# RETURN INNOVATION: THE KNOWLEDGE SPILLOVERS OF THE BRITISH MIGRATION TO THE UNITED STATES, 1870–1940 *

Davide M. Coluccia† Gaia Dossi‡

This Version: October, 2025
Latest Version

**Abstract**

The cross-country diffusion of technology is a central driver of productivity growth and economic convergence. This paper provides novel evidence on the role of out-migration as a channel of technology transfer from destination to origin countries. We construct an individual-level dataset linking four million British immigrants in the US to the UK census (1870–1930) and digitize the universe of British patents granted in the nineteenth century. Using a shift-share design that leverages shocks to US innovation, together with text analysis of patents, we document that migration ties facilitated the diffusion of US technology to the UK. Return migration was an important driver of this effect. However, migrants' social networks at home fostered technology transfer even in the absence of physical return.

**Keywords:** Age of Mass Migration, Innovation, Networks, Out-migration.
**JEL Classification:** F22, N73, N74, O15, O31, O33.

* We are grateful to Mara Squicciarini and Joel Mokyr for their continued guidance and support. For helpful discussions, we thank Hadar Avivi, Jan Bakker, Richard Blundell, Maristella Botticini, Jeremiah Dittmar, Joe Ferrie, Sofia Franchini, Carola Frydman, Simon Görlach, Allison Green, Rachel Griffith, Walker Hanlon, Leander Heldring, Alessandro Iaria, Xavier Jaravel, Wolfgang Keller, Matteo Leombroni, Nicola Limodio, Alan Manning, Paolo Masella, Filippo Mezzanotti, Thomas Monk, Nathan Nunn, Alessandro Nuvolari, Laura Ogliari, Sebastian Ottinger, Eleonora Patacchini, Marta Prato, Mounu Prem, Nancy Qian, Joshua Rosenbloom, Paola Sapienza, Carlo Schwartz, Marco Tabellini, John Van Reenen, Fabian Waldinger, Yanos Zylberberg, and numerous seminar and conference audiences. Antonin Bergeaud, Enrico Berkes, and Walker Hanlon kindly shared data. We acknowledge financial support from Bocconi, Fondazione Invernizzi, LSE, and POID. All errors are our own.

†School of Economics, University of Bristol. Email: davide.coluccia@bristol.ac.uk. Website: dcoluccia.github.io.

‡Einaudi Institute for Economics and Finance. Email: gaia.dossi@eief.it. Website: sites.google.com/view/gaiadossi.

## INTRODUCTION

The diffusion of new technologies across countries is a central driver of productivity growth and economic convergence (Comin and Hobijn, 2010). A notable example is Western Europe in the 1980s, where an estimated 70% of productivity growth relied on technologies developed in the United States and Japan (Eaton and Kortum, 1999).[1] Yet, the channels through which such diffusion occurs are not fully understood. In this paper, we study out-migration as a novel channel of technology transfer from migrants' destination countries to their regions of origin. We refer to this phenomenon as "return innovation."

Out-migration is a widespread feature of economic development (Clemens, 2020). Because migrants typically move to technologically advanced countries, they are ideally positioned to transfer technologies back to their *sending* regions, as qualitatively documented by Saxenian (2006). At the same time, the "brain drain" resulting from out-migration may deplete human capital and hinder innovation (Gibson and McKenzie, 2011). Despite this theoretical tension, existing research has primarily focused on the impact of migrants on innovation in their *destination* countries. This gap stems primarily from data constraints. Moving beyond cross-country analyses requires information on where migrants settle within the receiving country and on their areas of origin within the sending country, but such data are rarely available.

To overcome this limitation, we link individuals recorded as English or Welsh migrants in the US Census between 1870 and 1930 to their records in the UK Census.[2] This new dataset allows us to observe detailed information on migrants' origins in the UK and their destinations in the US. Our main finding is that higher rates of emigration to the US promoted innovation in the UK and increased the similarity between British and American patents, particularly in technological fields that were prevalent in the locations where migrants settled in the US. About half of this "return innovation" effect is driven by return migrants. However, migrants' social networks facilitated cross-border technology flows even in the absence of return migration.

Between 1850 and 1940, a period known as the "Age of Mass Migration," approximately 30 million European migrants settled across the Atlantic. Nearly four million came from England and Wales. During this time, the US reached the technology frontier in many industries, from engines to agricultural machinery (David, 1966; Rosenberg, 1970).[3] In this paper, we document that migration networks promoted the diffusion of these technologies back to Britain, the cradle of the Industrial Revolution.

We assemble two novel general-purpose datasets to overcome the limitations of existing sources. First, the available data do not contain information on the origin of British immigrants in the US beyond the country. To fill this gap, we use confidential individual-level UK and US census data to link the records of British immigrants in the US to the UK census (Schurer and Higgs, 2020; Ruggles et al., 2021). This dataset, which we validate extensively, enables us to track individual migration

[1] Recent endogenous growth models include cross-country technology diffusion dynamics and quantify its implications for economic growth (e.g., Alvarez et al., 2018; Buera and Oberfield, 2020; Perla et al., 2021).

[2] Throughout the paper, we focus on England and Wales. With a slight abuse of terminology, we use the terms "Britain" and "United Kingdom" (UK) as shortcuts to collectively refer to England and Wales. We refer to the United States as the "US."

[3] By the 1890s, the American technological primacy in several sectors was well established. Nelson and Wright (1992) note that, starting in the 1880s, American technology saw major advances in sectors such as textiles and firearms. From the 1890s, mass production led to innovations in consumer products, light machinery, electrical equipment, and industrial machinery.

between the UK and the US, as well as return migration. Second, to reconstruct the geography of innovation in the UK in the second half of the nineteenth century, we digitized 300,000 original patent documents issued in England and Wales between 1853 and 1899. Together with the existing sources, this dataset covers the universe of patents granted between 1853 and 1940.

We measure exposure to US innovation by combining historical emigration flows from the UK to the US with the geography of US patenting. Our approach builds on the empirical observation that the development and diffusion of new technologies are highly spatially concentrated (Kalyani et al., 2025). Consequently, exposure to US innovation depends not only on the number of emigrants from a given district but also on the intensity and specialization of innovative activity in their destination counties. Districts whose emigrants disproportionately settled in highly inventive areas were therefore more likely to gain access to new knowledge. Moreover, because US counties specialize in different technological fields, districts became differentially exposed to specific technologies depending on where their migrants settled. This motivates two complementary measures of exposure to US innovation: (i) a pooled measure, reflecting the overall strength of links to innovative US counties, and (ii) a technology-specific measure, reflecting exposure to the fields in which destination counties specialized.

Identifying the causal impact of migrants' exposure to US technology on innovation poses two key empirical challenges. First, British migrants may sort across US counties in ways correlated with both the innovation of their district of origin and that of their county of destination. Second, common shocks may simultaneously influence innovation in the receiving counties in the US and in the sending districts in the UK. Our technology-level specifications mitigate these concerns by including rich fixed effects that absorb unobserved time-varying heterogeneity at both the district and technology levels, as well as initial district-by-technology characteristics. However, we cannot conclusively rule out that omitted factors drive the association between out-migration and innovation.

To address these concerns, we construct a shift-share instrument that combines predetermined migration patterns with predicted US patenting. The "shares" are given by UK district-to-US county migration flows between 1870 and 1879. By fixing flows in the decade preceding our analysis period, we mitigate concerns that migrants self-selected into US counties in response to contemporaneous shocks. To construct the "shifts," we predict US county-level innovation by apportioning national patenting across technologies to each county in proportion to its pre-1880 patent share. This predicted patenting series is based solely on historical specialization and aggregate national trends. Hence, as we empirically verify, the shifts are independent of contemporaneous county-level shocks, such as changes in industry composition, wages, or local innovation, that could simultaneously influence migrants' location choices and affect innovation in their home districts.

Identification relies on the assumption that, conditional on fixed effects and controls, the shifts are independent of contemporaneous factors driving innovation in migrants' districts of origin (Borusyak et al., 2022). We assess the empirical plausibility of this assumption by showing that the instrument is uncorrelated with pre-determined UK district characteristics, and that the predicted shocks to US innovation are not systematically related to US county observables. We also show that our results are robust to the "recentering" approach of Borusyak and Hull (2023). This approach removes the instru-

ment's mechanical correlation with migrant location patterns, so that identification comes solely from exogenous shifts in US patenting. Consistent with this, no single US state disproportionately drives the identifying variation. Finally, we implement the instrumental-variable local projection method of Stock and Watson (2018) and find no evidence of statistically significant pre-trends, indicating that our results are unlikely to reflect pre-existing differences in patenting unrelated to migrants' exposure to US technology.

The main finding of this paper, which we refer to as "return innovation," is that technological change in the UK was shaped by exposure to US innovation through out-migration linkages. The instrumental variable estimates indicate that moving from the 25th to the 75th percentile of knowledge exposure leads to a 6.5% increase in patenting, equivalent to 10 additional patents issued over a decade in a given district in the UK. We construct a text-based measure of patent importance using the methodology developed by Kelly et al. (2021) applied to patent titles and show that exposure to US technology increased the number of breakthrough innovations—defined as patents in the top quintile of the importance distribution—in the UK. An increase in knowledge exposure from the 25th to the 75th percentile raises breakthrough patenting by 5.7%, equivalent to two additional patents over a decade. Finally, we measure technology transfer by computing textual similarity between UK and US patents. Districts with greater migrant exposure experience a 10% increase in similarity when moving from the 25th to the 75th percentile.

These findings also hold within technological fields. In these more demanding specifications, we control for time-varying unobserved heterogeneity at the district level (such as changes in population or sectoral employment), at the technology level (such as UK-specific demand shocks), and at the district–technology level (such as baseline specialization). These results suggest that exposure to US technology through migration ties increased UK innovation in the technologies to which migrants were most exposed. Moreover, exposure to a given technology generated positive spillovers to related technologies: a one-unit increase in exposure increased patenting by 6% in that technology and by 2.5% in the most closely related ones.

The return innovation effect follows an inverted U-shape with respect to the UK's specialization relative to the US. Patenting in the UK increases most in sectors where the two countries have similar specialization. The effect is substantially smaller in technologies where the US is more advanced (e.g., agriculture and metallurgy) or where the UK is more advanced (e.g., textiles). This pattern aligns with the theoretical predictions of Van Patten (2023), which posit that the gains from international technology diffusion are greatest when a country engages with partners whose technological development is higher but comparable.

In the second part of the paper, we examine the mechanisms behind the return innovation effect. While technology transfer may occur through migrants' physical return, migration ties can also foster diffusion without return migration. We find that return migration is an important driver of return innovation. However, the effect of exposure to US technology remains sizable even after accounting for return migration, suggesting that out-migration facilitates knowledge diffusion even when migrants do not return.

To examine the mechanisms underlying return innovation, we study the role of interactions be-

tween emigrants and their communities of origin. We focus on two channels that may facilitate the diffusion of innovation: family ties and geographical proximity. This analysis builds on previous work in development economics that links the diffusion of technology to network interactions (e.g., Bandiera and Rasul, 2006; Conley and Udry, 2010; Beaman et al., 2021).

First, we find that family members–defined as surname groups within small geographical areas–increase their patenting after their relatives emigrate to the US. The effect emerges after three to five years. Returnees have a larger impact than non-returnees, although the difference is not statistically significant. This implies that non-migrants innovate more even when their relatives do not return.

Second, we leverage the geographical proximity between emigrants and their former neighbors as an alternative proxy for local social networks. Linking UK patent data to geo-coded census records, we show that non-migrants' patenting increases after their neighbors migrate to the US. The effect remains positive and of similar magnitude even when excluding the neighbors of return migrants, suggesting that cross-country interactions are a key driver of return innovation even in the absence of physical return migration.

This paper provides new evidence on how knowledge diffuses across countries. We document that exposure to foreign technologies through migration ties contributes to the diffusion of innovation from migrants' destination countries back to their countries of origin. Our results suggest that out-migration can promote innovation by fostering the diffusion of knowledge into emigration countries.

***Related Literature.*** This paper is closely related to four streams of literature. First, previous literature examines technology diffusion within and across countries (among others, see Jaffe et al., 1993; Bahar et al., 2014). Early studies document the role of foreign direct investments (Aitken and Harrison, 1999), trade (Keller, 2002), and technology sourcing (Griffith et al., 2006) as drivers of international technology diffusion. More recent studies examine the relationship between human mobility and the cross-country diffusion of knowledge (e.g., see Kerr, 2008; Agrawal et al., 2011; Hornung, 2014; Bahar et al., 2022, 2024b; Prato, 2025). We contribute to this literature by focusing on the effects of emigration on sending countries, whereas previous studies have primarily focused on receiving countries. In addition, while prior work mainly focuses on migrant inventors, we analyze the universe of emigrants and show that technology also flows back through non-inventors. This suggests that focusing only on inventors may understate the contribution of out-migration to innovation. Finally, we document that migrants' social networks promote knowledge diffusion even in the absence of their return. In contrast, previous studies on sending countries primarily focus on return migration as the key mechanism (Choudhury, 2016; Bahar et al., 2024a).

Second, we contribute to the literature on the effects of out-migration on sending countries. Previous work shows that emigration shapes a range of economic and social outcomes, including economic development (Khanna et al., 2022), human capital formation (Fernandez-Sanchez, 2022), attitudes toward democracy (Spilimbergo, 2009; Batista and Vicente, 2011) and political change (Barsbai et al., 2017; Karadja and Prawitz, 2019), technology adoption (Coluccia and Spadavecchia, 2024), entrepreneurship (Anelli et al., 2023), and social norms (Beine et al., 2013; Bertoli and Marchetta, 2015; Tuccio and Wahba, 2018; Godlonton and Theoharides, 2025).[4] We contribute to this literature by

[4]Batista et al. (2025) provide an overview of this literature.

showing that emigration can shape the rate and direction of technical change in origin countries by increasing their exposure to knowledge developed abroad. Our findings reveal a previously unexplored channel through which migration affects economic development: rather than merely depleting the domestic stock of human capital, out-migration can foster knowledge creation at home through cross-border exposure and idea flows.

Third, we contribute to the literature on the determinants of the direction of technical change. Pioneering work was formalized by Acemoglu (2002, 2010). More recently, the question has been studied both theoretically (e.g., Bryan and Lemus, 2017; Hopenhayn and Squintani, 2021; Acemoglu, 2023) as well as empirically (e.g., Hanlon, 2015; Moscona, 2021; Einiö et al., 2023; Moscona and Sastry, 2023; Dossi, 2024; Truffa and Wong, 2025). We add to this literature by identifying a new determinant–international human mobility–through the return innovation effect. We show that out-migration fosters innovation in sectors and technologies similar to those migrants are exposed to abroad, thereby influencing the trajectory of technological development in the sending country.

Finally, by its setting, this paper contributes to the literature on innovation and technology diffusion during the Age of Mass Migration. A growing body of work examines the short-run (e.g., Moser et al., 2014; Arkolakis et al., 2020; Moser et al., 2020) and long-run (e.g., Akcigit et al., 2017; Burchardi et al., 2020) impact of immigration on innovation in the US. Our work is most closely related to Andersson et al. (2022), who show that mass out-migration from Sweden led to labor-saving innovation by increasing the relative cost of labor. We document a distinct phenomenon–the "return innovation" effect–through which emigrants generated knowledge flows from their destinations back to their regions of origin. Our findings provide the first evidence that the European mass migration to the US facilitated the diffusion of American technology back to Europe.

***Outline.*** The rest of the paper is structured as follows. Section I describes the historical and institutional context. Section II presents the data. Section III discusses the research design. Section IV documents the main findings. Section V discusses the mechanisms and possible alternative interpretations. Section VI concludes.

## I HISTORICAL BACKGROUND

This section provides details on the English and Welsh migration to the US (Section I.A) and on intellectual property protection in the UK and in the US over the period of our study (Section I.B).

### I.A The English and Welsh Migration to the United States

Between 1850 and 1920–the so-called Age of Mass Migration–more than 30 million Europeans migrated to the US (Abramitzky and Boustan, 2017). Migrants from England and Wales accounted for approximately 10% of the total (Willcox, 1928), and emigration rates in Britain were among the highest in Europe (Baines, 2002).[5]

[5]Only Ireland, Italy, and Norway had higher emigration rates, although, in England, massive out-migration lasted longer than in the other countries above.

### *I.A.1 Migration Policy in the United Kingdom and in the United States*

Until 1917, the US imposed minor restrictions on European immigration (Abramitzky and Boustan, 2017). As a result, the Age of Mass Migration was characterized by virtually no legal constraints on human mobility to the US. Until 1900, immigrants mostly originated from Northern Europe, particularly the UK, Ireland, Germany, Sweden, and Norway. This positive attitude towards immigration started to decline in the 1890s as flows from Eastern and Southern Europe increased. The restrictive immigration policies of the 1920s nevertheless allotted generous quotas to the UK, which were never filled (Abramitzky and Boustan, 2017).[6]

As in other European countries, UK emigration policy encouraged migration within the Empire (Baines, 2002).[7] Emigration to the US was neither subsidized nor discouraged. Attitudes toward outmigration remained positive after World War I, and policymakers viewed the post-War decline in emigrant flows with concern (Leak and Priday, 1933).

### *I.A.2 The Characteristics of British Emigrants*

The British migration to the US had two distinctive features compared to the broader European phenomenon.[8] First, unlike most continental countries, Britain was already highly urbanized and industrialized at the inception of the Mass Migration.

Second, the selection of British migrants differed from that in continental Europe (Erickson, 1957). Compared to the occupational structure of Great Britain, migrants were less likely to be employed in agriculture and more likely to be low and high-skilled industrial workers (Baines, 2002, p. 83). Until the 1880s, British emigrants generally came from rural areas and, consequently, many were farmers. However, as cities and smaller urban centers gained prominence, migrants started being increasingly employed in industrial manufacturing occupations. Starting in the 1880s, urban areas supplied most overseas migrants (Erickson, 1957, 1972; Thomas, 1954).

At the beginning of the 1860s, when the transatlantic migration was taking off, about 15% of emigrants were employed in agriculture, and only five percent were white-collar workers. By the early 1900s, this composition had shifted: agricultural workers accounted for 5% of the overall emigrant stock, while 25% were employed in white-collar occupations.

## I.B Intellectual Property Protection in the United Kingdom and in the United States

During the period of our analysis, patents in the UK and the US were granted by the respective patent offices, but the two countries did not mutually recognize patents. In this section, we describe patent protection in the two countries and international intellectual property protection.

---

[6] The 1921 (respectively, 1924) Act computed the quota for a given country as 3% (respectively, 2%) of the population from that country that was recorded in the US census in 1910 (respectively, 1880). This scheme favored first-wave immigration countries, such as the UK and Germany, at the expense of new ones (Higham, 1955).

[7] Emigration was promoted through subsidized fares and land allotments, coordinated by the Committee of the Emigrants' Information Office. These policies had limited success: Baines (2002) estimates that under 10% of emigrants received government assistance between 1814 and 1918, with Leak and Priday (1933) reporting similar figures after the War.

[8] Throughout the period, the US was the primary destination for English and Welsh migrants. Between 1850 and 1930, over 40% settled there, compared with 25% in Canada, 20% in Australia, and 15% elsewhere (Baines, 2002).

### *I.B.1 The Patent System in the United Kingdom*

The UK established the world's oldest continuously operating patent system in 1623-1624. However, access to intellectual property protection was difficult until 1850 (Bottomley, 2014). Fees amounted to approximately four times the average annual income in 1860, and the application process was lengthy and uncertain (Dutton, 1984). A large literature documents the poor performance of this regime during the Industrial Revolution (e.g., Macleod, 1988). The 1852 Patent Law Amendment Act reformed the system, drawing inspiration from the US model by lowering fees and streamlining procedures. The 1883 reform further reduced fees, allowed mail applications, and established litigation and professional examiners. Novelty examinations began in 1902, and until 1907, patents were granted only if the invention was produced in Britain (Coulter, 1991).

### *I.B.2 The Patent System in the United States*

The first article of the US Constitution gives Congress the power to grant inventors exclusive rights over their discoveries. In 1836, the Patent Act formally established the US Patent Office (USPTO). Two features distinguished it from the British system. First, professional examiners carried out a novelty examination to ascertain the originality of patent applications. Second, low application fees ensured that access to intellectual property protection was widespread. The USPTO has been credited as the first modern patent system in the world (Khan and Sokoloff, 2004).

### *I.B.3 International Intellectual Property Protection*

As national patent systems were established in Europe and in the US during the $19^{th}$ century, demands for international regulation increased. The Paris Convention–formally, the "Paris Convention for the Protection of Industrial Property"–of 1883 governed international patent protection (Penrose, 1951). This agreement emerged from a decade of multilateral confrontations that started with the World Exhibitions in Vienna (1873) and Paris (1878). The Paris Convention introduced two major principles. First, nationals and residents of subscribing countries were guaranteed equality of treatment with nationals. This concept, known as "national treatment," rejects the principle of "reciprocity," which maintains that nationals in subscribing countries would be granted the same protection as their origin country. The US strongly advocated for reciprocity (Penrose, 1951). Second, upon applying for a patent in a member country, inventors were granted a "right of priority" of six months. Patents filed in foreign countries during the priority period would not invalidate the inventor's claim for protection in other member countries. However, patents obtained in one member state were *not* automatically recognized by other countries. To effectively claim protection, inventors had to submit different patent applications, which represented a substantial bureaucratic and financial burden. The UK joined the Convention in 1884, and the US joined in 1887.

While the Paris Convention–and its numerous amendments–are still in operation today, international patents were only established in 1970. Since the UK and the US did not mutually recognize patents during the period of our analysis, we use patenting patterns as a proxy for knowledge flows between the two countries.

## II DATA

We construct two new datasets. The first links British migrants in the US to the UK census, enabling us to measure bilateral migration flows between UK districts and US counties. We describe this dataset in Section II.A. The second dataset comprises the universe of patents granted in England and Wales in the second half of the 19th century. We describe this dataset in Section II.B. In Section II.C, we outline the additional data sources used in the analysis.

### II.A An Individual-level Dataset of British Immigrants in the United States

#### *II.A.1 Motivation*

Our empirical analysis requires information on the detailed origin of English and Welsh immigrants *within* the UK. Available datasets do not contain this information, as neither the US nor the UK collected disaggregated records on the origin of immigrants or the destination of emigrants. We address this limitation by developing a new dataset that links British immigrants in the US to the British census. In this way, we can observe individuals in the US and track them to their UK census record prior to emigration. To the best of our knowledge, this is the first dataset that reconstructs migration flows at this granular level for a non-Scandinavian European country during the Age of Mass Migration.[9]

To construct the linked dataset, we leverage confidential individual-level data from population censuses in the UK (Schurer and Higgs, 2020) and in the US (Ruggles et al., 2021). We extract the universe of British male immigrants from US censuses in 1900, 1910, 1920, and 1930.[10] These records contain the name and surname, birth year, and immigration year of each migrant. Appendix A.II provides further details on the primary sources, the technical implementation of the linking algorithm, and the validation exercises discussed here.

#### *II.A.2 Linking Algorithm*

We use these variables to construct bilateral links between US and UK census records along the lines of Abramitzky et al. (2021). For each British immigrant $A$ in the US census, born in year $b_A$ and who migrated to the US in year $t_A$, we restrict potential matches to records in the UK census immediately preceding $t_A$ with birth year $b_A$.

Let $\mathcal{M}^A$ denote the resulting set of *possible* matches $m_i^A$. For each $m_i^A$, we compute the Jaro-Winkler similarity between $A$'s name and surname and those of $m_i^A$. We denote the similarity scores of the names and surnames as $d_N(A, m_i^A)$ and $d_S(A, m_i^A)$, respectively. To obtain the set of *candidate* matches, we implement a two-step string similarity maximization, keeping all candidate matches having the maximum possible name similarity among those with the highest possible surname similarity with record $A$. First, let $\overline{d}_S^A \equiv \max_{m_i^A \in \mathcal{M}^A} d_S(A, m_i^A)$ be the maximum surname similarity between $A$ and all possible matches. Within the possible matches $\mathcal{M}^A$, we select candidate matches as $\overline{\mathcal{M}}^A \subseteq \mathcal{M}^A$ such that for all elements $\bar{m}_i^A \in \overline{\mathcal{M}}^A$, we have $d_S(A, \bar{m}_i^A) = \overline{d}_S^A$. Second, let $\overline{d}_N^A \equiv \max_{\bar{m}_i^A \in \overline{\mathcal{M}}^A} d_N(A, m_i^A)$ be

[9] Our work adds to the dataset developed by Abramitzky et al. (2014) for Norway. England and Wales were larger in terms of overall population and their contribution to the US immigrant stock: in 1890, Norway's population was approximately 2 million, compared to 27 million in England and Wales.

[10] We cannot use information from the 1870 and 1880 censuses because the immigration year was not recorded. Individual-level data from the 1890 census have not survived. Following the standard practice in the literature, we only link men because women usually changed their surname upon marriage.

the maximum name similarity between $A$ and all the elements of $\overline{\mathcal{M}}^A$. The set of candidate matches $\overline{\mathcal{M}}^A$ is further refined so that for all $\bar{m}_i^A \in \overline{\mathcal{M}}^A$, it is $d_N(A, \bar{m}_i^A) = \bar{d}_N^A$; hence, $\overline{\mathcal{M}}^A$ is the result of a two-step refinement. Finally, we define the set of *accepted* matches $\widetilde{\mathcal{M}}_{\boldsymbol{\tau}}^A \subseteq \overline{\mathcal{M}}^A$ as the set of matches such that the name and the surname similarity between $\tilde{m}_i^A$ and $A$ are above a certain threshold $\boldsymbol{\tau} = (\tau_1, \tau_2) \in [0,1]^2$, i.e., $d_N(A, \tilde{m}_i^A) > \tau_1$ and $d_S(A, \tilde{m}_i^A) > \tau_2$ for all $\tilde{m}_i^A \in \widetilde{\mathcal{M}}_{\boldsymbol{\tau}}^A$.[11]

Following Abramitzky et al. (2021) and Green (2025), we set $\tau_1 = \tau_2 = 0.95$ to retain only high-quality matches. The accepted set may be empty (unmatched), contain one element (single match), or include multiple matches. In the latter case, we weight each migrant by the inverse of the number of matches to reduce the influence of potential false positives. Figure A.2 reports name and surname similarities for all candidate matches ($\overline{\mathcal{M}}^A$) and accepted matches ($\widetilde{\mathcal{M}}_{0.95}^A$).

We also construct a dataset of return migrants following the same procedure as above. In this case, however, migrants are matched to UK censuses conducted in the decades *after* their year of emigration. For example, an individual who migrated to the US in 1895 is matched to the 1901 and 1911 censuses. When multiple matches are found, we retain only those from the earliest census to avoid double-counting.

### *II.A.3 Match Rate and Representativeness*

In our preferred version of the data, we successfully match 53% of male British immigrants in the US census (Figure A.1). The match rate is stable over time, and approximately 35% of matches are unique. Throughout the empirical analysis, we confirm that our results hold when restricting the sample to unique matches.

The match rate is 10-12% higher than in previous studies that build on Abramitzky et al. (2021) to link the US census with external data using names and surnames (e.g., see Green, 2025) or to link censuses across generations (e.g., see Althoff et al., 2025). There are at least three possible explanations for this difference. First, British immigrants were positively selected urban, skilled manufacturing workers, whose census records were likely more accurate than those of other groups.[12] Second, linguistic distance posed no issue for English-speaking immigrants, unlike for other groups whose names may have been misrecorded by enumerators, potentially affecting intergenerational linkage quality. Finally, the high quality of UK census records further facilitated the matching process (Schürer and Day, 2019).

A residual concern is that our sample of linked migrants is not representative of the British migrant population because the matching probability is not random across migrants. Table A.3 reports the correlation between emigrants' observable characteristics and their linking probability. These associations are small and often statistically insignificant, indicating that while the sample of linked migrants is not fully representative of the population, the imbalances appear minor.

[11] This is equivalent to including in the linked sample only records $A$ such that $\bar{d}_S^A \geq \tau_1$ and $\bar{d}_N^A \geq \tau_2$. This equivalence holds because $d_N^A(A, \bar{m}_i^A) = \bar{d}_N^A$ and $d_S^A(A, \bar{m}_i^A) = \bar{d}_S^A$ for all $\bar{m}_i^A \in \overline{\mathcal{M}}^A$.

[12] Helgertz et al. (2022) documents that individuals with higher socioeconomic status have higher linking rates in the US census.

*II.A.4 Internal and External Validation*

The performance of our linking algorithm is not directly comparable to the rates found in the previous literature, as existing studies perform either longitudinal linking of US census records or match US census records to non-census records. In contrast, our algorithm links US to non-US census records. We employ three approaches to validate the quality of our algorithm.[13]

We draw two random samples of 1,000 records each: one of accepted and one of rejected matches. Our first validation approach manually verifies the algorithm's outcomes. Column (1) of Table A.1a reports the results. Of 1,000 accepted matches, 977 are confirmed manually, yielding a precision rate of 98%. Among the 1,000 rejected matches, 825 would also be rejected by a human researcher, while 175 would have been accepted–implying a 17% false negative rate. Overall, out of 1,152 correct matches across the two samples, the algorithm's recall rate is 85%.

A potential concern with manual verification is that it may be influenced by the checker's cognitive biases. To address this, we replicate the exercise using a semi-automated classifier based on a state-of-the-art large language model (`GPT 5`) to classify matches. Column (2) of Table A.1a reports the results. According to this classifier, the algorithm's precision remains very high at 98%, while the share of false negatives–matches rejected by the algorithm despite being plausibly correct–increases to 36%. We do not see this rise as a concern, as the literature prioritizes minimizing false positives over maximizing recall in developing reliable linkage algorithms.

These checks are based on a stratified random sample of 2,000 records. To extend the validation to the full dataset, we train a classifier using a bidirectional text representation model (`BERT`) on the manually verified subsample, and we apply it to the complete linked sample.[14] Table A.2 reports the results. Of 359,757 matches accepted by the algorithm, the `BERT` classifier confirms 339,135 (a 94% precision rate). Among 202,255 rejected matches, `BERT` rejects 182,162, a 10% false negative rate.

All three validation approaches confirm that the linking algorithm produces reliable matches with false positive rates below 5%. It also rejects only about 10% of true matches, indicating that the final sample of linked migrants has high coverage.

It is challenging to validate the data with external sources on the origin of British emigrants to the US because, to the best of our knowledge, they do not exist. Figure A.3 reports the correlation of our data with county-level estimates of aggregate transatlantic emigration for the period 1870–1910 assembled by Baines (2002) and confirms that the two are positively correlated.

*II.A.5 Summary Statistics*

The gray line (left $y$-axis) in Figure I reports the annual number of immigrants from England and Wales to the US, digitized from official statistics (Willcox, 1928). The blue line (right $y$-axis) shows immigrant counts in our linked dataset, which closely track the official series. Figure IIa displays the cross-district distribution of emigration rates in the final sample, revealing substantial spatial heterogeneity. By decade, we document pronounced shifts in the geographic origins of US-bound emigrants (see Figure A.4 and Appendix A.II). Panel B of Table I reports district-level counts of emigrants and

[13] Appendix A.II.4 provides additional details on the implementation of the validation methodologies.

[14] Table A.1a (columns 3–4) and Table A.1b (Panel B) show that the `BERT`-based classification closely aligns with both manual and `GPT`-based validations.

return migrants in the UK–US linked sample by decade and US region of destination.

### II.B A Dataset Covering British Patents in the Second Half of the Nineteenth Century

We measure innovation activity using patents, as is standard practice in the literature (Griliches, 1998).[15] We use data on the universe of US patents digitized by Berkes (2018). Using information on inventors' county of residence and technology class reported in patent documents, we construct a balanced panel at the county–technology class–grant year level.[16]

We collect all patents granted in the UK between 1900 and 1939 from PATSTAT, which provides bulk access to European Patent Office records. These data include information on inventors and CPC classes but lack detailed geographic information. To recover inventors' locations, we merge them with data from Bergeaud and Verluise (2024) and map them to registration districts using 1890 borders.

For 1853–1899, UK patent data were unavailable in digitized form. To address this limitation, we digitized the universe of patents granted in England and Wales between 1853 and 1899 from scans of original documents obtained from the UK Patent Office. We applied optical character recognition (OCR) to over 300,000 documents and used large language models to parse the extracted text into a structured dataset. [17] The resulting data include patent title, text, inventors' names, geo-referenced addresses, filing and issue dates, assignee status, and whether the patent was filed through an agent. Patents are mapped to districts using 1890 borders.

This dataset extends Hanlon (2016) (1855–1883) by adding information on patent titles, texts, and geographic locations, and by extending coverage to 1899. It also expands upon Nuvolari and Tartari (2011), which ends in 1853. Figure A.6 reports the annual number of patents granted in the UK (Figure A.6a), with blue dots representing our data and red dots those from PATSTAT. The two series align closely and show steady growth in patenting over time. Figure A.6b illustrates the evolution of patenting by technological class.

Patents vary in innovativeness and economic value. In modern settings, forward citations are the standard proxy for patent importance, but these are absent from historical records (Andrews, 2021). To measure patent importance, we apply the text-based method developed by Kelly et al. (2021) to the titles of all British patents.[18] We also adapt this approach to compute textual similarity between British and American patents, which we use to measure technology transfer from the US to the UK. Intuitively, two patents are similar if their titles contain semantically related terms. Appendix A.III.6 provides more details on the algorithm.

Finally, for the neighborhood-level analysis, we link patent data to census records. Inventors are matched to census entries based on the string similarity of names and surnames, conditional on geographic proximity. We describe the precise implementation in Appendix A.III.7. Figure A.10 displays

[15] Patents are an imperfect proxy for innovation, as non-patented inventions account for a meaningful share of technological progress (Moser, 2019). However, this limitation is unlikely to affect our results as they also hold *within* technologies: if the propensity to patent is similar across sectors in both countries, cross-sectoral differences should not bias our analysis.

[16] Patents are mapped to counties using 1900 borders. Three-digit Cooperative Patent Classification (CPC) codes are aggregated into twenty broad technological sectors.

[17] Compared to standard text-extraction methods, this approach leverages AI to handle OCR errors and non-standard phrasing in historical patents. Appendix A.III provides additional details on the OCR and parsing procedures.

[18] We use titles for two reasons. First, to maximize coverage, as the text is not available for British patents granted after 1899. Second, patent titles provide a direct and comparable indication of the invention's core technological contribution.

the geographic distribution of inventors across England and Wales using the geo-coded addresses of the inventors linked to the census.

Table I, Panels C and D, present descriptive statistics for this sample. Panel C shows the total number of patents granted across districts by year, broken down by the five most common fields. Panel D summarizes the inventors linked to the population census. Figure IIb displays the spatial distribution of patents per capita, normalized by district population in 1891.

### II.C Additional Variables

We compile district-level statistics from UK population censuses at decade-level intervals between 1851 and 1911. Districts serve as our main unit of analysis. These statistical units had an average population of about 40,000, making them broadly comparable to US counties.[19] District boundaries changed only slightly over time. To ensure geographic consistency, we harmonize all variables to 1890 district borders, as detailed in Appendix A.I.4.

## III EMPIRICAL STRATEGY

This section outlines our empirical strategy. We first describe how we measure exposure to US technology (Section III.A), then introduce the baseline estimating equations (Section III.B). We discuss identification challenges that limit causal interpretation (Section III.C) and present our instrumental variable approach (Section III.D).

### III.A Measuring Exposure to US Innovation Through Migration Ties

We measure exposure to US technology through migration ties, building on recent evidence from Kalyani et al. (2025), who document that technological development is geographically concentrated and diffuses only slowly beyond its place of origin. This slow diffusion implies that migrants' access to new knowledge depends critically on where they settle: those in counties at the technological frontier are exposed early, while those in peripheral regions remain disconnected for long periods. Our measure of "knowledge exposure" reflects this mechanism by weighting districts according to the frontier status of their migrants' destination counties.

For illustration, suppose the same number of migrants leave three UK districts $A$, $B$, and $C$, and settle in US counties $a$, $b$, and $c$. If county $a$ has no patenting, district $A$ has zero exposure; if $b$ and $c$ produce the same number of patents, $B$ and $C$ have the same exposure. Suppose that counties $b$ and $c$ innovate in sectors $s_b$ and $s_c$. Then, extending the previous logic by technology class, our measure reflects that $B$ is exposed to technology $s_b$ and $C$ to $s_c$. These examples motivate our two complementary measures of exposure: a pooled one and a technology-specific one.

The first measure quantifies the intensity of exposure to US technology through migration ties, irrespective of the specific technology. This is defined as:

$$\text{Knowledge Exposure}_{it} = \sum_{j \in J} \text{Patents}_{jt} \times \text{Migrants}_{i \to j,t} \ , \tag{1}$$

where $i$, $j$, and $t$ respectively denote UK districts, US counties, and decades, and $J$ denotes the universe of US counties. The term $(\text{Patents}_{jt})$ denotes the number of patents granted in county $j$ in decade $t$,

[19] Unlike US counties, UK districts had no budgetary or administrative authority.

and $(\text{Migrants}_{i\to j,t})$ is the number of migrants that leave district $i$ and settle in county $j$ over the same ten-year window. The first term in the summation captures the idea that more innovative counties will expose migrants to more knowledge. The second term implies that a district's exposure to innovation in a given county is proportional to the number of migrants that move from that district to that county. According to (1), districts' exposure to US technology increases with the outflow of migrants and with the innovativeness of the counties in which they settle.

The second measure follows the same intuition but exploits two margins of variation: across technological classes and across district–county pairs. The definition is analogous to (1), except that it now leverages variation across technologies. We estimate the following specification:

$$\text{Knowledge Exposure}_{ikt} = \sum_{j \in J} \text{Patents}_{jkt} \times \text{Migrants}_{i\to j,t} \ , \tag{2}$$

where $k$ denotes technology classes. The first term in the summation measures local specialization across technologies and indicates the types of technologies to which migrants are exposed in the US. The second term, analogous to the pooled case, reflects that a district's exposure is proportional to the total number of migrants settling in each county.

### III.B Baseline Estimation

We begin by studying the relationship between overall patenting activity in the UK and exposure to US technology through migration ties:

$$\log \mathbb{E}\left[y_{it}\right] = \alpha_i + \alpha_t + \mathbf{X}'_{it}\beta + \gamma \times \text{Knowledge Exposure}_{it} \ , \tag{3}$$

where $\alpha_i$ and $\alpha_t$ denote district and decade fixed effects, $\mathbf{X}_{it}$ is a set of district-level time-varying controls, and $\text{Knowledge Exposure}_{it}$ is the baseline measure of knowledge exposure. Standard errors are clustered at the district level. District fixed effects control for time-invariant unobserved heterogeneity, such as geography-based confounders like proximity to coal. Decade fixed effects account for time-varying factors, such as the state of the business cycle, that may influence patenting activity.

