	40,320	39,593

Notes: The table reports estimates of equation (14). In columns 1-5 we control for region-year FE and consider different cutoffs for the number of patents in each cell. In column 6 and 7 we control for a full set of region-year-nationality interactions.

the census data because they focus on legal immigration. We thus assume that the total number of immigrants reported in the DHS data (M_t^{DHS}) is related to the immigrant inflow in the Census and our model (M_t) according to $M_t^{DHS} = \kappa^{DHS} M_t$. We then calibrate the three numbers $(\eta_b, d, \kappa^{DHS})$ to jointly minimize a loss function with two components: the squared deviation of model-implied total population growth from the data (1880–2000), and the mean squared error between model-implied and observed immigrant shares across all available years. As seen in Figure 6, we fit the aggregate share of immigrants well. As seen in Panel B of Figure A.5, our model also replicates the overall rate of population growth well.

Spatial Sorting of Recent Arrivals We estimate equation (19) via pooled regression and report the results in Table A.8. Across all specifications, population and nationality share enter with positive and significant coefficients, with population elasticities ranging from 0.519 to 0.734 and nationality share coefficients ranging from 0.346 to 0.530, indicating that immigrants systematically sort toward larger destinations and destinations with a bigger population share of their own nationality. The interaction between urban status and immigrants' skill is also positive and statistically significant throughout, suggesting that skilled immigrants disproportionately select into urban areas upon arrival. We take the estimates from Column (5), which absorbs the most comprehensive set of fixed effects via nationality × decade × skill interactions, as our preferred specification and use these coefficients to compute immigrants' arrival according to (18).

Estimation of Relative Skill Supply In Figure 3 we have shown the relative skill supplies of immigrants of different nationalities and of the native population. As explained in the main text, these results are computed from $\lambda_{rt}^{\nu} = \kappa^{\nu} \lambda_{rt}^{US}$, where κ^{ν} is estimated from (21), which we for convenience replicate here:

$$\ln e_{rt}^{\nu} = \ln\left(\lambda_{rt}^{\nu}\left(\frac{\psi_{rt}}{\bar{h}_{rt}}\right)^{\theta}\right) = \ln\kappa^{\nu} + \ln\left(\lambda_{rt}^{US}\left(\frac{\psi_{rt}}{\bar{h}_{rt}}\right)^{\theta}\right) = \ln\kappa^{\nu} + \delta_{rt}. \tag{A.21}$$

In Table A.9 we report the results of estimating this regression. Column 1 contains our baseline estimation, where we control for (as implied by our theory) a year-region fixed effect δ_{rt}. In columns 2 and 3 we only consider region-year-nationality cells with at least 200 or 500 observations. Finally, in column 4 we use the entire sample by adding a small positive number to e_{rt}^{ν} in case $e_{rt}^{\nu} = 0$. Table A.9 shows that the estimates of κ^{ν} are insensitive to these choices.

Time-invariant fundamentals As explained in the main text, our calibration strategy delivers estimates of local fundamentals $\{\zeta_{rt}, A_{rt}, \mathcal{B}_{rt}\}$ for the decades 1880, 1900, 1910, and 1920. Because such fundamentals are time-invariant in our theory, we estimate a region fixed effect and then use this estimate in our calibrated model. For example, we estimate the regression $\ln\zeta_{rt} = \delta_r + u_{rt}$ and then use $\zeta_r = \exp\left(\widehat{\delta_r}\right)$. We proceed in the

TABLE A.8: POOLED REGRESSION FOR ARRIVALS

	(1)	(2)	(3)	(4)	(5)
Urban	-0.177	0.187	0.217*	0.160	0.228*
	(0.111)	(0.114)	(0.121)	(0.117)	(0.119)
Urban × High Skill	0.355***	0.303***	0.347***	0.362***	0.219*
	(0.111)	(0.110)	(0.110)	(0.114)	(0.122)
Population	0.519***	0.636***	0.689***	0.733***	0.734***
	(0.0547)	(0.0412)	(0.0450)	(0.0436)	(0.0439)
Nationality Share	0.346***	0.410***	0.456***	0.527***	0.530***
	(0.0186)	(0.0183)	(0.0172)	(0.0191)	(0.0192)
Skill FE	Yes	Yes	Yes	No	No
Decade FE	No	Yes	Yes	No	No
Nationality FE	No	No	Yes	No	No
Nat× Decade FE	No	No	No	Yes	No
Nat× Skill FE	No	No	No	Yes	No
Nat× Decade× Skill FE	No	No	No	No	Yes
N	3,782	3,782	3,782	3,782	3,781
R^2	.287	.454	.543	.647	.654

Notes: This table reports the estimates of equation (19) via pooled regression. Column (1) includes Skill fixed effects. Columns (2) and (3) additionally control for Decade fixed effects. Column (3) further includes Nationality fixed effects. Column (4) replaces these with Nationality × Decade and Nationality × Skill fixed effects. Column (5) absorbs the most comprehensive set of fixed effects via Nationality × Decade × Skill interactions. The preferred specification is Column (5), whose coefficients serve as baseline parameters for model prediction. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.9: ESTIMATION OF RELATIVE SKILL SUPPLY: κ^{ν}

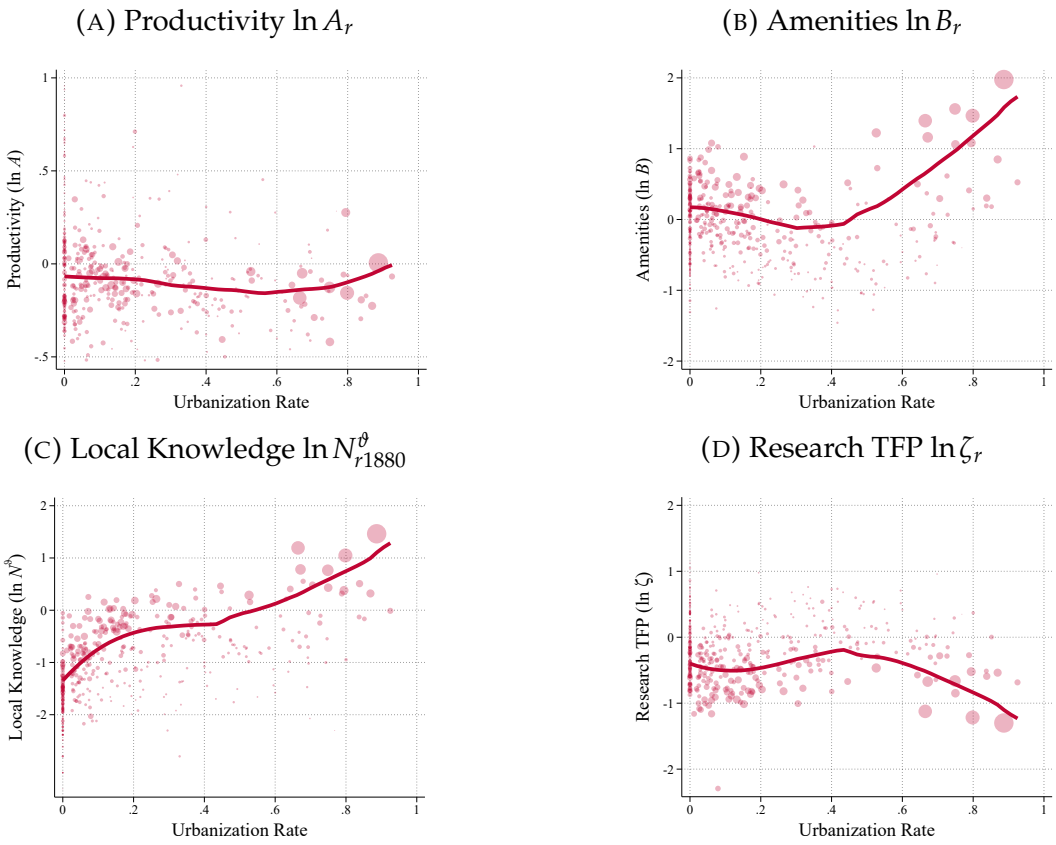
	(1)	(2)	(3)	(4)
British	0.279***	0.259***	0.222***	0.228***
	(0.025)	(0.025)	(0.026)	(0.024)
German	-0.269***	-0.277***	-0.294***	-0.271***
	(0.057)	(0.058)	(0.059)	(0.056)
Irish	-1.239***	-1.260***	-1.309***	-1.220***
	(0.093)	(0.091)	(0.087)	(0.088)
Italian	-1.472***	-1.482***	-1.506***	-1.409***
	(0.055)	(0.056)	(0.059)	(0.048)
Others	-0.524***	-0.530***	-0.541***	-0.521***
	(0.035)	(0.035)	(0.036)	(0.034)
Year-SEA FE	yes	yes	yes	yes
Cell size cutoff	-	200	500	-
Impute if zero?	no	no	no	yes
R^2	.887	.866	.854	.955
N	6,074	5,363	4,538	11,008

Notes: The table reports the results of estimating (21). Columns 1 uses the entire sample where $e^{\nu}_{rt} > 0$. In columns 2 and 3 we require all cells to have at least 200 or 500 observations respectively. In column 4, we add a small positive number to e^{ν}_{rt} in case $e^{\nu}_{rt} = 0$. All regressions are weighted by the number of observations in each cell.

same way for A_{rt} and $\mathcal{B}_{rt}$.

In Figure 5 we already plotted the resulting fundamentals for different deciles of urbanization. In Figure A.6 we plot the raw data. As also discussed around Figure 5, local productivity does not vary much with a region's urbanization rate, amenities are increasing at the very top, there is a steep urban gradient of local knowledge, and research TFP is relatively flat, except at the very top where the very large knowledge stock dictates a relatively low level of research TFP to rationalize the estimated innovation human capital ψ_r.

FIGURE A.6: LOCAL FUNDAMENTALS AND THE URBANIZATION RATE



Notes: Each panel plots a local fundamental against the urbanization rate in 1880: location productivity ($\ln A_r$), the exogenous component of amenities ($\ln B_r$), the knowledge component of local human capital ($\ln N^{\theta}_{r1880}$), and research TFP ($\ln \zeta_r$). Each dot is a State Economic Area (SEA) and the size of the dots reflects the SEA's population, overlaid with a local-linear fit.

Additional Results Figure A.5 also contains two additional implications of our estimated model. In Panel C we report the implied share of innovators as a function of the urbanization rate. As implied by Figure 6, our model predicts a higher share of innovators, reflecting the fact that not all innovation is patented. In terms of the cross-sectional predictions, our model predicts a non-monotone relationship. This is in contrast to the empirically observed pattern, which is increasing. Intuitively, our model implies that the return to innovation in very rural areas is quite high, given how little local knowledge (N_{rt}) these locations have accumulated. Between an urbanization rate of around 20%

and 100%, our model replicates the empirical relationship between the urbanization rate and the innovator share well. In Panel D we report our model's implication for research productivity, which (following Bloom et al. 2020) we define as idea growth per researcher, $g_{Nt}/\ell_{rt}e_{rt}$. Our model implies that

$$\frac{g_{Nt}}{\ell_{rt}e_{rt}} = \frac{\theta}{\theta-1}\frac{\zeta_r\left(\sum_\nu \lambda_{rt}^\nu \varpi_{rt}^\nu\right)^{1/\theta}}{N_{rt-1}^{1-\vartheta}e_{rt}^{\frac{1}{\theta}}}.$$

Research productivity is declining in the stock of knowledge N_{rt} (given that $\vartheta < 1$), reflecting the traditional logic of ideas getting harder to find. It is also decreasing in the share of innovators, because (given the aggregate skill supply and research TFP ζ_r) an increase in the share of innovators cuts deeper in the talent pool and reduces the average talent of researchers.

A.3.2 Quantification: Immigration and US Growth

A.3.2.1 Counterfactual I: The Era of Mass Migration: 1880-1920

This section contains additional results for our first quantitative exercise discussed in Section 5.1. Figure A.7 complements Figure 7 in the main text in that it shows the counterfactual path of the aggregate population (Panel A) and the aggregate flow of patents (Panel B), relative to the baseline for the three counterfactual migration paths considered in the main text. In the absence of international migration between 1880 and 1920, the overall population in the US would have declined by almost 30%. The decline in the flow of patents would have had almost the same magnitude. In Panels C and D, we depict the aggregate share of innovators and the population share in the urban SEAs (i.e., the SEAs in the top quintile of the urbanization distribution in 1880).