Second, we examine the relationship between district-level innovation in each technology and the corresponding technology-specific measure of knowledge exposure:

$$\log \mathbb{E}\left[y_{ikt}\right] = \alpha_{it} + \alpha_{kt} + \alpha_{ik} + \gamma \times \text{Knowledge Exposure}_{ikt} \ , \tag{4}$$

where $\alpha_{it}$, $\alpha_{kt}$, and $\alpha_{ik}$ respectively denote district-by-decade, technology-by-decade, and district-by-technology fixed effects. Standard errors are clustered at the district-technology level.[20] Compared to (3), equation (4) incorporates nonparametric controls for a broader set of potentially confounding factors that may influence innovation without being affected by migrants' knowledge exposure. District-by-decade fixed effects ($\alpha_{it}$) control for time-varying unobserved heterogeneity at the district level, including the vector of controls $\mathbf{X}_{it}$, which is therefore omitted. Hence, district-level time-varying factors that do not change at the technology level, such as population and manufacturing growth, are

[20] In Figure C.1–Figure C.2, we report the estimates of the analytical standard errors obtained using several estimators: White, clustering at different levels, and Conley (1999) standard errors. Additionally, we implement the wild-cluster bootstrap estimator proposed by MacKinnon and Webb (2017) at different clustering levels and a spatial block bootstrap that parallels the Conley (1999) correction to explicitly account for spatial autocorrelation. Statistical significance remains unaffected across all specifications.

partialled out of the identifying variation. Technology-by-decade fixed effects control for technology-specific trends and ensure that the estimates do not merely reflect the life-cycle of industries. Finally, district-technology fixed effects control for districts' initial specialization in the technology space. We come back to this in more detail in the next section.

Throughout the paper, we focus on three outcomes. First, we proxy patenting activity with the total number of patents. Second, to account for heterogeneity in economic significance, we restrict attention to the number of patents in the top 20% of the importance distribution as defined by Kelly et al. (2021) and computed on the corpus of UK patents. Third, we measure technology flows from the US to the UK using the average text-based similarity between US and UK patents (see Appendix A.III.6 for details). We estimate Poisson quasi-maximum likelihood (PQML) models, which are better suited than linear ones for count outcomes (Cameron and Trivedi, 2013).[21]

### III.C Potential Sources of Endogeneity

Our ideal experiment would randomly allocate British migrants across US counties and exogenously increase innovation in a subset of those counties. In practice, the setting departs from this ideal experiment for at least two reasons: assortative matching and the reflection problem.

Assortative matching arises if an unobserved factor both predicts migrants' destinations and correlates with innovation in their sending regions. As a result, the estimated $\hat{\gamma}$ coefficients in (3)–(4) may reflect this omitted factor rather than migrants' exposure to US technology. In the pooled specification (3), bias arises if this factor varies within districts over time, as district fixed effects absorb its time-invariant component. One plausible candidate is a district's evolving stock of human capital. If innovation relies on skilled labor, districts with faster educational growth are more likely to generate future innovations. At the same time, better-educated migrants might sort into US counties with higher returns to skill, which may have higher innovation rates. Consequently, changes in human capital could simultaneously predict both future innovation in the UK and migrants' exposure to US technology. In the technology-level specification (4), assortative matching biases the estimates if the omitted variable varies over time within district-technology pairs.[22] A plausible candidate is changes in a district's technology-specific specialization, which may also influence migrants' location choices across the US. As a result, the district's evolving specialization could correlate with the technologies to which migrants are exposed in the US simply through assortative matching.

A separate identification concern is the "reflection problem" (Manski, 1993; Griffith et al., 2006). This arises if common shocks affect UK districts and US counties in ways that are correlated with migration flows between them. In the pooled specification (3), for example, if migrants from high-innovation districts tend to settle in high-innovation US counties, a global shock could simultaneously

[21] We prefer PQML to zero-inflated log-linear models of patent counts (e.g., $\log(1 + \text{Patents})$) because, as shown by Chen and Roth (2024), such transformations yield estimates that are not unit-independent and average treatment effects that are not point identified. By contrast, PQML is consistent for non-negative dependent variables and does not require specifying the outcome's distribution (Correia et al., 2020). In Table C.1, however, we show that our results are robust to using a linear (OLS-2SLS) specification.

[22] This more stringent requirement follows from the granular fixed effects in our specification. District–decade and technology–decade fixed effects absorb all non-technology-specific, time-varying factors. District–technology fixed effects absorb any technology-specific but time-invariant district shocks. As a result, omitted variables would need to be both district–technology specific and time-varying. This distinction is important because Hanlon (2018) and Ottinger (2020) show that migrants' initial location choices were influenced by the industry composition of their origin districts.

raise patenting in both locations, even in the absence of any causal transmission through migration. In the technology-level specification (4), this concern arises if migrants sort into US counties whose industrial specializations mirror those of their home districts. For instance, migrants from textile-intensive regions may settle in US counties with a strong textile base. As a result, a global surge in demand for textiles could increase innovation in both areas independently of any migration-induced knowledge flows. Similarly, a global supply shock could confound the estimates. For example, a worldwide increase in the availability of cheap coal or steel–driven by technological advancements in extraction or international trade integration–could reduce production costs simultaneously in both the UK and the US. This would boost innovation in energy- or metal-intensive industries in both sending and receiving regions, even without any knowledge flows across locations.

In our setting, we find no systematic correlation between bilateral migration flows and the similarity of innovation between sending districts and receiving counties (Appendix B.III). This evidence mitigates concerns that districts and counties more closely connected through migration were already on parallel technological trajectories or jointly exposed to global shocks due to comparable patterns of specialization. Nevertheless, it is difficult to fully rule out assortative matching or reflection effects based solely on observational data.

### III.D Instrumental Variable Design

To address these concerns, we construct a shift-share instrument that uses historical migration patterns as exposure weights and exogenous county-level patent shocks as shifts. This design isolates variation in migrants' exposure that is plausibly unrelated to contemporaneous innovation shocks in the UK.

We construct an instrument for knowledge exposure that (i) fixes migration linkages between UK districts and US counties at the beginning of the period, and (ii) leverages plausibly random county-level variation in patenting activity arising from aggregate innovation trends at the national level. We present the "pooled" district-decade setting and defer the analogous district-technology-decade setting to Appendix C.I.

#### *III.D.1 Predicting Innovation in the United States*

We proceed in two steps. In the first step, we predict county-level patenting activity by interacting county technology specialization at baseline with aggregate innovation activity. We apply the logic of the classic Bartik (1991) instrument to generate a predicted innovation series that does not reflect either the contribution of British inventors to a given county's innovation nor any other county-specific innovation shock that may be correlated with innovation activity in the origin district of the migrants who subsequently settle there. Formally, predicted innovation is given by:

$$\widehat{\text{Patents}}_{jt} = \sum_{k} \omega_{jk} \times \text{Patents}_{kt}^{-j} \,, \tag{5}$$

where $\omega_{jk}$ is the number of patents granted to inventors in county $j$ and in technology $k$ in the decade 1860–1869 divided by the total number of patents granted in technology $k$ nationwide, and $(\text{Patents}_{kt}^{-j})$ is the total number of patents granted in technology $k$ and year $t$ across the US, net of those issued

in county $j$.[23] Similar leave-out strategies are common in the shift-share literature (e.g., see Burchardi et al., 2019; Tabellini, 2020). Figure C.3 displays the positive and statistically significant correlation between observed and predicted patenting at the county- and county-technology levels. Table C.2 provides descriptive statistics on the shifts constructed from (5).

*III.D.2 Instrument for Knowledge Exposure*

In the second step, we construct an instrument for knowledge exposure as:

$$\widehat{\text{Knowledge Exposure}}_{it} = \sum_j \widehat{\text{Patents}}_{jt} \times \text{Migrants}^0_{i \to j} \,, \tag{6}$$

where $(\widehat{\text{Patents}}_{jt})$ is the predicted innovation series described by equation (5), and $(\text{Migrants}^0_{i \to j})$ is the number of emigrants that leave UK district $i$ and settle in US county $j$ over the ten years before our analysis, i.e., between 1870 and 1879. As standard in shift-share designs, the instrument leverages a set of county-level "shifts" in the US $(\widehat{\text{Patents}}_{jt})$ weighted by UK district-level exposure "shares" $(\text{Migrants}^0_{i \to j})$.

Since the shares do not sum up to one, the design falls within the "incomplete shares" case (Borusyak et al., 2022). Accordingly, we always control for an interaction term between the sum of the shares $(\sum_j \text{Migrants}^0_{i \to j})$ and decade dummies in the instrumental variable regressions.[24] In our setting, the "shares" capture absolute, rather than relative, emigration outflows because exposure to US technology is plausibly driven by the number of migrants that are exposed to US technology. Hence, controlling for the sum of the shares implies that we compare districts with a similar level of overall out-migration to the US and leverage differences in location patterns across US counties.

The estimation approach follows the two-stage residual inclusion method, which allows for a linear first stage and a nonlinear second stage (Mullahy, 1997; Terza et al., 2008). In the first stage, we regress the endogenous knowledge exposure variable on the instrument, fixed effects, and the sum of shares. The second stage is a PQML regression of the outcome on the endogenous knowledge exposure variable, the first-stage residual, fixed effects, and the sum of shares.[25] Standard errors are bootstrapped, in line with common practice in control-function methods (Wooldridge, 2010).

Figure III reports the first-stage between the instrumented and observed knowledge exposure at the district-decade level (Figure IIIa) and at the district-technology-decade level (Figure IIIb). To construct each figure, we regress both variables on the relevant fixed effects (district and decade in the first case, and district-decade, district-technology, and technology-decade in the second), and plot the residuals in a binned scatter. Each panel reports the regression coefficient, the clustered standard error, and the within-fixed effects R-squared. In both cases, we find a positive and highly significant correlation between the instrument and the observed knowledge exposure.

---

[23] Formally, $\omega_{jk} = \left(\sum_{t=1860}^{1869} \text{Patents}_{jkt}\right) / \left(\sum_{t=1860}^{1869} \text{Patents}_{kt}\right)$ and $\text{Patents}^{-j}_{kt} = \sum_{j' \neq j} \text{Patents}_{j'kt}$.

[24] In the district-technology-level regressions, the sum of the shares is interacted with technology-decade dummies to reflect that knowledge exposure varies by technology as well as across districts and over time.

[25] Formally, the first stage is given by: $\mathbb{E}\left[\text{Knowledge Exposure}_{it}\right] = \alpha_i + \alpha_t + \mathbf{X}'_{it}\beta + \gamma \times \widehat{\text{Knowledge Exposure}}_{it} + \delta_t \times \sum_j \text{Migrants}^0_{i \to j}$, which yields the residuals $\hat{v}_{it}$. The second stage is $\log \mathbb{E}\left[y_{it}\right] = \alpha_i + \alpha_t + \mathbf{X}'_{it}\beta + \gamma \times \text{Knowledge Exposure}_{it} + \delta_t \times \sum_j \text{Migrants}^0_{i \to j} + \rho \hat{v}_{it}$, where $\delta_t$ codes decade-specific categorical variables. Implicitly, $\hat{\rho}$ yields a test against the null that the baseline knowledge exposure term is endogenous.

*III.D.3 Identification Assumption and Tests of Instrument Validity*

As shown in the recent econometric literature, identification in shift-share designs can rely on either exogenous shocks (Borusyak et al., 2022) or shares (Goldsmith-Pinkham et al., 2020). Because our predicted patenting series–i.e., the "shocks"–in (5) has a Bartik structure and the number of counties is large ($N \approx 3,000$), we interpret identification as coming from (quasi-)exogenous shocks. This approach does not require the baseline migration shares to be exogenous.

Predicted patenting is constructed as the interaction between national innovation trends and historically fixed county-level technological specialization. Our identifying assumption is that, conditional on fixed effects and controls, predicted patenting is orthogonal to unobserved determinants of innovation in UK districts. Under this assumption, the IV estimates can be interpreted as the causal effect of exposure to US innovation transmitted through migration ties.

While the identification assumption is not testable, we follow Borusyak et al. (2022) and conduct four exercises to assess its plausibility.[26]

First, we check for pre-trend balance at the UK district level by regressing the instrument on lagged observed district characteristics and the same set of fixed effects and controls we include in the estimating equation. Table II reports the results. In Panel A, we compute employment shares across sectors from the population census. In Panel B, we measure US emigration and the share of US emigrants who settle in Census divisions. In Panel C, we test for parallel trends in the outcomes at the district and the district-by-technology levels. The correlation between the various lagged outcomes and the instrument is almost always statistically insignificant, and in the few cases where it is–the employment share in textiles and transportation, the share of US emigrants in the West North Central division, and district-technology level similarity of UK and US patents–it is quantitatively small relative to the mean.[27]

Second, we test for shock-level balance by regressing lagged observable county characteristics on US county-level predicted patenting. Following Borusyak et al. (2022), we weight counties by the total sum of the shares ($\sum_i \text{Migrants}^0_{i\to j}$) to reflect their influence in the shift-share instrument. If shocks are as-good-as-randomly assigned across counties, we would expect them not to predict fixed observable characteristics. Table C.3 displays the correlation between shocks and lagged labor market indicators (Panel A), immigration at the country-of-origin-level (Panel B), and immigration from within the UK (Panel C).[28] Across all outcomes, we find no statistically significant correlation between the shocks and observable characteristics, except for a positive correlation with the employment rate and the share of workers employed in agriculture and metallurgy. In all cases, however, the correlation is economically small, confirming that shocks are plausibly as-good-as-randomly assigned across counties. Importantly, we detect no correlation between the shifts and the origin of migrants within the UK (Panel C), suggesting that our construction mitigates concerns that unobserved factors jointly

[26] Borusyak et al. (2025) provide a checklist of analyses to assess the validity of shift-share designs. We follow their recommendations closely.

[27] Figure C.4 shows that the results remain quantitatively unchanged when controlling for the variables that have a non-zero correlation with the instrument. In the district-technology regression, district-by-decade fixed effects implicitly control for these covariates. Figure C.5 shows that the estimates remain robust when controlling for an interaction term between each such variable and technology-decade categorical variables.

[28] Sample statistics for the shifts are reported in Table C.2.

shape migrant location choices and innovation in their origin areas.[29]

Third, Table C.4 reports the results of recentering the instrument by its expected value conditional on the shares, following Borusyak and Hull (2023).[30] This recentering approach removes the mechanical correlation between the instrument and the exposure shares. All results remain quantitatively unchanged. In a similar spirit, Figure C.6 reports estimates obtained when sequentially excluding one US state at a time from the set of counties $J$ used to construct both observed and instrumented knowledge exposure. This exercise assesses whether identification hinges disproportionately on particular migration corridors.[31] The stability of the estimated coefficients confirms that no single state drives the results and that identification stems from many plausibly exogenous shocks.

Finally, a possible concern is that our estimates reflect pre-existing differences across districts that correlate with the instrument. In Figure V, we use the instrumental variable local projection method developed by Stock and Watson (2018) to estimate the dynamic responses of innovation to instrumented knowledge exposure. We find no evidence of statistically significant pre-treatment differences across districts, indicating that our estimates are unlikely to reflect pre-existing differential trends.

## IV MAIN RESULTS

This section documents the "return innovation" effect. We begin with a case study (Section IV.A), then provide causal estimates (Section IV.B), examine how the effect varies with cross-country specialization (Section IV.C), assess spillovers across technologies (Section IV.D), and summarize robustness checks (Section IV.E).

### IV.A Case Study: The Diffusion of Bakelite from the United States to the United Kingdom

We begin by examining how one specific invention diffused from the US to the UK. Bakelite, invented in 1907 and patented in 1909 by Leo Baekeland, was the first plastic made from synthetic components. Marketed as the "material of a thousand uses," it enabled low-cost mass production of heat-resistant and electrically safe components, rapidly displacing wood, rubber, and shellac in sectors such as telephony and electrification (Meikle, 1995).

To track the diffusion of Bakelite-related innovation to the UK, we conduct a dictionary-based search in US and UK patent texts.[32] We restrict attention to US patents with inventors residing in the US and UK patents with inventors residing in the UK, excluding patents filed by agents on behalf of foreign inventors. Figure IVa plots the log count of Bakelite-related patents issued in the US and the UK over time. Prior to Baekeland's invention, plastic-related innovation was virtually absent in both countries. After Bakelite was granted patent protection, related patenting rose rapidly in the US

[29] It is worth noting that, while in the district-level regressions such omitted terms may affect overall patent intensity, in the district-decade specifications, only technology-specific factors would bias the estimates.

[30] To compute the conditional expected value of the instrument $\mu_{it}$–or $\mu_{ikt}$, for district-technology regressions–we draw 1,000 permutations of the shocks and assign them at random across counties, and average out the resulting values of the instrument.

[31] In particular, results are stable even when excluding the New England states, which, as shown in Table I, accounted for roughly 20% of the British immigrant stock.

[32] We identify Bakelite-related patents using the following terms: "Bakelite," "phenolic resin," "phenol formaldehyde," "phenolic condensation," "moulding compound," "molding compound," "synthetic resin," "phenolic plastic," and "phenol aldehyde." These terms were suggested by the large language model `GPT-5`.

and, with a lag of five to seven years, in the UK. While the UK trajectory closely tracks that of the US, patent volumes remained substantially lower throughout the period.

Figure IVb shows that Bakelite-related innovations were concentrated in UK districts highly exposed to US innovation. We divide districts into above- and below-median "Knowledge Exposure" following the definition provided in equation (1). Between 1900 and 1940, almost 95% of Bakelite-related patents originated in districts with above-median exposure (Panel A of Figure IVd).

Figure IVc shows that, even among districts with above-median exposure to US technology, those with stronger migrant ties to Bakelite-producing US locations adopted the technology earlier and innovated more. We measure exposure to Bakelite-specific innovation using the number of migrants from each district who settled in US counties with Bakelite-related patents (equation (2)). Districts are grouped into quartiles based on this measure. The red line plots the (log) number of Bakelite-related patents for districts in the top quartile, and the blue line plots those in the bottom quartile. Between 1900 and 1940, the top quartile accounted for nearly 50% of Bakelite-related patents issued in the UK, while the bottom quartile accounted for less than 12% (Panel B of Figure IVd).

The diffusion of Bakelite provides suggestive evidence that exposure to US technology through migration ties facilitated knowledge flows to the UK. The remainder of the paper tests whether this pattern generalizes beyond Bakelite and provides causal evidence on the effect of exposure to US innovation through migration ties on innovation in the UK.

### IV.B Exposure to Innovation in the United States and Innovation in the United Kingdom

This section moves beyond the Bakelite case study and provides empirical evidence of the broader "return innovation" effect. Table III reports estimates from specifications (3) (columns 1–3) and (4) (columns 4–6). Panel A presents correlations, and Panel B reports the first-stage and IV estimates using the instrument described in equation (6). We focus on Panel B as the IV estimates identify the causal effect of knowledge exposure on innovation.[33]

In columns (1) and (4), the dependent variable is the number of patents. In the district-level IV regression reported in column (1), a one-unit increase in knowledge exposure leads to a 3% rise in patenting activity. To put this magnitude in perspective, moving from the 25th to the 75th percentile of knowledge exposure increases patenting by 6.5%, equivalent to approximately twelve additional patents per decade relative to a mean of 151. Column (4) shows that exposure to US technology through migration ties continues to raise patenting in the more demanding district–technology specification, which controls for district-by-time unobserved factors, among others. In this setting, a one-unit increase in knowledge exposure leads to a 40% increase in patenting.

Columns (2) and (5) estimate equations (3)–(4) using patents in the top 20% of the importance distribution, which we label "breakthrough" innovations. We measure patent importance using the text-based metric of Kelly et al. (2021), which assigns higher importance to patents that depart from earlier technological language and become closer in language to subsequent patents.[34] At the dis-

[33] The OLS and IV estimates are always similar in magnitude, consistent with the earlier evidence that bilateral migration flows are not systematically correlated with pre-existing innovation similarities between sending districts and receiving US counties.

[34] We implement the Kelly et al. (2021) metric on the corpus of British patents using a five-year window. Results are robust to alternative bandwidths.

trict level, a one-unit increase in knowledge exposure raises breakthrough patenting by 5%, which corresponds to a 5.7% increase when moving from the 25th to the 75th percentile of exposure–about three additional breakthrough patents per decade relative to a mean of 47. In the district–technology specification, a one-unit increase in exposure increases breakthrough patenting by 65.8%. These results show that return innovation is not limited to incremental or low-value inventions but extends to novel and influential technologies.

In columns (3) and (6), we show that exposure to US technologies through migration ties increased the similarity between British and American patents. Quantitatively, the association is large. In the district-level regression, a unit increase in knowledge exposure yields a 3% increase in patent similarity, which translates into a 12% increase in the similarity between UK and US patents when moving from the 25th to the 75th percentile of its distribution. The within-district district-technology regression indicates that this pattern holds even when conditioning on time-varying unobserved factors at the district and technology levels, as well as when controlling for baseline district-technology time-invariant factors. A unit increase in knowledge exposure in this case yields a 50.8% increase in the similarity between US and UK patents. The text-based indicator of UK-US patent similarity allows us to refine our findings, as it strongly suggests that migration ties between the US and the UK facilitated the flow of novel knowledge into the UK.

A key concern is that our estimates might capture pre-existing differences in innovation trends across UK districts rather than the causal effect of migration-driven exposure to US technology ("assortative matching"). If migrants from faster-innovating districts were more likely to settle in more innovative US counties, contemporaneous regressions could mistakenly attribute these underlying trends to exposure through migration ties. To assess this concern, we estimate dynamic responses to foreign technology exposure using the local projection IV framework of Jordà (2005) and Stock and Watson (2018). We regress outcomes in past and future decades on contemporaneous knowledge exposure, instrumented with the shift–share measure in (6), and plot the resulting impulse response functions.[35] This approach serves two purposes: first, the inclusion of "lead" coefficients provides a direct pre-trend test, allowing us to verify whether exposure predicts outcomes before the shock.[36] Second, the sequence of "lag" coefficients measures the dynamic evolution of the causal effect of migrants' exposure to foreign technology on home innovation.

Figure V reports the estimates for district–decade (Figure Va) and district–technology–decade (Figure Vb) patenting. In both specifications, the pre-treatment coefficients are close to zero and not statistically significant, suggesting that the results are unlikely to be driven by differential pre-trends across districts. Following the shock, the effects are positive and significant for up to two decades before gradually returning to zero.

---

[35] Formally, for each horizon $k$ between $-20$ and 30 years, we estimate local projections of the form $\log \mathbb{E}[\text{Patents}_{it+k}] = \alpha_i + \alpha_t + \mathbf{X}'_{it}\beta + \gamma_k \times \text{Knowledge Exposure}_{it} + \delta_t \times \sum_j \text{Migrants}^0_{i\to j} + \rho\hat{v}_{it}$, where the control function $\hat{v}_{it}$ is obtained by regressing the endogenous knowledge exposure term against the instrument $\widehat{\text{Knowledge Exposure}}_{it}$ as in the baseline setting and the fixed effects and included controls are the same as in the baseline estimating equations. We estimate a separate regression for each time horizon $k$ and report the sequence of estimated $\hat{\gamma}_k$ coefficients. The framework for the district-technology-decade case is analogous.

[36] This validation approach is closely related to instrument balance regressions in Panel C of Table II, which report the correlation between the instrument and a one-decade lag of the outcome variables.

Another concern is that our instrument may capture global shocks that simultaneously affect innovation in both the US and the UK (the "reflection problem"). For example, suppose there is a positive global demand shock for technology $k$. The exclusion restriction would be violated if (i) migrants from UK districts specialized in $k$ disproportionately settled in similarly specialized US counties before 1880, and (ii) such specialization persisted over time. In that case, the shift–share predictor (5) would assign more $k$-related patents to linked US counties, while the same UK districts would also increase $k$-related innovation in response to the global shock. This would bias the IV coefficient upward by attributing a common global shock to migrant exposure.

To address this point, we estimate a version of equation (4) that interacts knowledge exposure with indicators for quintiles of pre-1880 district specialization by technology (Figure C.7). We assume that technology specialization proxies for a district's vulnerability to technology-specific global supply and demand shocks, which should disproportionately affect areas more concentrated in the relevant technology. By contrast, if the estimated effects reflect return innovation through migrant networks, we would expect them to scale with migrant exposure and not vary systematically across districts with comparable US connectivity, regardless of their initial specialization. Comparing migrant-based exposure within similarly specialized districts thus helps distinguish the impact of global shocks from that of migration-mediated knowledge flows.[37] We find statistically significant and similar effects across all quintiles, supporting the interpretation that our IV estimates measure migrant-mediated exposure rather than correlated global shocks. Further supporting the return innovation channel is the fact that UK patents became more similar to the US patents granted in the preceding five years. This pattern is consistent with the transmission of specific US technologies to the UK, rather than with a common global shock, such as one originating in a third country, causing parallel innovation trends in both countries independently.

How persistent is return innovation? The link between exposure to US technology through emigration and UK innovation appears to persist for several decades. Using data on patenting activity between 1940 and 2015, we estimate the long-run association between knowledge exposure in the 1930s and innovation in the later twentieth century. Figure B.1 reports the correlation between patenting and a series of interaction terms between knowledge exposure and decade dummies between the 1940s and the 2010s at the district level (Figure B.1a) and at the district-technology level (Figure B.1b). The relationship remains stable through the 1960s, begins to weaken in the 1970s, and declines more sharply in the 1980s. Appendix B.II provides a detailed discussion of this analysis.

### IV.C Cross-Country Specialization and Exposure to Foreign Technology

In this section, we show that return innovation varies with the relative comparative advantage of the UK and the US. Specifically, the effect is strongest in technologies where the two countries were similarly positioned along the technological frontier.

We build on the international trade literature and construct a measure of revealed comparative advantage (RCA). This index takes a value above one for technologies where the UK was more inno-

[37] Empirical evidence from trade and innovation supports this interpretation. For example, Autor et al. (2020) show that rising Chinese import competition led US firms in exposed manufacturing industries to reduce R&D investment and produce fewer patents.

vative than the US in terms of share of total patents, and below one otherwise (Balassa, 1965).[38] The historical evidence suggests that the US was relatively more specialized than the UK in agricultural and heavy machinery technologies, even in the early part of the sample period. In contrast, the UK was more specialized in chemical and textile innovation (David, 1966; Rosenberg, 1970).

To study how return innovation varies across technologies, we estimate equation (4), interacting knowledge exposure with a set of technology dummies. Figure VI shows the coefficients associated with these interaction terms. Each dot reports the coefficient of one technology class on the $y$-axis and the UK revealed comparative advantage for that class on the $x$-axis. We find an inverted U-shaped relationship: return innovation is strongest in technologies where the UK and the US had similar levels of specialization, such as blasting and food processing, and weaker when either country dominated. In sectors where the US was far ahead (e.g., mining) or the UK was at the frontier (e.g., electricity), the returns to exposure are statistically indistinguishable from zero. The figure also reports the coefficients of a regression between the coefficients, the RCA, and its squared values, and confirms the hump-shaped pattern. This finding is consistent with the theoretical argument of Van Patten (2023), who develops a model of endogenous technology adoption where the gains from international technology diffusion are highest when a country trades with partners with higher–but similar–levels of development.

### IV.D Knowledge Spillovers on Adjacent Technology Classes

We now study how technology-specific shocks in the US affect innovation across the technological spectrum. We find that exposure to US innovation in sector $k$ raises UK patenting not only within sector $k$ itself, but also in closely related technologies, while having no significant impact on more distant sectors.

In Figure VII, the red dot reports the IV estimate from the baseline estimating equation (4). To compute the blue dots, we first measure the technological distance between two classes as the probability that a patent is assigned to both classes. We then divide the resulting distribution into quintiles: the first contains the most similar technology classes, and the fifth the least similar. We then estimate a series of regressions along the lines of (4), where, in each case, UK innovation in a given technology is regressed on exposure to US innovation in other technologies within a specific similarity quintile, excluding the technology itself. The resulting coefficients are plotted as blue dots.[39]

Return innovation is largest in the same technology to which migrants were exposed, but also spills over to closely related ones. The coefficient for the most similar group (first quintile) is about half the size of the own-technology effect and is statistically significant. The coefficients for the more

[38] In the international trade literature, revealed comparative advantage is a metric that hinges on the observation that a country's comparative advantage is revealed by the country's relative exports (Balassa, 1965). In our setting, we define the revealed comparative advantage as

$$\text{RCA}_{ik} \equiv \frac{\text{Patents}_{ik} \big/ \sum_{k' \in K} \text{Patents}_{ik'}}{\sum_{i' \in I} \text{Patents}_{i'k} \big/ \sum_{i' \in I, k' \in K} \text{Patents}_{i'k'}},$$

where $i$ and $k$ denote countries and sectors within sets $I$ and $K$. Specifically, $I = \{\text{UK}, \text{US}\}$. Then, the UK is relatively more specialized in sectors with $\text{RCA}_{\text{UK},k}$ above one.

[39] For comparability, in each regression, we standardized the explanatory variable (knowledge exposure) to have a mean of zero and a standard deviation of one.

distant quintiles are generally insignificant, except for the fifth (and most dissimilar) quintile, which is negative and statistically significant at about one-third the size of the own-technology effect. This suggests that exposure to US innovation may crowd out innovation in distant technologies. However, this crowding-out is more than offset by increased innovation in the exposed and adjacent sectors. This result is consistent with our finding that return innovation reallocates inventive activity across technologies while also boosting total innovation.

### IV.E Robustness Checks

This section summarizes the robustness checks, which are discussed in detail in Appendix C. We first test the robustness of our findings to alternative definitions of patenting activity. In the baseline analysis, we use a Poisson quasi-likelihood estimator. Table C.1 shows that results are similar under a linear specification. One concern is that patent counts may not reflect economically relevant innovation–i.e., technologies actually adopted by firms. To address this, Table C.6 re-estimates the baseline using only patents with a firm assignee, a closer proxy for firm-level innovation.[40] Another concern is that results may be driven by inventors patenting the same innovation in both countries. To test this, we exclude UK patents with inventors who also appear on US patents. As shown in Table C.5, the results remain consistent, suggesting that cross-filing is not a main driver of our findings.

Second, we test the robustness of our findings to alternative definitions of the text-based dependent variables: importance and similarity. In the baseline, both measures are computed over a five-year window around a patent's issue year. Importance is computed using patents issued five years before and after, and similarity using US patents from the prior five years. Since the choice of the year-window is arbitrary, Table C.9 reports results using one- and ten-year windows, showing that the findings remain qualitatively unchanged. Similarly, we define "high-importance" patents as those in the top 20% of the importance distribution and show that results are robust to stricter thresholds, including the top 10% and 5% (Table C.10).

Third, we assess the robustness of our results to alternative definitions of exposure. Table C.11 shows that the baseline relationship between patenting and exposure to US technology remains qualitatively unchanged across several alternative measures. Additionally, Table C.8 shows that the effect of exposure through migration ties is stronger when exposure is measured using only the most important patents. This pattern is consistent with the idea that migrants were more likely to transmit knowledge back to the UK when the technologies they encountered were of higher importance. Additional details on the construction of these alternative measures are provided in Appendix C.II.3.

Fourth, we confirm that our results are robust when restricting the linked UK-US sample to exact census matches only. Our linkage algorithm matches British immigrants in the US to individuals in the UK census and may return multiple potential matches per immigrant. In the baseline analysis, we follow standard practice and weight each match by the inverse of the number of matches to prevent double-counting. The results remain unchanged when we limit the analysis to unique (single) matches, which represent about 40% of the sample. Table C.7 reports these results.

Fifth, we verify that the baseline results are robust to controlling for district-level observables. Specifically, we interact the 1880 census variables listed in Table II with decade dummies (for district-

[40] UK patent law requires the inventor to be an individual, but many filings indicate firm affiliation.

level regressions) or technology–decade dummies (for district–technology regressions). Figure C.4 and Figure C.5 show estimates obtained when controlling for each variable individually and, in the last row, jointly. The baseline coefficients remain highly stable, even when including characteristics that are modestly correlated with the instrument, as shown in Table II. This suggests that this minor imbalance is unlikely to compromise the validity of our identification strategy.

Finally, we verify that the statistical significance of our results is robust to alternative standard error estimators. In the baseline analysis, standard errors are clustered at the district level (for district-level regressions) and at the district-technology level (for district-technology regressions). Figure C.1 and Figure C.2 present results using alternative clustering methods and adjustments for spatial autocorrelation following Conley (1999). We also report wild- and block-bootstrap standard errors for the IV estimates.[41] Statistical significance remains stable across all estimators.

## V MECHANISMS

After documenting the "return innovation" effect, we examine its underlying mechanisms. Section V.A shows that exposure to US technology influences innovation even after controlling for return migration. Section V.B documents that emigrants affect the innovative activity of their social networks in the UK, even without returning. Finally, Section V.C discusses complementary mechanisms that we cannot test empirically.

### V.A The Role of Return Migration

Return migration is a leading candidate to explain our findings through two channels. First, return migrants may engage in innovation in fields they were exposed to abroad. Second, they may facilitate access to US knowledge without directly innovating. The literature offers mixed evidence on the impact of return migration on innovation.[42]

The baseline linked sample traces British emigrants back to their UK census records prior to migration. To measure return migration, we add one step: linking British migrants in the US to later UK censuses completed after their settlement abroad. We then aggregate return migration flows at the district–county level and by decade. Analogous to equation (1), we define:

$$\text{Return Knowledge Exposure}_{it} \equiv \sum_{j \in J} \text{Patents}_{jt} \times \text{Return Migrants}_{j \to i, 1880} \,, \tag{7}$$

where $\text{Return Migrants}_{j \to i, 1880}$ denotes the number of migrants returning from county $j$ to district $i$ at baseline (1870–1880). We construct a district–technology-level measure of return knowledge exposure following the same logic as in (2). Fixing return-migration flows at baseline has two advantages. First, it mitigates concerns about endogenous return migration. Second, since return migration data are available only for 1880–1910, this approach extends the coverage to 1880–1930, improving comparability with the rest of the analysis.

Table IV reports the results of estimating equations (3) (columns 1–3) and (4) (columns 4–6), adding

[41] Appendix C.II.4 provides additional details on the implementation of the standard errors.

[42] Several studies find modest effects of programs recruiting high-skilled nationals working abroad (Ash et al., 2022; Shi et al., 2023), whereas Giorcelli (2019) shows that individuals exposed to foreign (managerial) knowledge change their behavior upon returning home.

return knowledge exposure as an explanatory variable lagged one decade to capture that return migrants bring back technologies acquired abroad. In columns (1) and (4), the dependent variable is total patents; in columns (2) and (5), breakthrough patents; and in columns (3) and (6), similarity to US patents. Across all specifications, the coefficients on knowledge exposure remain statistically significant and comparable in magnitude to those in Panel A of Table III. Return migration is also positively associated with patenting and statistically significant in columns (1–3). Quantitatively, the effect of return knowledge exposure is similar to that of overall knowledge exposure in the district-level regressions (columns 1–3) and even larger in the district–technology specifications (columns 4–6). This pattern suggests that return migration plays an important role in the return innovation effect, consistent with anecdotal evidence on technology transfer by returnees discussed in Appendix B.IV. At the same time, out-migration continues to explain innovation even when accounting for return migration, implying that physical return is not required for cross-border technology diffusion.

### V.B The Role of Social Networks

In the remainder of this section, we present evidence consistent with knowledge exchange between emigrants and local communities in Britain, even in the absence of physical return migration. We examine the relationship between emigrants and innovative activity within their social networks in Britain. Our approach builds on experimental evidence from developing countries showing that technology diffuses through social networks (e.g., Bandiera and Rasul, 2006; Conley and Udry, 2010; Beaman et al., 2021). We focus on two channels through which emigrants may remain connected to non-migrants: family ties and geographic proximity.

#### *V.B.1 Transatlantic Emigration Within Families*

To reconstruct extended families using census data, which list only household members, we exploit the geographic distribution of surnames. We assume that individuals who share a surname and reside in the same county belong to the same extended family. This assumption is reasonable as long as surnames are not too common. We therefore exclude the top 5% most common surnames, as well as very rare ones. Results are robust to alternative cutoffs, such as excluding the top and bottom 1% or 10% of the surname distribution.[43]

We study the association between emigration to the US and the innovative activity of migrants' relatives who remained in Britain using a difference-in-differences framework. The treatment exploits variation across counties in exposure to emigrant surnames over time. We estimate the following specification:

$$\log \mathbb{E}\left[y_{ft}\right] = \alpha_f + \alpha_t + \sum_{h=-a}^{b} \beta_h \times I[D_{ft} = h] + \varepsilon_{ft}\ , \tag{8}$$

where $f$ denotes a family–i.e., a surname-by-county pair–and $t$ is a year. The term $D_{ft}$ denotes the number of periods since at least one member of family $f$ moved to the US. Since families are identified at the county level, standard errors are clustered by county. Under the standard parallel trends assumption, the coefficients $\beta_{h\geq 0}$ estimate the impact of emigration on patenting activity carried out

[43] A potential concern is that surnames may not reliably identify families in London, where internal migration had been substantial for centuries (e.g., Schürer and Day, 2019). Table C.12 shows that excluding Greater London from the sample leaves the results quantitatively unchanged.

by the emigrant's family. As in the rest of the paper, we employ the PQML (8) estimator because the number of patents is a count variable.

Table V reports the estimated average treatment effect associated with (8). Emigration within the family is associated with higher patenting activity (column 1). This effect is not limited to low-importance innovations. Total patenting rises by 22% among families with at least one emigrant, and high-importance patents increase by 18% (column 2). Patents granted to relatives of US emigrants also become 22% more textually similar to US patents (column 3). This pattern is consistent with knowledge transfer between migrants and their families in Britain. Figure VIIIa presents the corresponding event-study estimates. The figure shows no significant pre-treatment trends and a sharp rise in patenting that builds over the three years after emigration before reverting to zero.

Next, we distinguish between emigrants who return and those who remain abroad. Returnees can interact with their families in person, while non-returning migrants may maintain ties from a distance. In columns (4–6), we limit the treatment to emigrants who never return to Britain. The estimates show that non-returning emigrants are also associated with higher innovation within their families. Comparing columns (1) and (4), we find no statistically significant difference in the magnitude of the estimates. This result suggests that migrants remain connected to their origin communities and may facilitate knowledge exchange even without returning.

*V.B.2 Transatlantic Emigration Within Neighborhoods*

Geographic proximity between the origin of emigrants and stayers is an alternative proxy for local social networks. In this section, we show an increase in the innovation activity of stayers whose former neighbors migrate to the US.

We exploit the granularity of our data to perform an individual-level analysis. We first extract all men aged 18-50 in 1900 who did not emigrate, using the 1891 census. We then build a yearly balanced panel of the number of patents granted to each individual between 1880 and 1900, based on the linked inventor–census data described in Appendix A.III.7.[44] We complement these data with information on the geographic proximity between these stayers and US migrants. Specifically, we define $D_{pt}$ as the number of periods since at least one individual within $k$ kilometers of non-migrant $p$ emigrated to the US, and zero otherwise. In the baseline analysis, we consider a five-kilometer threshold.[45]

The regression specification mirrors (8), except that we replace $D_{ft}$ with $D_{pt}$. All regressions include individual and year-fixed effects. Standard errors are clustered at the neighborhood level because the treatment is assigned based on the location of each individual.

Table VI reports the results. Migration to the US within a neighborhood is associated with higher innovative activity among stayers (column 1). Neighborhood out-migration is also linked to an increase in high-importance patents (column 2). In column (3), we find that individuals exposed to US emigrants through neighborhood ties produce patents that are more textually similar to those

[44] Table C.13 shows that results are robust to excluding Wales, where the probability of linking patent and census records is lower than in England (see Table A.3).

[45] Figure C.8 shows that the results are robust to alternative distance thresholds. The estimated treatment effect remains positive, consistent with the baseline specification. Extremely narrow thresholds exclude many treated individuals, biasing the effect toward zero, while broader definitions introduce measurement error that reduces precision. We discuss these modeling choices in detail in Appendix C.III.

issued in the US. This pattern is consistent with knowledge exchange between emigrants and their origin communities. The magnitude of the neighborhood association is smaller than that of family ties. Having at least one emigrant nearby raises the probability of patenting by about 16%, which is approximately 80% of the within-family estimate. Figure VIIIb shows the corresponding event-study results. We observe no pre-treatment differences and a peak in patenting two to three years after emigration, followed by a gradual decline.