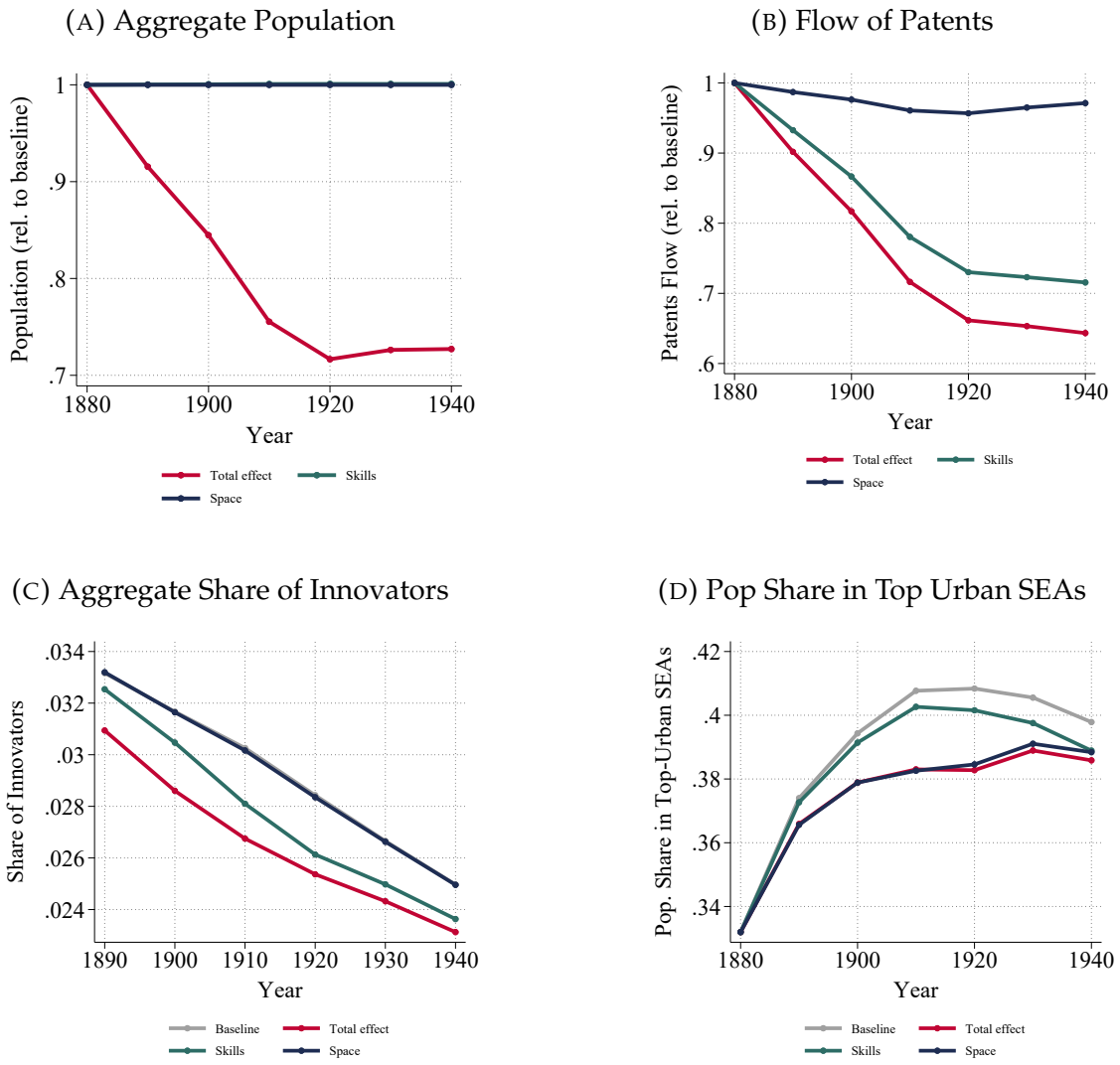
Figure A.8 complements Figure 8 and reports local population growth, relative to the baseline economy, in the absence of international arrivals. Consistent with migrants' urban bias, the population decline is particularly large in cities.

A.3.2.2 Counterfactual II: Removing the Immigration Restrictions in 1920s

In this section, we explain in more detail how we implement our second counterfactual of removing the immigration quotas in the early 20th century (Section 5.2).

Background The 1921 Emergency Quota Act and the subsequent 1924 Johnson–Reed Act represented a watershed in American immigration history. Prior to these statutes, the United States operated under a largely open-door regime: between 1880 and 1914, annual arrivals routinely exceeded one million, with Southern and Eastern Europeans accounting for the majority of the inflows. The 1921 Act imposed the first official ceiling on immigration, capping each nationality at three percent of its foreign-born population

FIGURE A.7: THE AGGREGATE IMPACT OF INTERNATIONAL MIGRATION

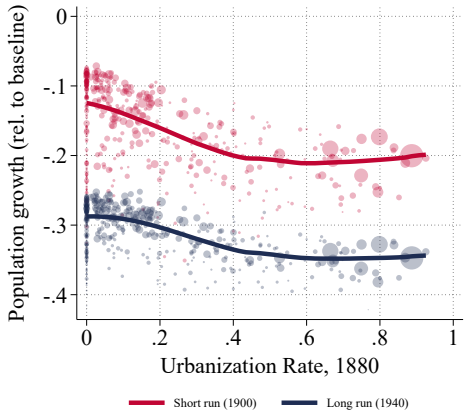


Notes: The figure shows the change in the aggregate population (panel A), the change in the flow of patents (panel B), the aggregate share of innovators (panel C), and the population share in the urban SEAs (panel D) relative to the baseline economy. We depict the baseline calibration in grey, the counterfactual with no immigrant inflows between 1880 and 1920 in red, the "no skilled migrants" counterfactual in teal and the "no urban bias" counterfactual in blue.

recorded in the 1910 Census. The 1924 Act further tightened these quotas, reducing them to two percent of the 1890 Census count, a benchmark chosen deliberately to favor Western European source countries. As a consequence, gross inflows from Italy, Poland, Russia, and much of Southern and Eastern Europe fell by more than 85 percent between the early 1920s and the 1930s, remaining suppressed until the Hart–Celler Act of 1965 dismantled the national-origins framework.

Constructing Counterfactual Inflows To quantify the growth consequences of these immigration quotas, we need to construct the counterfactual inflows in the absence of the policy. To do so, we exploit that the immigration restrictions fell unevenly across nationality groups. Drawing on the pre-restriction census, we decompose the group of "other immigrants" into three sub-populations: approximately 70% were other Europeans

FIGURE A.8: INTERNATIONAL MIGRATION AND LOCAL POPULATION GROWTH



Notes: The figure shows the change in the local population ℓ_{rt} of the economy without the arrival of international immigrants between 1880 and 1920 relative to the baseline economy. We show both the short-run effect in 1890 and the long-run effects in 1940.

(including Southern and Eastern Europeans), who faced the same severe quota restrictions as Italians; roughly 5% were Asians; and the remaining 25% were Western Hemisphere migrants (e.g., Canadians, Mexicans, and Latin Americans), who were explicitly exempt from numerical quotas and therefore continued to arrive largely unimpeded. Hence, approximately 75% of the group of "other immigrants" was subject to restrictions.

We anchor the total immigrant inflow for each decade in $t \in \{1930, \ldots, 1960\}$ to the 1920 aggregate level of $\bar{M}_{1920}$, because this year represents the last pre-restriction period unaffected by the quota legislation.

Western European groups. The British, German, and Irish groups received quotas that were never binding under either Act, so their observed inflows were not directly distorted by the restrictions. We therefore retain their actual historical values:

$$\widetilde{M}_{n,t} \;=\; M_{n,t}, \qquad n \in \{\text{Bri, Ger, Iri}\}, \quad t \in \{1930, \ldots, 1960\}. \tag{A.22}$$

Italians and the Other group. For the two restricted groups—Italians and Others—we allocate the residual between the counterfactual aggregate and the observed Western European total, preserving the relative shares of the two groups at their 1920 levels. Formally, define the residual available to restricted groups in decade t as

$$R_t \;=\; \bar{M}_{1920} \;-\; \sum_{n \in \{\text{Bri, Ger, Iri}\}} M_{n,t}. \tag{A.23}$$

The 1920 share of Italians among all restricted-group arrivals is $s_{\text{Ita}} = M_{\text{Ita},1920} \,/\, (M_{\text{Ita},1920} + M_{\text{Oth},1920})$, with the complementary share $s_{\text{Oth}} = 1 - s_{\text{Ita}}$ going to the Other group. Counterfactual inflows are then $\widetilde{M}_{\text{Ita},t} = s_{\text{Ita}} \times R_t$ and $\widetilde{M}_{\text{Oth},t} = s_{\text{Oth}} \times R_t$ for $t \in \{1930, \ldots, 1960\}$.

Post-1960 inflows revert to their actual values for all nationality groups, reflecting the

structural break introduced by the Hart–Celler Act.

Skill Composition of Counterfactual Immigrants Knowing the volume of counterfactual immigrants is insufficient; we also need to know their skill content, that is, κ^{ν}. We therefore extend the regression analysis of Section 4.4 to a finer nationality decomposition, splitting the Other group into four constituent sub-groups: *Other European*, *American*, *Asian*, and *African and Other*. Estimating the model on this expanded classification yields sub-group-specific κ^{ν} coefficients — see Table A.10. We then assign the counterfactual immigrant inflows the average of *Other Europeans* and *Asians*, because both groups were affected. In terms of their spatial allocation, we hold fixed the initial arrival probabilities at their 1920 levels.

TABLE A.10: ESTIMATION OF RELATIVE SKILL SUPPLY WITH NON-EUROPEAN: κ^{ν}

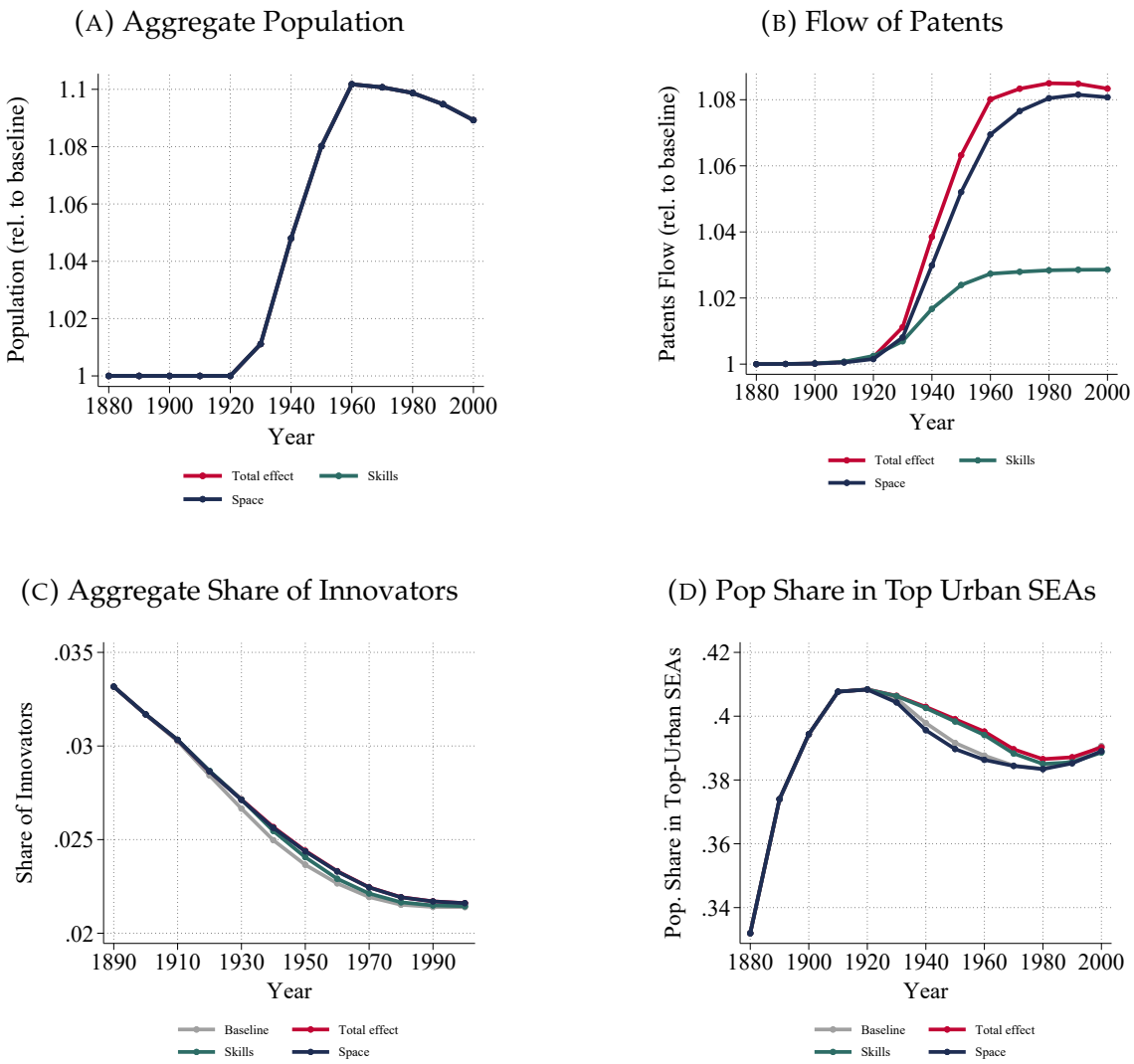
	(1)	(2)	(3)	(4)
British	0.281***	0.260***	0.223***	0.234***
	(0.025)	(0.025)	(0.026)	(0.024)
German	-0.268***	-0.277***	-0.294***	-0.268***
	(0.057)	(0.057)	(0.059)	(0.056)
Irish	-1.235***	-1.258***	-1.308***	-1.207***
	(0.092)	(0.090)	(0.086)	(0.087)
Italian	-1.479***	-1.491***	-1.516***	-1.399***
	(0.054)	(0.055)	(0.058)	(0.048)
Other European	-0.623***	-0.633***	-0.649***	-0.623***
	(0.041)	(0.042)	(0.043)	(0.041)
American	-0.215***	-0.235***	-0.272***	-0.255***
	(0.073)	(0.073)	(0.076)	(0.067)
Asian	-1.784***	-1.834***	-1.900***	-1.517***
	(0.417)	(0.415)	(0.417)	(0.262)
African and Other	0.539**	0.125	0.099	-0.607***
	(0.248)	(0.170)	(0.223)	(0.094)
Year-SEA FE	yes	yes	yes	yes
Cell size cutoff	-	200	500	-
Impute if zero?	no	no	no	yes
R^2	.871	.837	.828	.956
N	7,024	5,967	4,863	16,249

Notes: This table reports estimates of κ^{ν} from an extended nationality decomposition, splitting the Other group into four sub-groups: *Other European*, *American*, *Asian*, and *African and Other*. All specifications include Year-SEA fixed effects. In columns 2 and 3 we require all cells to have at least 200 or 500 observations respectively. In column 4, we add a small positive number to Θ_{rt}^{ns} in case $\Theta_{rt}^{ns} = 0$. All regressions are weighted by the number of observations in each cell.