Columns (4–6) repeat the analysis using only non-returning migrants. The specification is identical to the baseline but defines the treatment variable $D_{pt}$ using emigrants who never return to Britain. Neighborhood emigration is associated with higher innovation even when migrants do not return. As in the family setting, the estimates in columns (1–3) and (4–6) are similar in magnitude. This pattern suggests that return migration plays a limited role in explaining the results.

Figure C.9 reports the results of re-estimating the baseline regression interacting the treatment variable with dummies for inventors' occupational categories. We find substantial heterogeneity in responses to neighborhood out-migration: effects are strongest among positively selected inventors–entrepreneurs, utility workers (a category including independent inventors), chemists, metallurgists, and engineers. In contrast, inventors in lower-skilled occupations, such as agriculture and construction, show no response to neighborhood emigration.

Taken together, the results in this section highlight the importance of social networks in driving return innovation. As evidenced by the text-based measure of similarity between British and US patents, emigrants generate knowledge flows that benefit these communities. Importantly, these effects operate even in the absence of physical return migration.

### V.C Complementary Mechanisms

To conclude this section, we discuss two alternative channels–temporary migrations and monetary remittances–that may play a complementary role in shaping innovation outcomes, although they are unlikely to be the primary drivers of the return innovation effect.

***Temporary Migrations.*** In our study of the mechanisms behind the return innovation effect, we find that physical return is an important determinant but not its exclusive driver. Short-term migration may also influence innovation dynamics in the UK. However, because our migration flows are derived from decennial censuses, the data cannot capture temporary movements. For the same reason, we are unable to quantify the role of industrial espionage.

The quantitative impact of temporary migration and industrial espionage is likely modest. First, the concept of a "temporary migrant" in 19th-century transatlantic migration is poorly defined. Piore (1980) characterizes Southern and Eastern European migrants as temporary because they intended to return home eventually, though such returns often took decades. For instance, a one-way cabin ticket from New York to Liverpool–about $100–represented roughly 20% of the average annual US income, making short-term stays rare. Temporary migrations were especially true for UK migrants, whose return rates were substantially lower than those of Southern and Eastern Europeans (Bandiera et al., 2013), suggesting more permanent settlement in the US. Moreover, as Piore (1980) notes, "temporary" migrants were typically low-skilled and thus less likely to facilitate technological transfer. Industrial espionage also appears quantitatively insufficient to explain the return innovation effect

we document. Finally, our research design itself argues against these mechanisms: for temporary migration or espionage to drive the individual- and family-level difference-in-differences results, such activities would need to covary with family- or neighborhood-level emigration on an annual basis. This channel seems unlikely, although it cannot be directly tested due to the absence of data.

***Monetary Remittances.*** While the contribution of remittances to economic development appears modest, these capital inflows may have sustained increased innovation, perhaps by relaxing constraints in access to credit (Gorodnichenko and Schnitzer, 2013). However, remittances alone cannot explain why out-migration shapes the direction of innovation unless monetary and knowledge transfers move together. Because disaggregated data on financial remittances are unavailable, we cannot test this empirically. Nonetheless, the evidence suggests that financial and "knowledge" remittances likely complement, rather than substitute for, one another in fostering innovation.

## VI CONCLUSIONS

The diffusion of technology across countries is a key driver of long-run growth and economic convergence. However, evidence on the mechanisms underlying this diffusion remains limited. This paper provides the first causal evidence that out-migration facilitates the transfer of innovation from migrants' destination countries to their countries of origin.

Our analysis focuses on the mass migration from England and Wales to the US between 1870 and 1940. We link individual-level census records of British immigrants in the US to UK population censuses to construct spatially disaggregated emigration flows. We then combine these data with newly digitized patent records, which, together with existing sources, cover the universe of patents granted in England and Wales throughout the nineteenth and twentieth centuries.

Using a shift-share instrumental variable, we show that exposure to foreign technology through migration ties fosters its diffusion to migrants' countries of origin. First, greater exposure to US innovation through emigration increases overall inventive activity in the UK. Second, innovation shifts toward technologies to which emigrants are most exposed in the US. Textual similarity between UK and US patents further reveals that exposure to US innovation stimulates technology transfer, affecting both low- and high-impact inventions. We refer to this phenomenon as "return innovation."

We explore the channels behind this effect and find that return innovation occurs even in the absence of physical return migration. Interactions between emigrants and their home communities are a key conduit of technology diffusion, even in the absence of return.

Although our setting is historical, our results likely have implications for contemporary contexts. They echo qualitative studies by Saxenian (1999, 2006) and align with contemporary evidence on the importance of in-person interactions for innovation and diffusion (Boudreau et al., 2017; Atkin et al., 2017). Our findings challenge the view that emigration necessarily leads to underdevelopment or stagnation, as suggested by the "brain drain" hypothesis. Instead, migration can promote the diffusion of innovation, a key driver of economic growth. Return innovation and brain drain are likely to coexist, with their relative importance depending on the type of migration (skilled or unskilled) and the relative positions of the origin and destination countries on the technology frontier. Our results suggest that fostering economic integration between migrants' countries of origin and destination is

likely to promote knowledge exchange.

## REFERENCES


**Abramitzky, Ran and Leah Boustan**, "Immigration in American Economic History," *Journal of Economic Literature*, 2017, *55* (4), 1311–45.

**_ , _ , Katherine Eriksson, James Feigenbaum, and Santiago Pérez**, "Automated Linking of Historical Data," *Journal of Economic Literature*, 2021, *59* (3), 865–918.

**_ , Leah Platt Boustan, and Katherine Eriksson**, "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 2014, *122* (3), 467–506.

**Acemoglu, Daron**, "Directed Technical Change," *The Review of Economic Studies*, 2002, *69* (4), 781–809.

_ , "When Does Labor Scarcity Encourage Innovation?," *Journal of Political Economy*, 2010, *118* (6), 1037–1078.

_ , "Distorted Innovation: Does the Market Get the Direction of Technology Right?," *AEA Papers and Proceedings*, 2023, *113*, 1–28.

**Agrawal, Ajay, Devesh Kapur, John McHale, and Alexander Oettl**, "Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation," *Journal of Urban Economics*, 2011, *69* (1), 43–55.

**Aitken, Brian J and Ann E Harrison**, "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela," *American Economic Review*, 1999, *89* (3), 605–618.

**Akcigit, Ufuk, John Grigsby, and Tom Nicholas**, "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age," *NBER Working Paper*, 2017, (No. w23047).

**Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt**, "America's Rise in Human Capital Mobility," *Working Paper*, 2025.

**Alvarez, Fernando E, Francisco J Buera, and Robert E Lucas**, "Idea Flows, Economic Growth, and Trade," *Annual Review of Economics*, 2018, *10*, 315–345.

**Andersson, David E, Mounir Karadja, and Erik Prawitz**, "Mass Migration and Technological Change," *Journal of the European Economic Association*, 2022, *20* (5), 1859–1896.

**Andrews, Michael J**, "Historical Patent Data: A Practitioner's Guide," *Journal of Economics & Management Strategy*, 2021, *30* (2), 368–397.

**Anelli, Massimo, Gaetano Basso, Giuseppe Ippedico, and Giovanni Peri**, "Emigration and Entrepreneurial Drain," *American Economic Journal: Applied Economics*, 2023, *15* (2), 218–252.

**Arkolakis, Costas, Sun Kyoung Lee, and Michael Peters**, "European Immigrants and the United States' Rise to the Technological Frontier," *Working Paper*, 2020.

**Ash, Elliott, David Cai, Mirko Draka, and Shaoyu Liu**, "Bootstrapping Science? The Impact of a "Return Human Capital" Programme on Chinese Research Productivity," *Working Paper*, 2022.

**Atkin, David, Amit K Khandelwal, and Adam Osman**, "Exporting and Firm Performance: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 2017, *132* (2), 551–615.

**Autor, David, David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu**, "Foreign Competition and Domestic Innovation: Evidence from US Patents," *American Economic Review: Insights*, 2020, *2* (3), 357–74.

**Bahar, Dany, Andreas Hauptmann, Cem Özgüzel, and Hillel Rapoport**, "Migration and Knowledge Diffusion: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia," *The Review of Economics and Statistics*, 2024, pp. 1–50.

**_ , Prithwiraj Choudhury, Ernest Miguelez, and Sara Signorelli**, "Global Mobile Inventors," *Journal of Development Economics*, 2024, *171*, 103357.

**_ , _ , James Sappenfield, and Sara Signorelli**, "Talent Flows and the Geography of Knowledge Production: Causal Evidence from Multinational Firms," *Working Paper*, 2022.

**_ , Ricardo Hausmann, and Cesar A Hidalgo**, "Neighbors and the Evolution of the Comparative Advantage of Nations: Evidence of International Knowledge Diffusion?," *Journal of International Economics*, 2014, *92* (1), 111–123.

**Baines, Dudley**, *Migration in a mature economy: emigration and internal migration in England and Wales 1861-1900*, Cambridge University Press, 2002.

**Balassa, Bela**, "Trade Liberalisation and "Revealed" Comparative Advantage," *The Manchester School*, 1965, *33* (2), 99–123.

**Bandiera, Oriana and Imran Rasul**, "Social Networks and Technology Adoption in Northern Mozambique," *The Economic Journal*, 2006, *116* (514), 869–902.

**_ , _ , and Martina Viarengo**, "The Making of Modern America: Migratory Flows in the Age of Mass Migration," *Journal of Development Economics*, 2013, *102*, 23–47.

**Barsbai, Toman, Hillel Rapoport, Andreas Steinmayr, and Christoph Trebesch**, "The Effect of Labor Migration on the Diffusion of Democracy: Evidence from a Former Soviet Republic," *American Economic Journal: Applied Economics*, 2017, *9* (3), 36–69.

**Bartik, Timothy J**, "Who Benefits from State and Local Economic Development Policies?," *WE Upjohn Institute for Employment Research*, 1991.

**Batista, Catia and Pedro C Vicente**, "Do Migrants Improve Governance at Home? Evidence from a Voting Experiment," *The World Bank Economic Review*, 2011, *25* (1), 77–104.

**_ , Daniel Han, Johannes Haushofer, Gaurav Khanna, David McKenzie, Ahmed Mushfiq Mobarak, Caroline Theoharides, and Dean Yang**, "Brain Drain or Brain Gain? Effects of High-Skilled International Emigration on Origin Countries," *Science*, 2025, *388* (6749).

**Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak**, "Can Network Theory-Based Targeting Increase Technology Adoption?," *American Economic Review*, 2021, *111* (6), 1918–1943.

**Beine, Michel, Frédéric Docquier, and Maurice Schiff**, "International Migration, Transfer of Norms and Home Country Fertility," *Canadian Journal of Economics/Revue canadienne d'économique*, 2013, *46* (4), 1406–1430.

**Bergeaud, Antonin and Cyril Verluise**, "A New Dataset to Study a Century of Innovation in Europe and in the US," *Research Policy*, 2024, *53* (1), 104903.

**Berkes, Enrico**, "Comprehensive Universe of US patents (CUSP): Data and Facts," *Working Paper*, 2018.

**Bertoli, Simone and Francesca Marchetta**, "Bringing It All Back Home–Return Migration and Fertility Choices," *World Development*, 2015, *65*, 27–40.

**Borusyak, Kirill and Peter Hull**, "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

**_ , _ , and Xavier Jaravel**, "Quasi-Experimental Shift-Share Research Designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

**_ , _ , and _** , "A Practical Guide to Shift-Share Instruments," *Journal of Economic Perspectives*, 2025, *39* (1), 181–204.

**Bottomley, Sean**, *The British Patent System during the Industrial Revolution 1700–1852: From Privilege to Property*, Cambridge (UK): Cambridge University Press, 2014.

**Boudreau, Kevin J, Tom Brady, Ina Ganguli, Patrick Gaule, Eva Guinan, Anthony Hollenberg, and Karim R Lakhani**, "A Field Experiment on Search Costs and the Formation of Scientific Collaborations," *Review of Economics and Statistics*, 2017, *99* (4), 565–576.

**Bryan, Kevin A and Jorge Lemus**, "The Direction of Innovation," *Journal of Economic Theory*, 2017, *172*, 247–272.

**Buera, Francisco J and Ezra Oberfield**, "The Global Diffusion of Ideas," *Econometrica*, 2020, *88* (1), 83–114.

**Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan**, "Migrants, Ancestors, and Foreign Investments," *The Review of Economic Studies*, 2019, *86* (4), 1448–1486.

**_ , _ , Tarek Alexander Hassan, Lisa Tarquinio, and Stephen J Terry**, "Immigration, Innovation, and Growth," *NBER Working Paper*, 2020, (No. w27075).

**Cameron, Adrian C and Pravin K Trivedi**, *Regression Analysis of Count Data*, Cambridge University Press, 2013.

**Chen, Jiafeng and Jonathan Roth**, "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, 2024, *139* (2), 891–936.

**Choudhury, Prithwiraj**, "Return Migration and Geography of Innovation in MNEs: A Natural Ex-

periment of Knowledge Production by Local Workers Reporting to Return Migrants," *Journal of Economic Geography*, 2016, *16* (3), 585–610.

**Clemens, Michael A**, "The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries," *IZA Discussion Paper*, 2020.

**Coluccia, Davide M and Lorenzo Spadavecchia**, "Emigration Restrictions and Economic Development: Evidence from the Italian Mass Migration to the United States," *Working Paper*, 2024.

**Comin, Diego and Bart Hobijn**, "An Exploration of Technology Diffusion," *American Economic Review*, 2010, *100* (5), 2031–2059.

**Conley, Timothy G**, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

_ **and Christopher R Udry**, "Learning About a New Technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

**Correia, Sergio, Paulo Guimarães, and Tom Zylkin**, "Fast Poisson Estimation with High-Dimensional Fixed Effects," *The Stata Journal*, 2020, *20* (1), 95–115.

**Coulter, Moureen**, *Property in Ideas: The Patent Question in Mid-Victorian England*, Kirksville (MO): Thomas Jefferson Press, 1991.

**David, Paul A**, "The Mechanization of Reaping in the Ante-Bellum Midwest," in "Industrialization in Two Systems: Essays in Honor of Alexander Gershenkron," Harvard University Press Cambridge, MA, 1966.

**Dossi, Gaia**, "Race and Science," *Working Paper*, 2024.

**Dutton, H I**, *The Patent System and Inventive Activity during the Industrial Revolution, 1750-1852*, Manchester (UK): Manchester University Press, 1984.

**Eaton, Jonathan and Samuel Kortum**, "International Technology Diffusion: Theory and Measurement," *International Economic Review*, 1999, *40* (3), 537–570.

**Einiö, Elias, Josh Feng, and Xavier Jaravel**, "Social Push and the Direction of Innovation," *Working Paper*, 2023.

**Erickson, Charlotte**, *American industry and the European immigrant, 1860–1885*, Cambridge (MA): Harvard University Press, 1957.

_ , "Who were the English and Scots emigrants in the 1880s?," in "Population and Social Change," Arnold, 1972.

**Fernandez-Sanchez, Martin**, "Mass Emigration and Human Capital Over a Century: Evidence from the Galician Diaspora," *Working Paper*, 2022.

**Gibson, John and David McKenzie**, "Eight Questions About Brain Drain," *Journal of Economic Perspectives*, 2011, *25* (3), 107–28.

**Giorcelli, Michela**, "The Long-Term effects of Management and Technology Transfers," *American Economic Review*, 2019, *109* (1), 121–52.

**Godlonton, Susan and Caroline Theoharides**, "Diffusion of Reproductive Health Behavior Through International Migration: Effects on Origin-Country Fertility," *American Economic Review*, 2025, *115* (10), 3597–3637.

**Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift**, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 2020, *110* (8), 2586–2624.

**Gorodnichenko, Yuriy and Monika Schnitzer**, "Financial Constraints and Innovation: Why Poor Countries Don't Catch Up," *Journal of the European Economic Association*, 2013, *11* (5), 1115–1152.

**Green, A**, "Networks and Geographic Mobility: Evidence from World War II Navy ships," *Working Paper*, 2025.

**Griffith, Rachel, Rupert Harrison, and John Van Reenen**, "How Special is the Special Relationship? Using the Impact of US R&D Spillovers on UK Firms as a Test of Technology Sourcing," *American Economic Review*, 2006, *96* (5), 1859–1875.

**Griliches, Zvi**, "Patent Statistics as Economic Indicators: A Survey," in "R&D and productivity: the econometric evidence," University of Chicago Press, 1998, pp. 287–343.

**Hanlon, W Walker**, "Necessity is the Mother of Invention: Input supplies and Directed Technical Change," *Econometrica*, 2015, *83* (1), 67–100.

_ , "British Patent Technology Classification Database: 1855-1882," 2016. Unpublished [data collection].

_ , "Skilled Immigrants and American Industrialization: Lessons from Newport News Shipyard," *Business History Review*, 2018, *92* (4), 605–632.

**Helgertz, Jonas, Joseph Price, Jacob Wellington, Kelly J Thompson, Steven Ruggles, and Catherine A Fitch**, "A New Strategy for Linking US Historical Censuses: A Case Study for the IPUMS Multigenerational Longitudinal Panel," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2022, *55* (1), 12–29.

**Higham, John**, *Strangers in the land: Patterns of American nativism, 1860-1925*, Rutgers University Press, 1955.

**Hopenhayn, Hugo and Francesco Squintani**, "On the Direction of Innovation," *Journal of Political Economy*, 2021, *129* (7), 1991–2022.

**Hornung, Erik**, "Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia," *American Economic Review*, 2014, *104* (1), 84–122.

**Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson**, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 1993, *108* (3), 577–598.

**Jordà, Òscar**, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 2005, *95* (1), 161–182.

**Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Hassan, Josh Lerner, and Ahmed Tahoun**, "The Diffusion of New Technologies," *The Quarterly Journal of Economics*, 2025, *140* (2), 1299–1365.

**Karadja, Mounir and Erik Prawitz**, "Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States," *Journal of Political Economy*, 2019, *127* (4), 1864–1925.

**Keller, Wolfgang**, "Trade and the Transmission of Technology," *Journal of Economic Growth*, 2002, *7*, 5–24.

**Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy**, "Measuring Technological Innovation Over the Long Run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

**Kerr, William R**, "Ethnic Scientific Communities and International Technology Diffusion," *The Review of Economics and Statistics*, 2008, *90* (3), 518–537.

**Khan, B Zorina and Kenneth L Sokoloff**, *Institutions and Technological Innovation During the Early Economic Growth: Evidence from the Great Inventors of the United States, 1790-1930*, National Bureau of Economic Research Cambridge, Mass., USA, 2004.

**Khanna, Gaurav, Emir Murathanoglu, Caroline B Theoharides, and Dean Yang**, "Abundance from Abroad: Migrant Income and Long-Run Economic Development," *NBER Working Paper*, 2022, (29862).

**Leak, H and T Priday**, "Migration from and to the United Kingdom," *Journal of the Royal Statistical Society*, 1933, *96* (2), 183–239.

**MacKinnon, James G and Matthew D Webb**, "Wild Bootstrap Inference for Wildly Different Cluster Sizes," *Journal of Applied Econometrics*, 2017, *32* (2), 233–254.

**Macleod, Christine**, *Inventing the Industrial Revolution*, Cambridge (UK): Cambridge University Press, 1988.

**Manski, Charles F**, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

**Meikle, Jeffrey L**, *American Plastic: A Cultural History*, Rutgers University Press, 1995.

**Moscona, Jacob**, "Environmental Catastrophe and the Direction of Invention: Evidence from the American Dust Bowl," *Working Paper*, 2021.

_ **and Karthik A Sastry**, "Does Directed Innovation Mitigate Climate Damage? Evidence from US Agriculture," *The Quarterly Journal of Economics*, 2023, *138* (2), 637–701.

**Moser, Petra**, "Patents and Innovation in Economic History," in "Research Handbook on the Economics of Intellectual Property Law," Edward Elgar Publishing, 2019, pp. 462–481.

_ **, Alessandra Voena, and Fabian Waldinger**, "German Jewish Émigrés and US Invention," *American*

*Economic Review*, 2014, *104* (10), 3222–55.

_ **, Sahar Parsa, and Shmuel San**, "Immigration, Science, and Invention. Evidence from the Quota Acts," *Working Paper*, 2020.

**Mullahy, John**, "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economics and Statistics*, 1997, *79* (4), 586–593.

**Nelson, Richard R and Gavin Wright**, "The Rise and Fall of American Technological Leadership: The Postwar Era in Historical Perspective," *Journal of Economic Literature*, 1992, *30* (4), 1931–1964.

**Nuvolari, Alessandro and Valentina Tartari**, "Bennet Woodcroft and the Value of English Patents, 1617–1841," *Explorations in Economic History*, 2011, *48* (1), 97–115.

**Ottinger, Sebastian**, "Immigrants, Industries, and Path Dependence," *Working Paper*, 2020.

**Patten, Diana Van**, "International Diffusion of Technology: Accounting for Heterogeneous Learning Abilities," *Working Paper*, 2023.

**Penrose, Edith**, *The Economics of the International Patent System*, Baltimore (MD): Johns Hopkins University Press, 1951.

**Perla, Jesse, Christopher Tonetti, and Michael E Waugh**, "Equilibrium Technology Diffusion, Trade, and Growth," *American Economic Review*, 2021, *111* (1), 73–128.

**Piore, Michael J**, *Birds of Passage: Migrant Labor and Industrial Societies*, Cambridge (UK): Cambridge University Press, 1980.

**Prato, Marta**, "The Global Race for Talent: Brain Drain, Knowledge Transfer, and Growth," *The Quarterly Journal of Economics*, 2025, *140* (1), 165–238.

**Rosenberg, Nathan**, "Economic Development and the Transfer of Technology: Some Historical Perspectives," *Technology and Culture*, 1970, *11* (4), 550–575.

**Ruggles, Steven, Catherine Fitch, Ronald Goeken, J Hacker, M Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek**, "IPUMS ancestry full count data: Version 3.0 [dataset]," *Minneapolis, MN: IPUMS*, 2021.

**Saxenian, AnnaLee**, *Silicon Valley's New Immigrant Entrepreneurs*, Public Policy Institute of California, 1999.

_ , *The New Argonauts: Regional Advantage in a Global Economy*, Harvard University Press, 2006.

**Schurer, K and E Higgs**, "Integrated Census Microdata (I-CeM) Names and Addresses, 1851-1911: Special Licence Access," 2020. [data collection] Second Edition, UKDS.

**Schürer, Kevin and Joe Day**, "Migration to London and the Development of the North–South Divide, 1851–1911," *Social History*, 2019, *44* (1), 26–56.

**Shi, Dongbo, Weichen Liu, and Yanbo Wang**, "Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists?," *Science*, 2023, *379* (6627), 62–65.

**Spilimbergo, Antonio**, "Democracy and Foreign Education," *American Economic Review*, 2009, *99* (1), 528–43.

**Stock, James H and Mark W Watson**, "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 2018, *128* (610), 917–948.

**Tabellini, Marco**, "Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration," *The Review of Economic Studies*, 2020, *87* (1), 454–486.

**Terza, Joseph V, Anirban Basu, and Paul J Rathouz**, "Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling," *Journal of Health Economics*, 2008, *27* (3), 531–543.

**Thomas, B**, *Migration and economic growth. A study of Great Britain and the Atlantic Economy*, Cambridge (MA): NIESR and Cambridge University Press, 1954.

**Truffa, Francesca and Ashley Wong**, "Undergraduate Gender Diversity and the Direction of Scientific Research," *American Economic Review*, 2025, *Forthcoming*.

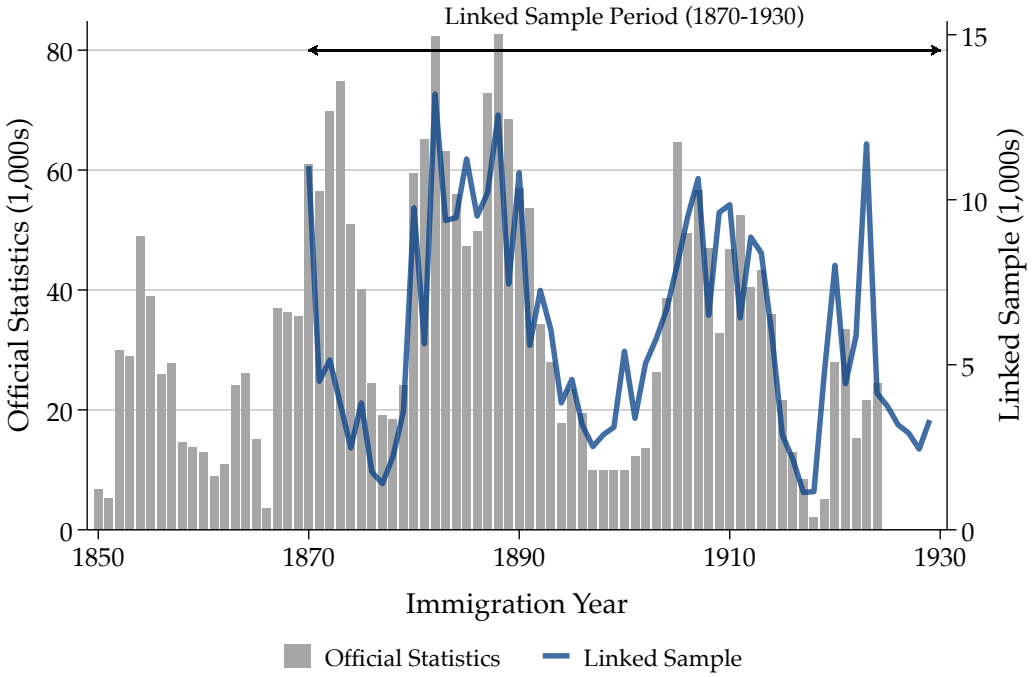
**Tuccio, Michele and Jackline Wahba**, "Return Migration and the Transfer of Gender Norms: Evidence from the Middle East," *Journal of Comparative Economics*, 2018, *46* (4), 1006–1029.

**Willcox, William F**, *International Migrations, Volume I: Statistics*, Cambridge (MA): National Bureau of Economic Research, 1928.

**Wooldridge, Jeffrey M**, *Econometric Analysis of Cross-Section and Panel Data*, MIT press, 2010.

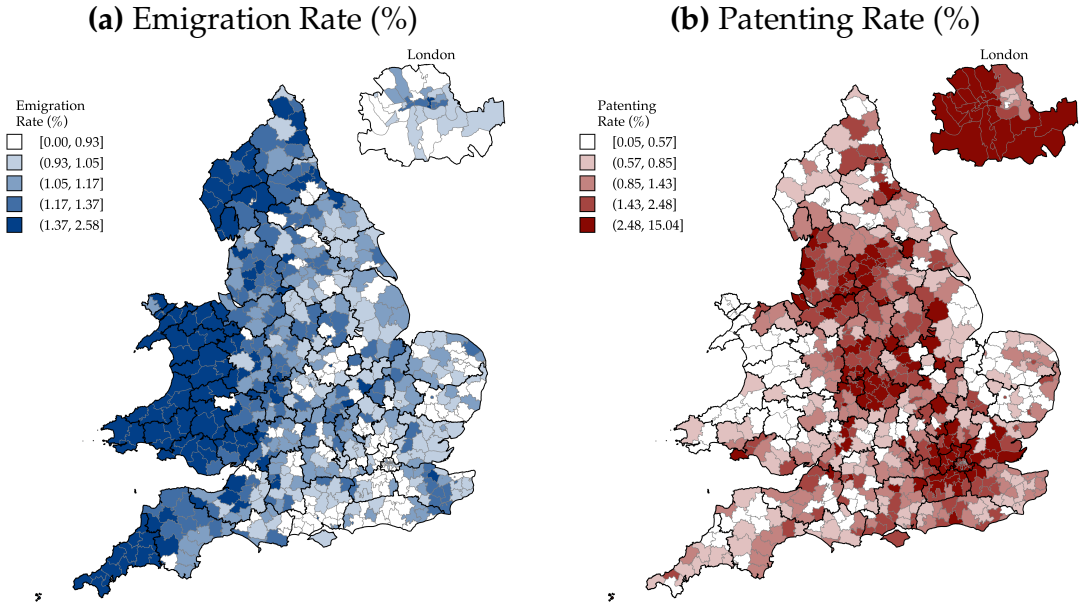
## FIGURES

**Figure I.** British Emigration to the US: Official Statistics and Linked Sample



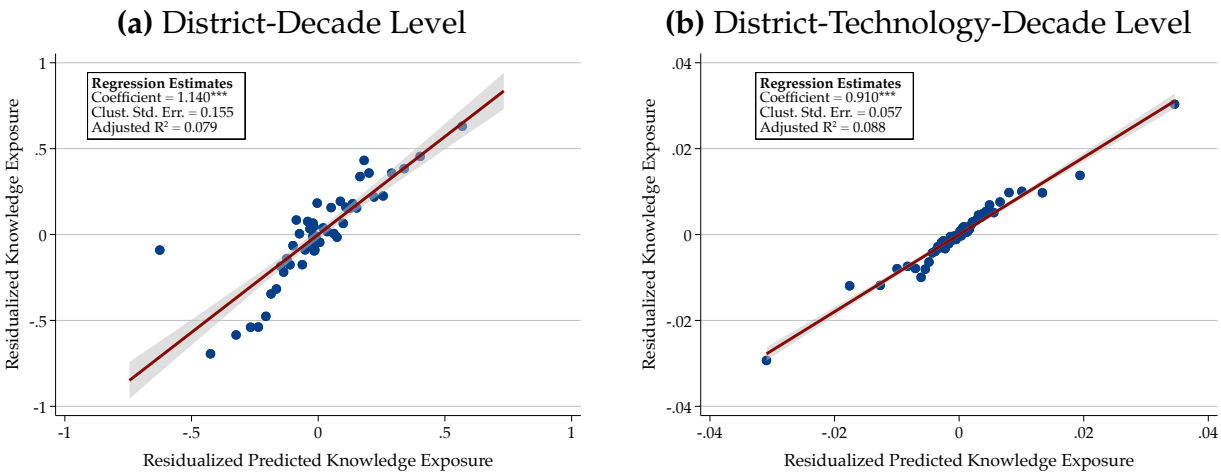

*Notes*. This Figure compares the total number of English and Welsh immigrants in the US as recorded in official statistics from Willcox (1928) with the linked emigrant sample assembled in this paper. The gray bars, plotted on the left $y$-axis, show the number of English and Welsh immigrants in the official statistics by year of recorded immigration to the US. The blue line, whose values are reported on the right $y$-axis, shows the total number of English and Welsh immigrants in the US that appear in our matched sample. By construction, we can only match men who appear at least once in a British census. Numbers are in thousands. The black two-sided arrow marks the period covered by the linked sample, i.e., the years between 1870 and 1930. Referenced on page: 10.

**Figure II.** Patenting and Emigration Rates in the UK



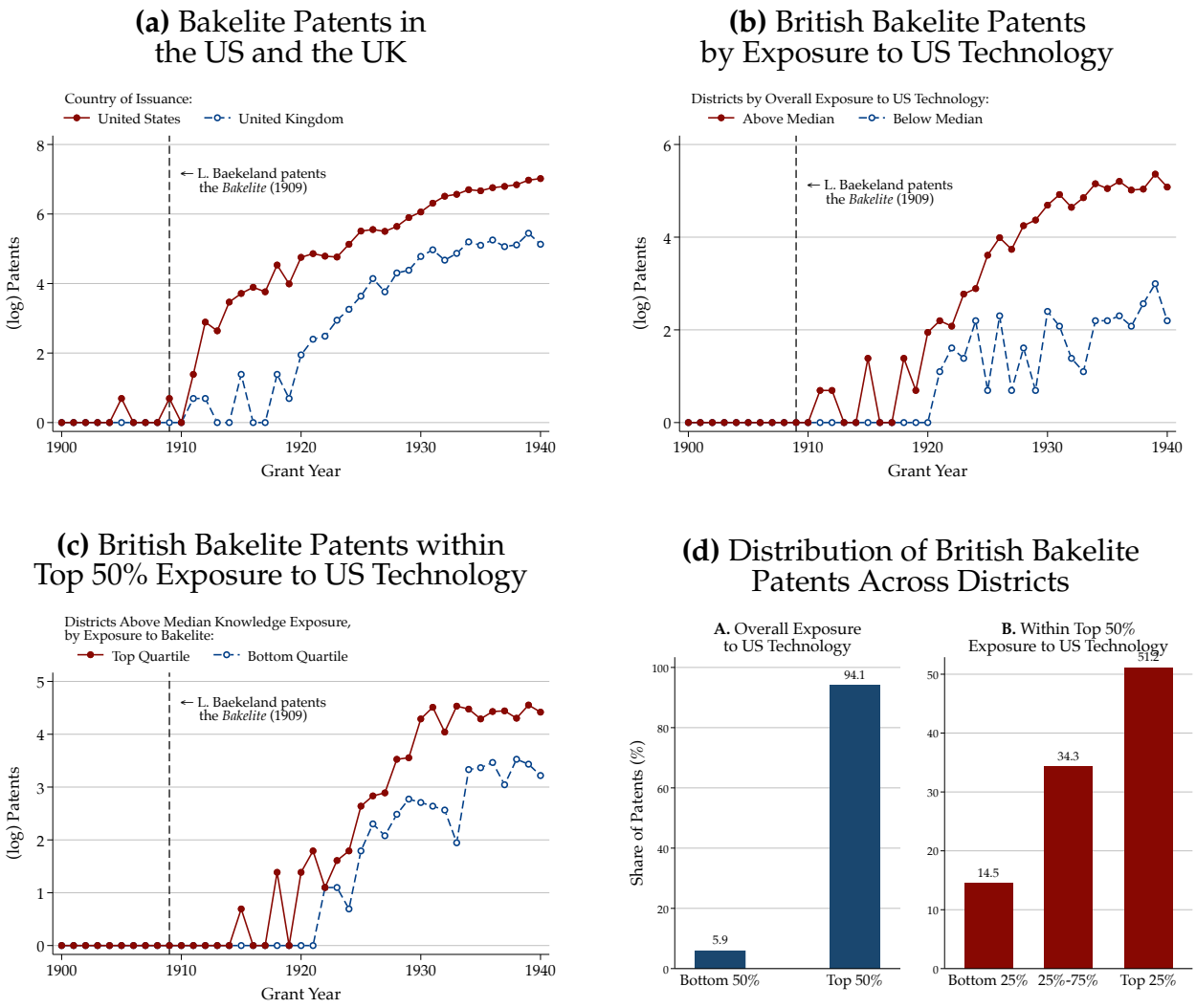

*Notes*. This Figure shows the spatial distribution of emigrants across English and Welsh districts over the period 1870–1930 (Panel IIa) and the number of patents issued over the same period (Panel IIb). Both series are normalized by district population in 1891 and multiplied by 100 for readability. Districts are displayed at 1891 historical borders, and the emigrant population is cross-walked to consistent borders. Lighter to darker tones indicate increasing quantiles of each variable. The London area is displayed separately. Referenced on pages: 10, 12, A12.

**Figure III.** Correlation Between Observed and Predicted Knowledge Exposure



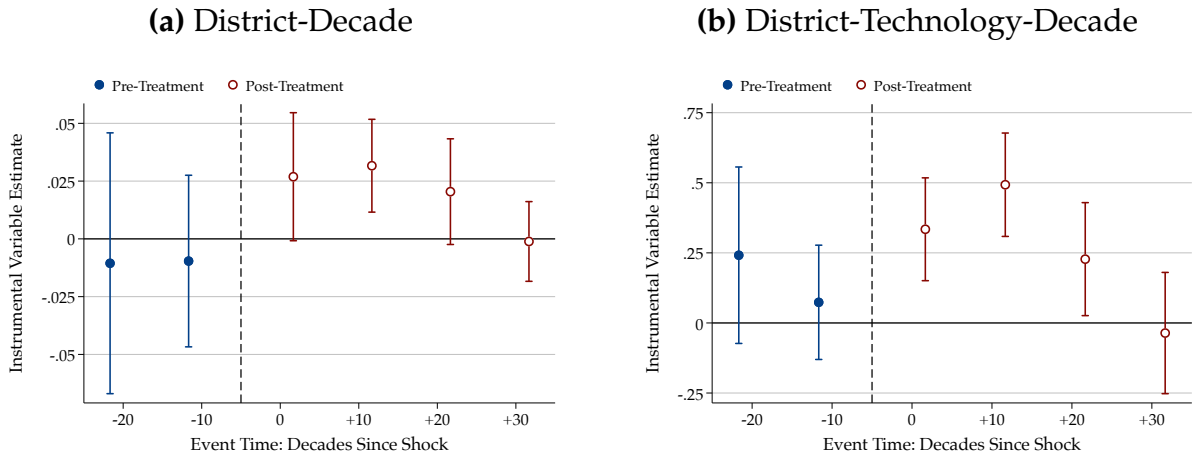

*Notes*. This Figure reports the correlation between observed and predicted knowledge exposure. In Panel IIIa, the unit of observation is a district; in Panel IIIb, the unit of observation is a district-technology pair. Units are observed at a decadal frequency between 1880 and 1930. Knowledge exposure is defined in equations (1)–(2). The instrument is defined in equation (6). To construct Panel IIIa (respectively, IIIb), we residualize both variables on district and decade (respectively, district-technology, district-decade, and technology-decade) fixed effects, controlling for the sum of shares as in the instrumental variable estimating equations, and display binned scatter plots of the corresponding residuals. Each panel overlays a linear fit and the associated 95% confidence bands, and reports the regression coefficient, clustered standard errors, and within-fixed effects $R^2$. Referenced on page: 16.

**Figure IV.** Case Study of the Diffusion of Synthetic Plastics ("Bakelite") Innovation



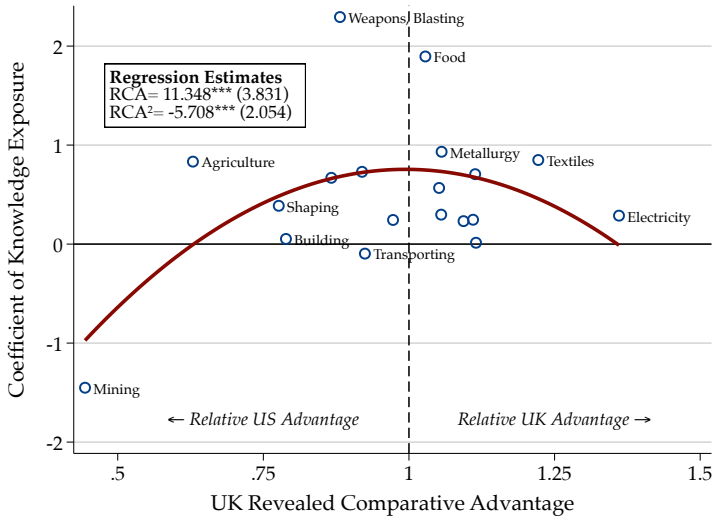

*Notes*. This Figure illustrates the diffusion of Bakelite, the first synthetic plastic material, from the US, where L. Baekeland invented it in 1909, to the UK. Panel IVa reports the total number of Bakelite-related patents issued in the US (red) and the UK (blue). Panel IVb reports patents issued to inventors living in districts below- (blue) and above-median (red) migrants' exposure to US technology. Panel IVc focuses on districts with above-median knowledge exposure. The red line reports patenting in districts in the top quartile of exposure to counties with Bakelite-related innovations; the blue line reports patenting in districts in the bottom quartile. Panel IVd displays the proportion of patents by district exposure to US technology (A) and by exposure to Bakelite-related technology within the top 50% of districts most exposed to US technology (B). Patenting is measured as $\log(1+\cdot)$. Referenced on pages: 18, 19.