Supplementary Results Figure A.9 reports additional aggregate outcomes for the counterfactual exercise that removes immigration restrictions. Panels A and B show that removing the restrictions raises aggregate population and patenting after 1920, with effects increasing through mid-century and remaining positive thereafter. The patenting effect is much smaller when the skill-composition channel is shut down, but remains close to the benchmark removing-ban counterfactual when the urban-bias channel is eliminated.

Panels C and D show that the aggregate share of innovators and the population share in the most urban SEAs change little relative to the baseline. Overall, the effects operate mainly through the scale of population and innovation rather than through large changes in innovator shares or urban concentration.

FIGURE A.9: THE AGGREGATE IMPACT OF REMOVING RESTRICTIONS



Notes: The figure shows the change in the aggregate population (panel A), the change in the flow of patents (panel B), the aggregate share of innovators (panel C), and the population share in the urban SEAs (panel D) relative to the baseline economy. We depict the baseline calibration in grey, the counterfactual removing the ban in red, the counterfactual where all previously banned migrants are unskilled in teal ("Skills") and the counterfactual where immigrants do not have an urban bias in blue ("Space").

SUPPLEMENTARY MATERIAL

"IMMIGRATION, INNOVATION, AND THE GEOGRAPHY OF GROWTH" by C. Arkolakis, S. Lee, and M. Peters

OA.1. ONLINE APPENDIX: DATA PROCESSING

OA.1.1 Geography Harmonization

In the main analysis, geographic location is defined at the level of **State Economic Areas (SEAs)**. SEAs are groups of contiguous counties within a state delineated by the U.S. Census Bureau prior to the 1950 Census to reflect integrated regional labor and economic markets.

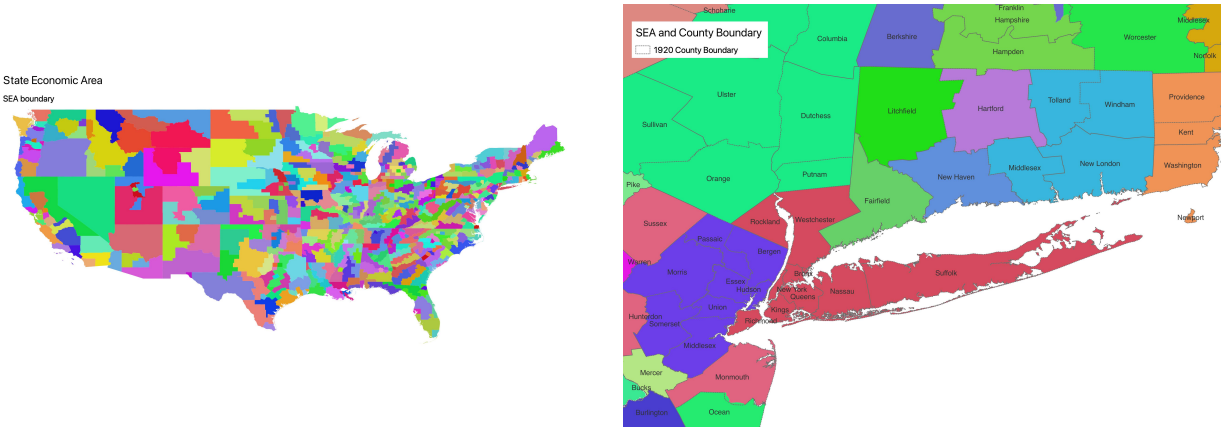
We aggregate data to the SEA level for two reasons. First, SEA boundaries remain consistent over our sample period, which avoids time-varying geographic definitions that would otherwise introduce measurement inconsistencies in panel regressions (county boundaries changed substantially between 1880 and 1920). Second, SEA-level aggregation reduces sparsity in patent counts. At the county level, patenting is frequently zero or missing, which generates instability in log specifications and inflates sampling variance. Aggregation to SEAs ensures strictly positive patent counts for the vast majority of observations.

Figure OA.1 illustrates the geographic construction. The left panel displays the 452 SEA boundaries used in the analysis. We overlay historical county shapefiles with 1940 SEA shapefiles obtained from the National Historical Geographic Information System (Manson et al., 2019) using Geographic Information System (GIS) software. For each county-SEA intersection, we compute area-based weights and use these weights to aggregate county-level variables to the SEA level. All patent and manufacturing data are then expressed at the SEA-year level.[30]

For record linkage between patent and population census data, we implement a hierarchical geographic matching procedure. We first attempt to match individuals within the recorded county of residence. If no match is found, we expand the search to the corresponding SEA. This procedure accommodates short-distance migration across adjacent counties.

[30]The left panel of Figure OA.1 shows the 452 SEA boundaries, with colors indicating distinct SEAs. The right panel of Figure OA.1 shows the New York SEA (in Red), which includes counties covering New York City and its surrounding areas.

FIGURE OA.1: State Economic Areas (SEA) and county boundaries



Notes: In the left panel, we present the State Economic Area (SEA) boundary, adhering to the definition provided by the US Census Bureau. The right panel depicts the New York SEA (in Red), which includes counties covering New York City and its surrounding areas. These overlays are used to construct spatial crosswalks that harmonize geographic identifiers across the Population Census, Census of Manufactures, and U.S. Patent records.

OA.1.2 Data Translation

Immigration Records I: Hamburg Passenger Lists

The Hamburg Passenger Lists (HPL) are recorded in German and stored in a non-structured format. To construct analyzable variables, we implement a structured pipeline consisting of parsing, translation, and harmonization. We focus on three fields central to the analysis: final destination (*Zielort*), nationality (*Nationalität*), and pre-migration occupation (*Beruf*).

Destinations (*"Zielort"*) We translate ship route and destination fields to identify passengers bound for the United States. While most departures from Hamburg were to North America, some passengers traveled to South America (e.g., Buenos Aires, Rio de Janeiro) and other European destinations. We restrict the sample to passengers whose final destination is recorded as the United States.

Nationality (*"Nationalitaet"*) Nationality information is used to facilitate linkage between immigration records and the U.S. Population Census. This field is directly recorded in the HPL beginning in 1898. For earlier records, nationality is imputed using the name-based classification tool *NamePrism* (also used by Diamond et al. (2019)), which predicts nationality and ethnicity based on first and last names. Because NamePrism provides probability-based reliability scores, we restrict the sample to observations with predicted nationality probabilities of at least 50 percent.

Occupation (*"Beruf"*) We translate and harmonize pre-migration occupations into the 1950 Census Bureau occupation classification (OCC1950) to ensure comparability across years. Approximately 53 percent of HPL records report occupation, with near-universal coverage for working-age males. The most commonly reported occupations include farm

laborers, general laborers, managers, and skilled manufacturing workers (e.g., tailors, shoemakers, cabinet makers). The translation from German to English and the subsequent harmonization into OCC1950 codes are described in detail in Section OA.1.3.

OA.1.3 Data Harmonization

Pre-Migration Occupations Both the Castle Garden records and the Hamburg Passenger Lists report immigrants' occupations prior to their arrival in the United States. We refer to these as pre-migration occupations. To ensure comparability across datasets and over time, we harmonize all reported occupations to the U.S. Census occupation classification system (OCC1950). For the Hamburg Passenger Lists, occupational entries are recorded in German and in free-text format. We first translate the original German occupational strings into English. We then standardize spelling variants and resolve ambiguous titles before mapping all occupations to the corresponding OCC1950 codes. This harmonization procedure yields a consistent occupational classification across immigration records and U.S. Census data, enabling direct comparison of pre-migration and post-arrival occupational outcomes.

Classifying pre-migration occupations to *OCC1950* The Castle Garden (CG) data contain 1,290 distinct pre-migration occupation strings, while the Hamburg Passenger Lists (HPL) contain 903 distinct German occupation strings. We construct a crosswalk mapping all pre-migration occupations to the U.S. Census occupation classification system (OCC1950), which consists of 245 occupational categories (we exclude non-occupational responses such as "keeping house" or "in school").

For HPL records, we first translate German occupation strings into English using the Google Translation API. All translated entries are manually reviewed to correct mistranslations and standardize terminology prior to classification.

We then implement a multi-step matching procedure to assign each distinct occupation string (both from CG and HPL) to an OCC1950 category:

1. **String Similarity Matching.** We perform initial matching based on exact and approximate string similarity, including the normalization of spelling variants, the removal of punctuation, and case standardization.

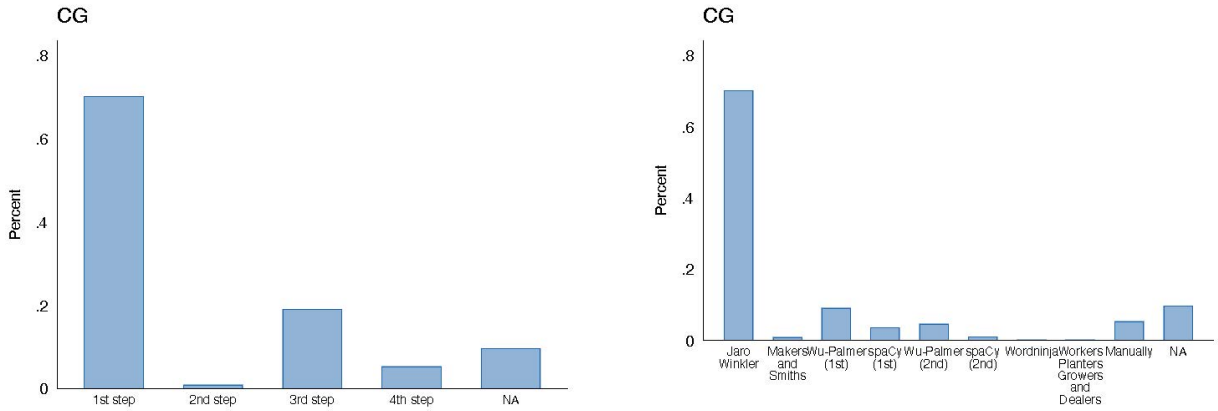
2. **Targeted Group Analysis.** For high-frequency occupational groups containing suffixes such as "maker" (e.g., shoemaker, bookmaker) or "smith," we conduct targeted semantic matching to ensure consistent classification within related occupational families.

3. **Semantic and Lexical Matching.** For unmatched strings, we apply iterative semantic matching using lexical databases (WordNet) and natural language processing tools (spaCy) to identify conceptually related OCC1950 categories.

4. **Manual Verification and Correction.** Remaining unmatched or ambiguous cases are manually reviewed and assigned to the most appropriate OCC1950 category.

Figures OA.2 and OA.3 report the share of distinct occupation strings matched at each stage for CG and HPL, respectively. Approximately 80 percent of CG occupation strings are matched in the initial string-similarity stage. For HPL, roughly 70 percent are matched through semantic and lexical procedures. Manual assignment accounts for approximately 5 percent of CG occupations and 20 percent of HPL occupations. This procedure yields a complete and consistent crosswalk from pre-migration occupations to OCC1950. Details for each step are discussed below.