**Figure V.** Exposure to US Technology and UK Innovation: Instrumental Variable Local Projection Estimates



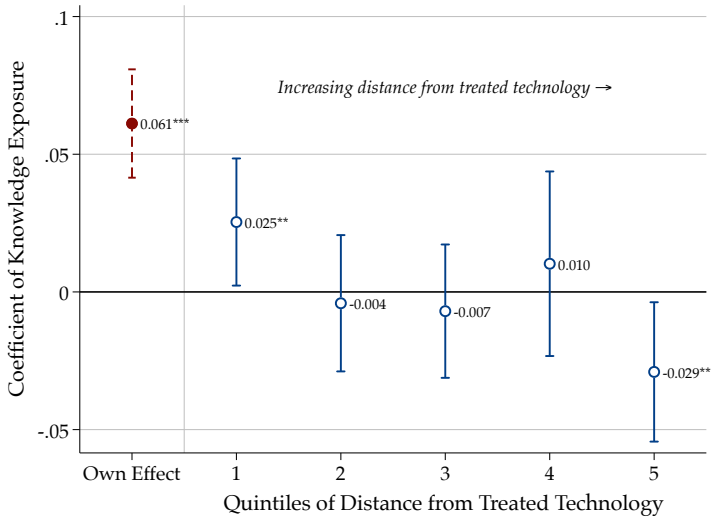

*Notes*. This Figure reports the effect of migrants' exposure to US technology on innovation in the UK. In Panel Va (respectively, Vb), the unit of observation is a district (respectively, district-technology pair) observed at a decadal frequency between 1880 and 1930. The dependent variable is the number of patents. The framework is the instrumental variable local projection approach of Stock and Watson (2018): each dot plots the two-stage least-squares estimated coefficient of a regression between the outcome at the $k$-th lead and lag on knowledge exposure (defined in (1)–(2)), obtained using the instrument defined in (6). Panel Va includes district and decade fixed effects. Panel Vb includes district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors are clustered at the district (Panel Va) and district-technology (Panel Vb) level and are bootstrapped. The bands report 95% confidence intervals. Referenced on pages: 18, 20.

**Figure VI.** Effects of Exposure to US Innovation by US-UK Relative Specialization




*Notes*. This Figure reports the heterogeneous effects of exposure to US technology on UK innovation across technology classes. The unit of observation is a district-technology pair observed at a decadal frequency between 1880 and 1930. The estimating equation follows (4), except that we interact the baseline knowledge exposure term with technology indicator variables; each dot reports the coefficient on one of these interaction terms. The dependent variable is the number of patents. Each regression includes district-by-decade, technology-by-decade, and district-by-technology fixed effects. The estimated coefficients are plotted against the measure of revealed comparative advantage discussed in the main text. Referenced on page: 22.

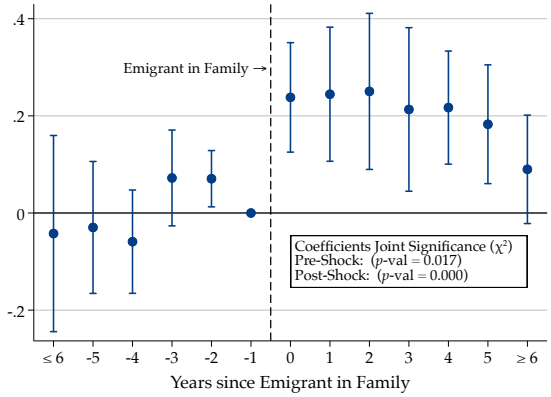
**Figure VII.** Knowledge Spillovers of Migrants' Exposure to US Innovation



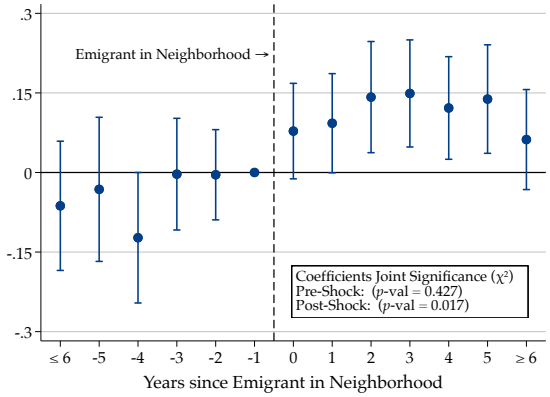

*Notes*. This Figure reports the cross-technology knowledge spillovers of exposure to US technology on UK innovation. The unit of observation is a district-technology pair observed at a decadal frequency between 1880 and 1930. The red dot reports the baseline treatment effect from regression (4). To estimate the blue dots, we compute the similarity between pairs of technology classes and divide the resulting measure into quintiles. Then, we estimate a set of regressions where each technology is paired with "Knowledge Exposure" in technologies in the $k$-th quintile of the similarity distribution. The dots report the resulting IV coefficients. To make the estimates comparable, knowledge exposure is standardized. Standard errors are clustered at the district-technology level and are bootstrapped. The bands report 95% confidence intervals. Referenced on page: 22.

**Figure VIII.** Dynamic Effects of Family and Neighborhood Emigration

**(a)** Within-Family Emigration




**(b)** Within-Neighborhood Emigration




*Notes*. This Figure reports the impact of emigration to the US on the innovation activity of migrants' relatives in the UK (Panel VIIIa) and of their neighbors (Panel VIIIb). In Panel VIIIa, the unit of observation is an extended family defined as all individuals sharing the same surname within a given county. Families are observed at a yearly frequency between 1880 and 1930. In Panel VIIIb, the unit of observation is an individual observed at a yearly frequency between 1880 and 1900. Each dot reports the coefficient on an indicator variable that captures the number of years since the first emigrant in the family (Panel VIIIa) or in the neighborhood (Panel VIIIb). The black dashed line indicates the first treatment period; the previous period serves as the baseline category. The dependent variable is the number of patents. All regressions control for unit and year fixed effects. Standard errors are clustered at the county (Panel VIIIa) and neighborhood (Panel VIIIb) level. The bands report 95% confidence intervals. All regressions are PQML. Referenced on page: 26.

## TABLES

### **Table I.** Descriptive Statistics

| | Mean | Std. Dev. | Min | Max | Units | Obs. |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A. Demographics** | | | | | | |
| Population$_{it}$ (1,000s) | 48.560 | 62.785 | 0.092 | 703.559 | 632 | 2,517 |
| Employment Share$_{it}$ (%) | 61.100 | 9.097 | 37.523 | 90.217 | 632 | 2,517 |
| Employment Share by Sector: (%) | | | | | | |
| Agriculture$_{it}$ | 22.387 | 11.545 | 2.592 | 63.152 | 632 | 2,517 |
| Metallurgy$_{it}$ | 1.334 | 2.341 | 0.000 | 23.205 | 632 | 2,517 |
| Other Manufacturing$_{it}$ | 4.230 | 3.254 | 0.666 | 27.014 | 632 | 2,517 |
| Textiles$_{it}$ | 7.340 | 7.416 | 0.852 | 60.358 | 632 | 2,517 |
| Transportation$_{it}$ | 4.581 | 2.487 | 0.000 | 23.652 | 632 | 2,517 |
| **Panel B. Emigration to the US and Knowledge Exposure** | | | | | | |
| US Emigrants$_{it}$ | 100.533 | 136.299 | 0.000 | 1457.429 | 632 | 2,499 |
| Emigrant Share by US Region: (%) | | | | | | |
| Pacific$_{it}$ | 8.295 | 4.144 | 0.000 | 50.418 | 632 | 2,492 |
| Mountain$_{it}$ | 6.936 | 4.072 | 0.000 | 100.000 | 632 | 2,492 |
| West North Central$_{it}$ | 8.642 | 4.503 | 0.000 | 55.752 | 632 | 2,492 |
| East North Central$_{it}$ | 20.428 | 5.795 | 0.000 | 82.096 | 632 | 2,492 |
| West South Central$_{it}$ | 1.407 | 1.471 | 0.000 | 20.208 | 632 | 2,492 |
| East South Central$_{it}$ | 0.828 | 0.927 | 0.000 | 11.076 | 632 | 2,492 |
| Middle Atlantic$_{it}$ | 34.837 | 6.834 | 0.000 | 73.620 | 632 | 2,492 |
| New England$_{it}$ | 16.387 | 7.085 | 0.000 | 51.545 | 632 | 2,492 |
| South Atlantic$_{it}$ | 2.239 | 1.710 | 0.000 | 29.554 | 632 | 2,492 |
| Return US Emigrants$_{it}$ | 11.830 | 17.051 | 0.000 | 185.591 | 632 | 2,517 |
| Knowledge Exposure$_{it}$ | 2.295 | 3.854 | 0.000 | 67.490 | 652 | 3,779 |
| Knowledge Exposure$_{ikt}$ | 0.186 | 0.363 | 0.000 | 4.080 | 652 | 4,429 |
| **Panel C. Patents** | | | | | | |
| Total Patents$_{it}$ | 216.040 | 756.065 | 0.000 | 21000.000 | 652 | 4,424 |
| Share of Patents by Technology: (%) | | | | | | |
| Electricity$_{it}$ | 4.240 | 8.075 | 0.000 | 100.000 | 632 | 4,271 |
| Instruments$_{it}$ | 6.515 | 10.460 | 0.000 | 100.000 | 632 | 4,271 |
| Lighting and Heating$_{it}$ | 6.286 | 9.264 | 0.000 | 100.000 | 632 | 4,271 |
| Personal Goods$_{it}$ | 8.718 | 12.026 | 0.000 | 100.000 | 632 | 4,271 |
| Transportation$_{it}$ | 16.691 | 15.079 | 0.000 | 100.000 | 632 | 4,271 |
| High-Impact Patents$_{it}$ | 43.217 | 160.263 | 0.000 | 3678.000 | 652 | 4,424 |
| US Patents Similarity$_{it}$ | 5.435 | 22.751 | 0.000 | 768.088 | 652 | 4,424 |
| **Panel D. Linked Inventor-Census Sample** | | | | | | |
| Number of Patents$_{it}$ | 0.121 | 0.830 | 0.000 | 274.000 | 116,709 | 2,334,180 |
| Age$_{i}$ | 37.053 | 13.730 | 9.000 | 79.000 | 116,709 | 116,709 |
| Neighborhood Emigrants$_{i}$ | 2.276 | 4.030 | 0.000 | 53.424 | 116,709 | 116,709 |
| Non-Return Neighborhood Emigrants$_{i}$ | 2.238 | 3.962 | 0.000 | 52.677 | 116,709 | 116,709 |

*Notes*. This Table presents descriptive statistics for the variables used in the main analysis at the district-decade (Panels A–C) and individual-year level (Panel D). Demographic data in Panel A are tabulated from the population censuses of 1881–1911 and cross-walked to consistent 1891 borders. Panel B is constructed from the linked migrant sample. Panel C covers patenting between 1880 and 1930. Panel D is constructed from the sample of inventors linked to the UK population census. The subscripts report the aggregation level: Panels A–C by district-decade $it$; Panel D by individual $i$ and year $t$. Referenced on pages: 10, 12, 18.

**Table II.** Instrument Balance

| | Correlation with Instrument | | | |
|---|---|---|---|---|
| | Coefficient | Standard Error | Std. Coefficient | Mean Dep. Var. |
| | (1) | (2) | (3) | (4) |
| **Panel A. Labor-Market Variables** | | | | |
| Employment Share | 0.001 | (0.004) | 0.013 | 0.613 |
| Employment Share in: | | | | |
| Agriculture | 0.005 | (0.005) | 0.041 | 0.225 |
| Metallurgy and Chemicals | 0.000 | (0.001) | 0.016 | 0.013 |
| Other Manufacturing | 0.001 | (0.002) | 0.023 | 0.042 |
| Textiles | 0.005** | (0.002) | 0.063 | 0.073 |
| Transportation | 0.002** | (0.001) | 0.085 | 0.045 |
| **Panel B. US Emigration** | | | | |
| Share of Emigrants | 0.001 | (0.000) | 0.160 | 0.002 |
| Share of Emigrants by US Region: | | | | |
| Pacific | -0.002 | (0.003) | -0.033 | 0.094 |
| Mountain | -0.000 | (0.003) | -0.005 | 0.066 |
| West North Central | 0.007** | (0.003) | 0.159 | 0.080 |
| East North Central | 0.004 | (0.005) | 0.048 | 0.214 |
| West South Central | -0.001 | (0.001) | -0.055 | 0.014 |
| East South Central | -0.001 | (0.001) | -0.076 | 0.008 |
| Middle Atlantic | -0.004 | (0.005) | -0.058 | 0.339 |
| New England | -0.003 | (0.004) | -0.040 | 0.162 |
| South Atlantic | 0.001 | (0.001) | 0.035 | 0.024 |
| Share of Return Migrants | 0.000 | (0.000) | 0.053 | 0.000 |
| **Panel C. Parallel Trends** | | | | |
| Patenting by District: | | | | |
| Patents | 0.015 | (0.091) | 0.000 | 129.097 |
| High-Impact Patents | -0.099 | (0.096) | -0.001 | 31.236 |
| US Patents Similarity | 0.038 | (0.099) | 0.000 | 25231.035 |
| Patenting by District-Technology: | | | | |
| Patents | -0.766 | (0.476) | -0.002 | 8.315 |
| High-Impact Patents | 0.645 | (0.534) | 0.005 | 2.909 |
| US Patents Similarity | -1.017* | (0.567) | -0.000 | 1623.654 |

*Notes*. This Table presents the correlation between the instrument for knowledge exposure and pre-determined district characteristics. The unit of observation is a district, or, in the three rows of Panel C, a district-technology pair, observed at a decadal frequency between 1880 and 1930. Each row reports the coefficient (column 1), standard error (column 2), and standardized coefficient (column 3) from a regression of the lagged dependent variable, listed in each row, on the instrument for knowledge exposure. Column (4) displays the mean of the dependent variable. Specifications in Panels A and B are OLS; specifications in Panel C are PQML. Each regression includes district and decade, or district-by-technology, technology-by-decade, and district-by-decade fixed effects. Standard errors are clustered at the district level. Referenced on pages: 17, 20, 23, C37, C38.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table III.** Exposure to Innovation in the US and Innovation in the UK

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| **Panel A. Observed Knowledge Exposure** | | | | | | |
| Knowledge Exposure | 0.036*** | 0.052*** | 0.042*** | 0.203** | 0.400*** | 0.266*** |
| | (0.013) | (0.016) | (0.013) | (0.084) | (0.122) | (0.098) |
| **Panel B. Instrumental Variable Estimates** | | | | | | |
| *Second-Stage Regressions* | | | | | | |
| Knowledge Exposure | 0.027* | 0.048*** | 0.038*** | 0.334*** | 0.506*** | 0.411*** |
| | (0.014) | (0.018) | (0.014) | (0.094) | (0.135) | (0.112) |
| *First-Stage Regressions* | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ | 1.155*** | 1.155*** | 1.155*** | 0.730*** | 0.730*** | 0.730*** |
| | (0.159) | (0.159) | (0.159) | (0.035) | (0.035) | (0.035) |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 632 | 622 | 632 | 631 | 596 | 631 |
| Observations | 3,147 | 3,098 | 3,147 | 48,553 | 32,940 | 48,414 |
| Mean Dep. Var. | 188.682 | 47.097 | 4.310 | 11.478 | 3.954 | 0.260 |

*Notes*. This Table reports the effect of migrants' exposure to US technology on innovation in the UK. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2). Panel A reports the association between observed knowledge exposure and innovation; Panel B reports the instrumental variable estimates and the associated first-stage regression. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. The specifications are estimated using PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. The standard errors in Panel B are bootstrapped. Referenced on pages: 19, 24.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table IV.** Return Migration and Innovation in the UK

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| Knowledge Exposure ($\hat{\beta}_1$) | 0.035*** | 0.048*** | 0.041*** | 0.203** | 0.394*** | 0.267*** |
| | (0.013) | (0.016) | (0.013) | (0.084) | (0.122) | (0.098) |
| Return Knowledge Exposure ($\hat{\beta}_2$) | 0.004*** | 0.004*** | 0.004*** | 0.018* | 0.011 | 0.014 |
| | (0.001) | (0.001) | (0.001) | (0.011) | (0.015) | (0.012) |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 632 | 622 | 632 | 631 | 596 | 631 |
| Observations | 3,147 | 3,098 | 3,147 | 48,553 | 32,940 | 48,414 |
| Mean Dep. Var. | 188.682 | 47.097 | 4.310 | 11.478 | 3.954 | 0.260 |
| Standardized $\hat{\beta}_1$ | 0.096 | 0.133 | 0.115 | 0.033 | 0.071 | 0.044 |
| Standardized $\hat{\beta}_2$ | 0.098 | 0.106 | 0.101 | 0.028 | 0.018 | 0.022 |

*Notes*. This Table compares the effect of out-migration and return migration on innovation in the UK. The unit of observation is a district (columns 1–3) and a district-technology (columns 4–6). Units are observed at decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. Knowledge exposure is defined in equations (1)–(2); return knowledge exposure is defined in (7). The coefficient of "Knowledge Exposure" is labeled as $\beta_1$, and the coefficient of "Return Knowledge Exposure" is labeled as $\beta_2$. We report the standardized coefficients at the bottom of the table. These are calculated by multiplying each estimated coefficient by the standard deviation of the corresponding right-hand-side variable. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-decade, technology-by-decade, and district-by-technology fixed effects. All specifications are estimated using PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. Referenced on page: 24.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table V.** Within-Family Migration to the US and Innovation in the UK

| | Overall Emigration | | | Non-Return Emigration | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| Post × Emigrant in Family | 0.184*** | 0.148* | 0.169** | | | |
| | (0.066) | (0.080) | (0.077) | | | |
| Post × Non-Return Emigrant in Family | | | | 0.177** | 0.120 | 0.166** |
| | | | | (0.076) | (0.103) | (0.077) |
| Family FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Families | 18,903 | 21,694 | 18,825 | 17,053 | 19,747 | 16,977 |
| Observations | 964,053 | 1,106,394 | 960,075 | 869,703 | 1,007,097 | 865,827 |
| Mean Dep. Var. | 0.189 | 0.033 | 39.960 | 0.153 | 0.027 | 32.426 |

*Notes*. This Table reports the impact of migration to the US on the innovation activity of migrants' relatives who remain in the UK. The unit of observation is a family, defined as all individuals with the same surname in the same county. Units are observed at a yearly frequency between 1880 and 1930. The dependent variable is the number of patents issued to family members in the UK in columns (1) and (4); the number of patents in the top 20% importance distribution in columns (2) and (5); the text-based similarity between UK patents and US patents issued in the previous five years in columns (3) and (6). "Post × Emigrant in Family" is an indicator equal to one after the first emigrant in the family moves to the US and zero otherwise. "Non-Return Emigrant in Family" is an indicator for non-return migrants. All columns include family (i.e., surname-county) and year fixed effects. All regressions are estimated using PQML. Standard errors, reported in parentheses, are clustered at the county level. Referenced on page: 26.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table VI.** Within-Neighborhood Emigration to the US and Innovation in the UK

| | Overall Emigration | | | Non-Return Emigration | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| Post × Emigrant in Neighborhood | 0.146*** | 0.157** | 0.141*** | | | |
| | (0.042) | (0.062) | (0.044) | | | |
| Post × Non-Return Emigrant in Neighborhood | | | | 0.124*** | 0.135** | 0.120*** |
| | | | | (0.038) | (0.059) | (0.041) |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Individuals | 127,357 | 127,357 | 127,357 | 111,055 | 111,055 | 111,055 |
| Observations | 2,546,280 | 2,546,280 | 2,546,280 | 2,220,345 | 2,220,345 | 2,220,345 |
| Mean Dep. Var. | 0.121 | 0.025 | 17.750 | 0.121 | 0.025 | 17.768 |

*Notes*. This Table reports the impact of emigration to the US on the innovation activity of former neighbors in the UK. The unit of observation is an individual inventor observed yearly between 1880 and 1900. The sample includes the universe of British inventors linked to the 1891 population census. The dependent variable is the number of patents granted to the individual inventor in columns (1) and (4); the number of patents in the top 20% of the importance distribution in columns (2) and (5); the text-based similarity between UK patents and US patents issued in the previous five years in columns (3) and (6). "Post × Emigrant in Neighborhood" is an indicator equal to one after the first neighbor of the inventor moves to the US and zero otherwise. "Non-Return Emigrant in Neighborhood" is an indicator for non-return migrants. All columns include individual and year fixed effects. All regressions are estimated using PQML. Standard errors, reported in parentheses, are clustered at the neighborhood level. Referenced on page: 26.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

# ONLINE APPENDIX

## Return Innovation: The Knowledge Spillovers of the British Migration to the United States, 1870-1940

Davide M. Coluccia & Gaia Dossi

October, 2025

## TABLE OF CONTENTS

**A DATA APPENDIX** **A2**

A.I Summary of Data Sources

A.II Linked Migrants Sample

A.III Patent Data

Tables

Figures

**B ADDITIONAL RESULTS** **B28**

B.I Selection of British Emigrants

B.II Long-Run Association between Emigration and Innovation

B.III Assortative Matching of Emigrants in the United States

B.IV Anecdotal Evidence of Return Innovation

Tables

Figures

**C ROBUSTNESS ANALYSES** **C36**

C.I District-Technology Instrumental Variable Framework

C.II Robustness Analyses: Main Results

C.III Robustness Analyses: Mechanisms

Tables

Figures

# A DATA APPENDIX

This section describes the data sources and the construction of the final sample (Appendix A.I), the linking of British immigrants in the US to the UK census (Appendix A.II), and the digitization of the UK patent data and their linking to the census (Appendix A.III).

## A.I Summary of Data Sources

### *A.I.1 Patent Data*

US patent data are from Berkes (2018), who digitizes the universe of patents granted between 1836, when the US patent and trademark office was established, and 2010. In this paper, we are interested in the CPC technology class, the issue year, and the coordinates of residence of each inventor. We then assign each patent to US counties at 1900 borders. Depending on the number of inventors, a single patent may be assigned to multiple counties. In the case of patents with more than one inventor, we weigh each by the inverse of the number of inventors to avoid multiple counting. English and Welsh patents after 1900 are available from the European Patent Office. To construct our dataset, we leverage bulk access to the PATSTAT dataset. Information contained in PATSTAT includes the CPC class and the issue year. To retrieve the location of each inventor, we merge the PATSTAT data with the PatCity repository, which contains geo-coded information on the universe of English and Welsh patents during this period (Bergeaud and Verluise, 2024). Data before 1900 are not available. In Appendix A.III, we describe how we digitize the universe of patent documents issued over the period 1853–1899 to fill this substantial gap.

Importantly, we map 3-digit CPC classes to a coarser taxonomy of classes. To do that, we reduce them to functional units using the CPC classification scheme. The scheme is publicly available at the following `link`. To accommodate the historical context, we divide the transportation category into two classes: "Transporting", which includes carriages, railways, and cars, and "Ships and Aeronautics". Moreover, we conflate the "Weapons and Blasting" and the "Mining" classes into the "Metallurgy" category because few patents were observed in those industries. We further augment patent data by defining a measure of "innovativeness" or "importance" following Kelly et al. (2021). This metric assigns higher importance to patents whose text is relatively dissimilar from the existing stock of patents at the time of grant but more similar to that of subsequent patents. We compute this metric using the abstracts of patents granted after 1900. We apply this sample restriction for consistency: in 1853–1899, we observe the full text of patents instead of their abstract.

### *A.I.2 Migration Data*

Disaggregated data on the origin of English and Welsh immigrants–and, more generally, of all other nationalities–do not exist as neither US authorities nor those in the sending countries collected them. As a result, we lack precise information on where British immigrants in the US came from *within* the UK. We fill this gap and link the individual-level UK and US censuses, as described in Appendix A.II. In the US Census, we observe virtually the universe of British emigrants to the US between 1870 and 1930. For these individuals, we observe all information contained in the US census and, after merging it with their records in the UK census prior to emigration, those recorded there as well, including their place of origin. As we discuss in more detail later, we also link return migrants. Since the last publicly

available UK census dates to 1911, we can only construct return migration flows for the period 1870–1910.

### *A.I.3 Data Constructed from the Population Censuses*

The main data sources we use are individual-level, non-anonymized UK and US population censuses. The US census features prominently in the economic history literature as a major source of detailed microdata, and we therefore do not discuss it further here (Ruggles et al., 2021). The UK census is relatively less well-known (Schurer and Higgs, 2020). Although not as rich as its US counterpart, the UK population census covers individuals residing in the UK since 1850. The first census was conducted in 1841, but only 1851, 1861, 1881, 1891, 1901, and 1911 are fully digitized.[1] The census records include name and surname, birth year, division, county, district, parish, precise address of residence, detailed occupation (coded using HISCO), and additional variables not used in this paper, such as dwelling characteristics and fertility information. We augment these data by geo-coding the universe of addresses that appear in the census to precise geographical coordinates.

### *A.I.4 GIS Shapefiles & Boundary Harmonization*

Patents are mapped to 1900 registration district borders using historical GIS files and their geographic coordinates.[2] All data from the population censuses appear at historical borders. Registration districts have not undergone major boundary changes over the period that we studied. Nevertheless, to ensure consistent geographic units over time, we adapt the method proposed by Eckert et al. (2020) to UK registration districts.

Constructing geographical crosswalks using this method requires assuming that variables are evenly distributed over the area of each geographical unit. The crosswalk is obtained by overlapping geographical units across time. Suppose that unit $x$ in decade $d$ is split, with 80% of its territory remaining in $x$ and 20% assigned to another district $y$. To construct a crosswalk to period $d + t_2$ for a generic variable between decades $d - t_1$ and $d + t_2$, for $t_1, t_2 > 0$, one needs to multiply the variable measured in district $x$ in $d - t_1$ by $4/5$ and add $1/5$ of the variable in $x$ to that measured in $y$ in the same decade. We map registration districts to their boundaries in 1901. Less than 5% of the overall area of England and Wales is reassigned in this way. We adopt the same methodology to map counties to their 1900 borders.

### *A.I.5 Geo-referencing the Historical British Census*

A notable feature of the UK census is that it contains precise information on the residential address of the universe of the British population. This information is extremely valuable because, in principle, it assigns the finest possible location to each individual. In practice, however, it is highly non-standardized and challenging to use. In this paper, we expand earlier work by Lan and Longley (2019), who analyze the 1901 census and geo-reference the 1851-1911 censuses. The geo-coded census sample is used in the neighborhood individual-level analysis, while all other results do not rely on these data.

---

[1] The 1921 census is currently being digitized and is partially available. We do not use it because its coverage is still not complete and because it is not available in bulk. All censuses after 1921 are subject to privacy restrictions.

[2] `GIS` data for the US are provided by `NHGIS`, while district boundaries have been digitized by the Great Britain Historical `GIS` Project.

There are two ways to geo-reference historical addresses. The first is to manually digitize historical locations, either streets or enumeration units, from historical maps. However, this method does not scale up and becomes rapidly unfeasible as the data grows. The second, automated approach uses text-based address matching between historical sources and modern or historical address databases that have already been georeferenced. We follow this latter method, as we need to geo-reference 5,464,578 unique addresses.

To perform the actual matching, we first carry out a preliminary manual cleaning of addresses. First, we remove house numbers because they undergo many changes and re-sequencing over time. Second, we remove uninformative locations, such as "village", "farm", and "rectory". Then, we input the resulting addresses as queries into the geo-referencing engine (`MapTiler API`). We discard the match if the resulting coordinates are not within the parish's boundaries where the address is recorded. This consistency check is necessary because homonyms are frequent. Since observing two addresses with the same name within a given parish is extremely rare, this procedure minimizes spurious matches.

We use the geo-coded 1891 census in the neighborhood analysis. We geo-code 85% of the addresses listed in the census, covering 86% of the population. The geo-coding ratio is relatively homogeneous across districts, with the exception of Wales, where many addresses are reported in Welsh. In one robustness exercise, we exclude Wales from the sample and confirm that the results hold.

## A.II Linked Migrants Sample

### *A.II.1 Data Sources*

We rely on two externally compiled data sources.[3] For the US, we use the IPUMS full-count, non-anonymized census (Ruggles et al., 2021). A census was taken in the US every ten years starting in 1790.[4] Until 1840, this was run at the household level. From 1850 onward, we have detailed *individual*-level information for the universe of the US population.[5] For confidentiality, these data are available until 1940. Our dataset contains snapshots of the entire US population at any given decade between 1850 and 1940, although we focus on the years 1870-1930 for the purpose of this paper. Because we use the non-anonymized version of the IPUMS data, we observe each individual's recorded name and surname in addition to the publicly available information.

In the UK, the I-CeM data mirrors the IPUMS content (Schurer and Higgs, 2020). More precisely, it contains information on the universe of individuals living in England, Scotland, and Wales. Similar to the US, the census was conducted at decadal frequency from 1851 to 1911, with no census taken in 1871. As with the IPUMS data, we use the full-count, non-anonymized version of the dataset. In addition to publicly available information, this includes full names and addresses for the universe of individuals living in the UK at each census year.

[3] We are deeply thankful to IPUMS and I-CeM for allowing us access to their confidential data. Without their help, this paper would not have been possible.

[4] The individual records of the 1890 census have been lost in a fire.

[5] By US population, we refer to the universe of individuals who *lived* in the US at a given point in time.

### *A.II.2 Linking Algorithm*

Our methodology follows Abramitzky et al. (2021). As neither the US nor the UK–nor any other European countries–recorded where British immigrants came from *within* the UK, we match British immigrants residing in the US with their entry in the UK census, which records where they come from at a granular geographical level.[6] More precisely, we take the stock of British immigrants residing in the US in a given census year–say, 1900–and match them with their entry in the preceding UK census–in this case, 1891.[7] This implies that we measure the *flow* of British immigrants over time rather than their stock.

We use three variables to link individuals: first name, surname, and birth year. The baseline linked sample consists of individuals who report England or Wales–or analogous denominations, such as Great Britain–as their country of origin in the US census. In the 1900 census, we take all those who immigrated between 1870 and 1899. In the subsequent censuses, until 1930, we retrieve the stock of those who immigrated in the preceding decade. Then, to match each unit in the sample–call the generic one $A$–to an entry in the UK census, we perform these steps:

1. Take the census that precedes the immigration year of $A$. Hence, for instance, we match all those who immigrated in 1896 to the 1891 census.

2. Select all records in that census with the same reported birth year and with the same initial name and surname letter as $A$. Call the resulting sample $\mathcal{M}^A = \{m_1^A, \ldots, m_N^A\}$.[8]

3. Compute a string-similarity measure between the name and surname of $A$ and that of all elements of $\mathcal{M}^A$. In other words, for every $m_i^A \in \mathcal{M}^A$, compute $d_N(A, m_i^A)$ as the Jaro-Winkler similarity between $i$'s and $m_i^A$'s name, and $d_S(A, m_i^A)$ as the surname Jaro-Winkler similarity.[9]

4. The set of candidate matches is defined iteratively. First, we select the matches within $\mathcal{M}^A$ with the highest surname similarity with $A$:

$$\overline{\mathcal{M}}^A = \left\{ m_i^A \in \mathcal{M}^A \mid d_S^A(A, m_i^A) = \max_{m_{i'}^A \in \mathcal{M}^A} d_S^A(A, m_{i'}^A) \right\}. \tag{A.1}$$

   Then, we refine the set of candidate matches by selecting the elements with the highest name similarity with $A$:

$$\overline{\mathcal{M}}^A = \left\{ m_i^A \in \overline{\mathcal{M}}^A \mid d_N^A(A, m_i^A) = \max_{m_{i'}^A \in \overline{\mathcal{M}}^A} d_N^A(A, m_{i'}^A) \right\}, \tag{A.2}$$

[6] Since women usually changed their name upon marriage, we cannot match them. This is a common problem in linking algorithms (Abramitzky et al., 2021).

[7] Since no census was conducted in the UK in 1871, we link the 1880 US census to the 1861 UK one. This is not overly problematic because we can still match all those aged ten or older in 1871.

[8] Restricting $\mathcal{M}^A$ to only include records with the same initial name and surname letter as $A$ is recommended by Abramitzky et al. (2021). It is computationally advantageous and unlikely to significantly affect the performance of the matching algorithm because errors in names and surnames are more common at the end than at the beginning.

[9] We cannot match on the exact same name and surname because coding errors are common in historical census data. Restricting to exact matches would thus reject possible true positive matches.

implying that we restrict the set of possible matches to include only those with the highest name similarity among those with the highest surname similarity with record $A$.

5. Finally, for a given threshold couple $\boldsymbol{\tau} = (\tau_1, \tau_2) \in [0,1]^2$, we select only the candidate matches whose similarity score is above $(\tau_1, \tau_2)$. The set of accepted matches thus boils down to:

$$\widetilde{\mathcal{M}}^A_{\boldsymbol{\tau}} = \left\{ m_i^A \in \overline{\mathcal{M}}^A \;\middle|\; d_N(A, m_i^A) \geq \tau_1 \;\wedge\; d_S(A, m_i^A) \geq \tau_2 \right\}. \tag{A.3}$$

Clearly, $\widetilde{\mathcal{M}}^A_{\boldsymbol{\tau}}$ can be empty, meaning that $A$ has no accepted matches. It can have one element, in which case we refer to it as a "single match," or it can have multiple elements. In our baseline exercise, we set $\tau_1 = \tau_2 = 0.95$ to ensure that the sample only includes high-quality matches.

We evaluate the distance between two strings $A$ and $j$ in terms of their Jaro-Winkler similarity $d(A, i)$:

$$d(A,i) \equiv \widehat{d}(A,i) + \ell p \left[1 - \widehat{d}(A,i)\right], \tag{A.4}$$

where

$$\widehat{d}(A,i) \equiv \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|A|} + \frac{m}{|i|} + \frac{m-t}{m}\right) & \text{else} \end{cases}, \tag{A.5}$$

where $m$ is the number of matching characters, $|A|$ is the length of string $A$, $t$ is half the number of transpositions, $\ell$ is the length of the common or eventual common prefix no longer than four characters between $A$ and $i$, and $p = 0.1$ is a constant scaling factor. Two characters are matching only if they are the same and are not farther than $\left\lfloor \frac{\max(|A|,|i|)}{2} \right\rfloor - 1$. Half the number of matching characters in a different sequence order is the number of transpositions.[10] The Jaro-Winkler distance has been shown to perform well in linking routines (Abramitzky et al., 2021). In our particular case, this metric outperforms simpler string dissimilarity metrics because coding errors are far more frequent at the end of names and surnames than at the beginning.

### *A.II.3 Matching Rate*

Figure A.1 reports the key matching rate statistics of the linked sample. In particular, Figure A.1a displays the crude matching rate over time, i.e., the share of English immigrants recorded in the US census linked to at least one record in the UK census. Since it is impossible to link British immigrants born after the last census before they migrated (i.e., are born and migrate within two census years), we distinguish between the full sample in blue and the "matchable" sample in red. The matching rate in the matchable sample is approximately 53%. It declines from 65% in the 1870s to 50-55% in the 1890s and remains constant after that. The large jump in 1880 in the crude matching rate is because the 1871 census is unavailable; hence, between 1870 and 1881, we cannot match anyone younger than 20–as opposed to 10 in the rest of the sample–when they migrated. The matching rate is relatively high compared to other algorithms, such as Abramitzky et al. (2021). In the next section, we thus ask whether matched and unmatched immigrants are balanced on observables. Figure A.1b replicates the previous plot but reports the absolute number of emigrants in the sample and those linked instead of the percentage shares.

[10] The Jaro-Winkler distance has been recently employed in the economic history literature for intergenerational linking purposes by, among others, Abramitzky et al. (2021)

Figure A.1c displays the distribution of linked immigrants by number of matches. A large fraction ($\approx$ 45%) are paired with a single entry in the UK census. Approximately 60% are matched to one or two entries. Beyond this threshold, multiple matches are more common, and the average number of matches is 16. To ensure that multiple matches do not confound our analysis, we always weigh emigrants by the inverse of the number of matches. Additionally, we check that the baseline results remain qualitatively unchanged when we include only single matches in the dataset.

In Figure A.2, we report the distribution of the name and surname similarities between the records of British immigrants in the US census and their links in the UK census. The dashed black lines plot the threshold below which we reject the link (.95). In the sample, we find that if an immigrant is linked to at least one record in the UK census, then the quality of the match is high, as more than 80% of the sample lies right to the .95 line. However, it is important to note that for a link to be accepted in the final sample, we require that *both* the name and the surname similarities be above 0.95.

### *A.II.4 Internal Consistency: Natural Language Processing Applied to Census Linking*

We validate the internal consistency of the linked migrants sample using manual and semi-automated methods. The spirit of these exercises is to verify that the matches produced by the linking algorithm would be delivered by an expert human agent operating with the same information available to the algorithm. We adopt three methods: a manual validation on a random subsample of the data, a validation based on a classification constructed by a state-of-the-art large language model (`GPT-5`) on the same random subsample, and a validation of the entire dataset using a bidirectional text representation model (`BERT`) trained on the manually validated random subsample. We now describe these three approaches in detail.

***Manual Validation*** The scale of the dataset precludes a manual verification of all the matches. We therefore take two random samples of 1,000 observations each. The first only comprises matches that we accept according to the Jaro-Winkler classification rule explained above, i.e., all the elements in the set $\widetilde{\mathcal{M}}^A_{0.95}$. The second only comprises matches that we reject according to the same classification rule, i.e., all the elements in the set $\overline{\mathcal{M}}^A \setminus \widetilde{\mathcal{M}}^A_{0.95}$. Our first validation approach is to manually verify that the matches in $\widetilde{\mathcal{M}}^A_{0.95}$ are indeed plausible, and that the matches in $\overline{\mathcal{M}}^A \setminus \widetilde{\mathcal{M}}^A_{0.95}$ are indeed to be rejected. Naturally, we expect some classification error, so we calculate the share of true positives and the share of false negatives as statistics to evaluate the overall performance of the algorithm.

Table A.1a (column 1) reports the results. Within the sample of 1,000 accepted matches (Panel A), we confirm 977. This figure corresponds to a 98% share of true positive matches. Panel B reports the corresponding exercise on the subsample of rejected matches. We find that out of 1,000 matches rejected by our algorithm, a human agent would have rejected 825, and 175 would have been accepted. These figures correspond to a 17% false negative rate. The spirit of setting a 0.95 threshold for the Jaro-Winkler classifier, as in Green (2025), is to minimize the share of false positives, thereby retaining only high-confidence matches. This choice is reflected in the 98% true positive rate. However, we find that this has only a moderate impact on the recall rate. Out of 977+175 = 1,152 true matches in the two datasets, our classifier correctly identifies 85%. We thus estimate a 175/1,152 = 15% share of missed matches.

***Semi-Automated (GPT) Validation*** A possible concern with this strategy is that the manual verification is conducted by a human researcher and is thus subject to cognitive biases. To address this issue, we adopt a semi-automated approach and employ GPT-5 to establish the matches in place of the human researcher using the prompt below.

**GPT-5 Prompt for Record Linkage Verification**

You are a native English speaker with expertise in record linkage. Consider the following two individual names:

- name_in_us_census
- name_in_uk_census

Determine if they refer to the same person, taking into account:

- Phonetic similarities despite minor spelling differences,
- Common abbreviations, nicknames, and misspellings,
- Possible OCR or transcription errors,
- If only one record has a middle name or initial middle name, disregard it. If both have it, consider it.