FIGURE OA.2: SHARE OF OCCUPATION MATCHES FOR EACH STEP (CG)

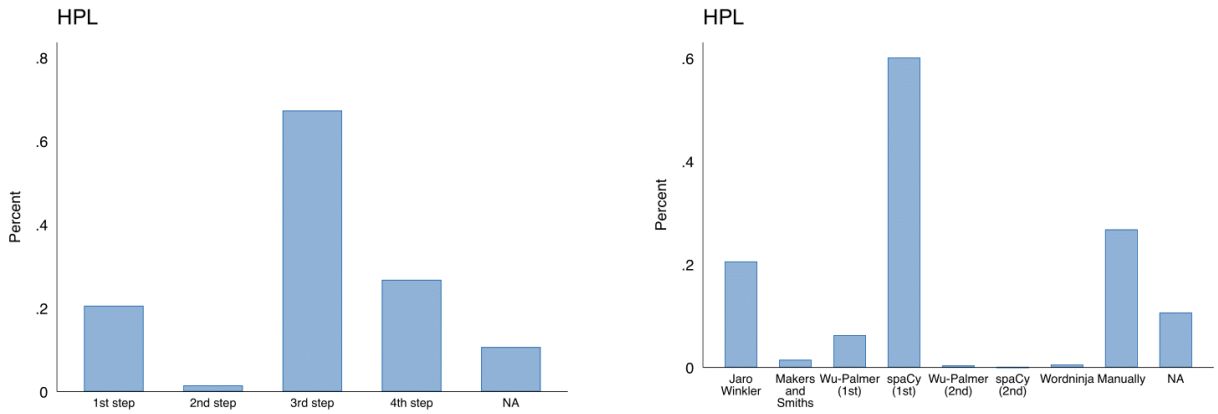


Notes: This graph illustrates the share of CG occupations matched at each step, indicating the contribution of each step to the matching process. The values are derived by summing the frequencies of CG occupation titles that correspond to Census occupation titles at each step. The left panel displays the overall share of occupations matched by four aggregate steps. In the right panel, we further dissect this match rate by different methods within the four steps.

Primer: Data cleaning In the data cleaning process (as the initial step before occupation matching), we employ standard text data pre-processing techniques, including lowercasing, punctuation removal, stop words removal, and lemmatization for occupation strings. Subsequently, to enhance matching rates, we group certain OCC1950 occupations into broader categories. For example, various types of mechanics and repairmen in OCC1950, such as airplane, automobile, and office machine mechanics, are aggregated into the category of "mechanics and repairmen."

To further refine matches, we also break down occupation strings containing multiple occupations. For instance, "mechanics and repairman" is broken down into "mechanics" and "repairmen," both used for linking to improve match accuracy. Additionally, we ensure the order of words within occupations is comparable. For example, OCC1950's "engineers, civil" is reported as "civil engineer" in the Castle Garden data.

FIGURE OA.3: SHARE OF OCCUPATION MATCHES FOR EACH STEP (HPL)



Notes: This graph illustrates the share of HPL occupations matched at each step, indicating the contribution of each step to the matching process. The values are derived by summing the frequencies of HPL occupation titles that correspond to Census occupation titles at each step. The left panel displays the overall share of occupations matched by four aggregate steps. In the right panel, we further dissect this match rate by different methods within the four steps.

Following this cleaning, we are left with 1,290 distinct occupations in the Castle Garden database and 245 in OCC1950.

Step 1: String Similarity Matching In the initial step, we match occupations directly based on the similarity of the occupation strings. This is done by calculating the Jaro–Winkler similarity between the occupational titles in the immigration data and the OCC1950 Census classification (for more details on the Jaro Winkler distance, see **?**, **?**, Winkler (2014)). Matches with a similarity measure below 0.98 are discarded, ensuring a stringent criterion. The OCC1950 classification with the highest Jaro–Winkler similarity measure is then assigned.

This approach is considered accurate due to its strict criteria. For example, the Jaro–Winkler similarity between "book keeper" and "bookkeeper" is 0.982. Out of 1,290 Castle Garden occupations, 154 occupations are matched using Jaro–Winkler similarity. Although this represents a relatively small number, it accounts for almost 50% of all individuals in the data, as such occupations are often larger in scale.

Step 2: Analysis of Specific Groups For occupations not matched in Step 1 using Jaro–Winkler Distance, our focus turns to occupations containing terms like "maker" and "smith." In the Castle Garden data, there are 220 occupations involving "maker" and 23 involving "smith."

To assign these occupations to OCC1950, we compare the occupational descriptions preceding "maker" or "smith" in terms of meaning. We utilize *Wordnet* and *spaCy* for this purpose. *Wordnet*, part of the NLTK library in Python, is a lexical database grouping words into *synsets* (sets of synonyms) based on conceptual-semantic and lexical relationships.

We match occupations by maximizing the Wu-Palmer similarity measure.[31] For instance, "furniture maker" in Castle Garden is matched to "cabinet maker" in OCC1950.

spaCy, an open-source library for Natural Language Processing, vectorizes words and determines similarity using word vectors. As a supplement to WordNet, *spaCy* is employed for occupations not matched using *Wordnet*, benefiting from its extensive word list.

Step 3: Semantic and Lexical Matching For the remaining occupations, we continue with a similar procedure. Initially, we aim to match occupations using *Wordnet* with a Wu-Palmer similarity score of 0.94, resulting in 149 occupation matches. Subsequently, we use *spaCy* for the remaining unmatched occupations, applying a similarity score of 0.6. In this iteration, we achieve 240 occupation matches. This process is repeated with a Wu-Palmer score of 0.9 and a *spaCy* score of 0.55, yielding an additional 120 matches. For the remaining unmatched occupations, we apply the same steps after breaking occupation titles using *Wordninja,* a Python package that separates words without spaces. For example, *Wordninja* separates "laundrymaid" into "laundry maid". This results in 73 additional occupational matches.

Among the unmatched occupations, we encounter multiple instances of "workers", "planters", "growers", and "dealers." We match these occupations to the Census using *spaCy*. For instance, instead of trying to match "cattle dealer", we find the best match for "dealer," which turns out to be "salesman." Similarly, "workers" are matched with "laborers", and "planter" is matched with "farmer."

Step 4: Manual Correction In the last step, we meticulously review the matches generated by the previous procedures, manually correcting any matches that seem incorrect. A total of 532 matches are corrected through this manual review process; this represents approximately 5% of all matches.

OA.2. Features and Measurement of the Patent Data

In this section, we outline the data harmonization and processing procedures for the Patent Database, emphasizing its role in measuring innovation across time and space. While patents may only capture specific aspects of innovative activities, their consistent and extensive coverage makes them a plausible proxy for measuring innovation during the study period.[32]

[31] Wu-Palmer similarity denotes words' pairwise similarity as determined by the depth at which a common ancestor word can be found for the pair's synset trees. In short, this similarity measure calculates the similarity in terms of the meaning of words.

[32] To the best of our knowledge, other innovation-related activities, such as trademark data during our study period, are deemed unsuitable for our study. This is attributed to the trademark database's lack of information on founders' names and its inability to capture the type of knowledge created.

Despite the extensive use of the current US patent database in economics literature, its current format poses challenges for our study due to limited and inconsistent coverage of key information, including patent inventors' and assignees' names, geographic locations, collaboration patterns, and nationality. To overcome this limitation, we collect patent-level documents from Google Patents and employ modern Natural Language Processing (NLP) tools to extract relevant information.[33] While similar efforts have been made by Google Patents (https://patents.google.com/), the PatentCity Database by Bergeaud and Verluise (2022), and HISTPAT (Petralia et al., 2016), we compile our own data tailored to our research questions, augmenting Google's Patent search engine results with data from PatentCity. The resulting unified dataset includes patent inventors' and assignees' names, geographic locations, collaboration patterns, nationality, and a flag for individual versus corporate patents, enabling a systematic study of the geography of innovative activities.

Co-Invention and the Number of Co-Inventors During our study period, the patenting process becomes increasingly organized, leading to a rise in the occurrence of multiple co-inventors for a given patent. To capture this collaborative innovation, we identify all co-inventors' names along with their corresponding residential locations. For a patent with 5 co-inventors, for example, we create a dummy variable indicating the co-invented status of the patent. Additionally, we generate a measure of fractional co-invention; in the mentioned case, the fractional co-invention would be 0.2. This fractional invention information proves valuable in assessing both the collaborative nature of innovative activities and the intensive margin of patenting.

Corporate Patenting During our study period, the organization of patenting became more structured, leading to an increased prevalence of corporations orchestrating innovative activities. In such cases, the patent document reflects a distinction between patent inventors (i.e., individuals who have contributed to the claimed invention) and patent assignees (i.e., organizations and/or individuals that have a legal ownership interest in the rights a patent offers). Often, the assignee is the organization employing the inventor.

By scrutinizing the patterns of patent inventors and assignees, we discern corporate patents throughout our study period. For example, if the assignee information includes terms like "Company", "Corp.," "Ltd", "Mfg", "Property", "System", "Gmbh", "Enterprise", and the patent inventor(s) is not the assignee, we classify such patents as corporate. This analysis helps identify patents associated with organizational entities during the specified period.

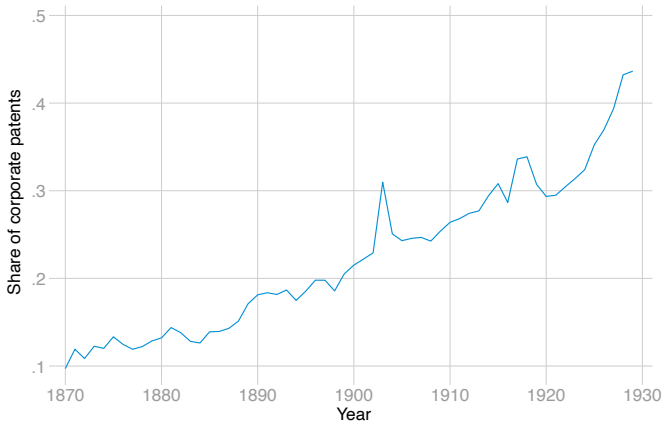
For our patent-census linking process, we include corporate patents by focusing on patent

[33] As discussed in Bergeaud and Verluise (2022), most historical patent documents are available as scanned images rather than a systematically harmonized database. Therefore, converting images into a database necessitates the use of modern techniques such as Natural Language Processing (NLP) and Geographic Information System (GIS).

inventors rather than assignees. In such cases, we utilize the patent inventor's name and residential location to establish links between patent and U.S. Population Census records.

Figure OA.4 plots the share of corporate patents during our study period. Each value is calculated as the number of corporate patents divided by the total number of patents granted in that year. As shown in Figure OA.4, a consistent upward trend in corporate patents is evident throughout the study period. The share of corporate patents starts at 10% in the early 1880s and steadily increases, surpassing 40% by the late 1920s.

FIGURE OA.4: SHARE OF CORPORATE PATENTS



Notes: The figure plots the share of corporate patents during our study period. Each value is calculated as the number of corporate patents divided by the total number of patents granted in that year.

Citizenship/Nationality of Patent Inventors As noted by Bergeaud and Verluise (2022), citizenship information of patentees is not available in standard patent datasets during the study period. To address this gap, we employ both Natural Language Processing (NLP) techniques and record linking methods to extract and enhance inventor citizenship information.

In terms of NLP techniques, we extract inventor citizenship information by analyzing the text of patents obtained from the Google Patents API. However, not all patent inventors declare their citizenship during the study period, with less than 40% of patents (granted between 1880 and 1929) providing such information, consistent with Bergeaud and Verluise (2022). Recognizing the limitations of NLP-based extraction alone, especially in capturing citizenship, which is crucial for our focus on studying immigration and innovation, we augment this information.

We augment patent inventors' citizenship by linking patent inventors (from the patent database) with federal US population censuses from 1880 to 1920. Modern machine-learning-based record-linking techniques are employed for this purpose. This record-linking-based citizenship classification aids in identifying immigrants, even when citizenship was not declared in the original patent documents — an occurrence in over 60% of patents during the study period.

Multiple Patents by the Same Inventor We observe instances where the same inventor holds multiple patents, determined by considering the patent inventor's full name, residential location, patent grant years, and the areas of claimed inventions. To capture this phenomenon, we generate a quasi-inventor ID and quantify the number of patents attributed to each inventor. This information on the "number of patents" proves valuable in measuring the intensive margin of patenting — shedding light on the extent of an inventor's involvement in patent activities.

International Patents Utilizing Natural Language Processing techniques on patent-level documents sourced from Google Patents, we have pinpointed inventors residing abroad who were credited with US patents. We refer to these as "international inventors" and their associated patents as "international patents," a classification that allows us to study their impact on the patent landscape.

International patents constitute between 5 and 10% of the overall patent portfolio, indicating that inventors abroad made a meaningful contribution to US innovation.

OA.3. Coverage, Quality, and Potential Biases of the Linked Records

This section provides additional detail on the linked records used throughout the paper, organized in three parts: (i) match counts and coverage, (ii) the quality of our matches, and (iii) potential biases and selection concerns in the matching procedure. These results are referenced in Appendix Section A.2.