Scoring rules:

- 1 = Certainly or almost certainly the same person: exact match or very minor variation,
- 0 = Otherwise.

Additional guidelines:

- First priority is to minimize false positives,
- Output only valid JSON, with no explanation or extra text.

Respond strictly in this format: {{ score: < 0 | 1 > }}

We prompt the language model to ensure that it has the same objective function as the researcher, i.e., to minimize the probability of delivering false positive matches.

Column (2) of Table A.1a reports the results. The GPT-powered classifier confirms 978 matches out of the 1,000 accepted by the Jaro-Winkler algorithm, which corresponds to a 98% accuracy rate. The LLM classifier thus exactly replicates the accuracy derived through manual classification. In Panel B, we find that GPT would have accepted 356 out of the 1,000 matches discarded by the Jaro-Winkler classifier, yielding a false-negative rate of 36%. This figure is larger than the 17% we obtained through manual verification, indicating that the LLM is plausibly more lenient than the researcher. However, the Jaro-Winkler classifier still displays a 978/(978+356) = 1,334 = 73% recall rate when considering the GPT-classifier as the ground truth. This is a highly satisfactory performance, given that maximizing

the accuracy rate is the main objective.

In principle, it is possible that while the manual and the GPT classifiers give similar accuracy rates, they are actually confirming and rejecting different matches. To check that this is not the case, in Table A.1b, we display the confusion matrix of the manual and the GPT classifiers on the subsample of accepted (columns 1–3) and rejected (columns 4–6) matches. A confusion matrix tabulates the joint distribution of predicted and actual classifications, displaying the number of instances correctly and incorrectly assigned to each category. It separates correct identifications (true positives and true negatives) from misclassifications (false positives and false negatives). The confusion matrix indicates that the human and the GPT classifiers are essentially concordant on the accepted-matches sample (columns 1–3). On the rejected-matches sample (columns 4–6), they reject and accept the same observations, except that the GPT classifier accepts 189 matches that the manual validation rejects. Overall, however, the two classification approaches yield substantially aligned predictions.

***Full Sample Automated Validation*** These validation approaches are performed on a 2,000 (stratified) random sample of the linked dataset. To validate the entire dataset, we employ a bidirectional text representation model (BERT).[11] The BERT model yields vector representations of the names in the US and the UK censuses. Then, we compute the cosine similarity between two vectors to measure how close they are, and consider them a match if their similarity is above a certain threshold $\tau^*$. This threshold is thus the only hyperparameter of the BERT-based classifier. To obtain $\tau^*$, we take an 80% random sample of the 2,000 matches that we annotated manually for the previous validation exercise. We select $\tau^*$ to maximize the F1 score of the BERT classifier on this training set.[12]

To evaluate the performance of the BERT classifier, we compare it to the manual and the GPT classifiers on the test set, i.e., the 20% random sample that we annotated manually and using GPT that we did *not* use to compute $\tau^*$. Columns (3–5) of Table A.1a report the results. Panel A of columns (3) and (4) reports the predictions of the manual and the GPT classifiers, whereas column (5) displays the BERT classifier, on the sample of accepted matches according to the Jaro-Winkler classifier. All models deliver a true positive rate in excess of 95%, with BERT yielding a slightly higher number of false positives (6 out of 182). Panel B replicates the exercise on the sample of rejected matches. In this case, BERT yields the lowest rate of false negative matches, which is less than 50% of that of GPT. Panel B of Table A.1b reports the confusion matrix comparing the manual classification outcomes with those obtained from BERT. The two methods deliver virtually the same classification outcomes on both samples of accepted (columns 1–3) and rejected (columns 4–6) matches.

Since the performance of the BERT-based classifier is nearly identical to that of the manual classifier, we can scale it up to the full dataset to evaluate the predictions of the Jaro-Winkler linking algorithm on the entire sample of linked migrants. Table A.2 reports the results. Column (1) shows the number of matches that are accepted by the Jaro-Winkler classifier that the BERT classifier rejects (first row) and accepts (second row). Out of 359,757 accepted matches in the dataset, the BERT classifier accepted

[11] Technically, we employ `all-MiniLM-L6-v2`, a distilled sentence-BERT from the `SentenceTransformers` library with six transformer layers and a 384-dimension embeddings. This model is pre-trained on one billion sentence pairs and fine-tuned for semantic similarity tasks.

[12] The F1 score of a classifier is the harmonic average of its precision and recall. We BERT classifier attains a 0.91 F1 score on the training set.

339,135. The implied share of true positive links (taking the outcome of the BERT classification as the ground truth) generated by the Jaro-Winkler classifier is 94%. Column (2) reports the matches that are rejected by the Jaro-Winkler classifier. Out of 202,255 rejected matches, the BERT classifier rejects 182,162 and accepts 20,093, corresponding to a true negative rate of 90% and a false positive rate of 10%. Overall, the Jaro-Winkler classifier thus correctly identifies 94% of 339,135+20,093=359,228 true matches.

### *A.II.5 Balance of Linked Emigrants*

Table A.3 reports the correlation between the probability that a British immigrant in the US is linked to an entry in the UK census and a set of individual-level characteristics observed in the US census. Column (1) reports the sample value of each variable for unmatched immigrants, and column (2) refers to matched immigrants. Column (3) reports the difference between the two groups. The linking probability does not correlate with either literacy, for which one may expect a correlation since educated immigrants could present systematically more precise census records, or income, another proxy for educational standing. Matched immigrants are less likely to work as professionals, in clerical occupations, or as sales workers. These differences, however, are small in magnitude and affect a small portion of the population. On the contrary, manufacturing workers (skilled "craftsmen" and unskilled "operatives") account for almost 60% of the population and do not appear to be selected on their linking probability. In terms of their region of residence, emigrants living in the Northeast are more likely to be linked, although the difference is small (less than 2%). They are also less likely to reside in the South, but again, this difference affects only approximately 1% of the sample. Overall, Table A.3 provides reassuring evidence that linked emigrants seem not to be selected on observable characteristics, as any statistically significant differences are small in magnitude.

### *A.II.6 External Validation*

It is challenging to validate our sample with external data because, to the best of our knowledge, primary data on the origin of British emigrants within the UK does not exist. Baines (2002) attempts to estimate decade-level data for British counties between 1880 and 1910. To the best of our knowledge, these estimates are the only alternative source we can use to gauge the plausibility of our data. They nonetheless present important limitations. First, they are based on interpolations from population data tabulated from the population census. Second, the author tries to account for internal migration, which remains an important confounding factor. Third, the estimates refer to aggregate emigration outflows, whereas our data focuses only on emigration to the US. Fourth, the estimates are computed at the county level, which are coarse geographical units in the UK.

Keeping these limitations in mind, in Figure A.3, we aggregate our data at the county-decade level and plot it against the estimates produced by Baines (2002). Both series are shown in logs to reduce the influence of extreme values, and the data is weighted by county population. We find a positive and statistically significant correlation between our data and the Baines series. Our numbers are generally lower than those provided by Baines, which plausibly reflects that our data do not cover emigration to countries other than the US. Overall, we view this graph as providing supporting evidence of the plausibility of our dataset.

### *A.II.7 Return Migration Data*

Following the same linking algorithm described before, we construct a linked sample of return migrants. This identifies English and Welsh immigrants in the US in year $t$ and looks for possible matches in the UK census in year $t + 10$ using the algorithm previously described. Since the last available UK census is in 1911, we face a hard upper bound for the coverage of return migration, as we can only construct return migrants linked samples spanning the period 1870–1910.

## A.III Patent Data

### *A.III.1 Motivation*

Despite its historical significance, we lack comprehensive patent data for the Second Industrial Revolution period (1850–1900) in the United Kingdom. In particular, it is impossible to reconstruct the geographical distribution of innovation activity during this period. This data limitation sharply contrasts the effort undertaken to document. While data on UK patents between the inception of the English patent law in 1617 up until the end of the First Industrial Revolution in the 1840s (Nuvolari and Tartari, 2011; Nuvolari et al., 2021) is available in digitized form, we lack comprehensive patent data for the Second Industrial Revolution period (1850–1900). In particular, it is impossible to reconstruct the geographical distribution of innovation activity during this period. To fill this gap, we digitize the universe of English and Welsh patents for the period 1853- 1900. These data contain detailed information on the text, geographical location, inventors' personal information, and grant date.

### *A.III.2 Data Sources*

We access restricted full-page scans of original patent documents granted in England and Wales between 1617 and 1899 from the UK Intellectual Property Office. This paper focuses on the period 1853-1899 for two reasons. First, Nuvolari and Tartari (2011) already digitized patents before 1853 from Bennet Woodcroft's index, although patent documents contain additional information compared to the index. Second, in 1853, a reform dramatically lowered patent application prices. This makes it challenging to compare patents before and after the reform. Patent documents contain a wealth of unstructured information. We provide two examples in Figure A.5: Figure A.5a shows the patent granted to Henry Bessemer for the eponymous process to produce steel, and Figure A.5b displays the patent granted to John Starley for the first modern safety bicycle. Both patents are in our dataset. The rectangles identify the location of the textual data that we extract. These include (i) a short title, (ii) a long title, (iii) the author(s)'s name(s), (iv) the author(s)'s address(es), (v) the author(s)'s profession(s), (vi) the filing date, (vii) the issue date, (viii) the type of protection, (ix) an indicator of whether the application was filed by an agent on behalf of someone living abroad, and (x) the full text of the patent. Not all (i-x) are available throughout the sample. In particular, (i), (vi), and (viii) are available only until 1873. After that date, a short title is no longer reported, the filing date is reported only sporadically, and the type of protection becomes uninformative as only granted patents are included in the sample.

### *A.III.3 Digitization*

We perform optical character recognition (OCR) on each patent to structure the data in a machine-readable dataset. To ensure state-of-the-art performance, we OCR the first page of each document, where all the (i–ix) variables are located, using Amazon's commercial `textract` engine. To retrieve the rest of the text not used in this paper, we use the open-source engine `tesseract`. An OCR-ed document is a text file. To extract the relevant variables, we harness the flexibility of large language models–specifically, `GPT 3.5`–to parse all variables (i–ix). The text of patents is relatively standardized, but inconsistencies arise relatively frequently due to idiosyncratic writing, phrasing, and OCR errors. The flexibility of large language models allows us to consistently and precisely extract all the relevant information for a large (95%) share of all patent applications. Running the same model on the title of each patent, we assign the technological class to patent grants. The final database comprises approximately 800,000 patents granted between 1853 and 1899.

### *A.III.4 Geo-Coding*

To retrieve each patent's location, we geocode each inventor's listed address using the commercial geocoding engine provided by MapTiler AG. To geocode an address, if a coarse geographical unit is listed on the patent (e.g., the county), we condition the outcome coordinates to lie within that unit. Figure IIb reports the resulting distribution of patents per capita and Figure A.7 reports the spatial distribution of patent grants across technologies. These maps are consistent with underlying population and economic development indicators and with the historical evidence (e.g., there is a substantial clustering of textile patents in the Lancashire districts).

### *A.III.5 External Validation*

We validate the data with the two existing series covering a portion of the period 1853-1899. Hanlon (2016) digitized an index of patents issued between 1855 and 1883. This data list, for each patent, the inventor(s) and their profession(s), the technology class, and the issue year. The second dataset is the "A Cradle of Invention" (COI) series, published by Finishing Publications (2018). These data, too, were digitized from indices and therefore only list authors, issue year, and, often, titles. In principle, this series spans the years 1617–1895. However, after 1883, patent applications that were eventually denied protection were also listed. Absent a way to identify granted patents, we do not report statistics for the COI series after 1883.

Table A.4 reports the aggregate number of patents issued and recorded in our series (columns 2 and 6), COI (columns 3 and 7), and Hanlon (2016) (columns 4 and 8). The three are highly consistent. Our series is closest to Hanlon (2016), but the COI figures are not too far off either. The table strongly suggests that our series is as complete as the Hanlon (2016) database. We cannot, however, externally validate it for the later part of the period because there is no data available.

### *A.III.6 Measuring the Similarity Between US and UK Patents and their Importance*

This section describes the construction of the patent similarity metric to measure "copying" and "originality" of UK innovation. The approach builds on Kelly et al. (2021). We adapt their methodology to our context by using information from patent titles only. Although we do not have access to the

full text of US patents, titles are typically highly informative about technological content. We previously showed that a title-based machine learning algorithm predicts the technological classification of the patent with nearly 90% accuracy. Titles for UK patents are embedded in the digitized text for 1870–1899 and collected from PATSTAT for later years, while titles for US patents are collected from PATSTAT for the full sample period.

We define the backward inverse-document frequency associated with each word $w$ as the inverse frequency with which the word $w$ appears in US patents $p$ issued until year $t$. Formally, we have

$$\text{BIDF}_{w,t} \equiv \log\left(\frac{\text{Number of Patents Issued Before } t}{1 + \text{Number of Patents Issued Before } t \text{ that contain word } w}\right) \tag{A.6}$$

Then, to each patent-word pair, we associate the term frequency variable that counts the number of instances word $w$ appears in patent $p$, normalized by the length of the patent. With a slight abuse of notation, let $p$ denote the patent's index and the set of words it contains. We then compute

$$\text{TF}_{wp} \equiv \frac{\sum_{c \in p} 1(c = w)}{\sum_{c \in p} 1(c)} \tag{A.7}$$

where the numerator returns how many times word $w$ appears in patent $p$, and the denominator is the number of words in patent $p$. Then, we define the TF-BIDF associated with word $w$, patent $p$ at time $t$ as the product between these two terms:

$$\text{TF-BIDF}_{wp,t} \equiv \text{TF}_{wp} \times \text{BIDF}_{w,t} \tag{A.8}$$

where the vector $\text{TF-BIDF}_{p,t}$ collects the term frequency-backward inverse document frequency for all words $w$ in $p$. For comparability, the vector $\text{TF-BIDF}_{p,t}$ is normalized by its norm to have unit length.

We compute $\text{TF-BIDF}_{p,t}$ vectors for US and UK patents, but $\text{BIDF}_{w,t}$ are computed on the corpus of US patents only. Then, we compute the cosine similarity $\rho_{i,j}$ between each UK patent $i$ and each US patent $j$. This allows us to define two variables. First, we seek to measure the similarity between British innovation and previous American patents. This yields a measure of backward similarity that, for the paper's narrative, we define as "copying." Formally, define

$$\text{Backward Similarity}_i^{\tau} \equiv \sum_{j \in \mathcal{F}_i^{-\tau}} \rho_{i,j} \tag{A.9}$$

where the set $\mathcal{F}_i^{-\tau}$ denotes the set of US patents issued within $\tau$ years from the issue year of patent $i$. This measures the degree of similarity between a given patent in the UK and previous patents in the US. Second, we define a measure of "originality" of UK patents compared to previous US patents. This leverages the insight of Kelly et al. (2021) that innovative and influential patents are those that are most dissimilar from existing innovation while at the same time retaining semantic proximity with subsequent patents. Formally, we have

$$\text{Excess Forward Similarity} \equiv \frac{\sum_{j \in \mathcal{F}_j^{+\tau}} \rho_{i,j}}{\sum_{j \in \mathcal{F}_i^{-\tau}} \rho_{i,j}} \tag{A.10}$$

where $\mathcal{F}_i^{+\tau}$ denotes the set of US patents issued within $\tau$ years after the issue year of patent $i$. In the baseline analysis, we set a symmetric window of $\tau = 5$ years around each patent's issue date.

Table C.9 reports the results using an alternative threshold of ten years. The same table reports the results obtained by netting out year and technology class fixed effects at the patent level. As noted by Kelly et al. (2021), this ensures that linguistic drift over time does not contaminate the measurement of technological similarity.

### *A.III.7 Census Linking*

To perform the individual-level analysis on neighborhood emigration, we link the inventors listed on patent documents issued between 1881 and 1900 to the 1891 population census. The availability of census data determines the sample restriction. First, we aim to follow the inventing activity of individuals over time. Therefore, we link patent records to a single census. When linking inventors to the census, we use the location listed on the document to restrict the pool of potential matches in the census. This assumes that the inventor did not move between the time when the patent was granted and the closest census year, which is a plausible assumption only for patents issued not too far from the closest census year. Since census records for 1871 are not available, we restrict our attention to patents granted no earlier than ten years before the closest census, i.e., 1891.

***Matching Algorithm*** Given a patent $p$, define the set of its inventors as $\mathcal{A}_p = \{A_1, \ldots, A_{n_p}\}$. Most patents are solo-authored in this period, meaning $|\mathcal{A}_p| = 1$. Call $\mathcal{L}_p = \{\ell_1, \ldots, \ell_{m_p}\}$ the set of locations patent $p$ is associated to. Each $\ell$ is a pair of latitude-longitude coordinates. Let $\mathcal{L}_p^{\text{parish}}$ be the set of parishes associated with each coordinate. Analogously, let $\mathcal{L}_p^{\text{neighbor parishes}}$, $\mathcal{L}_p^{\text{district}}$, $\mathcal{L}_p^{\text{neighbor districts}}$, and $\mathcal{L}_p^{\text{county}}$ be the set of, respectively, neighboring parishes, districts, neighboring districts, and counties where each coordinate locates. These are progressively coarser units: parishes are contained within districts, which are nested within counties. Because we do not observe the inventor-location pair directly, we proceed as follows to match a generic $A_p$:

1. With a slight abuse of notation, let $\mathcal{L}_p^{\text{parish}}$–and, analogously, $\mathcal{L}_p^{\text{district}}$ and $\mathcal{L}_p^{\text{county}}$–denote the set of census records in each parish, district, and county within the respective sets.

2. Take all entries $i$ within the set of parishes $\mathcal{L}_p^{\text{parish}}$ that are at least 18 when the patent $p$ is filed. Let $\text{year}_i$ and $t_p$ denote, respectively, the birth year of $i$ and the issue date:

$$\mathcal{M}_{A_p}^{\text{parish}} = \left\{ i \in \mathcal{L}_p^{\text{parish}} \,\middle|\, t_p - \text{year}_i \geq 18 \right\} \tag{A.11}$$

3. For each $i \in \mathcal{M}_{A_p}^{\text{parish}}$, compute the distance between the name and surname of $i$ and that of $A_p$:

$$\delta_i^{A_p} = \alpha \times d_N(A_p, i) + (1 - \alpha) \times d_S(A_p, i) \tag{A.12}$$

for some $\alpha \in [0, 1]$. In our baseline setting, we pick $\alpha = 0.3$ as in the migrant linked algorithm to assign a larger weight to the surname. The terms $d_N(A_p, i)$ and $d_S(A_p, i)$ denote the name and surname Jaro-Winkler similarities between record $A_p$ and $i$.

4. Define the set of acceptable matches as those with the highest similarity with the given $A_p$:

$$\overline{\mathcal{M}}_{A_p}^{\text{parish}} = \left\{ i \in \mathcal{M}_{A_p}^{\text{parish}} \,\middle|\, \delta_i^{A_p} = \overline{\delta}^{A_p} \equiv \max_{i' \in \mathcal{M}_{A_p}^{\text{parish}}} \delta_{i'}^{A_p} \right\} \tag{A.13}$$

where $\overline{\delta}^{A_p}$ denotes the similarity between all elements in $\overline{\mathcal{M}}_{A_p}^{\text{parish}}$ and $A_p$. Notice that this is the same across all $i \in \overline{\mathcal{M}}_{A_p}^{\text{parish}}$.

5. Set a threshold $\tau$ such that if $\overline{\delta}^{A_p} < \tau$, $\overline{\mathcal{M}}_{A_p}^{\text{parish}} = \emptyset$, otherwise pass. As in the UK-US census-linked sample, we select $\tau = 0.95$.

6. If $\overline{\mathcal{M}}_{A_p}^{\text{parish}}$ is not empty, then inventor $A_p$ is matched to all records in $\overline{\mathcal{M}}_{A_p}^{\text{parish}}$. If empty, repeat steps 2–4 conditioning on records in the coarser matching set.

Patent data have the clear advantage of providing geographical information on the location of inventors. However, inventors are mobile, and there may be a considerable time lag between the patent grant date and the 1891 census. For these reasons, we incrementally exploit geographical information on the inventor's location. First, we look for high-quality matches within the same parish where the patent is filed. We then progressively expand the set of records by coarsening their geographic location to their neighboring parishes, districts, neighboring districts, and counties.

***Evaluation of the Matching*** In Figure A.8, we report the matching rate from this procedure. We successfully match approximately 75% of inventors in the sample, and this share remains stable over time. Almost 50% of the matches are attained at the parish level–the smallest geographical unit we consider–and about half of the remaining matches are obtained within neighboring parishes of the location indicated in the patent document. These figures indicate that mobility remained relatively low within the relatively narrow time frame around the census year we analyze.

A plausible concern is that the probability of obtaining a link is not random. This may be the case if, for instance, more successful inventors were more educated and therefore more likely to report their names correctly in the census. On the other hand, if successful inventors were relatively more mobile, we may fail at linking them because obtaining a match would require expanding the search to a national level, which would likely result in multiple potential matches and lead us to drop them. These hypotheses are difficult to test directly. Figure A.9 reports the overall distribution of the number of matches in the sample. Approximately 60% of the inventors listed on patent documents are linked to one single census record, and the share of inventors linked to more than 10 records is negligible. Indeed, restricting the analysis to single matches yields results that are very similar to the baseline. Table A.5 reports the correlation between the number of matches and a set of individual-level characteristics. Panel A displays the age and the residence divisions, while Panel B lists the occupation reported in the census. The number of matches is correlated with some of these characteristics, but the magnitude of this association is small, except for the fact that inventors living in Wales have substantially more matches in the sample. To ensure that Wales does not drive our results, we replicate the main results excluding them from the sample and confirm that the results hold.

## Tables

**Table A.1.** Validation of the Linked Migrants on Random Subsample

**(a)** Classification Report

| | Full Random Sample | | Test Set Random Sample | | |
|---|---|---|---|---|---|
| | (1) Manual | (2) GPT | (3) Manual | (4) GPT | (5) BERT |
| **Panel A. Accepted Matches (Jaro-Winker Classifier)** | | | | | |
| Rejected | 23 | 22 | 4 | 6 | 10 |
| Accepted | 977 | 978 | 178 | 176 | 172 |
| True Positive Rate | 0.98 | 0.98 | 0.98 | 0.97 | 0.95 |
| **Panel B. Not Accepted Matches (Jaro-Winker Classifier)** | | | | | |
| Rejected | 825 | 644 | 180 | 141 | 188 |
| Accepted | 175 | 356 | 38 | 77 | 30 |
| False Negative Rate | 0.17 | 0.36 | 0.17 | 0.35 | 0.14 |

**(b)** Confusion Matrix

| | Jaro-Winkler Classifier | | | | | |
|---|---|---|---|---|---|---|
| | Accepted Matches | | | Rejected Matches | | |
| | (1) Rejected | (2) Accepted | (3) Total | (4) Rejected | (5) Accepted | (6) Total |
| **Panel A. GPT Classifier (Full Random Sample)** | | | | | | |
| Rejected | 11 | 11 | 22 | 636 | 8 | 644 |
| Accepted | 12 | 966 | 978 | 189 | 167 | 356 |
| Total | 23 | 977 | 1000 | 825 | 175 | 1000 |
| **Panel B. BERT Classifier (Test Set Random Sample)** | | | | | | |
| Rejected | 1 | 9 | 10 | 170 | 18 | 188 |
| Accepted | 3 | 169 | 172 | 10 | 20 | 30 |
| Total | 4 | 178 | 182 | 180 | 38 | 218 |

*Notes*. Table A.1a reports the number of accepted and rejected matches on two random 1,000 samples (columns 1–2) and on the test set of the BERT classifier (columns 3–5). In columns (1) and (3), we report the results of the manual verification; columns (2) and (4) refer to the classification generated by GPT, and column (5) refers to the BERT classifier. Panel A reports the results on the 1,000 random sample of matches that we accept in the linked dataset following the Jaro-Winkler-based classifier described in Appendix A.III.7. Panel B refers to the other 1,000 sample of matches that we reject according to the same Jaro-Winkler classifier. Table A.1b reports the confusion matrix between the baseline Jaro-Winkler classification and the GPT (Panel A) and BERT classifiers. Panel B refers to the test set of the BERT classifier. Columns (1–3) refer to the subsample of matches accepted by the Jaro-Winkler classifier; columns (4–6) refer to the matches rejected by the Jaro-Winkler classifier. Referenced on pages: 10, A7, A8, A9.

**Table A.2.** Validation of the Linked Migrants Full Sample

| | Jaro-Winkler Classifier | |
|---|---|---|
| BERT Classifier | (1) Accepted Matches | (2) Rejected Matches |
| Rejected | 20622 | 182162 |
| Accepted | 339135 | 20093 |
| Total | 359757 | 202255 |
| True Positive Rate | 0.94 | |
| False Negative Rate | | 0.10 |

*Notes*. This Table reports the results of validating the full sample of linked migrants using the BERT classifier trained on the manually annotated random subsample. In column (1), in the first row, we report the number of matches that the Jaro-Winkler classifier identifies as valid that the BERT classifier rejects. In the second row, we report the number of matches that both the Jaro-Winkler and the BERT classifier classify as valid. In column (2), we replicate on the sample of matches that the Jaro-Winkler algorithm rejects: in the first row, we report those that are also rejected by the BERT classifier, while in the second row, we report those that are accepted by the BERT classifier. The Table also reports the share of true positive matches identified by the Jaro-Winkler classifier, when taking the results of the BERT classifier as ground truth. Referenced on pages: 10, A9.

**Table A.3.** Balance of Linked Emigrants Sample

| | Correlation with Matching Status | | | |
|---|---|---|---|---|
| | = 1 if Not matched | = 1 if Matched | Diff. | Std. Err. |
| | (1) | (2) | (3) | (4) |
| **Panel A. Individual Characteristics** | | | | |
| Literacy | 0.898 | 0.965 | 0.066 | (0.057) |
| Income | 3.146 | 3.141 | -0.005 | (0.009) |
| **Panel B. Occupation** | | | | |
| Professional | 0.052 | 0.041 | -0.011*** | (0.002) |
| Farmer | 0.087 | 0.091 | 0.004 | (0.008) |
| Manager | 0.065 | 0.059 | -0.006* | (0.003) |
| Clerical | 0.056 | 0.044 | -0.012*** | (0.003) |
| Sales | 0.060 | 0.046 | -0.014*** | (0.001) |
| Craftsman | 0.263 | 0.281 | 0.018** | (0.008) |
| Operative | 0.310 | 0.333 | 0.023* | (0.012) |
| Service | 0.050 | 0.049 | -0.001 | (0.003) |
| Laborer | 0.057 | 0.056 | -0.001 | (0.003) |
| **Panel C. Region of Residence** | | | | |
| North East | 0.518 | 0.543 | 0.025*** | (0.006) |
| Midwest | 0.278 | 0.269 | -0.009*** | (0.003) |
| South | 0.056 | 0.044 | -0.012*** | (0.001) |
| West | 0.147 | 0.144 | -0.004 | (0.007) |

*Notes*. This Table reports the correlation between the matching probability in the migrants-linked sample and a set of individual-level characteristics observed in the US census. The dependent variable equals one if the immigrant is linked and zero otherwise. We report the average value of the row variable for unmatched (column 1) and matched immigrants (column 2), as well as the difference between the two groups (column 3) along with its standard error clustered at the census year level (column 4). All row variables are indicators except income, which is the log of the occupational income score. Standard errors are clustered at the census year level and are displayed in parentheses. Referenced on pages: 26, 9, A10.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table A.4.** External Validation of Newly Digitized Patent Data

| Years 1853–1876 | | | | Years 1877–1899 | | | |
|---|---|---|---|---|---|---|---|
| Year (1) | New Data (2) | Hanlon (3) | COI (4) | Year (5) | New Data (6) | Hanlon (7) | COI (8) |
| 1853 | 2926 | | 3042 | 1877 | 4937 | 4940 | 5133 |
| 1854 | 2737 | | 2715 | 1878 | 5338 | 5333 | 5258 |
| 1855 | 2868 | 2955 | 2883 | 1879 | 5312 | 5325 | 5442 |
| 1856 | 3204 | 3102 | 3003 | 1880 | 5531 | 5509 | 5211 |
| 1857 | 3175 | 3197 | 3108 | 1881 | 5762 | 5745 | 5760 |
| 1858 | 3071 | 2999 | 2988 | 1882 | 6244 | 6233 | 6187 |
| 1859 | 3031 | 2998 | 3008 | 1883 | 6118 | 5981 | 6075 |
| 1860 | 3231 | 3190 | 3161 | 1884 | 9809 | | |
| 1861 | 3279 | 3272 | 3303 | 1885 | 8695 | | |
| 1862 | 3479 | 3486 | 3485 | 1886 | 8913 | | |
| 1863 | 3312 | 3308 | 3411 | 1887 | 9070 | | |
| 1864 | 3245 | 3257 | 3286 | 1888 | 9283 | | |
| 1865 | 3391 | 3378 | 3436 | 1889 | 10315 | | |
| 1866 | 3422 | 3452 | 3481 | 1890 | 10376 | | |
| 1867 | 3720 | 3720 | 3723 | 1891 | 10768 | | |
| 1868 | 3976 | 3984 | 3955 | 1892 | 11454 | | |
| 1869 | 3837 | 3781 | 3781 | 1893 | 11986 | | |
| 1870 | 3459 | 3405 | 3323 | 1894 | 11664 | | |
| 1871 | 3574 | 3525 | 3542 | 1895 | 12243 | | |
| 1872 | 3951 | 3967 | 4013 | 1896 | 13619 | | |
| 1873 | 4261 | 4282 | 4336 | 1897 | 14304 | | |
| 1874 | 4419 | 4491 | 4533 | 1898 | 13105 | | |
| 1875 | 4576 | 4557 | 4537 | 1899 | 13267 | | |
| 1876 | 5048 | 5064 | 5085 | | | | |

*Notes*. This Table reports the total number of patents in England and Wales between 1853 and 1899. Columns (2) and (6) report the series constructed from our novel dataset; columns (3) and (7) tabulate data from *A Cradle of Inventions* (Finishing Publications, 2018); columns (4) and (8) report data from Hanlon (2016). The *A Cradle of Inventions* series potentially stretches until 1899. However, after 1883, there is no way to distinguish between patents granted and applications. Hence, we do not report figures for these later years (Nicholas, 2014). Data from Hanlon (2016) only cover the years 1855–1883. Referenced on page: A12.

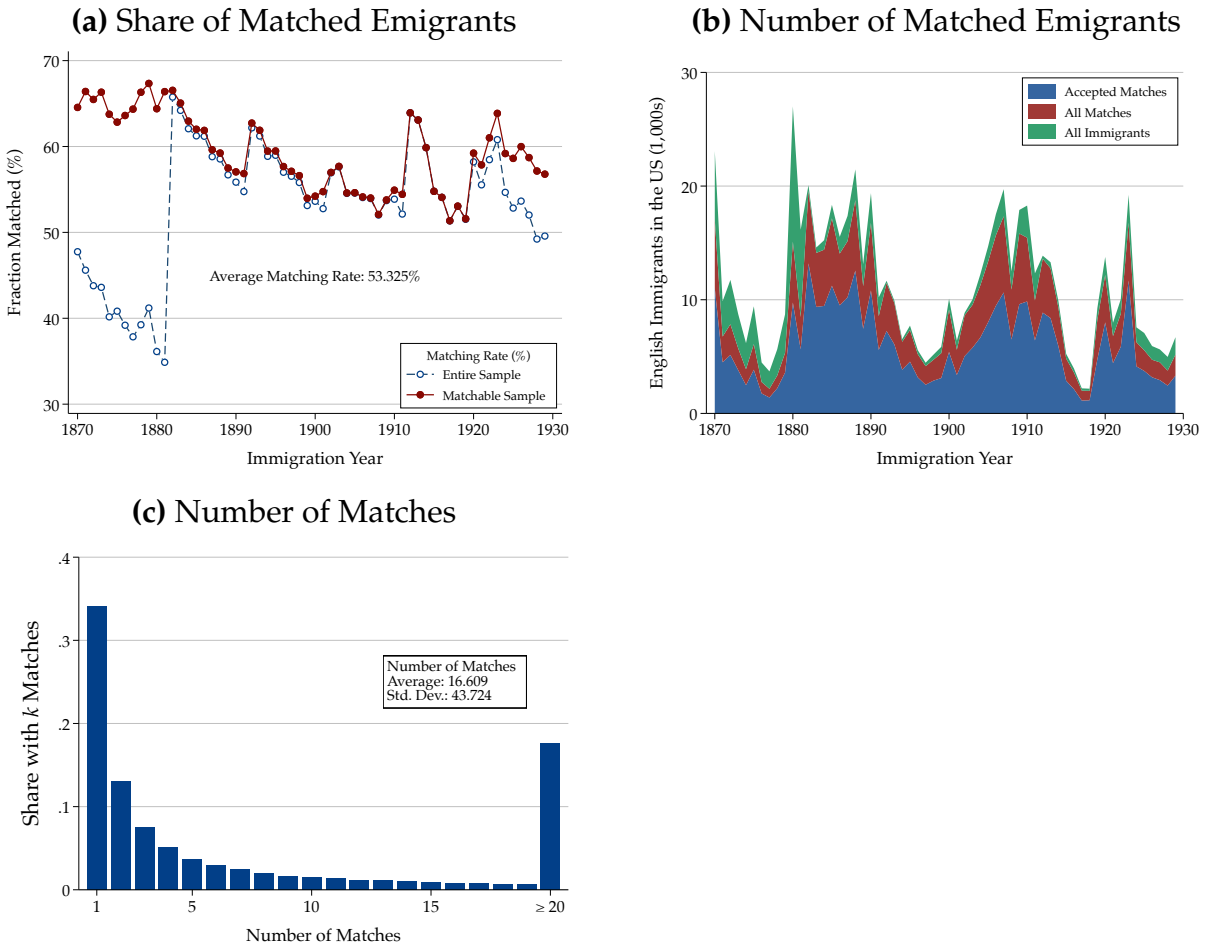
**Table A.5.** Balance of Linked Inventor Sample

| | Unconditional | | Year FE | |
|---|---|---|---|---|
| | Mean (1) | Std. Err. (2) | Mean (3) | Std. Err. (4) |
| **Panel A. Age and Place of Residence** | | | | |
| Age | -0.002 | (0.002) | -0.002 | (0.001) |
| London | -0.197 | (0.144) | -0.177 | (0.128) |
| South East | -0.362*** | (0.132) | -0.342*** | (0.114) |
| East | -0.246* | (0.149) | -0.242* | (0.136) |
| Yorkshire | -0.324** | (0.137) | -0.308** | (0.122) |
| South West | -0.412*** | (0.124) | -0.404*** | (0.114) |
| West Midlands | 0.022 | (0.199) | 0.022 | (0.186) |
| East Midlands | -0.246* | (0.135) | -0.230* | (0.119) |
| North East | -0.260** | (0.133) | -0.234** | (0.110) |
| North West | -0.165 | (0.159) | -0.169 | (0.154) |
| Wales | 1.809*** | (0.227) | 1.752*** | (0.186) |
| **Panel B. Occupation** | | | | |
| Agriculture | 0.528 | (0.365) | 0.494 | (0.320) |
| Chemicals | -0.209** | (0.105) | -0.189** | (0.083) |
| Construction | -0.061 | (0.076) | -0.055 | (0.068) |
| Engineering | -0.070 | (0.062) | -0.063 | (0.052) |
| Liberal Professions | -0.186*** | (0.071) | -0.180*** | (0.064) |
| Metallurgy | 0.114 | (0.081) | 0.109 | (0.075) |
| Public Administration | -0.150 | (0.094) | -0.148* | (0.089) |
| Textiles | -0.119 | (0.093) | -0.109 | (0.083) |
| Trade | -0.164* | (0.084) | -0.155** | (0.072) |
| Transports | 0.016 | (0.027) | 0.018 | (0.024) |
| Utilities | -0.178** | (0.080) | -0.170** | (0.071) |

*Notes*. This Table reports the correlation between the number of matches in the linked inventor sample and a set of individual-level co-variates observed in the census. In columns (1–2), we display the unconditional correlation and the associated standard error. In columns (3–4), we repeat the exercise but control for the issue year of the patent. Standard errors are clustered at the county level and are displayed in parentheses. Referenced on pages: A15, C41.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

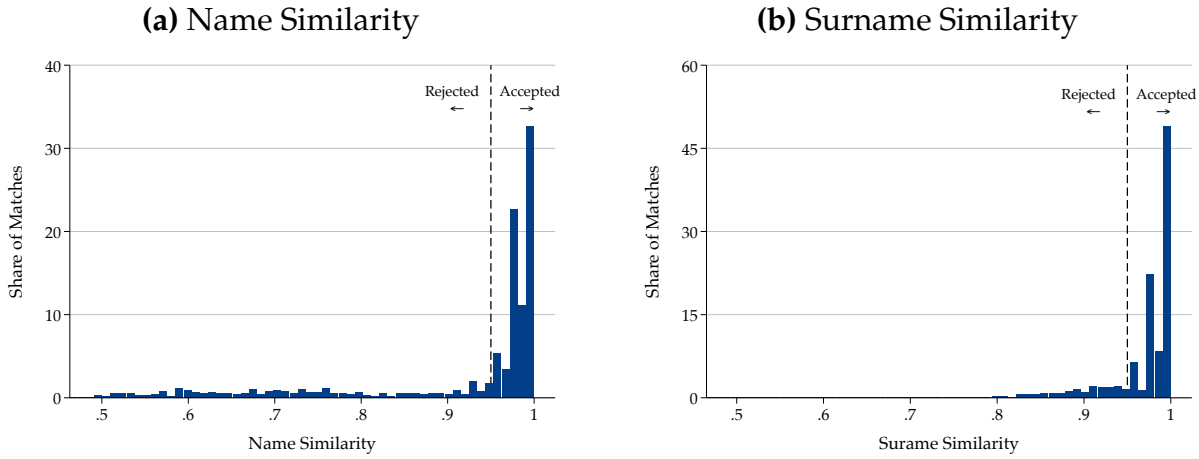
## Figures

**Figure A.1.** Matching Rate and Number of Matches



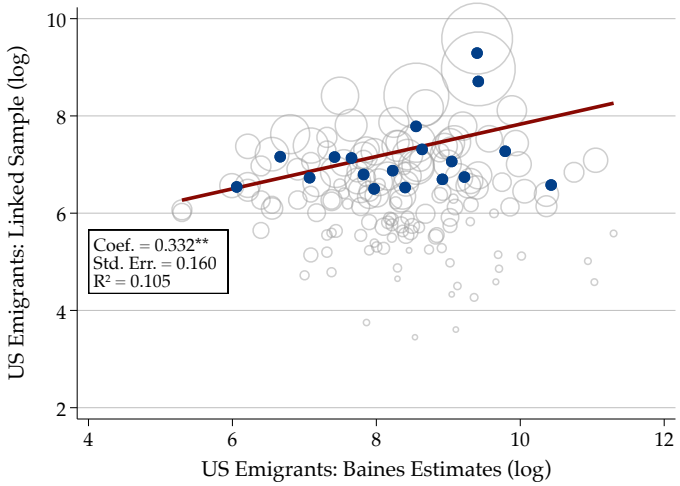

*Notes*. This Figure reports information on the matching performance of the emigrants' linked sample. In Panel A.1a, we display the matching rate, i.e., the share of emigrants recorded in the US census that are successfully linked to at least one entry in the UK census. The blue dots refer to the overall sample, and the red dots refer to the emigrants who could be recorded in the UK census (see main text for more precise information). In Panel A.1b, we plot the overall number of immigrants in the US census (in green), the number of those that are matched (in red), and the number of matches that we accept (in blue). Panel A.1c reports the distribution of the number of matches, where the last bin collects all those with more than 20 matches. Referenced on pages: 9, A6.

**Figure A.2.** Quality of Matches



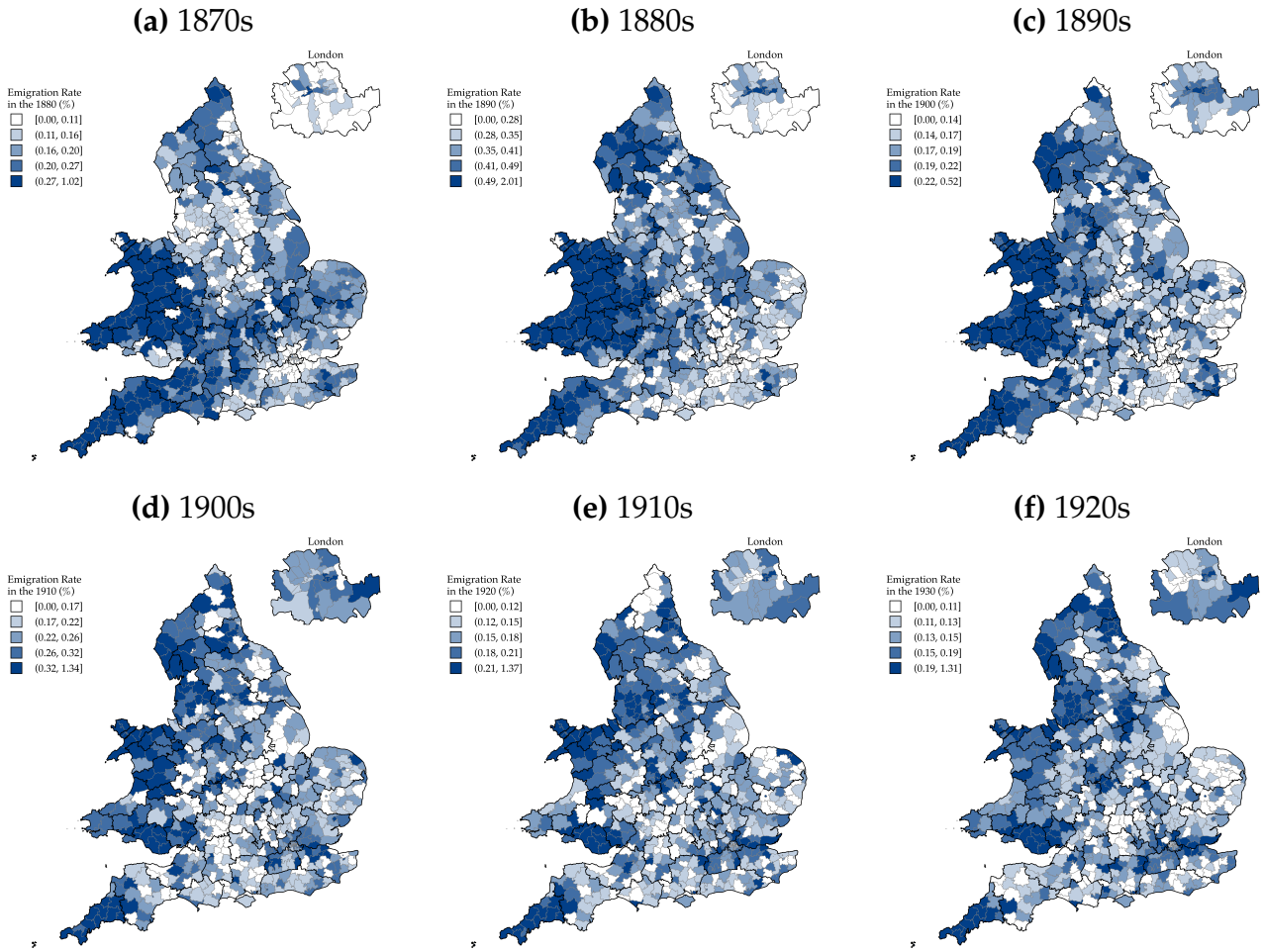

*Notes*. This Figure reports the distribution of the match quality in terms of name and surname similarity for the set of records with no more than two matches in the baseline sample. The similarity measure we use to construct the links is the Jaro-Winkler. This string metric measures the edit distance between the name and surname of the British immigrant recorded in the US census and their match(es) in the UK census. Panel A.2a reports the distribution of the name similarity; Panel A.2b refers to surnames. The vertical lines mark the quality thresholds we impose for a match to be part of the final linked sample. Referenced on pages: 9, A7.

**Figure A.3.** External Validation with Out-Migration Estimates




*Notes*. This Figure reports the correlation between out-migration in our linked inter-census sample and the estimates produced by Baines (2002). These estimates are at the county level and span 1880–1910. We thus aggregate our data by county. Furthermore, we exclude the London area because we cannot map our data to Baines' geographical divisions. Both variables are expressed in log terms. The gray dots display county-decade observations. Counties are weighted by their population. The blue dots report binned means. The red line overlays a linear fit between the variables. The Figure also reports the regression coefficient, its robust standard error, and the $R^2$ of the regression. Referenced on pages: 10, A10.

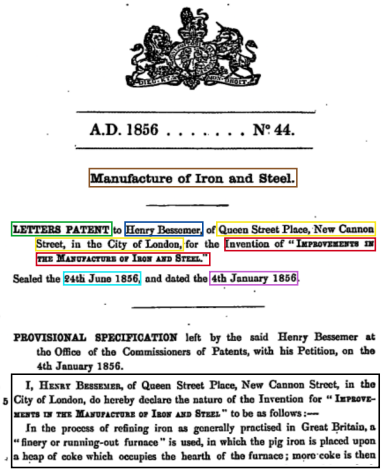
**Figure A.4.** Geographic Distribution of Emigration Rates Over Time




*Notes*. This Figure reports the distribution of US emigrants across districts in England and Wales over the period 1870–1940 by decade. Data are from the matched emigrants sample. The number of emigrants in each decade is normalized by population in 1891 and is expressed in percentage terms. Districts are displayed at their 1891 historical borders. Black edges also display historical county borders. Out-migration is cross-walked to consistent historical borders. Lighter to darker shades of blue indicate increasing quantiles of the emigration rate. The London area is displayed separately. Referenced on page: 10.

**Figure A.5.** Sample Annotated Patent Documents

**(a)** Henry Bessemer's 1856 Patent



A.D. 1856 . . . . . Nº 44.

Manufacture of Iron and Steel.

LETTERS PATENT to Henry Bessemer, of Queen Street Place, New Cannon Street, in the City of London, for the Invention of "IMPROVEMENTS IN THE MANUFACTURE OF IRON AND STEEL."

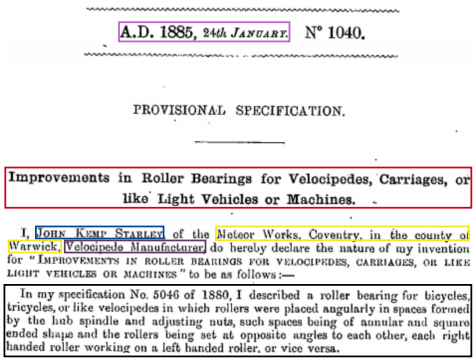
Sealed the 24th June 1856, and dated the 4th January 1856.

PROVISIONAL SPECIFICATION left by the said Henry Bessemer at the Office of the Commissioners of Patents, with his Petition, on the 4th January 1856.

I, HENRY BESSEMER, of Queen Street Place, New Cannon Street, in the City of London, do hereby declare the nature of the Invention for "IMPROVEMENTS IN THE MANUFACTURE OF IRON AND STEEL" to be as follows:—

In the process of refining iron as generally practised in Great Britain, a "finery or running-out furnace" is used, in which the pig iron is placed upon a heap of coke which occupies the hearth of the furnace; more coke is then

**(b)** John K. Starley's 1885 Bicycle Patent



A.D. 1885, 24th JANUARY. Nº 1040.

PROVISIONAL SPECIFICATION.

Improvements in Roller Bearings for Velocipedes, Carriages, or like Light Vehicles or Machines.

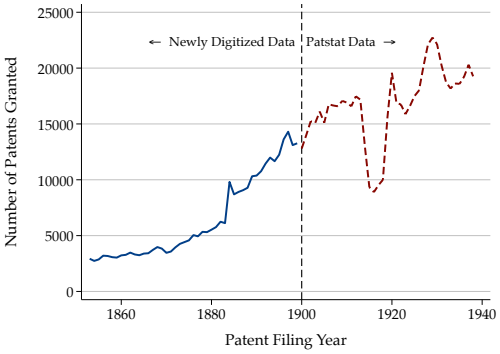
I, JOHN KEMP STARLEY, of the Meteor Works, Coventry, in the county of Warwick, Velocipede Manufacturer, do hereby declare the nature of my invention for "IMPROVEMENTS IN ROLLER BEARINGS FOR VELOCIPEDES, CARRIAGES, OR LIKE LIGHT VEHICLES OR MACHINES" to be as follows:—

In my specification No. 5046 of 1880, I described a roller bearing for bicycles, tricycles, or like velocipedes in which rollers were placed angularly in spaces formed by the hub spindle and adjusting nuts, such spaces being of annular and square ended shape and the rollers being set at opposite angles to each other, each right handed roller working on a left handed roller, or vice versa.

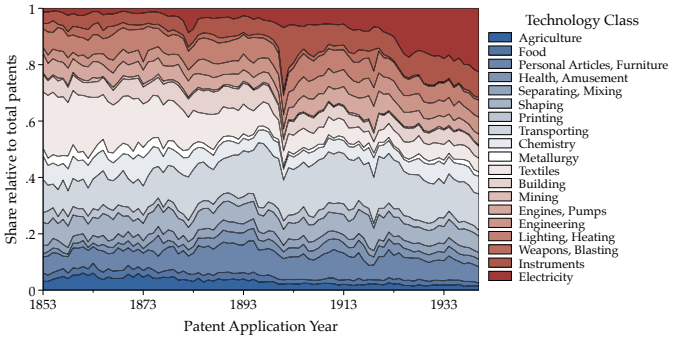
*Notes.* This figure displays two sample patent documents in our dataset. Panel A.5a was granted to Henry Bessemer in 1856 to invent the famous eponymous process for the mass production of steel from molten pig iron. Panel A.5b was granted to John Starley in 1885 to invent the first modern bicycle, which would soon revolutionize mobility in Europe and the US. Colors mark different variables that we structure in the dataset: (i) in brown, the short title; (ii) in red, the complete title (iii) in green, the type of protection granted; (iv) in blue, the author(s) name(s); (v) in yellow, the author(s)'s address(es); (vi) in light blue, the application date; (vii) in purple, the issue date; (viii) in black, the patent text that continues in the rest of the patent document; (ix) in dark purple, the author(s) profession(s). Not all (i–ix) data are available on every patent and in each year. Referenced on page: A11.

**Figure A.6.** Number and Composition of Patents Over Time

**(a)** Time Series of Patents

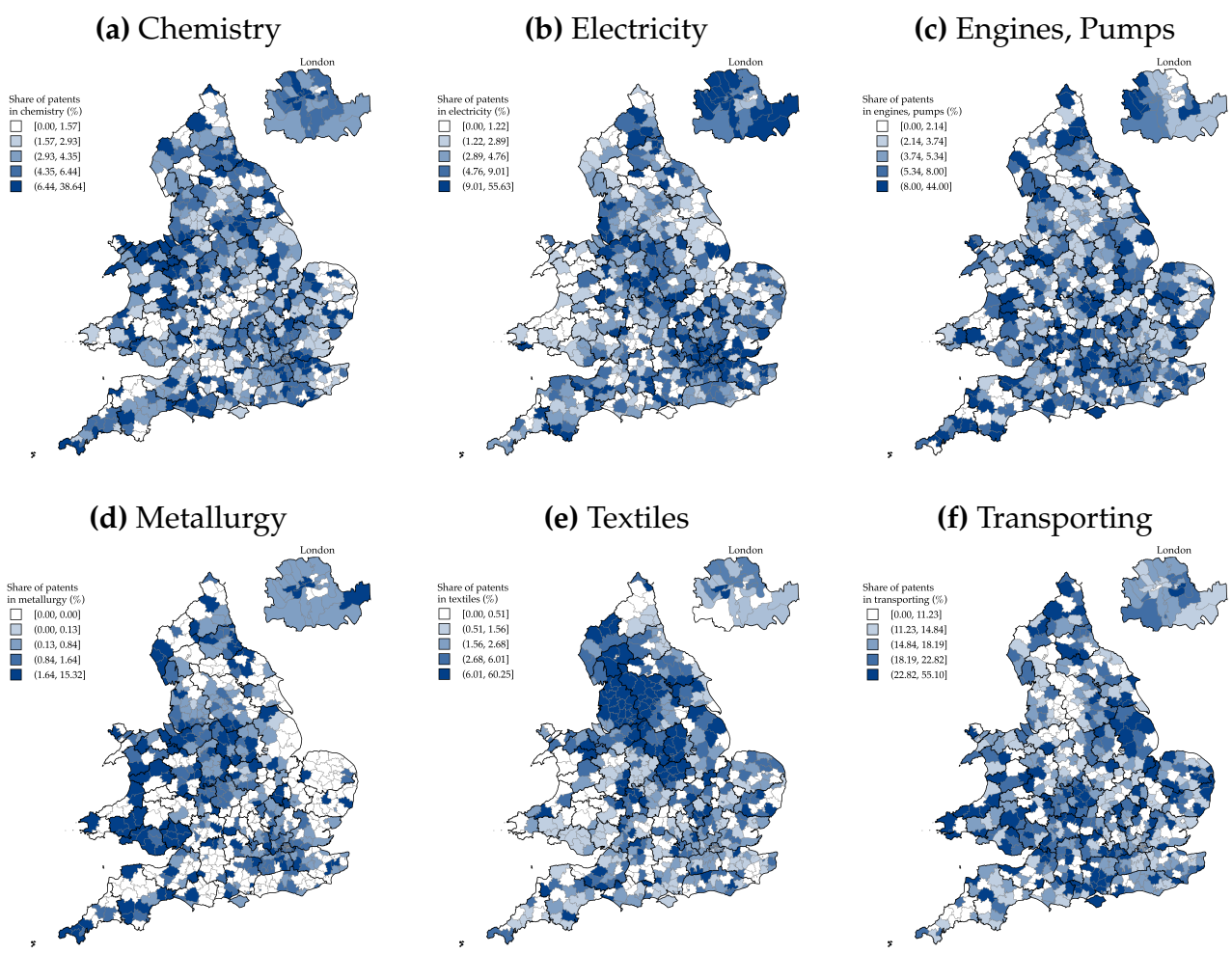



**(b)** Composition of Patents Across Technologies



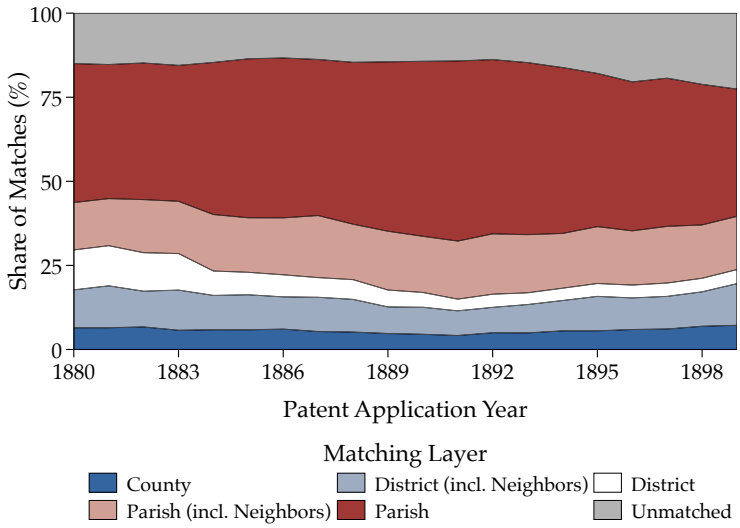

*Notes.* This Figure reports time-series information on the innovation activity in Britain between 1853 and 1939. In Panel A.6a, we report the total number of patents granted in the UK over the period. The blue dots report the newly digitized data that we assembled for this paper; the red dots report tabulations from the Patstat repository. In Panel A.6b, we plot the share of patents granted over time across technology classes. Referenced on page: 11.

**Figure A.7.** Geographic Distribution of Patents Across Technologies



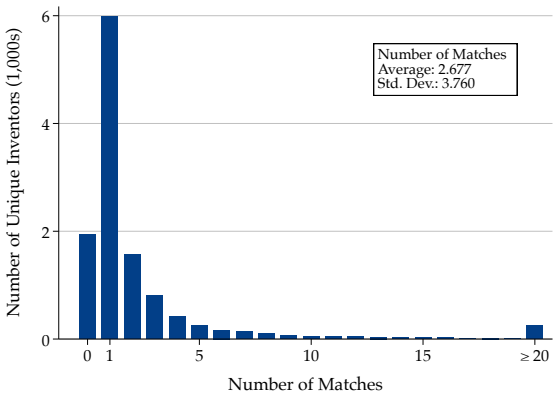

*Notes*. This Figure reports the intensity of patenting activity across districts over 1880–1939 for selected technology classes. Districts are displayed at 1891 borders. To assign patents to districts, we geo-reference the address of each author listed in the patent document and assign districts based on historical district borders. Black edges display county borders. The London area is displayed separately. Darker shades of blue indicate increasing quantiles of the patenting rate, defined as the percentage ratio between the number of patents in a given technology class and the overall number of patents produced. Referenced on page: A12.

**Figure A.8.** Matching Rate of Linked Inventors Sample



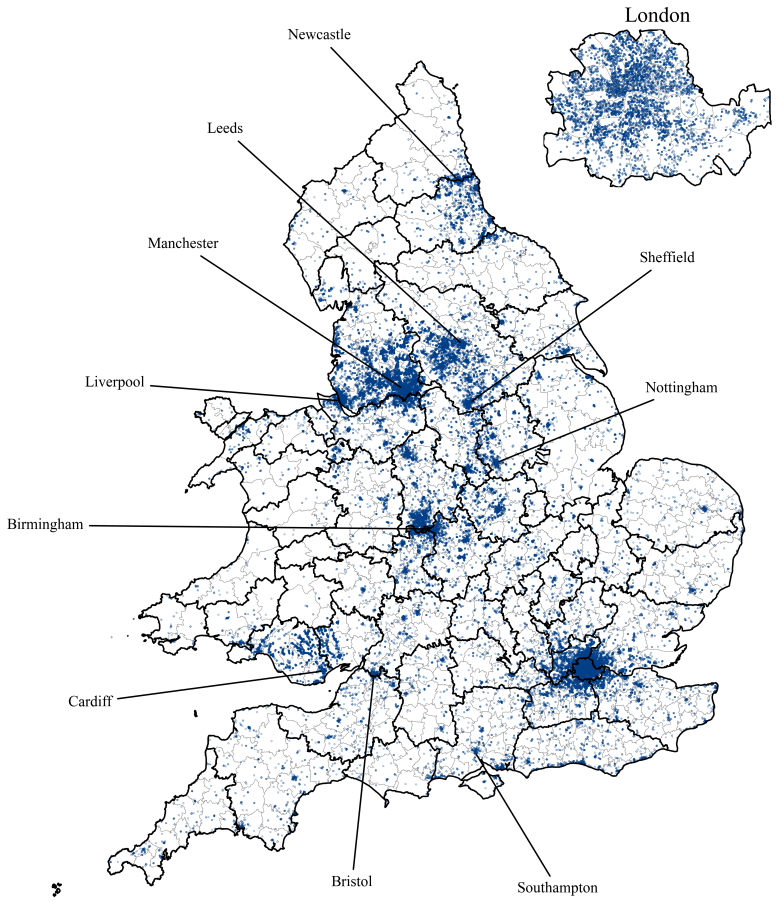

*Notes*. This Figure reports the matching rate of the linked sample of inventors. The records of the inventors who obtained a patent between 1880 and 1899 are linked to the 1891 population census, as detailed in the main text. The matching rate, i.e., the share of inventors successfully matched to the census, is reported on the $y$-axis in percentage points. The matching rate is broken down by the geographic layer of aggregation where the match is attained. Hence, we match over 75% inventors to the census throughout the period, and among those, the census record of slightly more than 50% of them is found in the same parish where the inventor is recorded living on the patent document. Referenced on page: A15.

**Figure A.9.** Number of Matches in Linked Inventor Sample




*Notes*. This Figure reports the number of census entries each inventor in the linked sample is matched to, including the share of inventors that we cannot match. We separately report the average number of matches and its standard deviation. Referenced on page: A15.

**Figure A.10.** Geographic Distribution of Linked Inventors




*Notes*. This Figure displays the spatial distribution of inventors across districts between 1880 and 1900. Each marker reports one inventor, defined as an individual who obtains at least one patent over the sample period. To retrieve the coordinates of the inventors, we first link population censuses, whose entries are, in turn, geo-referenced. The background map displays black counties and gray districts at historical borders in 1891. We highlight the ten largest urban centers at the time. Referenced on page: 11.

## B ADDITIONAL RESULTS

This section presents more details on several additional results referenced in the main text: the selection of British emigrants (Appendix B.I), the long-run association between emigration and innovation (Appendix B.II), and the possibility that immigrants sort into counties that are similar to their areas of origin (Appendix B.III).

### B.I Selection of British Emigrants

The historical scholarship argues that mass emigration from England and Wales to the US differed markedly from that of other countries (Berthoff, 1953; Baines, 2002). Unlike other European countries such as Germany, Sweden, or Italy, emigration from the UK in the second half of the nineteenth century was not predominantly a low-skilled rural phenomenon. Most emigrants left urban and industrial areas, and they did not come from the bottom of the human capital distribution, as was the case in Italy (Spitzer and Zimran, 2018) or Norway (Abramitzky et al., 2014). This distinction is important for our analysis, as it is unlikely that illiterate farmers would facilitate the flow of novel knowledge back to their origin areas. Even if this were the case, it would be unlikely that those rural areas would have the ability to reproduce US patents. However, much of the existing discussion relies on anecdotal evidence or partial census sources. In this section, we provide systematic evidence on the selection of British emigrants to the US and returning migrants to the UK relative to those who remained. To do so, we use the linked US-UK migrant sample and the UK population census.

Table B.1 presents the results of this analysis. Column (1) refers to non-migrants, and columns (2) and (5) refer to emigrants and return migrants, respectively. Columns (3) and (6) report the difference between non-migrants and emigrants and non-migrants and return migrants, respectively. Emigrants are more likely than stayers to work as engineers and as manufacturing workers in metallurgy and textiles. Unsurprisingly, they are less likely to work in public administration and as liberal professionals, since those occupations could not be transferred overseas. These patterns confirm historical evidence by, among others, Baines (2002), who describes transatlantic emigrants as a positively selected group of entrepreneurial individuals well-versed in the manufacturing crafts, especially in the second half of the nineteenth century. Return migrants appear somewhat different. They are more likely to work in construction, a lower-skilled sector compared to emerging industrial jobs, and are more likely to work as professionals, public officers, in transport (which in this period mainly comprised railway workers), and as utility workers. Taken together, this preliminary evidence suggests that the decision to return to the UK was relatively more common among those who were less successful in their American enterprise.

Individuals who migrated to the US were more likely to originate from the North West (including the industrial Lancashire districts), the South West (especially the rural areas of Cornwall and Devon), and Wales. As noted in the main text, the origins of emigrants shifted over time from predominantly rural areas in Southern England to the industrial regions of the North and the Midlands. The probability of return was not homogeneous across sending regions. Return migrants were less likely to reside in the East, in Wales, and in the West Midlands, while they were more likely to reside in the London area and in the South West. The different geographic distribution of emigrants and return

migrants is crucial for disentangling their respective contributions to innovation activity in the UK.

### B.II Long-Run Association between Emigration and Innovation

We now investigate the persistence of the association between exposure to foreign knowledge through migration ties and patenting activity.

We estimate the following regression:

$$\log \mathbb{E}\left[\text{Patents}_{it}\right] = \alpha_i + \alpha_t + \sum_{\substack{\tau=1940 \\ \tau \neq 2020}}^{2020} \beta^{\tau} \times \left[\text{Knowledge Exposure}_i \times 1\left(t = \tau\right)\right] \tag{B.1}$$

where $i$ and $t$ denote a district and year, respectively. In this setting, we have $t \in [1940, 2020]$. The term $\text{Knowledge Exposure}_i$ refers to knowledge exposure in the years 1920–1930, i.e., before the sample period. To reduce noise in the estimated $\beta^{\tau}$ coefficients, years in $\mathcal{T}$ are aggregated to decades. The estimated set of $\beta^{\tau}$ reports the conditional correlation between historical exposure to knowledge exposure and innovation activity in the two-year window indexed by $\tau$. Analogously, we estimate (B.1) at the district-technology-decade level by including the interactions between the technology-specific knowledge exposure term $\text{Knowledge Exposure}_{ik}$ and decade dummies.

Figure B.1 plots the set of estimated $\beta^{\tau}$ over time. The correlation between historical knowledge exposure and patenting activity remained positive and significant until the early 1980s, although it decreased over time. We interpret this as suggestive evidence that exposure to foreign knowledge through migration ties has a potentially long-lasting effect on the composition of innovation activity over time. In Table B.2, we re-estimate model (B.1), sector-by-sector, by decade. Compared to (B.1), we can therefore only include district and decade-fixed effects. The estimated correlation between historical exposure and patenting decreases over time in all sectors and, by the 1990s, it is no longer significant in many.

### B.III Assortative Matching of Emigrants in the United States

In this section, we lay out a simple framework to test whether British immigrants sort into US counties depending on the innovation similarity between the settlement location and their origin district. Let $\mathbf{P}_{j,t} = \{p_{1jt}, \ldots, p_{Njt}\}$ denote the patent portfolio of county $j$ in decade $t$, whose generic entry $p_{kjt}$ returns the number of patents in technology class $k$. Analogously, let $\mathbf{P}_{it}$ be the portfolio of district $i$. We define innovation similarity as:

$$\text{Innovation Similarity}_{ijt} \equiv \frac{\mathbf{P}_{it}^{\intercal}\mathbf{P}_{jt}}{\|\mathbf{P}_{it}\| \cdot \|\mathbf{P}_{jt}\|} = \frac{\sum_k p_{kit} p_{kjt}}{\sqrt{\sum_k p_{kit}^2}\sqrt{\sum_k p_{kjt}^2}} \leq 1 \tag{B.2}$$

which is a simple cosine similarity. The similarity measure returns a value of one if the patent portfolios of district $i$ and county $j$ are identical, meaning their composition across classes is the same. The index is normalized between zero and one.

We then estimate variations on the following OLS specification:

$$\mathbb{E}\left[\text{Emigrants}_{i \to j,t}\right] = \alpha_{i \times j} + \alpha_t + \beta \times \text{Innovation Similarity}_{ijt} + X_{ijt}'\Gamma \tag{B.3}$$

where the dependent variable is the number of emigrants from district $it$ to county $j$ in decade $d$, and

$\alpha_{i\times j}$ denotes county-by-district fixed effects. Controlling for dyadic fixed effects $\alpha_{i\times j}$ is important because it ensures that identification of $\hat{\beta}$ comes from within district–county pairs over time, rather than from time-invariant differences in pre-existing district–county characteristics. The coefficient $\beta$ therefore reports the correlation between innovation similarity and migration flows. Standard errors are two-way clustered by district and county. Under assortative sorting of migrants toward economically or technologically aligned locations, one would expect $\hat{\beta} > 0$.

We test this prediction in Table B.3. We do not find evidence that the innovation similarity between origin and destination areas explains migration patterns. The sign of the regression coefficients is unstable and, once we control for time-varying shifters at the county or district level, negative and statistically significant. In columns (4–6), we estimate the same regressions on the smaller sample of county-district pairs with positive migration flows. The results remain largely unchanged. We do not wish to over-emphasize these results as our measure of innovation similarity may be subject to measurement error, which may reduce the precision of the estimates. Taken together, however, we interpret these results as evidence that it is unlikely that assortative matching played a pivotal role in determining the location choice of immigrants.

### B.IV Anecdotal Evidence of Return Innovation

Who were the immigrants who contributed to the diffusion of US technology in Britain? Historical accounts highlight numerous cases of skilled artisans, entrepreneurs, and factory workers who were exposed to novel technologies where they settled and promoted their diffusion, or in some cases, appropriated them, in the UK.

In this section, we provide three examples, all involving return migration. Historical records typically focus on successful migrants who, upon returning, brought new technologies back to their origin areas and promoted local economic development. Our statistical analysis, however, suggests that this was only part of the story. In fact, we find that emigrants interacted with their origin communities even without returning.

#### *B.IV.1 British Puddlers and the Bessemer Process*

An 1856 article published in Scientific American described a new patent granted in the UK to Henry Bessemer (Wagner, 2008). Bessemer had discovered a new process, the would-be eponymous Bessemer process, that, for the first time, allowed the production of inexpensive steel from molten pig iron.[13] American inventor William Kelly complained:

> "I have reason to believe my discovery was known in England three or four years ago, as a number of English puddlers visited this place to see my new process. Several of them have since returned to England and may have spoken of my invention there."
> (Wagner, 2008, p. 363)

The veracity of Kelly's allegations remains unverified. They nonetheless indicate three important elements. First, American inventors knew that British immigrants posed a threat to the secrecy of their

[13] The Bessemer process was one of the most transformative technological developments of the nineteenth century (Rosenberg and Trajtenberg, 2004).

inventions. Second, technology transfer did not necessitate the very upper tail of the human capital distribution. Skilled workers, such as puddlers, could be agents of technology diffusion. Finally, the precise mechanism that emerges is return migration. Kelly expects British puddlers to speak of "his" invention upon returning to England.

### *B.IV.2 Henry Marsden and the Industrialization of Leeds*

Henry Rowland Marsden was born in Leeds to poor parents in 1823 (Curtis, 1875). At age twenty-five, he emigrated to the US, first to New York and then to Connecticut. There, he took on apprenticeships in engineering and metalworking firms. He obtained several engineering patents–mainly related to steam engines and pumps, including a "stone-crusher" which is still in use today. In 1862, Marsden returned to Leeds, where he set up a flourishing business centered around his newly patented inventions. He died in 1878 and is credited as one of the most prominent figures in the industrial development of Leeds.

### *B.IV.3 Migrants as Agents of Technology Transfer: Wellstood & Smith Ltd.*

The case of Stephen Wellstood and John Smith illustrates how international migration spurs technology transfers across countries. At age 16, James Smith (1811–1886) left Bonnybridge, Scotland, and migrated to the US. There, he established himself selling cooking stoves and married. However, as his wife got ill, Smith returned to Bonnybridge and started re-selling imported stoves from the US. He soon realized, however, that he could manufacture stoves directly in Britain. He then partnered with his long-time friend Stephen Wellstood and opened a foundry. They patented the exact same cooking stove Smith had been selling in the US and started a business that remained active until 1983.

**Tables**

**Table B.1.** Comparison between English Emigrants and Stayers

| | Non-Migrants | Emigrants | | | Return Migrants | | |
|---|---|---|---|---|---|---|---|
| | Mean (1) | Mean (2) | Difference (3) | Std. Err. (4) | Mean (5) | Difference (6) | Std. Err. (7) |
| **Panel A. Employment by Sector** (Dependent variable = 100 if individual employed in:) | | | | | | | |
| Agriculture | 29.354 | 29.100 | -0.007 | (0.075) | 26.255 | -1.369*** | (0.244) |
| Chemicals | 0.949 | 0.899 | -0.073*** | (0.016) | 1.006 | -0.035 | (0.055) |
| Construction | 15.686 | 15.969 | 0.035 | (0.061) | 17.029 | 0.444** | (0.210) |
| Engineering | 14.501 | 14.661 | 0.263*** | (0.059) | 14.280 | 0.185 | (0.195) |
| Entrepreneur | 0.007 | 0.004 | -0.003*** | (0.001) | 0.004 | -0.001 | (0.004) |
| Liberal Professions | 3.925 | 3.150 | -0.823*** | (0.030) | 4.516 | 0.247** | (0.118) |
| Metallurgy | 3.112 | 3.729 | 0.539*** | (0.032) | 3.247 | -0.068 | (0.097) |
| Public Administration | 3.498 | 3.361 | -0.198*** | (0.030) | 4.007 | 0.220** | (0.110) |
| Textiles | 9.485 | 10.058 | 0.717*** | (0.051) | 8.135 | -0.226 | (0.154) |
| Trade | 7.907 | 7.788 | -0.135*** | (0.046) | 8.534 | -0.038 | (0.158) |
| Transports | 10.771 | 10.471 | -0.317*** | (0.050) | 11.759 | 0.482*** | (0.180) |
| Utilities | 0.805 | 0.810 | 0.000 | (0.015) | 1.226 | 0.159*** | (0.061) |
| **Panel B. Region of Origin** (Dependent variable = 100 if individual resides in:) | | | | | | | |
| East | 10.487 | 8.357 | -2.128*** | (0.034) | 10.017 | -0.328** | (0.149) |
| East Midlands | 6.384 | 5.889 | -0.514*** | (0.029) | 6.212 | -0.062 | (0.119) |
| Greater London | 13.633 | 11.833 | -1.921*** | (0.041) | 15.358 | 1.733*** | (0.182) |
| North East | 6.583 | 7.088 | 0.538*** | (0.032) | 6.436 | -0.210* | (0.121) |
| North West | 17.433 | 19.798 | 2.676*** | (0.049) | 17.287 | -0.407** | (0.184) |
| South East | 12.400 | 10.329 | -2.023*** | (0.038) | 12.899 | -0.139 | (0.167) |
| South West | 6.442 | 7.298 | 0.847*** | (0.035) | 6.521 | 0.932*** | (0.126) |
| Wales | 6.469 | 8.579 | 2.111*** | (0.031) | 6.379 | -1.204*** | (0.115) |
| West Midlands | 11.188 | 11.506 | 0.293*** | (0.039) | 10.413 | -0.565*** | (0.149) |
| Yorkshire | 8.596 | 8.883 | 0.071** | (0.036) | 7.874 | 0.098 | (0.133) |

*Notes*. This Table displays the selection of emigrants and returning migrants to and from the US relative to the rest of the British population. The unit of observation is an individual. In each row, the dependent variable is equal to 100 if the individual belongs to the given category (e.g., if they are employed in agriculture) and zero otherwise. In columns (1), (2), and (5), we report the averages for non-migrants, migrants, and return migrants. In columns (3) and (6), we display the difference between non-migrants and, respectively, migrants and return migrants, controlling for the census they appear in. The associated robust standard errors are displayed in parentheses in columns (4) and (7). Referenced on page: B28.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table B.2.** Long-Run Correlation between Exposure to US Technology and Innovation in the UK

| | Patents by Technology Class | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1940s (1) | 1950s (2) | 1960s (3) | 1970s (4) | 1980s (5) | 1990s (6) | 2000s (7) | 2010s (8) |
| Agriculture | 2.401*** | 2.017*** | 1.278* | 0.797 | 0.605 | -0.308 | -0.293 | 0.163 |
| | (0.751) | (0.723) | (0.662) | (0.757) | (0.809) | (0.756) | (0.649) | (0.598) |
| Building | 5.165*** | 5.123*** | 4.649*** | 3.336** | 3.180* | -0.567 | -1.371 | -1.375 |
| | (1.580) | (1.593) | (1.637) | (1.673) | (1.722) | (1.441) | (1.682) | (1.571) |
| Chemistry | 1.472*** | 1.543*** | 1.510*** | 1.232*** | 0.838*** | 0.444* | 0.303 | 0.114 |
| | (0.271) | (0.284) | (0.314) | (0.296) | (0.299) | (0.247) | (0.213) | (0.155) |
| Electricity | 1.694*** | 1.379*** | 1.585*** | 1.171*** | 0.796** | 0.022 | -0.239 | -0.330 |
| | (0.338) | (0.340) | (0.349) | (0.369) | (0.390) | (0.283) | (0.249) | (0.234) |
| Engineering | 5.801*** | 4.939*** | 4.824*** | 4.500*** | 4.059*** | 2.726*** | 2.001** | 1.819** |
| | (1.058) | (1.152) | (1.179) | (1.131) | (1.140) | (0.960) | (0.932) | (0.785) |
| Engines, Pumps | 1.562*** | 1.196*** | 1.205*** | 1.141*** | 0.714*** | 0.165 | -0.089 | -0.305* |
| | (0.251) | (0.258) | (0.288) | (0.265) | (0.256) | (0.193) | (0.195) | (0.164) |
| Food | 3.320*** | 2.593*** | 2.983*** | 3.129*** | 2.270*** | 1.874*** | 1.753** | 1.130* |
| | (0.844) | (0.842) | (0.834) | (0.818) | (0.820) | (0.723) | (0.754) | (0.590) |
| Health, Amusement | 1.506*** | 1.316*** | 1.324*** | 1.172*** | 0.779*** | 0.278 | 0.046 | -0.263* |
| | (0.200) | (0.227) | (0.253) | (0.243) | (0.216) | (0.180) | (0.176) | (0.160) |
| Instruments | 2.970*** | 2.281*** | 2.586*** | 2.583*** | 2.211*** | 1.189** | 1.184** | 0.739 |
| | (0.569) | (0.582) | (0.616) | (0.611) | (0.640) | (0.576) | (0.535) | (0.480) |
| Lighting, Heating | 9.711*** | 9.262*** | 8.915*** | 10.117*** | 7.695** | 4.772* | 4.575* | 1.671 |
| | (2.811) | (2.936) | (3.063) | (3.155) | (3.261) | (2.611) | (2.422) | (2.487) |
| Metallurgy | 1.669*** | 1.558** | 1.435** | 1.453* | 1.177 | 0.009 | 0.033 | 0.170 |
| | (0.602) | (0.648) | (0.638) | (0.760) | (0.775) | (0.653) | (0.593) | (0.613) |
| Mining | 1.601*** | 1.271*** | 1.353*** | 1.068*** | 0.690*** | 0.396*** | 0.244* | -0.089 |
| | (0.196) | (0.239) | (0.259) | (0.224) | (0.209) | (0.147) | (0.137) | (0.124) |
| Personal Articles, Furniture | 9.374*** | 9.301*** | 7.789*** | 7.064** | 6.633** | 5.868*** | 1.841 | -3.366** |
| | (2.122) | (2.099) | (2.268) | (3.119) | (2.787) | (2.062) | (2.360) | (1.697) |
| Printing | 1.523*** | 1.295*** | 1.697*** | 1.495*** | 0.916* | 0.331 | -0.307 | -0.297 |
| | (0.404) | (0.434) | (0.506) | (0.515) | (0.517) | (0.501) | (0.421) | (0.369) |
| Separating, Mixing | 1.432*** | 1.149*** | 1.148*** | 1.021*** | 0.829*** | 0.353 | -0.011 | -0.243 |
| | (0.243) | (0.271) | (0.309) | (0.285) | (0.269) | (0.263) | (0.218) | (0.206) |
| Shaping | 2.332*** | 2.067*** | 1.954*** | 1.973*** | 1.418*** | 0.973*** | 0.873** | 0.441 |
| | (0.370) | (0.406) | (0.434) | (0.434) | (0.431) | (0.363) | (0.343) | (0.334) |
| Textiles | 19.326*** | 16.195*** | 15.938*** | 18.796*** | 12.469** | 10.755** | 8.304 | 8.443* |
| | (5.528) | (5.874) | (5.506) | (5.993) | (6.080) | (5.378) | (5.285) | (4.461) |
| Transporting | 2.023*** | 1.785*** | 1.818*** | 1.558*** | 1.040*** | 0.558*** | 0.134 | -0.046 |
| | (0.199) | (0.236) | (0.250) | (0.221) | (0.217) | (0.173) | (0.159) | (0.139) |
| Weapons, Blasting | 2.011*** | 1.920*** | 2.017*** | 1.668*** | 1.391*** | 0.651*** | -0.115 | -0.201 |
| | (0.254) | (0.258) | (0.285) | (0.280) | (0.256) | (0.213) | (0.202) | (0.158) |

*Notes*. This Table reports the long-run association between exposure to US technology and innovation in the UK. The unit of observation is a district-technology pair observed at a decadal frequency between 1940 and 2020. The dependent variable is the (log) number of patents. The independent variable is an interaction term between exposure to US technology over the 1930s, decade dummies, and technology dummies. The 2010 decade dummy serves as the baseline category. The regression includes district-by-technology and technology-by-time fixed effects. Standard errors are shown in parentheses and are clustered two-way by district and technology. Referenced on page: B29.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

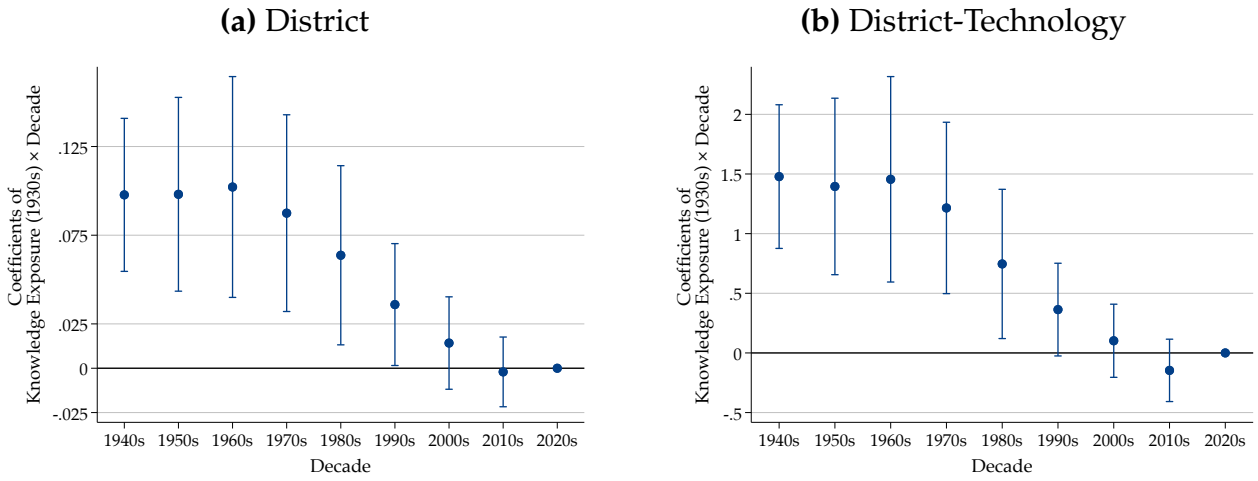
**Table B.3.** Assortative Matching of British Immigrants in the US

| | Dependent Variable: Number of Migrants | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Innovation Similarity | -0.342 | -0.304 | -1.213 | -58.477* | -48.414 | 10.963 |
| | (1.872) | (1.736) | (1.867) | (31.253) | (30.537) | (40.179) |
| County-District FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Decade FE | Yes | – | – | Yes | – | – |
| County-Decade FE | No | Yes | Yes | No | Yes | Yes |
| District-Decade FE | No | No | Yes | No | No | Yes |
| N. of County-District Dyads | 1,795,188 | 1,795,188 | 1,795,188 | 247,543 | 247,027 | 247,027 |
| Observations | 10,616,676 | 10,616,676 | 10,616,676 | 840,477 | 839,349 | 839,349 |
| Mean Dep. Var. | 0.032 | 0.032 | 0.032 | 0.384 | 0.384 | 0.384 |

*Notes*. This Table tests the hypothesis that British immigrants settled in US counties that innovated in the same fields as their district of origin. The unit of observation is a district-county pair. Units are observed at a decadal frequency between 1880 and 1930. The dependent variable is the number of migrants between the (UK) district and the (US) county. The independent variable is the similarity of the patent portfolios between the district and the county. The details of the similarity metric are explained in the main text. In columns (1–3), we include all district-county dyads; in columns (4–6), we include only the pairs with positive migration ties. Columns (1) and (3) include dyad and decade fixed effects; in columns (2) and (4), we include county-by-time fixed effects; columns (3) and (6) further include district-by-time fixed effects. Standard errors are clustered two-way by district and county and are displayed in parentheses. Referenced on page: B30.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

## Figures

**Figure B.1.** Exposure to US Innovation Activity and Innovation in the UK in the Long Run




*Notes*. This Figure reports the long-run association between exposure to US technology and innovation in the UK. The unit of observation is a district (Panel B.1a) or a district-technology pair (Panel B.1b) observed at a decadal frequency between 1940 and 2020. The dependent variable is the number of patents. The independent variable is an interaction term between exposure to US technology in the 1930s and decade dummies, as detailed in equation (B.1). The 2020 decade dummy serves as the baseline category. Each dot reports the two-stage least squares estimate obtained from the instrument in (6). The regression includes district and decade (Panel B.1a) and district-by-time and technology-by-time fixed (Panel B.1b) effects. Standard errors in Panel B.1a are clustered by district, and in Panel B.1b by district-technology. Bands report 95% confidence intervals. Referenced on pages: 21, B29.