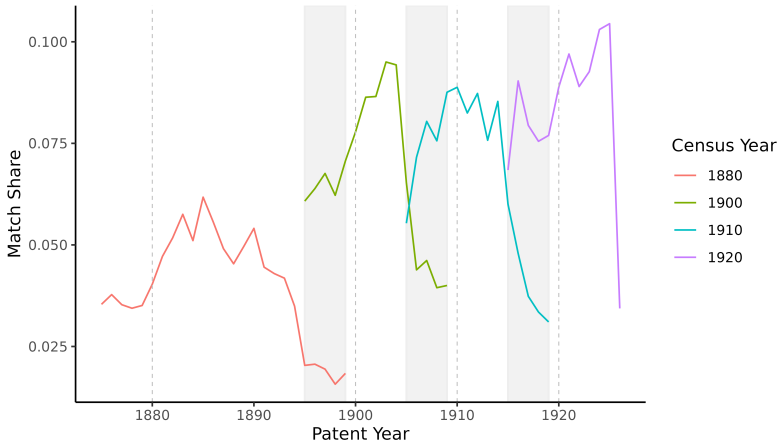
OA.3.1 Match Summary and Coverage

TABLE OA.1: MATCHING SUMMARY BY IMMIGRATION DATASET

		1880	1900	1910	1920
CG: Total	—	127,884	210,076	206,376	147,461
CG: British	No. Used for Matching	470,035	418,675	248,346	115,400
	Match Count	29,643	36,867	28,226	20,307
	Match Rate	0.057	0.055	0.044	0.039
CG: German	No. Used for Matching	1,086,610	1,506,990	1,110,002	675,860
	Match Count	63,169	136,390	114,182	75,719
	Match Rate	0.050	0.062	0.054	0.041
CG: Irish	No. Used for Matching	716,614	415,929	200,713	83,477
	Match Count	25,181	1,727	23,818	14,731
	Match Rate	0.031	0.002	0.029	0.023
CG: Italian	No. Used for Matching	8,901	500,057	608,741	492,375
	Match Count	260	16,676	23,708	22,780
	Match Rate	0.028	0.032	0.034	0.032
HPL: Total	—	30,879	84,072	112,593	94,404
HPL: British	No. Used for Matching	30,281	113,293	99,626	68,058
	Match Count	928	1,441	1,759	1,540
	Match Rate	0.029	0.011	0.013	0.012
HPL: German	No. Used for Matching	207,901	525,007	553,280	467,915
	Match Count	27,094	73,970	84,545	57,869
	Match Rate	0.121	0.116	0.111	0.072
HPL: Italian	No. Used for Matching	652	3,451	5,020	4,372
	Match Count	15	68	285	336
	Match Rate	0.022	0.018	0.048	0.051

Notes: This table reports the number of uniquely matched observations between the immigration datasets (CG and HPL) and the Census, along with the corresponding match rate by nationality over the study period.

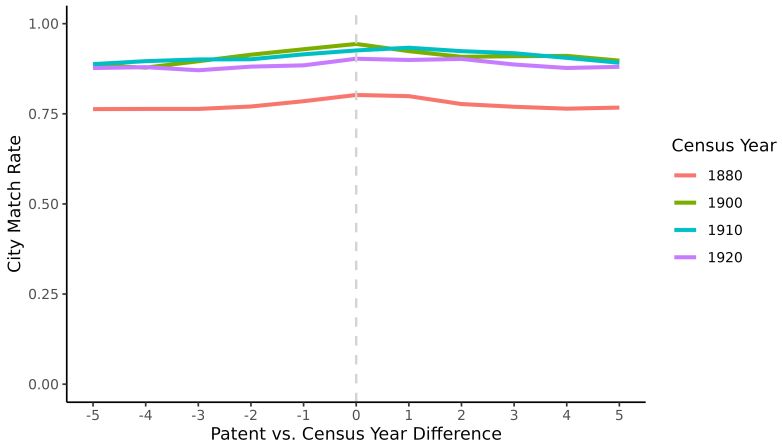
FIGURE OA.5: PATENT YEAR DISTRIBUTIONS



Note: The above figure illustrates the distribution of patent grant years of patents uniquely matched to each census. For example, of the 120,298 patents matched to the 1910 Census, approximately 8% were granted in 1907. Vertical dotted lines indicate census years. Shaded regions denote patent years eligible for matching to both adjacent censuses (1895–1899, 1905–1909, and 1915–1919). For patents granted in 1907, 4,379 records are matched to the 1900 Census and 9,673 records to the 1910 Census.

OA.3.2 Match Quality

FIGURE OA.6: CITY MATCH RATES BY CENSUS YEAR



Notes: The above figure illustrates the match rates between city information for census individuals and inventor city information. The x-axis represents the difference between patent year and the matched census year.

Potential Biases and Selection Concerns

We next examine whether systematic selection arises along geographic, demographic, or name-based dimensions that could affect match probabilities.

Geographic Bias in Patent-Census Matches Table OA.2 reports the distribution of patents across State Economic Areas (SEAs) in both the full patent dataset and the patent-census matched dataset. Match rates are lower in more populous SEAs, consistent with

increased name collisions in densely populated regions. All human-capital regressions therefore include SEA fixed effects to account for geographic variation in match probability (e.g., see Table OA.4).

TABLE OA.2: MATCH FREQUENCY BY SEA REGION

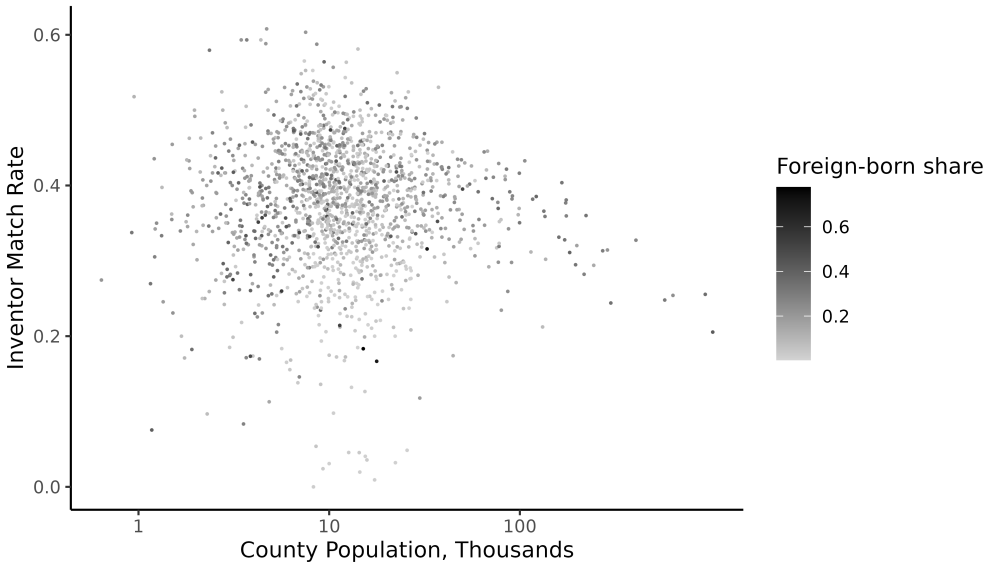
SEA	Record Count	Match Count	% All Patents	% Matched Patents
NYC Metro	128,075	30,558	8.9	6.8
Chicago Metro	92,512	24,802	6.5	5.5
Boston Metro	59,783	20,328	4.2	4.5
Jersey City/Newark	52,300	19,361	3.6	4.3
Philadelphia Metro	48,035	13,909	3.4	3.1
Pittsburgh Metro	27,668	8,129	1.9	1.8
San Francisco Metro	22,606	7,459	1.6	1.7
St. Louis Metro	21,483	7,084	1.5	1.6
Cleveland Metro	21,415	7,583	1.5	1.7
Los Angeles Metro	16,409	5,521	1.1	1.2

Notes: The table reports patent shares by SEA in the full and matched datasets. SEAs are ordered by total patent filings.

County Population, Immigrant Share, and Match Rate Figure OA.7 plots inventor match rates by county against county population. Match rates decline with population size, again consistent with increased name collisions in large counties. Substantial dispersion remains conditional on population. Larger counties also tend to have higher immigrant shares and lower match rates, suggesting that differences in name distributions across nativity groups may affect match probabilities.

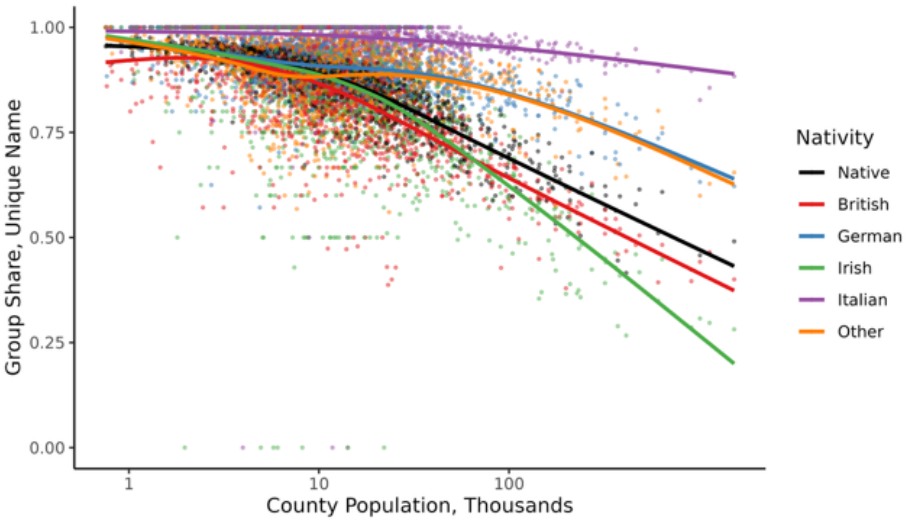
Name Uniqueness by Nativity A potential concern is that immigrants may have more distinctive names, mechanically increasing their match probability relative to natives. To evaluate this channel, Figure OA.8 plots name uniqueness against county population for patent-matching-eligible individuals in the 1920 Census. Within each county, a name is defined as unique if no other individual shares the same standardized given name and surname combination. For counties with fewer than approximately 22,000 residents (about 90 percent of all counties), name uniqueness is broadly similar across nativity groups, with differences emerging primarily in the largest counties. These patterns indicate that differential name uniqueness across nativity groups is unlikely to explain match-rate differences in the majority of counties.

FIGURE OA.7: MATCH RATES FOR INVENTOR RECORDS BY COUNTY POPULATION



Notes: This figure plots inventor match rate by county against county population. Each point represents a county. Darker shading indicates a higher share of foreign-born residents.

FIGURE OA.8: NAME UNIQUENESS BY COUNTY POPULATION



Notes: The figure reports the share of unique standardized given name–surname combinations by county and nativity group for patent-matching-eligible individuals in the 1920 Census. The sample is restricted to counties with at least 100 eligible individuals.

We next examine whether inventor name characteristics predict the likelihood of a unique match. Table OA.3 reports summary statistics for four name-based measures constructed from patent records: surname frequency, given name frequency (per 10,000 individuals), and surname and given name length. Name frequencies are computed at the county level using patent-matching eligible individuals from the 1900 Population Census. Given names are standardized prior to computing frequency and length to align with the matching procedure. These name characteristics are then used to assess their explanatory power in

matching probability regressions.

TABLE OA.3: SUMMARY OF NAME CHARACTERISTICS

	Mean	SD	Min.	Max.
Last Name Frequency (per 10k)	0.070	0.530	0.017	74.050
Given Name Frequency (per 10k, stnd.)	0.450	11.471	0.017	986.904
Surname Length	6.613	1.779	1	16
Given Name Length (stnd.)	5.954	1.402	1	15

Notes: Summary statistics for name frequency and length. Given names are standardized prior to computation.

To quantify these patterns formally, we estimate regressions of match probability on name characteristics and nativity composition.

In Table OA.4, we regress a unique match indicator variable for each inventor record against the characteristics of the inventor's given name and surname. The shares reported for natives and immigrants in the table are calculated from the share of potential matches with the associated birthplace. The first column reports the result when regressing only on nativity shares from potential matches. We find that records with a relatively large share of native potential matches are more likely to be matched. The second column reports results when regressing on name frequency and county population. The difference in explanatory power between the models in columns 1 and 2 suggests that match likelihoods are better explained by name characteristics rather than nativity. In addition, we note that the relative importance of county population is very high compared to name frequency, given the low magnitude of the name frequency variables listed in Table OA.3. Finally, in column 3, we observe that highly native names are still more likely to be matched than immigrant names, even when controlling for further name characteristics such as name length.