## C ROBUSTNESS ANALYSES

In this section, we first provide the analytical details of the district-technology-level instrumental variable framework (Appendix C.I). We then describe in detail the robustness of the main results presented in the paper (Appendix C.II) and of the neighborhood and family analyses (Appendix C.III).

### C.I District-Technology Instrumental Variable Framework

This section is a companion to Section III.D and presents the analytical derivations and identification assumptions of the instrumental variable design when the unit of observation is a district-technology pair, as opposed to the district-level framework described in the main text.

#### *C.I.1 Instrument for Knowledge Exposure by Technology*

In the first step, we construct predicted innovation by county-technology as follows:

$$\widehat{\text{Patents}}_{jkt} = \omega_{jk} \times \text{Patents}_{kt}^{-j}, \tag{C.1}$$

where $j$, $k$, and $t$ denote a county, technology, and decade. The share term is a time-invariant exposure measure of the importance of technology $k$ for county $j$'s portfolio and reads out as $\omega_{jk} = (\sum_{t=1860}^{1869} \text{Patents}_{jkt}) / (\sum_{t=1860}^{1869} \text{Patents}_{kt})$. $\omega_{jk}$ is the share of patents granted to inventors in county $j$ before the beginning of the analysis relative to the total number of patents granted in that technology sector. The shift term ($\text{Patents}_{kt}^{-j} = \sum_{j' \neq j} \text{Patents}_{j'kt}$) is the total number of patents granted in technology $k$ and year $t$, except for those issued to inventors located in county $j$.

In the second step, we construct the instrument as follows:

$$\widehat{\text{Knowledge Exposure}}_{ikt} = \sum_j \widehat{\text{Patents}}_{jkt} \times \text{Migrants}_{i \to j}^{0}, \tag{C.2}$$

where, as in the district-level case, the term ($\text{Migrants}_{i \to j}^{0}$) measures the time-invariant exposure of district $i$ to innovations produced in county $j$ and is equal to the number of migrants that moved from district $i$ to county $j$ before the beginning of the analysis (between 1870 and 1879).

As in the baseline framework, the instrument leverages innovation predicted by interacting initial specialization with aggregate patenting trends in each technology as shocks. The number of migrants between each county-district pair at baseline gives the shares. The estimation follows the two-stage residual inclusion method of Mullahy (1997). Specifically, in the first stage, we regress observed knowledge exposure on district-decade, technology-decade, and district-technology fixed effects, and an interaction term between the sum of the shares and technology-decade categorical variables. The residual of this regression is then plugged into the second-stage regression (4), where we further control for the same set of fixed effects and for the sum of shares interacted with technology-decade dummies that we include in the first-stage regression. Standard errors are bootstrapped and are clustered at the district-technology level.[14]

[14] Formally, the first stage is:

$$\mathbb{E}\left[\text{Knowledge Exposure}_{ikt}\right] = \alpha_{it} + \alpha_{kt} + \alpha_{ik} + \gamma \times \widehat{\text{Knowledge Exposure}}_{ikt} + \delta_{kt} \times \sum_j \text{Migrants}_{i \to j}^{0},$$

### C.I.2 *Identification Assumption*

The identification assumption in the district-technology setting is conceptually analogous to the district-level specification, but the more granular fixed effects reduce the scope for unobserved confounders that may bias the instrumental variable estimates. Specifically, the validity of the instrument requires that, conditional on district-decade, technology-decade, and district-technology fixed effects, the interaction of baseline migration ties and predicted county-level innovation shocks is orthogonal to unobserved factors that vary simultaneously by district, technology, and decade, and that influence innovation outcomes.

This identification assumption requires two conditions. First, baseline migration patterns from each UK district to US counties must not predict future unobserved determinants of innovation that vary within district-technology-decade cells. Given the rich set of fixed effects included, this assumption is substantially weaker than in the district-level case: it only requires that historical migrant destination choices do not correlate with unobserved shocks varying simultaneously by district, technology, and decade. Second, the predicted US innovation shocks at the county-technology-decade level must reflect exogenous variation unrelated to contemporaneous technology-specific unobserved determinants of UK innovation at the district level. In other words, the predicted shocks cannot be systematically driven by the same unobserved factors influencing innovation outcomes within district-technology-decade units in the UK.

In terms of the assortative matching and the reflection problem endogeneity taxonomy, the more granular specification implies that any unobservable factor that predicts either immigrant sorting or the joint innovation dynamics in the UK and in the US induces bias in the instrumental variable estimates only if it also predicts innovation within district-technology cells. This identifying assumption is therefore considerably weaker than the district-level identification assumption. Consistent with this, the district-technology results are always qualitatively similar to the district-level estimates, which we interpret as corroborating evidence of the causal nature of our findings.

Following Borusyak et al. (2022), Panel C Table II reports a balance test of district-technology patenting activity, where we regress the instrument on lagged patenting outcomes. The specification includes all fixed effects and controls of the baseline estimating equation. We find no statistically significant association between the instrument and pre-determined patenting, whether it be measured through the total number of patents, the number of high-importance patents, or the similarity with US innovation. The correlation, besides not being statistically significant, is also very small in magnitude. Moreover, the estimated coefficients are economically small in magnitude. These findings provide additional support for the empirical plausibility of the identification assumption.

---

which yields the residuals $\hat{v}_{ikt}$. The second stage is

$$\log \mathbb{E}\left[y_{ikt}\right] = \alpha_{it} + \alpha_{kt} + \alpha_{ik} + \gamma \times \text{Knowledge Exposure}_{ikt} + \delta_{kt} \times \sum_{j} \text{Migrants}^{0}_{i \to j} + \rho \hat{v}_{ikt},$$

where $\delta_{kt}$ codes technology-decade-specific categorical variables.

## C.II Robustness Analyses: Main Results

### *C.II.1 Alternative Dependent Variables*

In the main analysis, we use two indicators of patent novelty and similarity based on Kelly et al. (2021). Both measures compare UK patents to patents issued in the US within a given time window around their grant date. As in the original paper, we use a five-year window in the main analysis. In Table C.9, we show that the results remain similar when employing a one- or a ten-year window instead.

The baseline definition of a "high-importance" patent is a patent that lies within the top 20% of the importance distribution of Kelly et al. (2021). In Table C.10, we show that the results remain qualitatively unchanged when looking at patents in the top 10% and 5% of the distribution.

In Table C.6, we restrict the dependent variable to patents with at least one firm assignee. This restriction focuses on patents that plausibly reflect technology adoption and diffusion within firms rather than the activity of independent inventors. The baseline results are confirmed using this stricter definition of patenting.

The baseline estimating equation is a Poisson quasi-likelihood specification. We adopt this approach because patents are a count variable and because Chen and Roth (2024) show that applying the zero-inflated log-transformation yields scale-dependent estimates and precludes point identification of the average treatment effect. Table C.1 shows that the results remain quantitatively unchanged when employing a linear (OLS) specification.

### *C.II.2 Horserace Regressions*

Figure C.4–Figure C.5 report the estimates obtained from equations Equation (3)–(4) when controlling for district-level observable characteristics measured in the 1880 census (at baseline) and interacted with decade or technology-decade categorical variables. The red dots report the instrumental variable estimates. The last row reports the estimates obtained when including all controls in the same regression.

The results are remarkably stable irrespective of the set of controls. In particular, including the controls that display a non-zero correlation with the instrument, shown in Table II, does not alter either the magnitude or the statistical significance of the baseline estimates.

### *C.II.3 Alternative Definitions of Knowledge Exposure*

In Table C.11, we employ four alternative measures of knowledge exposure. First, we construct a measure that fixes bilateral emigrant flows:

$$\text{Knowledge Exposure}^{2}_{ik,t} = \sum_{j} \left( \text{Patents}_{jk,t} \times \text{Emigrants}_{i \to j,1880} \right), \tag{C.3}$$

which, compared to the main measure, restricts assortative matching to the first decade of the analysis. Second, we define the mirror measure that holds fixed specialization patterns across counties:

$$\text{Knowledge Exposure}^{3}_{ik,t} = \sum_{j} \left( \text{Patents}_{jk,t-10} \times \text{Emigrants}_{i \to j,t} \right). \tag{C.4}$$

Compared to the baseline one, this fixes county specialization one decade back relative to migrant flows. We use this measure to verify that previous knowledge influences innovation dynamics in the UK, in addition to contemporaneous patenting activity. Third, we define an alternative measure that leverages the *stock*, instead of the *flow* of patents issued:

$$\text{Knowledge Exposure}^{4}_{ik,t} = \sum_{j} \left[ \sum_{\tau \leq t} \left( \text{Patents}_{jk,\tau} \right) \times \text{Emigrants}_{i \to j,t} \right]. \tag{C.5}$$

The idea behind (C.5) is that specialization can be defined in terms of the cumulative number of patents filed before the given period. Fourth, we construct the baseline measure, but using the share of patents instead of the level:

$$\text{Knowledge Exposure}^{5}_{ik,t} = \sum_{j} \left( \frac{\text{Patents}_{jk,t}}{\text{Patents}_{j,t}} \times \text{Emigrants}_{i \to j,t} \right). \tag{C.6}$$

This last metric normalizes the number of patents by the total number granted in each county-decade to construct a measure of specialization in each technology. We construct the alternative knowledge exposure metrics (C.3)–(C.5) for both district- and district-technology-level specifications, while (C.6) is only defined in the district-technology-level setting. In Table C.11, we show that all these measures yield quantitatively similar results.[15]

Additionally, in Table C.8, we construct variations of the baseline knowledge exposure measure which include only patents in the top $\ell$% of the importance distribution. Specifically, the generic $k$-Knowledge Exposure variable is defined as

$$\text{Knowledge Exposure}^{6}_{ik,t}(\ell) = \sum_{j} \left( \text{Patents in Top } \ell\% \text{ Importance}_{jk,t} \times \text{Emigrants}_{i \to j,t} \right). \tag{C.7}$$

We consider three $\ell$ thresholds: top 20%, 10%, and 5%. The association between innovation and knowledge exposure is larger for more restrictive $\ell$-Knowledge Exposure metrics. This is consistent with the hypothesis that migrants' incentive to transmit knowledge to the UK was larger for patents with a higher economic return.

### *C.II.4 Alternative Standard Errors*

In the baseline analysis, we cluster standard errors at the district level for district-level regressions and at the district-technology level for district-technology regressions. The two-stage least squares standard errors are obtained via block bootstrap, following Wooldridge (2010). Figure C.1–Figure C.2 (Panels a–c) report the confidence bands on the baseline estimates obtained using alternative standard error estimators: the White heteroskedasticity-robust estimator, alternative clustering levels, and the Conley (1999) estimator that adjusts for spatial autocorrelation, at different bandwidths between 10 and 200 kilometers.

Figure C.1–Figure C.2 (Panels d–f) display the standard errors obtained via bootstrap. Specifically, we employ the pairs bootstrap approach to account for heteroskedasticity (Davison and Hinkley, 1997), wild-cluster bootstrap following (MacKinnon and Webb, 2017), and a spatial block bootstrap

[15] We apply the inverse hyperbolic sine to each knowledge exposure term so the coefficients are comparable across specifications.

where the resampling blocks are defined using the same distance kernel of (Conley, 1999). The estimates are obtained from 500 replications.

We find that the statistical significance of the baseline estimates remains intact across these approaches.

#### *C.II.5 Alternative Cut to Linked Sample*

In the baseline analysis, we retain all sufficiently high-quality migrant matches. Some British immigrants recorded in the US census are matched with multiple UK census records, as detailed in Appendix A.II. In such cases, we follow the standard practice in the literature and weigh each such match by the inverse of the number of matches to avoid multiple counting.

In Table C.7, we take an alternative route and compute the knowledge exposure terms (1)–(2) using records with a single match in the linked sample. An advantage of only considering single matches is that one may be worried that the probability of obtaining multiple matches correlates with unobserved individual characteristics. In practice, however, the estimates remain qualitatively and quantitatively unchanged relative to the baseline.

#### *C.II.6 Exclusion of Inventors Possibly Appearing on UK and US Patent Records*

Our results may be driven by inventors applying for the same patent in the US and the UK. It is worth noticing that this would impact our estimates only if there were a correlation between county-district migration and the inventors' location. Second, this behavior may still signify technology transfer operated by the inventors and, thus, contribute to return innovation.

To quantify its relevance, we compute the number of British patents–as well as their average similarity and the number of high-importance patents–excluding those filed by inventors who also appear in some US patents over a ten-year window around the patent issue year. This approach is very conservative because we cannot verify whether the inventor listed on the British patent is, in fact, the same individual appearing in the UK patent. In practice, this procedure excludes all patents with at least one inventor whose name also appears on any US patents.

Table C.5 reports the results. All baseline estimates are confirmed using this approach, suggesting that our findings are unlikely to be driven by inventors "double-filing" patents in both the US and the UK.

### C.III Robustness Analyses: Mechanisms

#### *C.III.1 Family Analysis*

The UK census does not contain information on relatives who do not live in the same household. To identify extended family members, we focus on individuals who share the same surname and live in the same county. During the period of our analysis, England and Wales were divided into 42 counties. By restricting surname clusters to individuals living in the same county, we substantially restrict the scope for false positives without excessively inflating the number of observations (say, at the surname-district or surname-parish level). This approach, however, is likely not well-suited for London. Internal migration to London has been a persistent feature of the British demographic and economic geography, both in pre-industrial (Landers, 2006) and industrial (Schürer and Day, 2019)

periods. As a result, surnames in London may not convey family relationships. Table C.12 repeats the baseline exercise and excludes families in London from the estimation sample. The results remain similar even though we lose some precision in columns (2) and (5).

### *C.III.2 Neighborhood Analysis*

We consider two departures from the baseline scenario described in the paper. First, we exclude Wales from the estimation sample in Table C.13. We apply this sample cut because, as shown in Table A.5, the number of matches in the sample of inventors linked to the census is larger for inventors residing in Wales. The higher number of matches likely reflects the lower probability of true matches in Wales, which is itself plausibly due to more common spelling mistakes of Welsh names and surnames by English census enumerators. By excluding them from the estimation sample, we therefore ensure that this imbalance does not drive the results. The estimates presented in the Table suggest that this does not appear to be the case.

Second, in the paper, we consider an emigrant to be in the neighborhood of an inventor if he lived within five kilometers of the inventor before emigrating. In Figure C.8, we consider four alternative threshold distances: two, three, ten, and twenty kilometers. We then estimate the baseline regression for these thresholds and separately report the estimated average treatment effects. The coefficients remain positive for all thresholds above two kilometers, but they are largest in magnitude–and statistically significant at the 1% level–for three- and ten-kilometer neighborhoods. The estimate for the twenty-kilometer definition is similar, albeit noisier. From an economic perspective, it is plausible that as the intensity of the social connections vanishes with distance, including larger neighborhoods in the treatment introduces noise, which reduces the precision of the estimates. In turn, a very low threshold discards a large number of emigrants, thus artificially dampening the treatment effect.

**Tables**

**Table C.1.** Exposure to Innovation in the US and Innovation in the UK: Linear Model

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| **Panel A. Observed Knowledge Exposure** | | | | | | |
| Knowledge Exposure | 34.639*** | 11.500*** | 1.119*** | 14.995*** | 3.439*** | 0.463*** |
| | (4.730) | (1.270) | (0.134) | (1.251) | (0.336) | (0.033) |
| **Panel B. Instrumental Variable Estimates** | | | | | | |
| *Second-Stage Regressions* | | | | | | |
| Knowledge Exposure | 80.235*** | 9.396*** | 2.839*** | 20.007*** | 2.356 | 0.737*** |
| | (22.896) | (3.470) | (0.733) | (7.025) | (1.609) | (0.184) |
| *First-Stage Regressions* | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ | 1.155*** | 1.155*** | 1.155*** | 0.730*** | 0.730*** | 0.730*** |
| | (0.159) | (0.159) | (0.159) | (0.035) | (0.035) | (0.035) |
| K-P $F$-statistic | 52.445 | 52.445 | 52.445 | 433.854 | 433.854 | 433.854 |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 614 | 614 | 614 | 614 | 614 | 614 |
| Observations | 3,057 | 3,057 | 3,057 | 58,083 | 58,083 | 58,083 |
| Mean Dep. Var. | 147.022 | 32.172 | 3.274 | 6.302 | 1.280 | 0.134 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (5), the dependent variable is the number of patents. In columns (2) and (4), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2). Panel A reports the association between observed knowledge exposure and innovation; Panel B reports the instrumental variable estimates and the associated first-stage regression. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are linear: Panel A is estimated via Ordinary Least Squares; Panel B is estimated via Two-Stage Least Squares. The K-P $F$-statistic reports the Kleibergen-Paap statistic for weak instrument. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. Standard errors in Panel B are bootstrapped. The sample is larger than in the baseline analysis because the PQML regression automatically drops observations with no variation within the fixed-effect groups (and thus do not contribute to the estimation), whereas 2SLS does not. Referenced on pages: 14, 23, C38.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.2.** Shock-Level Descriptive Statistics

| | Mean | Std. Dev. | Min | Max | Units | Obs. |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A. County-Decade Shifts** | | | | | | |
| All Counties | 91.844 | 557.604 | 0.000 | 25000.000 | 2,848 | 14,240 |
| Shifts by US Region: | | | | | | |
| Pacific | 6.771 | 126.877 | 0.000 | 7164.546 | 2,848 | 14,240 |
| Mountain | 1.835 | 30.818 | 0.000 | 1999.667 | 2,848 | 14,240 |
| West North Central | 8.199 | 90.613 | 0.000 | 5096.970 | 2,848 | 14,240 |
| West South Central | 2.376 | 23.951 | 0.000 | 1003.871 | 2,848 | 14,240 |
| East North Central | 27.108 | 405.982 | 0.000 | 25000.000 | 2,848 | 14,240 |
| East South Central | 1.994 | 21.407 | 0.000 | 1085.017 | 2,848 | 14,240 |
| Middle Atlantic | 24.764 | 293.655 | 0.000 | 11000.000 | 2,848 | 14,240 |
| New England | 12.865 | 187.169 | 0.000 | 6365.980 | 2,848 | 14,240 |
| South Atlantic | 5.188 | 71.965 | 0.000 | 3565.302 | 2,848 | 14,240 |
| **Panel B. County-Technology-Decade Shifts** | | | | | | |
| All Counties | 4.834 | 39.432 | 0.000 | 5074.354 | 54,112 | 270,560 |
| Shifts by US Region: | | | | | | |
| Pacific | 0.356 | 8.322 | 0.000 | 1087.014 | 54,112 | 270,560 |
| Mountain | 0.097 | 2.052 | 0.000 | 363.024 | 54,112 | 270,560 |
| West North Central | 0.432 | 6.561 | 0.000 | 1162.129 | 54,112 | 270,560 |
| West South Central | 0.125 | 1.737 | 0.000 | 341.703 | 54,112 | 270,560 |
| East North Central | 1.427 | 28.655 | 0.000 | 5074.354 | 54,112 | 270,560 |
| East South Central | 0.105 | 1.549 | 0.000 | 248.360 | 54,112 | 270,560 |
| Middle Atlantic | 1.303 | 20.652 | 0.000 | 2226.138 | 54,112 | 270,560 |
| New England | 0.677 | 13.330 | 0.000 | 1002.106 | 54,112 | 270,560 |
| South Atlantic | 0.273 | 5.067 | 0.000 | 768.487 | 54,112 | 270,560 |

*Notes*. This Table presents descriptive statistics for the shifts of the instrumental variable. The construction of the shifts is described in (5) (for the county-decade case) and in (C.1) (for the county-technology-decade case). The unit of observation is a county (Panel A) or a county-technology pair (Panel B); units are observed at decadal frequency between 1880 and 1930. In each panel, we report the descriptives for the shifts for all counties (in the first row) and by region (remaining rows). Referenced on pages: 15, 17.

## **Table C.3.** Shock-Level Balance

| | Correlation with Shock | | | |
|---|---|---|---|---|
| | Coefficient | Standard Error | Std. Coefficient | Mean Dep. Var. |
| | (1) | (2) | (3) | (4) |
| **Panel A. Labor-Market Variables** | | | | |
| Income per Capita | -0.769 | (6.739) | -0.002 | 1903.878 |
| Employment Share | 0.712*** | (0.274) | 0.179 | 35.096 |
| Employment Share in: | | | | |
| Agriculture | 0.980*** | (0.240) | 0.072 | 54.687 |
| Metallurgy and Chemicals | 0.336*** | (0.134) | 0.175 | 1.109 |
| Other Manufacturing | 0.041 | (0.080) | 0.011 | 4.623 |
| Textiles | 0.023 | (0.055) | 0.009 | 1.158 |
| Transportation | -0.038 | (0.047) | -0.014 | 4.994 |
| **Panel B. Country-Level Immigration** | | | | |
| Share of Immigrants | -0.252 | (0.249) | -0.040 | 10.823 |
| Share of Immigrants from: | | | | |
| the United Kingdom | 0.088 | (0.106) | 0.012 | 15.306 |
| Western Europe (Except UK) | -1.061 | (0.777) | -0.126 | 19.838 |
| Southern Europe | 0.659 | (0.432) | 0.103 | 4.107 |
| Northern Europe | -0.008 | (0.067) | -0.001 | 9.451 |
| Central and Eastern Europe | 0.807* | (0.439) | 0.067 | 33.407 |
| **Panel C. Within-UK Immigration** | | | | |
| Share of Immigrants | 0.015 | (0.013) | 0.391 | 0.083 |
| Share of Immigrants from: | | | | |
| London | 0.154 | (0.132) | 0.023 | 11.750 |
| South East | 0.201 | (0.152) | 0.028 | 11.064 |
| East of England | -0.139 | (0.165) | -0.021 | 9.661 |
| Yorkshire and the Humber | -0.031 | (0.070) | -0.009 | 2.991 |
| South West | 0.231 | (0.159) | 0.037 | 8.674 |
| West Midlands | -0.184 | (0.127) | -0.029 | 11.327 |
| East Midlands | 0.003 | (0.135) | 0.001 | 6.703 |
| North West | 0.294 | (0.205) | 0.029 | 22.811 |
| North East | -0.083 | (0.116) | -0.014 | 6.634 |
| Wales | -0.445* | (0.241) | -0.063 | 8.385 |

*Notes*. This Table presents the correlation between the "shocks," i.e., predicted patenting according to equation (5) and predetermined US county characteristics. The unit of observation is a county observed at a decadal frequency between 1880 and 1930. Each row reports the coefficient (column 1), standard error (column 2), and standardized coefficient (column 3) from a regression of the lagged dependent variable, shown on the rows, on predicted innovation. Column (4) displays the mean of the dependent variable. All regressions are estimated via OLS. Each regression includes county and decade fixed effects. Standard errors are clustered at the county level. Referenced on page: 17.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.4.** Exposure to Innovation in the US and Innovation in the UK: Recentered Instrument

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| *Second-Stage Regressions* | | | | | | |
| Knowledge Exposure | 0.027* | 0.048*** | 0.038*** | 0.335*** | 0.507*** | 0.411*** |
| | (0.014) | (0.018) | (0.014) | (0.094) | (0.135) | (0.112) |
| *First-Stage Regressions* | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ (Re-centered) | 1.150*** | 1.150*** | 1.150*** | 0.729*** | 0.729*** | 0.729*** |
| | (0.159) | (0.159) | (0.159) | (0.035) | (0.035) | (0.035) |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 613 | 604 | 613 | 612 | 579 | 612 |
| Observations | 3,052 | 3,008 | 3,052 | 46,972 | 31,628 | 46,838 |
| Mean Dep. Var. | 183.327 | 45.403 | 4.186 | 11.186 | 3.866 | 0.253 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2). The instrument for knowledge exposure is recentered, i.e., it is the difference between the variable obtained in (6) and its expected value, obtained from 1000 random permutations, following Borusyak and Hull (2023). Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level and bootstrapped. Referenced on page: 18.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.5.** Exposure to Innovation in the US and Innovation in the UK, Excluding Inventors Listed on both US and UK Patents

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| **Panel A. Observed Knowledge Exposure** | | | | | | |
| Knowledge Exposure | 0.029** | 0.050*** | 0.032** | 0.183* | 0.233* | 0.200* |
| | (0.012) | (0.016) | (0.013) | (0.095) | (0.140) | (0.111) |
| **Panel B. Instrumental Variable Estimates** | | | | | | |
| *Second-Stage Regressions* | | | | | | |
| Knowledge Exposure | 0.023* | 0.043** | 0.029** | 0.295*** | 0.308** | 0.318** |
| | (0.013) | (0.018) | (0.014) | (0.108) | (0.155) | (0.127) |
| *First-Stage Regressions* | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ | 1.155*** | 1.155*** | 1.155*** | 0.730*** | 0.730*** | 0.730*** |
| | (0.159) | (0.159) | (0.159) | (0.035) | (0.035) | (0.035) |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 632 | 613 | 632 | 628 | 569 | 628 |
| Observations | 3,147 | 3,053 | 3,147 | 44,713 | 28,052 | 44,543 |
| Mean Dep. Var. | 123.688 | 31.607 | 2.881 | 8.195 | 3.036 | 0.188 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK, excluding from the patent count all inventors who appear on patents both in the US and in the UK. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2). Panel A reports the association between observed knowledge exposure and innovation; Panel B reports the instrumental variable estimates and the associated first-stage regression. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. Standard errors in Panel B are bootstrapped. Referenced on pages: 23, C40.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.6.** Exposure to Innovation in the US and Firm Innovation in the UK

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| **Panel A. Observed Knowledge Exposure** | | | | | | |
| Knowledge Exposure | 0.051*** | 0.054*** | 0.065*** | 0.180 | 0.078 | 0.274** |
| | (0.015) | (0.021) | (0.018) | (0.123) | (0.206) | (0.137) |
| **Panel B. Instrumental Variable Estimates** | | | | | | |
| *Second-Stage Regressions* | | | | | | |
| Knowledge Exposure | 0.045*** | 0.046** | 0.059*** | 0.328** | 0.029 | 0.428*** |
| | (0.016) | (0.023) | (0.020) | (0.142) | (0.233) | (0.157) |
| *First-Stage Regressions* | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ | 1.155*** | 1.155*** | 1.155*** | 0.730*** | 0.730*** | 0.730*** |
| | (0.159) | (0.159) | (0.159) | (0.035) | (0.035) | (0.035) |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 594 | 537 | 594 | 411 | 298 | 411 |
| Observations | 2,587 | 2,423 | 2,587 | 14,646 | 11,068 | 14,646 |
| Mean Dep. Var. | 58.636 | 12.194 | 1.079 | 10.227 | 2.532 | 0.185 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK. To construct UK patenting, we only include patents with a firm assignee. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2). Panel A reports the association between observed knowledge exposure and innovation; Panel B reports the instrumental variable estimates and the associated first-stage regression. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. Standard errors in Panel B are bootstrapped. Referenced on pages: 23, C38.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.7.** Exposure to Innovation in the US and Innovation in the UK: Single-Matches Sample

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| **Panel A. Observed Knowledge Exposure** | | | | | | |
| Knowledge Exposure | 0.099*** | 0.110*** | 0.110*** | 0.477* | 0.897** | 0.520* |
| | (0.029) | (0.037) | (0.031) | (0.246) | (0.382) | (0.291) |
| **Panel B. Instrumental Variable Estimates** | | | | | | |
| *Second-Stage Regressions* | | | | | | |
| Knowledge Exposure | 0.048* | 0.065* | 0.066** | 0.573** | 0.718* | 0.519* |
| | (0.028) | (0.038) | (0.031) | (0.249) | (0.386) | (0.302) |
| *First-Stage Regressions* | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ | 0.929*** | 0.929*** | 0.929*** | 0.616*** | 0.616*** | 0.616*** |
| | (0.136) | (0.136) | (0.136) | (0.023) | (0.023) | (0.023) |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 632 | 622 | 632 | 631 | 596 | 631 |
| Observations | 3,160 | 3,110 | 3,160 | 48,768 | 33,135 | 48,627 |
| Mean Dep. Var. | 189.983 | 47.505 | 4.348 | 11.538 | 3.973 | 0.262 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK, when the linked US-UK migrants sample only includes single matches. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2). Panel A reports the association between observed knowledge exposure and innovation; Panel B reports the instrumental variable estimates and the associated first-stage regression. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. Standard errors in Panel B are bootstrapped. Referenced on pages: 23, C40.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.8.** Exposure to High-Importance Innovation in the US and Innovation in the UK

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| **Panel A. Exposure to Top 20% High-Impact US Patents** | | | | | | |
| Knowledge Exposure (Top 20%) | 0.153*** | 0.123** | 0.226*** | 1.279*** | 0.458 | 1.576*** |
| | (0.049) | (0.062) | (0.057) | (0.308) | (0.410) | (0.364) |
| **Panel B. Exposure to Top 10% High-Impact US Patents** | | | | | | |
| Knowledge Exposure (Top 10%) | 0.272*** | 0.214* | 0.405*** | 1.885*** | 0.538 | 2.271*** |
| | (0.091) | (0.114) | (0.105) | (0.565) | (0.763) | (0.665) |
| **Panel C. Exposure to Top 5% High-Impact US Patents** | | | | | | |
| Knowledge Exposure (Top 5%) | 0.467*** | 0.364* | 0.704*** | 2.886*** | 1.073 | 3.508*** |
| | (0.165) | (0.202) | (0.189) | (0.998) | (1.350) | (1.161) |
| District FE | Yes | Yes | Yes | – | – | – |
| Decade FE | Yes | Yes | Yes | – | – | – |
| District-Technology FE | – | – | – | Yes | Yes | Yes |
| District-Decade FE | – | – | – | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | Yes | Yes | Yes |
| Number of Districts | 632 | 622 | 632 | 631 | 596 | 631 |
| Observations | 3,147 | 3,098 | 3,147 | 48,553 | 32,940 | 48,414 |
| Mean Dep. Var. | 151.563 | 33.781 | 3.372 | 8.000 | 2.395 | 0.170 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK, for varying levels of "importance" of US patents. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2), except that, in each panel, we only include US patents at the top of the importance distribution. Panel A displays the effect of exposure to US patents in the top 20% of the importance distribution; Panel B restricts the sample to patents in the top 10% of the importance distribution; Panel C further restricts it to patents in the top 5% of the importance distribution. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. Referenced on pages: 23, C39.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.9.** Exposure to Innovation in the US and Innovation in the UK: Alternative Time Horizon for Text-Based Measures

| | Patents by District | | | | | | Patents by District-Technology | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | High-Impact Patents Horizon: | | | US Patents Similarity Horizon: | | | High-Impact Patents Horizon: | | | US Patents Similarity Horizon: | | |
| | (1) 1 Yr. | (2) 5 Yrs. | (3) 10 Yrs. | (4) 1 Yr. | (5) 5 Yrs. | (6) 10 Yrs. | (7) 1 Yr. | (8) 5 Yrs. | (9) 10 Yrs. | (10) 1 Yr. | (11) 5 Yrs. | (12) 10 Yrs. |
| **Panel A. Observed Knowledge Exposure** | | | | | | | | | | | | |
| Knowledge Exposure | 0.036** | 0.052*** | 0.051*** | 0.042*** | 0.042*** | 0.044*** | 0.109 | 0.400*** | 0.191 | 0.231** | 0.266*** | 0.275*** |
| | (0.017) | (0.016) | (0.017) | (0.013) | (0.013) | (0.013) | (0.121) | (0.122) | (0.129) | (0.098) | (0.098) | (0.099) |
| **Panel B. Instrumental Variable Estimates** | | | | | | | | | | | | |
| *Second-Stage Regressions* | | | | | | | | | | | | |
| Knowledge Exposure | 0.029 | 0.048*** | 0.051*** | 0.037*** | 0.038*** | 0.040*** | 0.249* | 0.506*** | 0.295** | 0.376*** | 0.411*** | 0.416*** |
| | (0.018) | (0.018) | (0.019) | (0.014) | (0.014) | (0.014) | (0.133) | (0.135) | (0.147) | (0.112) | (0.112) | (0.112) |
| *First-Stage Regressions* | | | | | | | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ | 1.155*** | 1.155*** | 1.155*** | 1.155*** | 1.155*** | 1.155*** | 0.730*** | 0.730*** | 0.730*** | 0.730*** | 0.730*** | 0.730*** |
| | (0.159) | (0.159) | (0.159) | (0.159) | (0.159) | (0.159) | (0.035) | (0.035) | (0.035) | (0.035) | (0.035) | (0.035) |
| District FE | Yes | Yes | Yes | Yes | Yes | Yes | – | – | – | – | – | – |
| Decade FE | Yes | Yes | Yes | Yes | Yes | Yes | – | – | – | – | – | – |
| District-Technology FE | – | – | – | – | – | – | Yes | Yes | Yes | Yes | Yes | Yes |
| District-Decade FE | – | – | – | – | – | – | Yes | Yes | Yes | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | – | – | – | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Districts | 625 | 622 | 620 | 632 | 632 | 632 | 597 | 596 | 578 | 631 | 631 | 631 |
| Observations | 3,113 | 3,098 | 3,088 | 3,147 | 3,147 | 3,147 | 33,141 | 32,940 | 31,649 | 48,380 | 48,414 | 48,414 |
| Mean Dep. Var. | 44.009 | 47.097 | 40.902 | 0.897 | 4.310 | 8.189 | 3.726 | 3.954 | 3.485 | 0.054 | 0.260 | 0.495 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK. In columns (1–6) (resp. 7–12), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1–3) and (7–9), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (4–6) and (10–12), the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. The regressors are defined in equations (1)–(2). For each dependent variable, we display the regression coefficient using various time thresholds between one and ten years; see the main text for additional details. Panel A reports the association between observed knowledge exposure and innovation; Panel B reports the instrumental variable estimates and the associated first-stage regression. Columns (1–6) include district and decade fixed effects. Columns (7–12) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–6) and district-technology (columns 7–12) level. Standard errors in Panel B are bootstrapped. Referenced on pages: 23, A13, C37.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.10.** Exposure to Innovation in the US and Innovation in the UK: Alternative High-Importance Thresholds

| | Patents by District | | | Patents by District-Technology | | |
|---|---|---|---|---|---|---|
| | (1) Top 20% | (2) Top 10% | (3) Top 5% | (4) Top 20% | (5) Top 10% | (6) Top 5% |
| **Panel A. Observed Knowledge Exposure** | | | | | | |
| Knowledge Exposure | 0.052*** | 0.064*** | 0.049** | 0.400*** | 0.571*** | 0.366* |
| | (0.016) | (0.020) | (0.022) | (0.122) | (0.160) | (0.207) |
| **Panel B. Instrumental Variable Estimates** | | | | | | |
| *Second-Stage Regressions* | | | | | | |
| Knowledge Exposure | 0.048*** | 0.067*** | 0.050** | 0.506*** | 0.665*** | 0.544** |
| | (0.018) | (0.023) | (0.025) | (0.135) | (0.177) | (0.236) |
| *First-Stage Regressions* | | | | | | |
| $\widehat{\text{Knowledge Exposure}}$ | 1.155*** | 1.155*** | 1.155*** | 0.730*** | 0.730*** | 0.730*** |
| | (0.159) | (0.159) | (0.159) | (0.035) | (0.035) | (0.035) |
| District FE | – | – | – | – | – | – |
| Decade FE | – | – | – | – | – | – |
| District-Technology FE | Yes | Yes | Yes | Yes | Yes | Yes |
| District-Decade FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Technology-Decade FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Districts | 622 | 601 | 569 | 596 | 541 | 472 |
| Observations | 3,098 | 2,993 | 2,834 | 32,940 | 23,812 | 16,617 |
| Mean Dep. Var. | 47.097 | 24.676 | 13.854 | 3.954 | 2.728 | 1.994 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK. In columns (1–3) (resp. 4–6), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. In columns (1) and (4), the dependent variable is the number of patents in the top 20% of the importance distribution; in columns (2) and (5), the dependent variable is the number of patents in the top 10% of the importance distribution; in columns (3) and (6), the dependent variable is the number of patents in the top 5% of the importance distribution. The regressors are defined in equations (1)–(2), except that we only include patents in the top 20% (columns 1 and 4), 10% (columns 2 and 5), and 5% (columns 4 and 6) of the patent importance distribution to compute the knowledge exposure term. Panel A reports the association between observed knowledge exposure and innovation; Panel B reports the instrumental variable estimates and the associated first-stage regression. Columns (1–3) include district and decade fixed effects. Columns (4–6) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–3) and district-technology (columns 4–6) level. Standard errors in Panel B are bootstrapped. Referenced on pages: 23, C38.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.11.** Exposure to Innovation in the US and Innovation in the UK: Alternative Definitions of Knowledge Exposure

| | Patents by District | | | | Patents by District-Technology | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Baseline Knowledge Exposure | 0.211*** | | | | 0.161*** | | | | |
| | (0.070) | | | | (0.039) | | | | |
| Knowledge Exposure$^2$ | | 1.478*** | | | | 0.117* | | | |
| | | (0.309) | | | | (0.063) | | | |
| Knowledge Exposure$^3$ | | | 0.161*** | | | | 0.075* | | |
| | | | (0.062) | | | | (0.043) | | |
| Knowledge Exposure$^4$ | | | | 0.212*** | | | | 0.172*** | |
| | | | | (0.070) | | | | (0.041) | |
| Knowledge Exposure$^5$ | | | | | | | | | 0.026 |
| | | | | | | | | | (0.045) |
| District FE | Yes | Yes | Yes | Yes | – | – | – | – | – |
| Decade FE | Yes | Yes | Yes | Yes | – | – | – | – | – |
| District-Technology FE | – | – | – | – | Yes | Yes | Yes | Yes | Yes |
| District-Decade FE | – | – | – | – | Yes | Yes | Yes | Yes | Yes |
| Technology-Decade FE | – | – | – | – | Yes | Yes | Yes | Yes | Yes |
| Number of Districts | 632 | 617 | 632 | 632 | 631 | 616 | 631 | 631 | 631 |
| Observations | 3,779 | 3,690 | 3,779 | 3,779 | 58,534 | 57,032 | 58,534 | 58,534 | 58,534 |
| Mean Dep. Var. | 163.328 | 159.754 | 163.328 | 163.328 | 9.894 | 9.696 | 9.894 | 9.894 | 9.894 |

*Notes*. This Table reports the effect of migrants' exposure to US technology and innovation in the UK. In columns (1–4) (resp. 5–9), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. The dependent variable is the number of patents. The first row displays the coefficient of the baseline measure of knowledge exposure defined in equations (1)–(2); the other rows report the coefficients of alternative definitions ((C.3)–(C.6)). Columns (1–4) include district and decade fixed effects. We apply the inverse hyperbolic sine transformation to each knowledge exposure term to reduce the influence of outlier observations. Columns (5–9) include district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district (columns 1–4) and district-technology (columns 5–9) level. Referenced on pages: 23, C38, C39.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.12.** Within-Family Migration to the US and Innovation in the UK: Excluding London

| | Overall Emigration | | | Non-Return Emigration | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| Post × Emigrant in Family | 0.177** | 0.141 | 0.183** | | | |
| | (0.076) | (0.097) | (0.086) | | | |
| Post × Non-Return Emigrant in Family | | | | 0.134* | 0.054 | 0.150* |
| | | | | (0.076) | (0.108) | (0.085) |
| Family FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Families | 16,203 | 18,654 | 16,137 | 14,579 | 16,940 | 14,515 |
| Observations | 826,353 | 951,354 | 822,987 | 743,529 | 863,940 | 740,265 |
| Mean Dep. Var. | 0.164 | 0.028 | 34.663 | 0.138 | 0.024 | 29.121 |

*Notes*. This Table reports the impact of migration to the US on the innovation activity of migrants' relatives who remain in the UK. The unit of observation is a family defined by all individuals with the same surname in the same county. Units are observed at a yearly frequency between 1880 and 1930. The sample excludes families in London. In columns (1) and (4), the dependent variable is the number of patents issued to family members in the UK. In columns (2) and (5), the dependent variable is the number of patents in the top 20% importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK patents and US patents issued in the previous five years. "Post × Emigrant in Family" is an indicator equal to one in the year the first emigrant in the family moves to the US and in later years, and equal to zero in the previous years. "Post × Non-Return Emigrant in Family" is an indicator equal to one in the year the first emigrant in the family moves to the US (without returning to the UK) and in later years, and equal to zero in the previous years. All columns include family (i.e., surname-county) and year fixed effects. All regressions are estimated through PQML. Standard errors, reported in parentheses, are clustered at the county level. Referenced on pages: 25, C40.
*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

**Table C.13.** Within Neighborhood US Emigration and Innovation in the UK: Excluding Wales

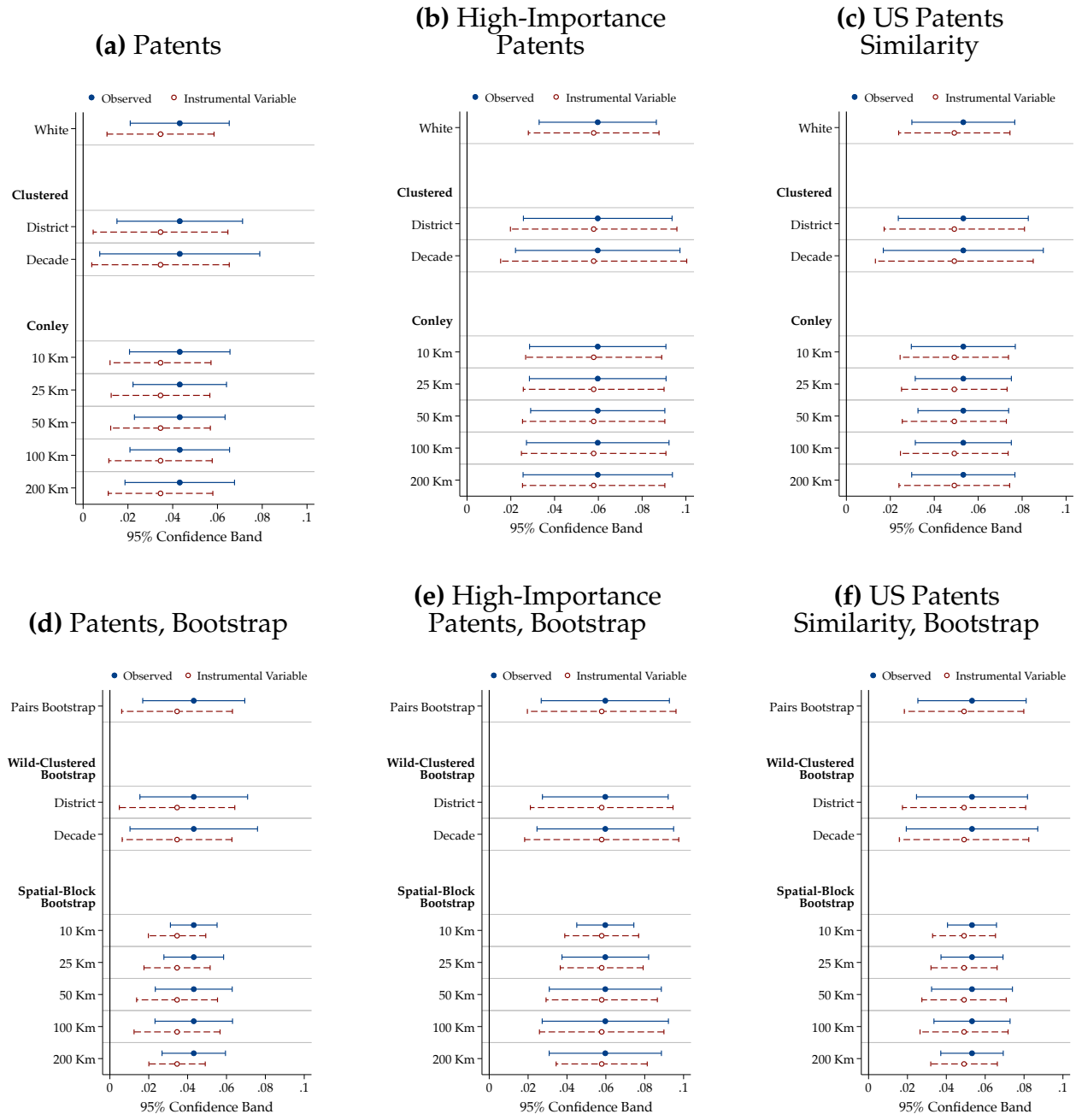
| | Overall Emigration | | | Non-Return Emigration | | |
|---|---|---|---|---|---|---|
| | (1) Number | (2) High Impact | (3) US Patents Similarity | (4) Number | (5) High Impact | (6) US Patents Similarity |
| Post × Emigrant in Neighborhood | 0.144*** | 0.152** | 0.138*** | | | |
| | (0.043) | (0.062) | (0.045) | | | |
| Post × Non-Return Emigrant in Neighborhood | | | | 0.121*** | 0.129** | 0.117*** |
| | | | | (0.039) | (0.060) | (0.041) |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of Individuals | 116,975 | 116,975 | 116,975 | 102,302 | 102,302 | 102,302 |
| Observations | 2,338,648 | 2,338,648 | 2,338,648 | 2,045,292 | 2,045,292 | 2,045,292 |
| Mean Dep. Var. | 0.124 | 0.026 | 18.178 | 0.124 | 0.026 | 18.171 |

*Notes*. This Table reports the impact of emigrants to the US on the innovation activity of their former neighbors in the UK. The unit of observation is an individual inventor observed yearly between 1880 and 1900. The sample includes the universe of British inventors linked to the 1891 population census, as detailed in the main text. The sample excludes all individuals geo-referenced in Wales. In columns (1) and (4), the dependent variable is the number of patents granted to each inventor. In columns (2) and (5), the dependent variable is the number of patents in the top 20% of the importance distribution. In columns (3) and (6), the dependent variable is the text-based similarity between UK patents and US patents issued in the previous five years. "Post × Emigrant in Neighborhood" is an indicator equal to one in the year the first neighbor of the inventor moves to the US and in subsequent years, and equal to zero in the previous years. "Post × Non-Return Emigrant in Neighborhood" is an indicator equal to one in the year the first neighbor of the inventor moves to the US (without returning to the UK) and in subsequent years, and equal to zero in the previous years. All columns include individual and year fixed effects. All regressions are estimated through PQML. Standard errors, reported in parentheses, are clustered at the neighborhood level. Referenced on pages: 26, C41.

*: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

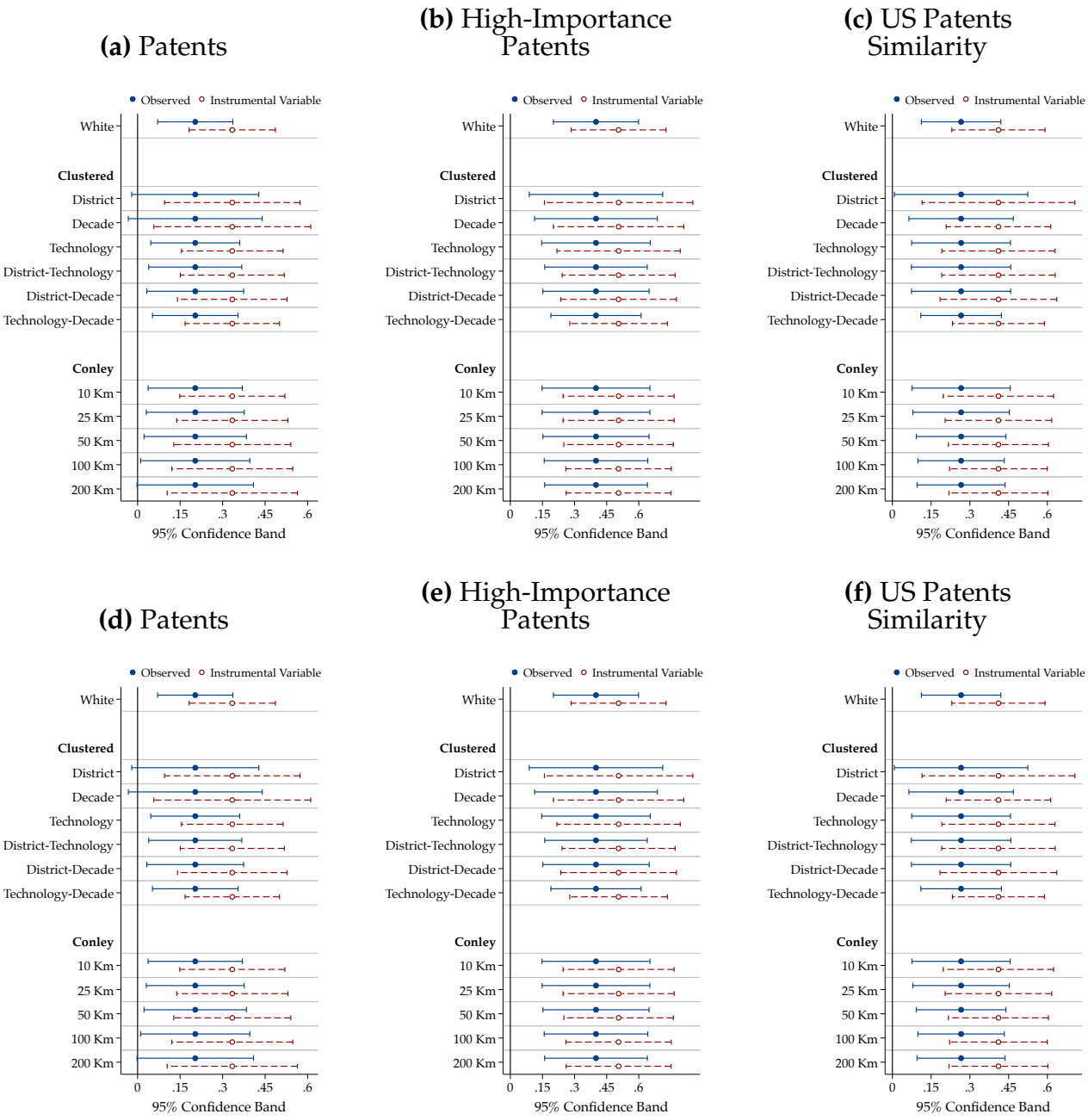
## Figures

**Figure C.1.** Exposure to Innovation in the US and Innovation in the UK: District-Level Alternative Standard Errors



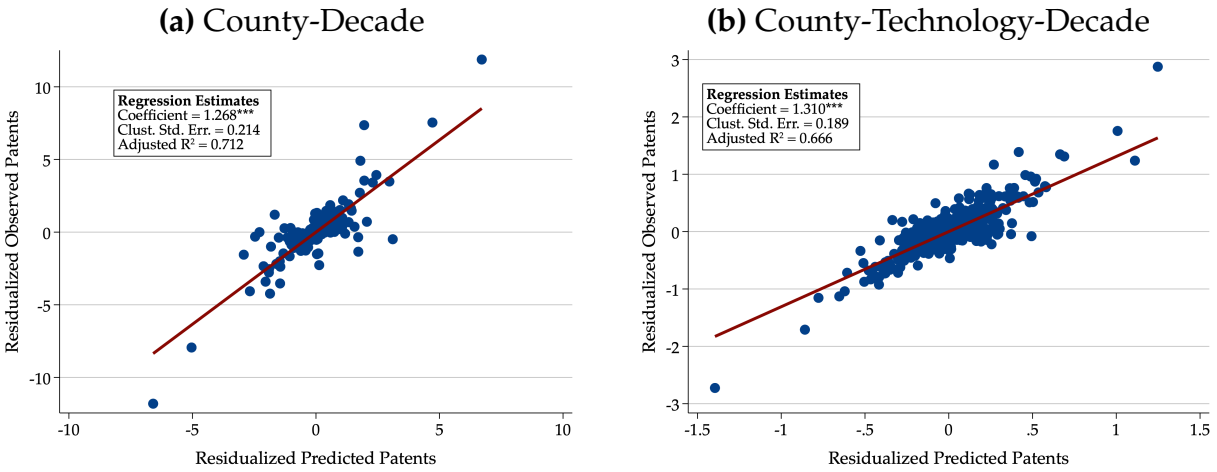

*Notes*. This Figure reports the effect of migrants' exposure to US technology and innovation in the UK, using alternative standard error estimators. The unit of observation is a district at a decadal frequency between 1880 and 1930. The dependent variable is: in Panels C.1a and C.1d, the number of patents; in Panels C.1b and C.1e, the number of patents in the top 20% of the importance distribution; in Panels C.1c and C.1f, the text-based similarity between UK and US patents. Each dot reports the coefficient of knowledge exposure as defined in equation (1). In each row, we report 95% confidence intervals for the knowledge exposure coefficient obtained from six alternative resampling methods: (a) heteroskedasticity-robust standard errors, (b) alternative clustering layers, (c) Conley (1999) standard errors, (d) pairs bootstrap (Davison and Hinkley, 1997), (e) wild cluster bootstrap MacKinnon and Webb (2017), and (f) spatial block bootstrap with resampling blocks defined using the distance of Conley (1999). Bootstrap estimates are obtained from 500 replications. The blue dots report the coefficient of observed knowledge exposure; the red dots report the instrumental variable estimate. All regressions include district and time fixed effects. All specifications are PQML. Referenced on pages: 13, 24, C39.

**Figure C.2.** Exposure to Innovation in the US and Innovation in the UK: District-Technology-Level Alternative Standard Errors



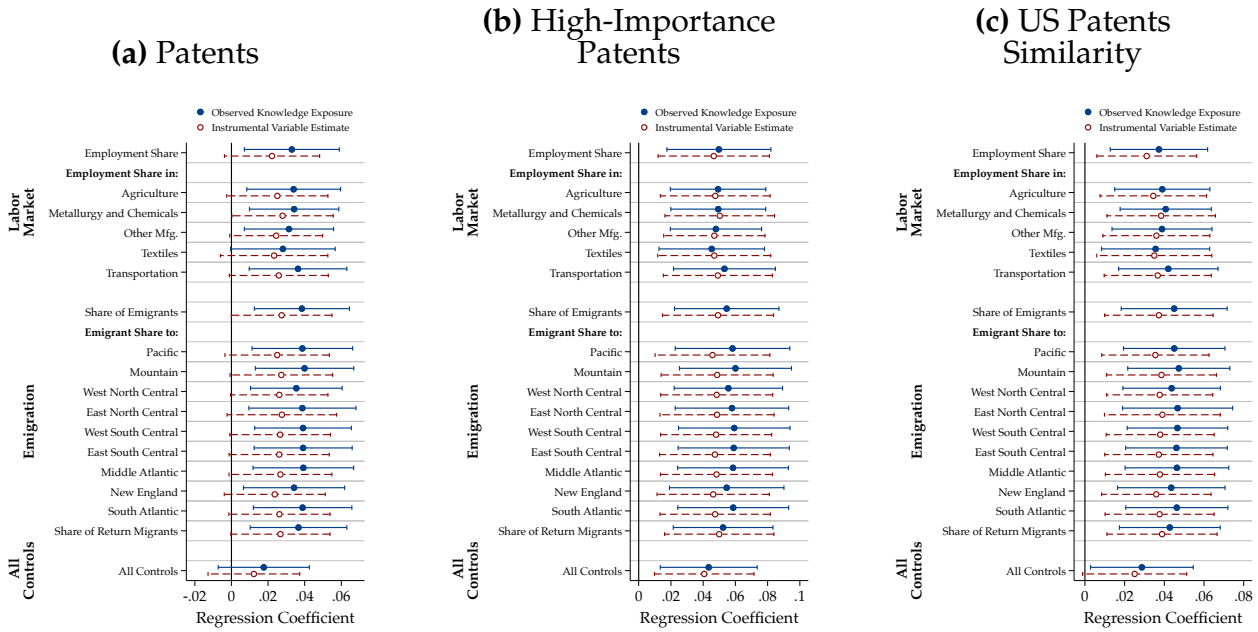

*Notes*. This Figure reports the effect of migrants' exposure to US technology and innovation in the UK, using alternative standard error estimators. The unit of observation is a district-technology pair observed at a decadal frequency between 1880 and 1930. The dependent variable is: in Panels C.2a and C.2d, the number of patents; in panels C.2b and C.2e, the number of patents in the top 20% of the importance distribution; in Panels C.2c and C.2f, the text-based similarity between UK and US patents. Each dot reports the coefficient of knowledge exposure as defined in equation (2). In each row, we report 95% confidence intervals for the knowledge exposure coefficient obtained from six alternative resampling methods: (a) heteroskedasticity-robust standard errors, (b) alternative clustering layers, (c) Conley (1999) standard errors, (d) pairs bootstrap (Davison and Hinkley, 1997), (e) wild cluster bootstrap MacKinnon and Webb (2017), and (f) spatial block bootstrap with resampling blocks defined using the distance of Conley (1999). Bootstrap estimates are obtained from 500 replications. The blue dots report the coefficient of observed knowledge exposure; the red dots report the instrumental variable estimate. All regressions include district and time fixed effects. All specifications are PQML. Referenced on pages: 13, 24, C39.

**Figure C.3.** Correlation Between Actual and Predicted US Patenting



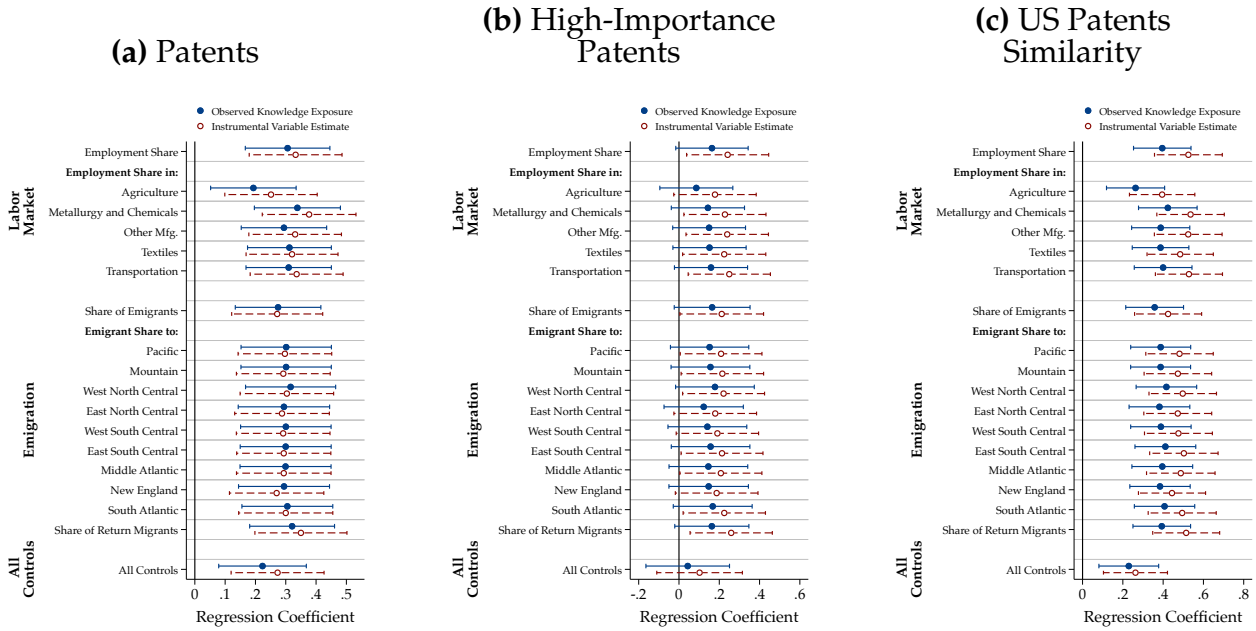

*Notes*. This Figure reports the correlation between observed and predicted patents. In Panel C.3a, the unit of observation is a county in the US; in Panel C.3b, the unit of observation is a county-technology pair. Units are observed at a decadal frequency between 1880 and 1930. Predicted patenting is defined in equation (5). To construct Panel C.3a (resp. C.3b), we residualize both variables on county and decade (resp. county-technology, county-decade, and technology-decade) fixed effects and display the corresponding residuals. Variables are scaled by 1,000 for readability. Each panel overlays a linear fit and the associated 95% confidence bands, and reports the regression coefficients, clustered standard errors, and within-fixed effects $R^2$. Referenced on page: 15.

**Figure C.4.** Exposure to Innovation in the US and Innovation in the UK: District-Level Horse-race Regressions




*Notes*. This Figure reports the effect of migrants' exposure to US technology and innovation in the UK when controlling for district-level observable characteristics. The unit of observation is a district observed at a decadal frequency between 1880 and 1930. In Panel C.4a, the dependent variable is the number of patents. In Panel C.4b, the dependent variable is the number of patents in the top 20% of the importance distribution. In Panel C.4c, the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. Each dot reports the coefficient of knowledge exposure as defined in equation (1). In each regression, we control for one variable measured in 1880 and interacted with decade dummies. In the last row, we include all the controls displayed above. The blue dots report the coefficient of observed knowledge exposure; the red dots report the instrumental variable estimate. All regressions include district and time fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district level. Standard errors of the IV estimates are bootstrapped. Bands report 95% confidence intervals. Referenced on pages: 17, 23, C38.

**Figure C.5.** Exposure to Innovation in the US and Innovation in the UK: District-Technology-Level Horse-race Regressions



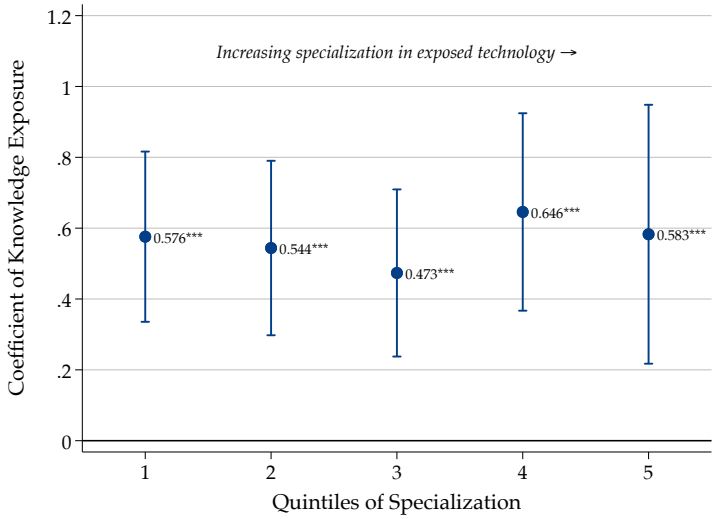

*Notes*. This Figure reports the effect of migrants' exposure to US technology and innovation in the UK when controlling for district-level observable characteristics. The unit of observation is a district-technology pair observed at a decadal frequency between 1880 and 1930. In Panel C.4a, the dependent variable is the number of patents. In Panel C.4b, the dependent variable is the number of patents in the top 20% of the importance distribution. In Panel C.4c, the dependent variable is the text-based similarity between UK and US patents issued in the previous five years. Each dot reports the coefficient of knowledge exposure as defined in equation (2). In each regression, we control for one variable measured in 1880 and interacted with technology-decade dummies. In the last row, we include all the controls displayed above. The blue dots report the coefficient of observed knowledge exposure; the red dots report the instrumental variable estimate. All regressions include district-by-decade, technology-by-decade, and district-by-technology fixed effects. All specifications are PQML. Standard errors, reported in parentheses, are clustered at the district-technology level. Standard errors of the IV estimates are bootstrapped. Bands report 95% confidence intervals. Referenced on pages: 17, 23, C38.

**Figure C.6.** Exposure to Innovation in the US and Innovation in the UK: Excluding One State at a Time from Exposure Computation



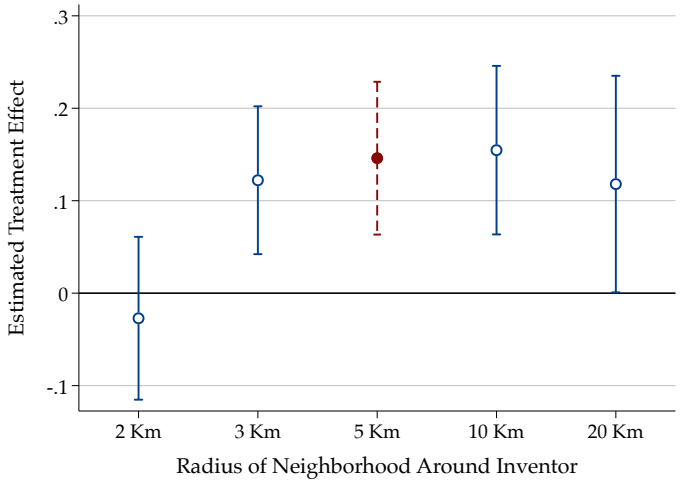

*Notes*. This Figure reports the effect of migrants' exposure to US technology and innovation in the UK. In Panel C.6a (resp. C.6b), the unit of observation is a district (resp. district-technology pair) observed at a decadal frequency between 1880 and 1930. The dependent variable is the number of patents. The regressors are defined in equations (1)–(2). Each dot reports the coefficient obtained when dropping one state from the sample of counties $J$ included in the construction of observed and instrumented knowledge exposure. The blue dots report the correlation regression coefficient of observed knowledge exposure; red dots report the instrumental variable estimates. Panel C.6a includes district and decade fixed effects. Panel C.6b includes district-by-technology, district-by-decade, and technology-by-decade fixed effects. All specifications are PQML. Standard errors are clustered at the district (Panel C.6a) and district-technology (Panel C.6b) level. Confidence bands report 95% confidence intervals. Referenced on page: 18.

**Figure C.7.** Heterogeneous Effects of Exposure to US Technology by Local Specialization



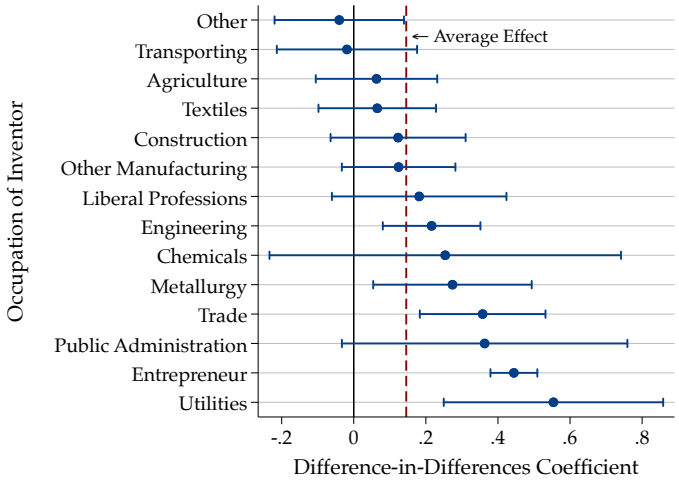

*Notes*. The Figure reports the effect of exposure to US technology through migration ties on UK innovation, by quintile of technology specialization. The dependent variable is the number of patents. We estimate a variant of the instrumental variable regression (C.2) where we interact the baseline knowledge exposure term with quintiles of the specialization of each district into each technology. Specialization is computed as the number of patents issued in each district and technology relative to the total number of patents issued in the district between 1870 and 1879. The dots report the coefficients of these interactions. The model is PQML and includes district-by-technology, district-by-decade, and technology-by-decade fixed effects. Standard errors are clustered at the district-technology level and are bootstrapped. Bands report 95% confidence levels. Referenced on page: 21.

**Figure C.8.** Within-Neighborhood US Emigration: Alternative Neighborhood Thresholds




*Notes*. This Figure reports how emigrants to the US impact the innovation activity fulfilled by their neighbors who remain in the UK. The unit of observation is an individual inventor observed at a yearly frequency between 1880 and 1900. The analysis sample is the universe of inventors linked to the 1891 population census, as detailed in the main text. The dependent variable is the number of patents produced by each inventor. Each dot reports the coefficient of a treatment variable equal to one after the first neighbor of the inventor moves to the US, and zero otherwise. A neighborhood is defined as a $k$-kilometer radius area around each inventor. Each dot refers to a different definition of neighborhood with $k$ equal to 2, 3, 5, 10, and 20 kilometers. The baseline case reported in the main text assumes $k = 5$ and is reported in red. All regressions include inventor and year-fixed effects. Standard errors are clustered at the neighborhood level. Bands report 95% confidence intervals. Referenced on pages: 26, C41.

**Figure C.9.** Heterogeneous Effects of Within-Neighborhood Emigration on Innovation by Occupation of the Stayers




*Notes*. This Figure reports how emigrants to the US impact the innovation activity fulfilled by their neighbors who remain in the UK. The unit of observation is an individual inventor observed at a yearly frequency between 1880 and 1900. The analysis sample is the universe of inventors linked to the 1891 population census, as detailed in the main text. The dependent variable is the number of patents produced by each inventor. Each coefficient refers to an interaction term between the baseline treatment–an indicator equal to one after the first neighbor of the inventor moves to the US, and zero otherwise– and a dummy variable that codes the occupation of the inventor. The dashed red line indicates the average treatment effect reported in the main text. All regressions include inventor and year-fixed effects. Standard errors are clustered at the neighborhood level. The bands report 95% confidence intervals. Referenced on page: 27.

## APPENDIX REFERENCES


**Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez**, "Automated Linking of Historical Data," *Journal of Economic Literature*, 2021, *59* (3), 865–918.

**_ , Leah Platt Boustan, and Katherine Eriksson**, "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 2014, *122* (3), 467–506.

**Baines, Dudley**, *Migration in a mature economy: emigration and internal migration in England and Wales 1861-1900*, Cambridge University Press, 2002.

**Bergeaud, Antonin and Cyril Verluise**, "A New Dataset to Study a Century of Innovation in Europe and in the US," *Research Policy*, 2024, *53* (1), 104903.

**Berkes, Enrico**, "Comprehensive Universe of US patents (CUSP): Data and Facts," *Working Paper*, 2018.

**Berthoff, Rowland**, *British Immigrants in Industrial America 1790-1950*, Harvard University Press, 1953.

**Borusyak, Kirill and Peter Hull**, "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

**_ , _ , and Xavier Jaravel**, "Quasi-Experimental Shift-Share Research Designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

**Chen, Jiafeng and Jonathan Roth**, "Logs with Zeros? Some Problems and Solutions," *The Quarterly Journal of Economics*, 2024, *139* (2), 891–936.

**Conley, Timothy G**, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

**Curtis, Sidney J**, *The story of the Marsden mayoralty: with sketch of the mayor's life*, Leeds (UK): Express Office, 1875.

**Davison, Anthony C. and David V. Hinkley**, *Bootstrap Methods and Their Application* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press, 1997.

**Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters**, "A Method to Construct Geographical Crosswalks with an Application to US Counties Since 1790," *NBER Working Paper*, 2020, (No. w26770).

**Green, A**, "Networks and Geographic Mobility: Evidence from World War II Navy ships," *Working Paper*, 2025.

**Hanlon, W Walker**, "British Patent Technology Classification Database: 1855-1882," 2016. Unpublished [data collection].

**Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy**, "Measuring Technological Innovation Over the Long Run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

**Lan, Tian and Paul Longley**, "Geo-Referencing and Mapping 1901 Census Addresses for England and Wales," *ISPRS International Journal of Geo-Information*, 2019, *8* (8), 320.

**Landers, John**, *Death and the Metropolis: Studies in the Demographic History of London, 1670-1830* number 20, Cambridge University Press, 2006.

**MacKinnon, James G and Matthew D Webb**, "Wild Bootstrap Inference for Wildly Different Cluster Sizes," *Journal of Applied Econometrics*, 2017, *32* (2), 233–254.

**Mullahy, John**, "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economics and Statistics*, 1997, *79* (4), 586–593.

**Nicholas, Tom**, "Technology, Innovation, and Economic Growth in Britain since 1870," in Roderick Floud, Jane Humphries, and Paul Johnson, eds., *The Cambridge Economic History of Modern Britain*, Cambridge (UK): Cambridge University Press, 2014, pp. 181–204.

**Nuvolari, Alessandro and Valentina Tartari**, "Bennet Woodcroft and the Value of English Patents, 1617–1841," *Explorations in Economic History*, 2011, *48* (1), 97–115.

**_ , _ , and Matteo Tranchero**, "Patterns of Innovation During the Industrial Revolution: A Reappraisal Using a Composite Indicator of Patent Quality," *Explorations in Economic History*, 2021, *82*, 101419.

**Publications, Ltd. Finishing**, "Early British Patents. A Cradle of Inventions. British Patents 1617 to 1895," 2018. MFIS [data collection].

**Rosenberg, Nathan and Manuel Trajtenberg**, "A General-Purpose Technology at Work: The Corliss

Steam Engine in the Late-Nineteenth-Century United States," *The Journal of Economic History*, 2004, *64* (1), 61–99.

**Ruggles, Steven, Catherine Fitch, Ronald Goeken, J Hacker, M Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek**, "IPUMS ancestry full count data: Version 3.0 [dataset]," *Minneapolis, MN: IPUMS*, 2021.

**Schurer, K and E Higgs**, "Integrated Census Microdata (I-CeM) Names and Addresses, 1851-1911: Special Licence Access," 2020. [data collection] Second Edition, UKDS.

**Schürer, Kevin and Joe Day**, "Migration to London and the Development of the North–South Divide, 1851–1911," *Social History*, 2019, *44* (1), 26–56.

**Spitzer, Yannay and Ariell Zimran**, "Migrant Self-Selection: Anthropometric Evidence from the Mass Migration of Italians to the United States, 1907–1925," *Journal of Development Economics*, 2018, *134*, 226–247.

**Wagner, Donald**, *Science and Civilisation in China*, Cambridge University Press (UK), 2008.

**Wooldridge, Jeffrey M**, *Econometric Analysis of Cross-Section and Panel Data*, MIT press, 2010.