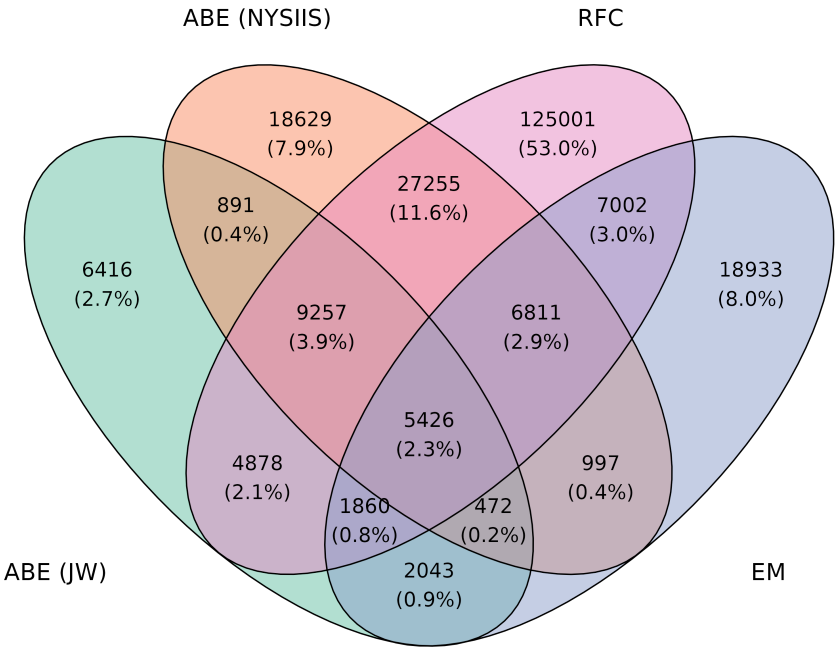
TABLE OA.4: MATCH LIKELIHOODS BY INVENTOR CHARACTERISTICS

	Dependent variable:		
	Inventor Match		
	(1)	(2)	(3)
Share British	−0.232***		−0.086***
	(0.002)		(0.002)
Share German	−0.179***		−0.064***
	(0.001)		(0.001)
Share Irish	−0.392***		−0.126***
	(0.002)		(0.002)
Share Italian	−0.219***		−0.245***
	(0.009)		(0.009)
Share Other Non-Native	−0.165***		−0.044***
	(0.001)		(0.001)
Given Name Frequency (per 10k, stnd.)		−0.0002***	−0.0002***
		(0.00000)	(0.00000)
Last Name Frequency (per 10k)		−0.003***	−0.003***
		(0.00001)	(0.00001)
Given Name Length (stnd.)			0.006***
			(0.0001)
Surname Length			0.015***
			(0.0001)
Log County Population		−0.094***	−0.087***
		(0.0001)	(0.0002)
Observations	2,465,446	2,465,446	2,465,446
Adjusted R^2	0.031	0.174	0.186

*p<0.05; **p<0.01; ***p<0.001

Notes: This table reports the results of regressing a unique match indicator variable on characteristics of the inventor record. Immigrant shares represent the share of potential census matches with the listed birthplace from 0 to 1. For example, a British share of 1 indicates that all potential matches are British. If all shares are 0, then all potential matches are native. Because given names are standardized before matching, we use the standardized given names to compute name frequency and length. We note that records with low shares of potential immigrant matches would be most likely to match.

FIGURE OA.9: MATCH OVERLAP BY LINKING METHOD



Notes: The Venn diagram shows the overlap of record pairs across linking methods for 1910 CG-Census linking. We color RFC-linked matches in pink.

OA.4. ROBUSTNESS

In this section we provide additional details for the robustness exercises contained in Section 6.

Robustness with respect to structural parameters

In the first six rows of Table 8 we test the robustness of our results with respect to different structural parameters. In rows 2 and 3 we allow for aggregate knowledge spillovers, that is we assume that $\psi_{rt} = \zeta_r N_{rt-1}^{\vartheta} (\sum_j N_{jt-1})^{\gamma}$. The higher γ, the stronger the sensitivity of local human capital ϕ_{rt} to the aggregate amount of knowledge, $\sum_j N_{jt-1}$. As such, γ operates as an additional scale elasticity and long-run growth is given by $g_y = g_A + \frac{1}{\sigma-1}\frac{1}{1-\vartheta-\gamma}\eta$. The baseline model is nested for the case of $\gamma = 0$. Because $\sum_j N_{jt-1}$ is a pure aggregate spillover that is common across locations, γ is not identified from cross-sectional data. In fact, our estimate of ϑ is independent of γ, because $\sum_j N_{jt-1}$ would be subsumed in the time fixed effect. To test to the importance of this channel, we consider two alternatives: $\gamma = 0.15$ and $\gamma = 0.3$.

In rows 4 and 5 we consider the elasticity of substitution across varieties σ. σ is an important parameter, because it determines the strength of variety gains. And because long-run growth is given by $g_y = g_A + \frac{1}{\sigma-1}\frac{1}{1-\vartheta}\eta$, a lower value of σ amplifies the effects

of population growth on growth. We consider two alterative values: $\sigma = 4$ and $\sigma = 7$.

Finally, in row 6 we turn to the capitalists' discount rate β. Rather then assuming that capitalists are long-lived, we assume that they are myopic and set $\beta = 0$. By putting more value on current profits, innovation incentives are more sensitive to current changes in policy.

For all cases, we fully recalibrate our model, including all location fundamentals, by matching the exact same moments as in our baseline model.[34] We then perform the counterfactuals of Section 5 in each model. In Table 8 we report a summary of these results.

Robustness: Dynamic Migration

In the last row of Table 8 we report the results of an estimation where we allow workers to be forward-looking. Consider an agent of type ν who currently resides in area o at the beginning of period t. The agent chooses destination $j' \in J$ to solve

$$v^{\nu}_{o,t} = \max_{j \in J} \ln\left(u_{j,t} B^{\nu}_{j,t}\right) - \ln \mu_{jo} + \beta V^{\nu}_{j,t+1} + \tfrac{1}{\varepsilon^D} \varepsilon_{j,t+1}, \tag{OA.1}$$

where $u_{j,t}$ is the indirect utility flow from real income in j, $B^{\nu}_{j,t}$ is the local amenity, μ_{jo} is the moving cost from o to j, and β is the discount factor. The preference shock $\varepsilon_{j,t+1}$ follows a Type-I Extreme Value distribution, and ε^D denotes the dynamic migration elasticity. The model delivers the closed-form choice probability

$$m^{\nu}_{jrt} = \frac{\exp\left(\ln u_{j,t} B^{\nu}_{j,t} - \ln \mu_{jr} + \beta V^{\nu}_{j,t+1}\right)^{\varepsilon^D}}{\sum_{j' \in J} \exp\left(\ln u_{j',t} B^{\nu}_{j',t} - \ln \mu_{j'r} + \beta V^{\nu}_{j',t+1}\right)^{\varepsilon^D}}. \tag{OA.2}$$

To bring Equation (OA.2) to the data, we follow Traiberman (2019) and combine the contemporaneous log-flow $\ln m^{\nu}_{jrt}$ with the one-period-ahead log-flow $\ln m^{\nu}_{oj,t+1}$ across pairs of destinations (j, j'). Using the closed-form expression of $V^{\nu}_{j,t+1}$, the value terms cancel and we obtain

$$\ln \frac{m^{\nu}_{jrt}}{m^{\nu}_{j'rt}} + \beta \ln \frac{m^{\nu}_{oj,t+1}}{m^{\nu}_{oj',t+1}} = \varepsilon^D \left[\ln u_{j,t} B^{\nu}_{j,t} - \ln u_{j',t} B^{\nu}_{j',t}\right] - \varepsilon^D \left(\ln \mu_{jr} - \ln \mu_{j'r}\right) - \beta \varepsilon^D \left(\ln \mu_{oj} - \ln \mu_{oj'}\right). \tag{OA.3}$$

Replacing $B^{\nu}_{j,t}$ with $\mathcal{B}_j \ell^{-\varrho}_{j,t} \left(\varpi^{\nu}_{j,t}\right)^{\iota}$, and using the equilibrium expression for the indirect-

[34] The calibration results for these different specifications are available upon request.

utility flow as in Equation (23), the final form of the equation can be written as

$$\begin{aligned}(\text{OA.4}) \quad \ln\frac{m^{\nu}_{jrt}}{m^{\nu}_{j'rt}} + \beta\ln\frac{m^{\nu}_{oj,t+1}}{m^{\nu}_{oj',t+1}} &= (C_j - C_{j'}) + \iota\varepsilon^D\left(\ln\varpi^{\nu}_{j,t} - \ln\varpi^{\nu}_{j',t}\right) \\ &+ \varepsilon^D\left[\ln\left(1+\tfrac{1}{\theta-1}e^{\nu}_{j,t}\right) - \ln\left(1+\tfrac{1}{\theta-1}e^{\nu}_{j',t}\right)\right] - \varepsilon^D\left(\ln\mu_{jrt} - \ln\mu_{j'rt}\right) \\ &- \beta\varepsilon^D\left(\ln\mu_{oj,t+1} - \ln\mu_{oj',t+1}\right),\end{aligned}$$

where the destination-specific component are defined as

$$C_j \equiv \varepsilon^D\ln(w_j/P_j) - \varrho\varepsilon^D\ln\ell_j + \varepsilon^D\ln\mathcal{B}_j.$$

Estimation Results The only structurally new parameter introduced in the dynamic model is the labor discount factor, β, which we set to 0.35, the same value as the capitalist discount factor in the main text.[35] We hold the migration elasticity ε^D and the labor-congestion elasticity ϱ fixed at their values in the static model reported in Table 6. The parameter of interest is the own-nationality homophily ι, which enters through the coefficient $\iota\varepsilon^D$ on the local nationality share difference $\ln\varpi^{\nu}_{j,t} - \ln\varpi^{\nu}_{j',t}$. To address the endogeneity of local nationality shares, we instrument ϖ^{ν}_{rt} using (i) the shift-share instrument of Card (2001), and (ii) the predicted ancestry distribution instrument following Burchardi et al. (2026).

Table OA.5 reports the estimates of Equation (OA.4). The second column using Card instrument is our preferred specification, since the pre-1920 sample is too short for a reliable first stage of the instrument proposed in Burchardi et al. (2026). We get an estimate of 0.115 for ι, which is not significantly different from the static estimation 0.13. Therefore, we keep ι the same value, 0.13, as in the static migration model for the further analysis.

Calibration of $\mathcal{B}$ Because the dynamic specification introduces a future expectation value term into the original Equation (23), the cross-sectional calibration of $\mathcal{B}_j$ used in the static model can no longer be applied; we therefore exploit the panel structure of the data together with the estimated fixed effects in Equation (OA.4) to recover $\mathcal{B}_j$ in an internally consistent way. From the definition of C_j, the local amenity is recovered as [36]

$$\mathcal{B}^{\text{DYN}}_j = \exp\left(\tfrac{1}{\varepsilon^D}C_j - \ln(w_j/P_j) + \varrho\ln\ell_j\right),$$

where superscript DYN indicates that $\mathcal{B}_j$ is consistent with agents' dynamic migration problem. In Figure OA.10 we compare $\mathcal{B}^{\text{DYN}}_j$ with its counterpart from the static model

[35] We also consider an alternative calibration with $\beta = 0.6$. After recalibrating all other parameters, the results are quantitatively similar. Interested readers may request these results from the author or reproduce them by modifying the relevant parameter values in the replication package.

[36] This recovers $\mathcal{B}$ as a residual that is, by construction, common across types ν. In the counterfactual exercises we allow the differential urban pull of immigrants and skilled workers as in Equation (16).

TABLE OA.5: Estimation of Own-nationality Homophily ι in dynamic model

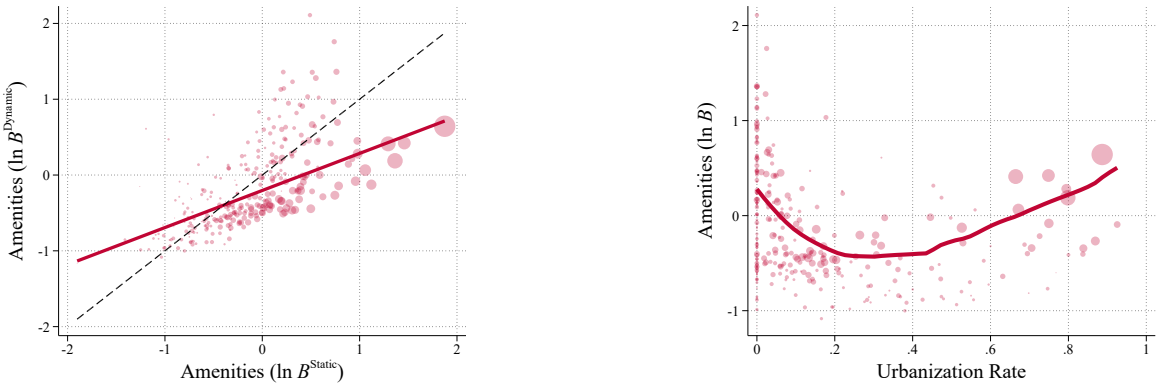
	(1)	(2)	(3)
Log Nationality Share Diff	0.167***	0.220***	0.388***
	(0.000)	(0.000)	(0.000)
Innovator Share Diff	52.038***	60.780***	90.370***
	(0.094)	(0.105)	(0.119)
Log Trade Cost Diff	-0.766***	-0.778***	-0.763***
	(0.000)	(0.000)	(0.000)
Forward Log Trade Cost Diff	-0.559***	-0.575***	-0.557***
	(0.000)	(0.000)	(0.000)
J-K FE	Yes	Yes	Yes
IV	No	Card	Burchardi et al.
R^2		.307	.296
N	104,004,134	99,838,374	102,468,612

Notes: This table reports estimates of equation (OA.4). All specifications include origin-destination (J-K) fixed effects. Column (1) presents OLS estimates. Column (2) presents estimates using the instrument from Card (2001). Column (3) presents estimates using the Burchardi et al. (2026) instrument, extended to include the contemporaneous year (no leave-one-out). Column (4) uses the Burchardi instrument in leave-one-out form, which excludes the own-country share when constructing predicted immigrant stocks. Column (5) uses the Burchardi leave-one-out instrument but omits the recursive term in its construction. Standard errors are reported in parentheses.

and plot it as a function of the urbanization rate. Panel A shows that the dynamic estimates are positively correlated with the static ones. Panel B shows that they remain positively related to urbanization, even though amenities are a bit larger at the very low levels of urbanization.

FIGURE OA.10: LOCAL AMENITIES: COMPARISON TO THE STATIC MODEL

(A) Dynamic vs. Static $\mathcal{B}$ (B) $\mathcal{B}$ vs. Urbanization Rate

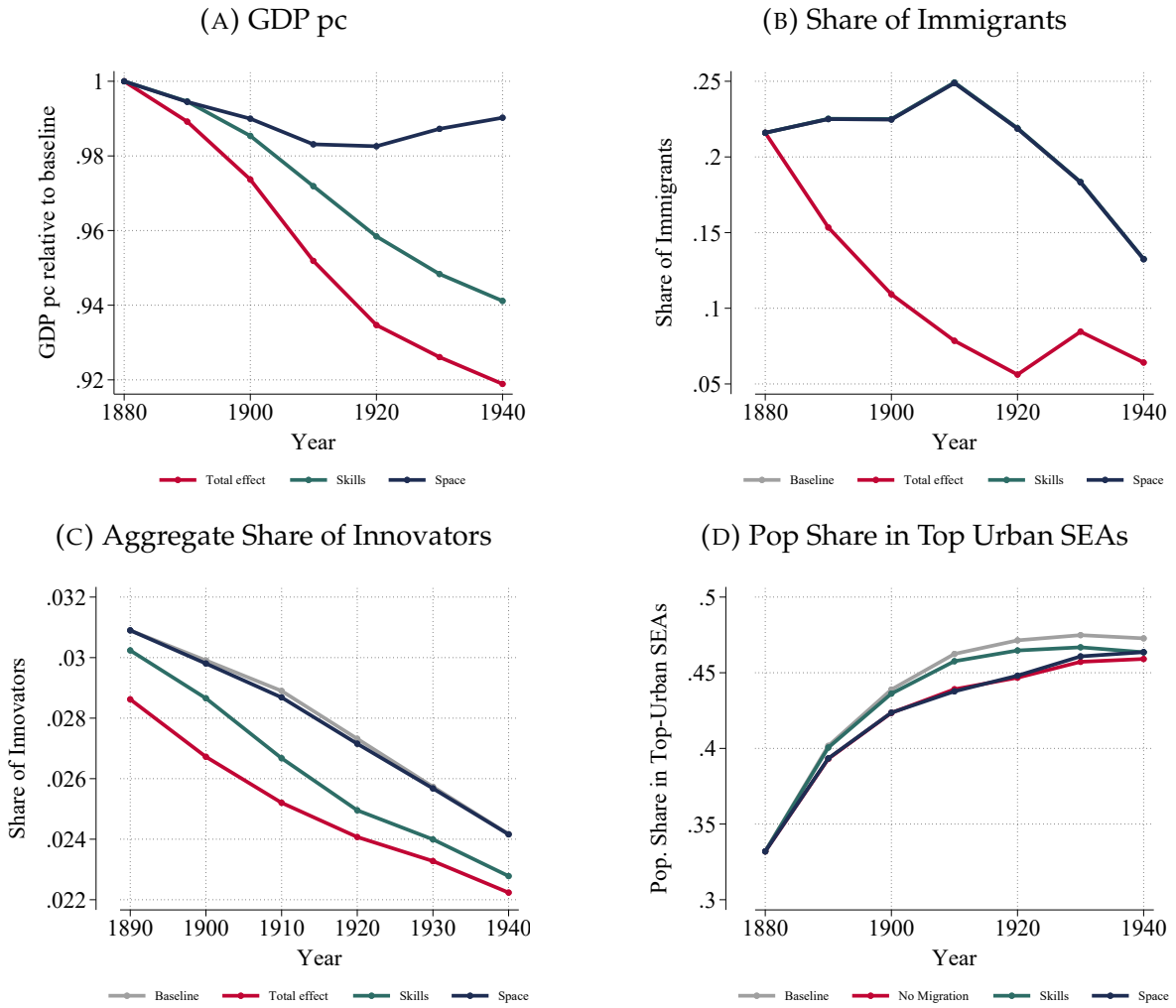


Notes: Each marker is an SEA. We normalize the average value of $\mathcal{B}$ to be 1. In panel (A) the horizontal axis reports $\ln \mathcal{B}$ from the static model and the vertical axis reports $\ln \mathcal{B}$ recovered from (OA.4); the dash line is the 45-degree line. In panel (B) the horizontal axis is the urbanization rate and the vertical axis reports the $\ln \mathcal{B}$ from (OA.4).

Counterfactual Analysis We then replicate the counterfactual exercises of the main text using the parameters calibrated in the dynamic model. As Figure OA.11 shows, the dynamic results closely mirror their static counterparts across all four aggregate margins: GDP per capita, the immigrant share, the share of innovators, and the population share in

top-urbanized SEAs. This is consistent with the last row of Table 8, that also showed very similar results between our baseline model and the model with forward-looking workers.

FIGURE OA.11: THE AGGREGATE IMPACT OF INTERNATIONAL MIGRATION (DYNAMIC MIGRATION)



Notes: The figure shows the change in aggregate income per capita, relative to the baseline economy (Panel A), the aggregate share of immigrants (Panel B), the overall share of people engaged in innovation (Panel C), and the share of the population living in the most urbanized locations. We depict the baseline calibration in grey, the counterfactual with no immigrant inflows between 1880 and 1920 in red, the "no skilled migrants" counterfactual in yellow and the "no urban bias" counterfactual, where immigrants initially settle in proportion to the spatial allocation of natives in blue.

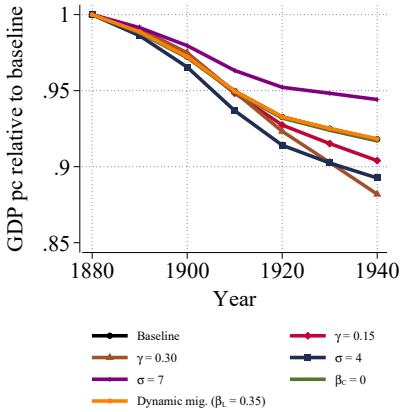
OA.4.1 Robustness Summary

Figure OA.12 examines the robustness of the counterfactual GDP per capita effects across alternative parameters. The qualitative patterns are stable across specifications: shutting down migration or skilled migration lowers GDP per capita in the migration-restriction counterfactuals, while removing the 1921/24 quotas raises GDP per capita over the longer horizon. The magnitudes are more sensitive to the parameters governing global knowledge spillovers, γ, and the elasticity of substitution, σ, than to the two discount-

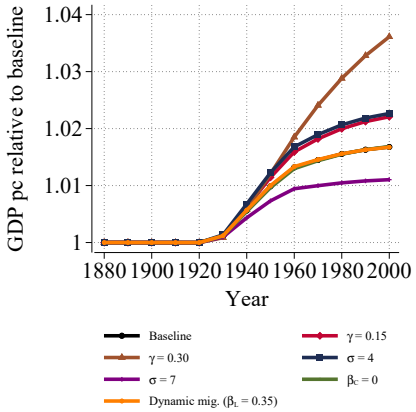
factor assumptions. In particular, changing γ or σ generates visibly larger variation in the counterfactual paths, whereas the no-capitalist-discounting case and the forward-looking migration specification remain close to the baseline in most panels.

FIGURE OA.12: ROBUSTNESS OF THE COUNTERFACTUAL GDP PER CAPITA EFFECTS

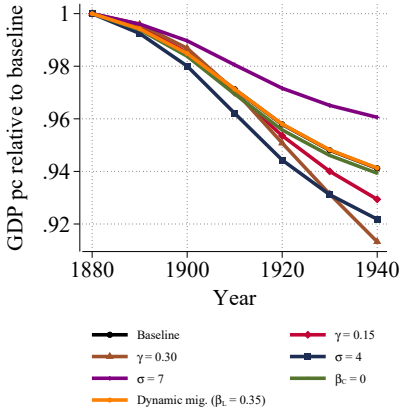
(A) No Migration (Counterfactual I)



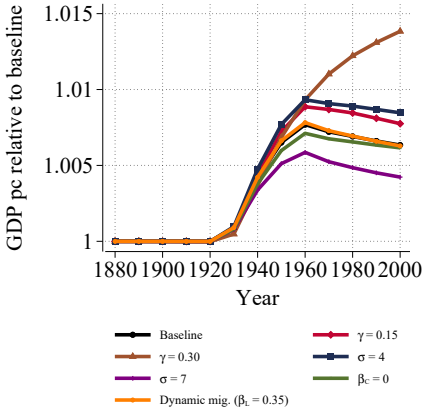
(B) No Migration (Counterfactual II)



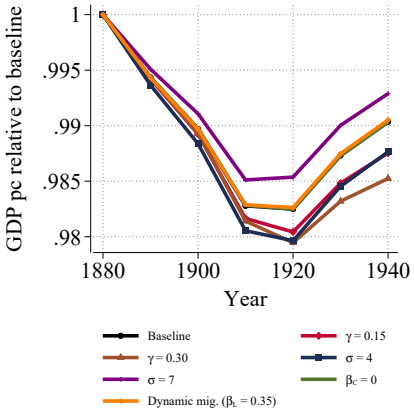
(C) No Skilled Migration (Counterfactual I)



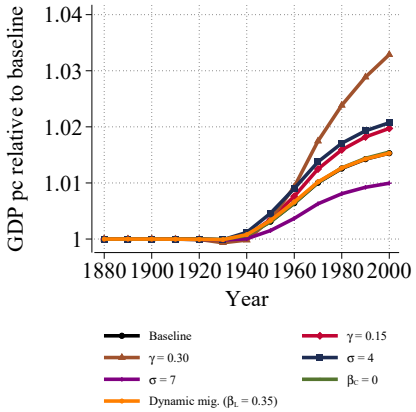
(D) No Skilled Migration (Counterfactual II)



(E) No Urban Bias (Counterfactual I)



(F) No Urban Bias (Counterfactual II)



Notes: Each panel plots model GDP per capita under a counterfactual relative to the baseline. Rows correspond to the three counterfactual families: shutting down all immigration (Panels A–B), shutting down skilled immigration (Panels C–D), and removing the urban settlement bias of immigrants (Panels E–F). The left column (Panels A, C, E) covers the 1880–1940 migration-restriction counterfactuals; the right column (Panels B, D, F) covers the 1880–2000 horizon that removes the 1921/24 quotas. Within each panel we overlay the baseline specification together with all robustness specifications: global knowledge spillovers $\gamma \in \{0.15, 0.30\}$, elasticity of substitution $\sigma \in \{4, 7\}$, no capitalist discounting $\beta_C = 0$, and forward-looking labor migration with $\beta_L = 0.35$.

OA.5. Construction of Instruments for Nationality Shares

This appendix describes the construction of the instruments for $\ln \varpi_{rt}^n$ in equation (17). The parameter of interest is $\iota\varepsilon$, which is the coefficient on $\ln \varpi_{rt}^n$, the share of nationality n in the immigrant stock of region r at time t. Since this variable may be endogenous to local conditions, we use two alternative shift-share instruments: a Card-style instrument and a Burchardi-style predicted-ancestry instrument. The Card instrument is our preferred specification, while the Burchardi-style instrument is used as a robustness check.

Card instrument. Let A_{rt}^n denote the immigrant stock of nationality n in region r at time t. Since in the region-level data we observe immigrant stocks rather than flows, we construct immigrant flows as

$$M_{rt}^n = A_{rt}^n - A_{r,t-1}^n. \tag{OA.5}$$

For 1880, we assume that the stock before 1880 is zero, so that

$$M_{r1880}^n = A_{r1880}^n. \tag{OA.6}$$

The total inflow of nationality n to the United States at time t is

$$M_t^n = \sum_r M_{rt}^n. \tag{OA.7}$$

The Card shift-share term is constructed as

$$Z_{rt}^{n,C} = M_t^n \frac{A_{r,t-1}^n}{A_{t-1}^n}, \qquad A_{t-1}^n = \sum_r A_{r,t-1}^n. \tag{OA.8}$$

The first-stage equation is

$$A_{rt}^n = \delta_t + \delta_{rt} + \delta_{nt} + \beta_1 A_{r,t-1}^n + \beta_2 M_t^n \frac{A_{r,t-1}^n}{A_{t-1}^n} + \epsilon_{rt}^n. \tag{OA.9}$$

Using the fitted value $\widehat{A}_{rt}^n$ from (OA.9), we construct the predicted nationality share as

$$\widehat{\varpi}_{rt}^n = \frac{\widehat{A}_{rt}^n}{\sum_{n'} \widehat{A}_{rt}^{n'}}. \tag{OA.10}$$

The instrument for $\ln \varpi_{rt}^n$ is then $\ln \widehat{\varpi}_{rt}^n$.

Burchardi-style instrument. The Burchardi-style instrument, proposed by Burchardi et al. (2026), uses the interaction between an origin-specific push factor and a destination-

specific pull factor to generate predicted immigrant stocks. In the original construction, the push factor from o to d is the total number of migrants arriving in the United States from o at time t, excluding migrants from o who settled in the same region as d. The pull factor from o to d is the fraction of migrants from other origins who settle in d at time t, excluding migrants from the same continent as o.

The starting point is the following linear formulation of immigrant stock:

$$A_{od}^{t} = a_t + a_{ot} + a_{dt} + b_t A_{od}^{t-1} + I_o^t \left(c_t \frac{I_d^t}{I^t} + d_t \frac{A_{od}^{t-1}}{A_o^{t-1}} \right) + \nu_{od}^t. \tag{OA.11}$$

Solving this equation recursively gives

$$A_{od}^{t'} = \sum_{t=initial}^{t'} \left(a_t + a_{ot} + a_{dt} + c_t I_o^t \frac{I_d^t}{I^t} + \nu_{od}^t \right) \prod_{\tau=t+1}^{t'} (b_\tau + d_{o\tau} I_o^\tau), \tag{OA.12}$$

with zero initial stock for all origin-destination pairs.

Following the original leave-out logic, the aggregate origin push and destination pull terms are replaced by leave-out versions:

$$I_o^t \quad \text{is replaced by} \quad I_{o,-r(d)}^t, \tag{OA.13}$$

and

$$\frac{I_d^t}{I^t} \quad \text{is replaced by} \quad \frac{I_{-c(o),d}^t}{I_{-c(o)}^t}. \tag{OA.14}$$

This adjustment avoids constructing the push or pull factor from flows that may be directly affected by the same origin-destination shocks.

In our implementation, to produce the shift-share, we use a slightly different leave-out structure:

$$Z_{rt}^{n,B} = M_{(-r)t}^{n} * \frac{M_{rt}^{-n}}{M_t^{-n}}. \tag{OA.15}$$

That is, for the push factor we exclude the same destination region r rather than the broader region containing r, and for the pull factor we exclude the same nationality n rather than the same continent as n. This is different from the original construction and may be problematic if there are common shocks for a broad origin group or for a particular U.S. region.

As in the original paper, we assume that the stocks of immigrants before 1880 were zero, and we treat the immigrant stock in 1880 as the immigrant flow for 1880.

Let $\mathcal{R}^n_{rt}$ denote the vector of recursive interaction terms generated by the recursive solution of the immigrant-stock equation. In our specification, the first-stage equation for the decade 1920 is given by

$$A^n_{r1920} = \beta_0 + \delta_{rt} + \delta_n + \beta_1 Z^{n,B}_{r1880} + \beta_2 Z^{n,B}_{r1900} + \beta_3 Z^{n,B}_{r1910} + \Gamma'_{1920}\mathcal{R}^n_{r1920} + \epsilon^n_{rt}, \tag{OA.16}$$

where $\mathcal{R}^n_{r1920}$ contains the 11 recursive interaction terms.

For 1910, the first-stage equation is

$$A^n_{r1910} = \beta_0 + \delta_{rt} + \delta_n + \beta_1 Z^{n,B}_{r1880} + \beta_2 Z^{n,B}_{r1900} + \Gamma'_{1910}\mathcal{R}^n_{r1910} + \epsilon^n_{rt}, \tag{OA.17}$$

where $\mathcal{R}^n_{r1910}$ contains the 4 recursive interaction terms.

For 1900, the first-stage equation is

$$A^n_{r1900} = \beta_0 + \delta_{rt} + \delta_n + \beta_1 Z^{n,B}_{r1880} + \Gamma'_{1900}\mathcal{R}^n_{r1900} + \epsilon^n_{rt}, \tag{OA.18}$$

where $\mathcal{R}^n_{r1900}$ contains the recursive interaction term

$$M^n_{(-r)1900} M^n_{(-r)1880} * \frac{M^{-n}_{r1880}}{M^{-n}_{1880}}. \tag{OA.19}$$

Similar to the Card IV, we use the fitted value $\widehat{A}^n_{rt}$ from the first stage to construct the predicted nationality share:

$$\widehat{\omega}^n_{rt} = \frac{\widehat{A}^n_{rt}}{\sum_{n'} \widehat{A}^{n'}_{rt}}. \tag{OA.20}$$

The instrument for $\ln \omega^n_{rt}$ is then $\ln \widehat{\omega}^n_{rt}$.

REFERENCES

ABRAMITZKY, R. AND L. BOUSTAN (2022): *Streets of gold: America's untold story of immigrant success*, Hachette UK.

ABRAMITZKY, R., L. BOUSTAN, K. ERIKSSON, J. FEIGENBAUM, AND S. PÉREZ (2021): "Automated Linking of Historical Data," *Journal of Economic Literature*, 59, 865–918.

ABRAMITZKY, R., L. BOUSTAN, K. ERIKSSON, S. PÉREZ, AND M. RASHID (2025): "Census Linking Project: Version 3.0 [dataset]," https://censuslinkingproject.org.

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2012): "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102, 1832–56.

——— (2014): "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 122, 467–506.

ABRAMITZKY, R., R. MILL, AND S. PÉREZ (2020): "Linking individuals across historical sources: A fully automated approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53, 92–111.

AKCIGIT, U. (2017): "Economic Growth: The Past, the Present, and the Future," *Journal of Political Economy*, 125, 1736–1747.

AKCIGIT, U., J. GRIGSBY, AND T. NICHOLAS (2017): "Immigration and the Rise of American Ingenuity," *American Economic Review*, 107, 327–31.

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

ALTONJI, J. G. AND D. CARD (1991): "The effects of immigration on the labor market outcomes of less-skilled natives," in *Immigration, trade, and the labor market*, University of Chicago Press, 201–234.

BERGEAUD, A. AND C. VERLUISE (2022): "PatentCity: a dataset to study the location of patents since the 19th century," .

——— (2024): "A New Dataset to Study a Century of Innovation in Europe and in the US," *Research Policy*, 53, 104903.

BERNSTEIN, S., R. DIAMOND, A. JIRANAPHAWIBOON, T. MCQUADE, AND B. POUSADA (2022): "The contribution of high-skilled immigrants to innovation in the United States," Tech. rep., National Bureau of Economic Research.

BILAL, A. (2023): "The geography of unemployment," *The Quarterly Journal of Economics*, 138, 1507–1576.

BILAL, A. AND E. ROSSI-HANSBERG (2021): "Location as an Asset," *Econometrica*, 89, 2459–2495.

BLOOM, N., C. I. JONES, J. VAN REENEN, AND M. WEBB (2020): "Are Ideas Getting Harder to Find?" *American Economic Review*, 110, 1104–1144.

BRODA, C. AND D. WEINSTEIN (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121, 541–585.

BURCHARDI, K. B., T. CHANEY, T. A. HASSAN, L. TARQUINIO, AND S. J. TERRY (2026): "Immigration, Innovation, and Growth," *American Economic Review*, 116, 828–861.

BURSTEIN, A., G. HANSON, L. TIAN, AND J. VOGEL (2020): "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence From the United States," *Econometrica*, 88, 1071–1112.

CARD, D. (1990): "The Impact of the Marial Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, 43, 245–257.

——— (2001): "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 19, 22–64.

CREWS, L. G. (2023): "A Dynamic Spatial Knowledge Economy," Working Paper.

DESMET, K., D. K. NAGY, AND E. ROSSI-HANSBERG (2018): "The Geography of Development," *Journal of Political Economy*, 126, 903–983.

DIAMOND, R., T. MCQUADE, AND F. QIAN (2019): "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review*, 109, 3365–94.

DUSTMANN, C., U. SCHÖNBERG, AND J. STUHLER (2017): "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment," *The Quarterly Journal of Economics*, 132, 435–483.

ECKERT, F. AND M. PETERS (2022): "Spatial Structural Change," Working Paper.

FAJGELBAUM, P. D., E. MORALES, J. C. SUÁREZ SERRATO, AND O. M. ZIDAR (2019): "State Taxes and Spatial Misallocation," *The Review of Economic Studies*, 86, 333–376.

FEIGENBAUM, J. J. (2016): "Machine Learning Approach to Census Record Linking," Tech. rep.

FUSHIKI, T. (2011): "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing*, 21, 137–146.

HASTIE, T., R. TIBSHIRANI, AND J. FRIEDMAN (2009): in *The elements of statistical learning*, Berlin, Germany: Springer.

HORNBECK, R. AND M. ROTEMBERG (2024): "Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies," *Journal of Political Economy*, 132, 3547–3602.

JARO, M. A. (1989): "Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida," *Journal of the American Statistical Association*, 84, 414–420.

——— (1995): "Probabilistic linkage of large public health data files," *Statistics in medicine*, 14, 491–498.

JONES, C. (2005): "Growth and Ideas," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1, Part B, chap. 16, 1063–1111, 1 ed.

JONES, C. I. (1995): "R&D-based Models of Economic Growth," *Journal of Political Economy*, 103, 759–784.

——— (2022): "The past and future of economic growth: A semi-endogenous perspective," *Annual Review of Economics*, 14, 125–152.

KERR, S. P., W. KERR, ÇAĞLAR ÖZDEN, AND C. PARSONS (2016): "Global Talent Flows," *Journal of Economic Perspectives*, 30, 83–106.

KERR, W. (2018): *The Gift of Global Talent: How Migration Shapes Business, Economy & Society*, Stanford University Press.

KLEINMAN, B., E. LIU, AND S. J. REDDING (2023): "Dynamic spatial general equilibrium," *Econometrica*, 91, 385–424.

KONERU, K., V. S. V. PULLA, AND C. VAROL (2016): "Performance Evaluation of Phonetic Matching Algorithms on English Words and Street Names - Comparison and Correlation," in *Proceedings of the 5th International Conference on Data Management Technologies and Applications - Volume 1: DATA*, Insticc, SciTePress, 57–64.

KORTUM, S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65, 1389–1419.

KRUGMAN, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70, 950–959.

LEE, S. K. (2019): "Essays in History and Spatial Economics with Big Data," Working Paper.

MANSON, S., J. SCHROEDER, D. VAN RIPER, AND S. RUGGLES (2019): "IPUMS National Historical Geographic Information System: Version 14.0 [Database]," .

NAGY, D. K. (2023): "Hinterlands, City Formation and Growth: Evidence from the U.S. Westward Expansion," *The Review of Economic Studies*, 90, 3238–3281.

OBERFIELD, E. AND D. RAVAL (2021): "Micro Data and Macro Technology," *Econometrica*, 89, 703–732.

PEDREGOSA, F., G. VAROQUAUX, A. GRAMFORT, V. MICHEL, B. THIRION, O. GRISEL, M. BLONDEL, P. PRETTENHOFER, R. WEISS, V. DUBOURG, J. VANDERPLAS, A. PASSOS, D. COURNAPEAU, M. BRUCHER, M. PERROT, AND E. DUCHESNAY (2011): "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 12, 2825–2830.

PERI, G. (2016): "Immigrants, Productivity, and Labor Markets," *The Journal of Economic Perspectives*, 30, 3–30.

PETERS, M. (2022): "Market size and spatial growth—evidence from Germany's post-war population expulsions," *Econometrica*, 90, 2357–2396.

PETRALIA, S., P.-A. BALLAND, AND D. RIGBY (2016): "HistPat Dataset," .

RAJKOVIC, P. AND D. JANKOVIC (2007): "Adaptation and Application of Daitch-Mokotoff Soundex Algorithm on Serbian Names," Tech. rep.

REDDING, S. J. AND E. ROSSI-HANSBERG (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9, 21–58.

ROMER, P. M. (1990): "Endogenous Technological Change," *The Journal of Political Economy*, 98, S71–s102.

RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, AND M. SOBEK (2020): "Integrated Public Use Microdata Series: Version 10.0 [Machine-readable database]," .

SEQUEIRA, S., N. NUNN, AND N. QIAN (2020): "Immigrants and the Making of America," *The Review of Economic Studies*, 87, 382–419.

TRAIBERMAN, S. (2019): "Occupations and import competition: Evidence from Denmark," *American Economic Review*, 109, 4260–4301.

WALSH, C. (2025): "The Entry Multiplier," Working paper, available at SSRN: https://ssrn.com/abstract=5468409.

WINKLER, W. E. (2014): "Advanced Methods for Record Linkage," Tech. rep.